Diode Lasers and Photonic Integrated Circuits

WILEY SERIES IN MICROWAVE AND OPTICAL ENGINEERING

KAI CHANG, Editor
Texas A&M University

A complete list of the titles in this series appears at the end of this volume.

Diode Lasers and Photonic Integrated Circuits

Second Edition

LARRY A. COLDREN
University of California, Santa Barbara

SCOTT W. CORZINE
Infinera

MILAN L. MAŠANOVIĆ
Freedom Photonics



A JOHN WILEY & SONS, INC., PUBLICATION

Copyright © 2012 by John Wiley & Sons. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400, fax 978-750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, 201-748-6011, fax 201-748-6008, or online at ttp://www.wiley.com/go/permission.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at 877-762-2974, outside the United States at 317-572-3993 or fax 317- 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data:

Printed in the United States of America

Coldren, L. A. (Larry A.)
Diode lasers and photonic integrated circuits / Larry A. Coldren, Scott Corzine, Milan Mashanovitch. – 2nd ed.
p. cm. – (Wiley series in microwave and optical engineering ; 218)
Includes index.
ISBN 978-0-470-48412-8 (hardback)
1. Semiconductor lasers. 2. Integrated circuits. I. Corzine, S. W. (Scott W.) II. Mashanovitch, Milan, 1974- III. Title.
TA1700.C646 2012
621.382'7–dc23

2011022712

10 9 8 7 6 5 4 3 2 1

Contents

Preface xvii

Acknowledgments xxi

List of Fundamental Constants xxiii

1 Ingredients 1

1.1 Introduction 1

1.2 Energy Levels and Bands in Solids 5

1.3 Spontaneous and Stimulated Transitions: The Creation of Light 7

1.4 Transverse Confinement of Carriers and Photons in Diode Lasers: The Double Heterostructure 10

1.5 Semiconductor Materials for Diode Lasers 13

1.6 Epitaxial Growth Technology 20

1.7 Lateral Confinement of Current, Carriers, and Photons for Practical Lasers 24

1.8 Practical Laser Examples 31

References 39

Reading List 40

Problems 40

2 A Phenomenological Approach to Diode Lasers 45

2.1 Introduction 45

2.2 Carrier Generation and Recombination in Active Regions 46

2.3 Spontaneous Photon Generation and LEDs 49
2.4 Photon Generation and Loss in Laser Cavities 52
2.5 Threshold or Steady-State Gain in Lasers 55
2.6 Threshold Current and Power Out Versus Current 60
2.6.1 Basic P–I Characteristics 60
2.6.2 Gain Models and Their Use in Designing Lasers 64
2.7 Relaxation Resonance and Frequency Response 70
2.8 Characterizing Real Diode Lasers 74
2.8.1 Internal Parameters for In-Plane Lasers: $\langle \alpha_i \rangle$, η_i, and g versus J 75
2.8.2 Internal Parameters for VCSELs: η_i and g versus J, $\langle \alpha_i \rangle$, and α_m 78
2.8.3 Efficiency and Heat Flow 79
2.8.4 Temperature Dependence of Drive Current 80
2.8.5 Derivative Analysis 84
References 86
Reading List 87
Problems 87

3 Mirrors and Resonators for Diode Lasers 91

3.1 Introduction 91
3.2 Scattering Theory 92
3.3 S and T Matrices for Some Common Elements 95
3.3.1 The Dielectric Interface 96
3.3.2 Transmission Line with No Discontinuities 98
3.3.3 Dielectric Segment and the Fabry–Perot Etalon 100
3.3.4 S-Parameter Computation Using Mason's Rule 104
3.3.5 Fabry–Perot Laser 105
3.4 Three- and Four-Mirror Laser Cavities 107
3.4.1 Three-Mirror Lasers 107
3.4.2 Four-Mirror Lasers 111
3.5 Gratings 113
3.5.1 Introduction 113
3.5.2 Transmission Matrix Theory of Gratings 115
3.5.3 Effective Mirror Model for Gratings 121

3.6 Lasers Based on DBR Mirrors 123
3.6.1 Introduction 123
3.6.2 Threshold Gain and Power Out 124
3.6.3 Mode Selection in DBR-Based Lasers 127
3.6.4 VCSEL Design 128
3.6.5 In-Plane DBR Lasers and Tunability 135
3.6.6 Mode Suppression Ratio in DBR Laser 139
3.7 DFB Lasers 141
3.7.1 Introduction 141
3.7.2 Calculation of the Threshold Gains and Wavelengths 143
3.7.3 On Mode Suppression in DFB Lasers 149
References 151
Reading List 151
Problems 151

4 Gain and Current Relations 157

4.1 Introduction 157
4.2 Radiative Transitions 158
4.2.1 Basic Definitions and Fundamental Relationships 158
4.2.2 Fundamental Description of the Radiative Transition Rate 162
4.2.3 Transition Matrix Element 165
4.2.4 Reduced Density of States 170
4.2.5 Correspondence with Einstein's Stimulated Rate Constant 174
4.3 Optical Gain 174
4.3.1 General Expression for Gain 174
4.3.2 Lineshape Broadening 181
4.3.3 General Features of the Gain Spectrum 185
4.3.4 Many-Body Effects 187
4.3.5 Polarization and Piezoelectricity 190
4.4 Spontaneous Emission 192
4.4.1 Single-Mode Spontaneous Emission Rate 192
4.4.2 Total Spontaneous Emission Rate 193
4.4.3 Spontaneous Emission Factor 198
4.4.4 Purcell Effect 198

4.5 Nonradiative Transitions 199
4.5.1 Defect and Impurity Recombination 199
4.5.2 Surface and Interface Recombination 202
4.5.3 Auger Recombination 211
4.6 Active Materials and Their Characteristics 218
4.6.1 Strained Materials and Doped Materials 218
4.6.2 Gain Spectra of Common Active Materials 220
4.6.3 Gain versus Carrier Density 223
4.6.4 Spontaneous Emission Spectra and Current versus Carrier Density 227
4.6.5 Gain versus Current Density 229
4.6.6 Experimental Gain Curves 233
4.6.7 Dependence on Well Width, Doping, and Temperature 234
References 238
Reading List 240
Problems 240

5 Dynamic Effects 247

5.1 Introduction 247
5.2 Review of Chapter 2 248
5.2.1 The Rate Equations 249
5.2.2 Steady-State Solutions 250
Case (i): Well Below Threshold 251
Case (ii): Above Threshold 252
Case (iii): Below and Above Threshold 253
5.2.3 Steady-State Multimode Solutions 255
5.3 Differential Analysis of the Rate Equations 257
5.3.1 Small-Signal Frequency Response 261
5.3.2 Small-Signal Transient Response 266
5.3.3 Small-Signal FM Response or Frequency Chirping 270
5.4 Large-Signal Analysis 276
5.4.1 Large-Signal Modulation: Numerical Analysis of the Multimode Rate Equations 277
5.4.2 Mode Locking 279
5.4.3 Turn-On Delay 283
5.4.4 Large-Signal Frequency Chirping 286

5.5 Relative Intensity Noise and Linewidth 288
5.5.1 General Definition of RIN and the Spectral Density Function 288
5.5.2 The Schawlow–Townes Linewidth 292
5.5.3 The Langevin Approach 294
5.5.4 Langevin Noise Spectral Densities and RIN 295
5.5.5 Frequency Noise 301
5.5.6 Linewidth 303
5.6 Carrier Transport Effects 308
5.7 Feedback Effects and Injection Locking 311
5.7.1 Optical Feedback Effects—Static Characteristics 311
5.7.2 Injection Locking—Static Characteristics 317
5.7.3 Injection and Feedback Dynamic Characteristics and Stability 320
5.7.4 Feedback Effects on Laser Linewidth 321
References 328
Reading List 329
Problems 329

6 Perturbation, Coupled-Mode Theory, Modal Excitations, and Applications 335

6.1 Introduction 335
6.2 Guided-Mode Power and Effective Width 336
6.3 Perturbation Theory 339
6.4 Coupled-Mode Theory: Two-Mode Coupling 342
6.4.1 Contradirectional Coupling: Gratings 342
6.4.2 DFB Lasers 353
6.4.3 Codirectional Coupling: Directional Couplers 356
6.4.4 Codirectional Coupler Filters and Electro-optic Switches 370
6.5 Modal Excitation 376
6.6 Two Mode Interference and Multimode Interference 378
6.7 Star Couplers 381
6.8 Photonic Multiplexers, Demultiplexers and Routers 382
6.8.1 Arrayed Waveguide Grating De/Multiplexers and Routers 383

6.8.2 Echelle Grating based De/Multiplexers and Routers 389
6.9 Conclusions 390
References 390
Reading List 391
Problems 391

7 Dielectric Waveguides 395

7.1 Introduction 395
7.2 Plane Waves Incident on a Planar Dielectric Boundary 396
7.3 Dielectric Waveguide Analysis Techniques 400
7.3.1 Standing Wave Technique 400
7.3.2 Transverse Resonance 403
7.3.3 WKB Method for Arbitrary Waveguide Profiles 410
7.3.4 2-D Effective Index Technique for Buried Rib Waveguides 418
7.3.5 Analysis of Curved Optical Waveguides using Conformal Mapping 421
7.3.6 Numerical Mode Solving Methods for Arbitrary Waveguide Profiles 424
7.4 Numerical Techniques for Analyzing PICs 427
7.4.1 Introduction 427
7.4.2 Implicit Finite-Difference Beam-Propagation Method 429
7.4.3 Calculation of Propagation Constants in a z-invariant Waveguide from a Beam Propagation Solution 432
7.4.4 Calculation of Eigenmode Profile from a Beam Propagation Solution 434
7.5 Goos–Hanchen Effect and Total Internal Reflection Components 434
7.5.1 Total Internal Reflection Mirrors 435
7.6 Losses in Dielectric Waveguides 437
7.6.1 Absorption Losses in Dielectric Waveguides 437
7.6.2 Scattering Losses in Dielectric Waveguides 438
7.6.3 Radiation Losses for Nominally Guided Modes 438
References 445
Reading List 446
Problems 446

8 Photonic Integrated Circuits 451

8.1 Introduction 451

8.2 Tunable, Widely Tunable, and Externally Modulated Lasers 452
8.2.1 Two- and Three-Section In-plane DBR Lasers 452
8.2.2 Widely Tunable Diode Lasers 458
8.2.3 Other Extended Tuning Range Diode Laser Implementations 463
8.2.4 Externally Modulated Lasers 474
8.2.5 Semiconductor Optical Amplifiers 481
8.2.6 Transmitter Arrays 484

8.3 Advanced PICs 484
8.3.1 Waveguide Photodetectors 485
8.3.2 Transceivers/Wavelength Converters and Triplexers 488

8.4 PICs for Coherent Optical Communications 491
8.4.1 Coherent Optical Communications Primer 492
8.4.2 Coherent Detection 495
8.4.3 Coherent Receiver Implementations 495
8.4.4 Vector Transmitters 498

References 499

Reading List 503

Problems 503

APPENDICES

1 Review of Elementary Solid-State Physics 509

A1.1 A Quantum Mechanics Primer 509
A1.1.1 Introduction 509
A1.1.2 Potential Wells and Bound Electrons 511

A1.2 Elements of Solid-State Physics 516
A1.2.1 Electrons in Crystals and Energy Bands 516
A1.2.2 Effective Mass 520
A1.2.3 Density of States Using a Free-Electron (Effective Mass) Theory 522

References 527

Reading List 527

2 Relationships between Fermi Energy and Carrier Density and Leakage **529**

A2.1 General Relationships 529

A2.2 Approximations for Bulk Materials 532

A2.3 Carrier Leakage Over Heterobarriers 537

A2.4 Internal Quantum Efficiency 542

References 544

Reading List 544

3 Introduction to Optical Waveguiding in Simple Double-Heterostructures **545**

A3.1 Introduction 545

A3.2 Three-Layer Slab Dielectric Waveguide 546

A3.2.1 Symmetric Slab Case 547

A3.2.2 General Asymmetric Slab Case 548

A3.2.3 Transverse Confinement Factor, Γ_x 550

A3.3 Effective Index Technique for Two-Dimensional Waveguides 551

A3.4 Far Fields 555

References 557

Reading List 557

4 Density of Optical Modes, Blackbody Radiation, and Spontaneous Emission Factor **559**

A4.1 Optical Cavity Modes 559

A4.2 Blackbody Radiation 561

A4.3 Spontaneous Emission Factor, β_{sp} 562

Reading List 563

5 Modal Gain, Modal Loss, and Confinement Factors **565**

A5.1 Introduction 565

A5.2 Classical Definition of Modal Gain 566

A5.3 Modal Gain and Confinement Factors 568

A5.4 Internal Modal Loss 570

A5.5 More Exact Analysis of the Active/Passive Section Cavity 571
A5.5.1 Axial Confinement Factor 572
A5.5.2 Threshold Condition and Differential Efficiency 573
A5.6 Effects of Dispersion on Modal Gain 576

6 Einstein's Approach to Gain and Spontaneous Emission 579

A6.1 Introduction 579
A6.2 Einstein *A* and *B* Coefficients 582
A6.3 Thermal Equilibrium 584
A6.4 Calculation of Gain 585
A6.5 Calculation of Spontaneous Emission Rate 589
Reading List 592

7 Periodic Structures and the Transmission Matrix 593

A7.1 Introduction 593
A7.2 Eigenvalues and Eigenvectors 593
A7.3 Application to Dielectric Stacks at the Bragg Condition 595
A7.4 Application to Dielectric Stacks Away from the Bragg Condition 597
A7.5 Correspondence with Approximate Techniques 600
A7.5.1 Fourier Limit 601
A7.5.2 Coupled-Mode Limit 602
A7.6 Generalized Reflectivity at the Bragg Condition 603
Reading List 605
Problems 605

8 Electronic States in Semiconductors 609

A8.1 Introduction 609
A8.2 General Description of Electronic States 609
A8.3 Bloch Functions and the Momentum Matrix Element 611
A8.4 Band Structure in Quantum Wells 615
A8.4.1 Conduction Band 615

A8.4.2 Valence Band 616
A8.4.3 Strained Quantum Wells 623
References 627
Reading List 628

9 Fermi's Golden Rule 629
A9.1 Introduction 629
A9.2 Semiclassical Derivation of the Transition Rate 630
A9.2.1 Case I: The Matrix Element-Density of Final States Product is a Constant 632
A9.2.2 Case II: The Matrix Element-Density of Final States Product is a Delta Function 635
A9.2.3 Case III: The Matrix Element-Density of Final States Product is a Lorentzian 636
Reading List 637
Problems 638

10 Transition Matrix Element 639
A10.1 General Derivation 639
A10.2 Polarization-Dependent Effects 641
A10.3 Inclusion of Envelope Functions in Quantum Wells 645
Reading List 646

11 Strained Bandgaps 647
A11.1 General Definitions of Stress and Strain 647
A11.2 Relationship Between Strain and Bandgap 650
A11.3 Relationship Between Strain and Band Structure 655
References 656

12 Threshold Energy for Auger Processes 657
A12.1 CCCH Process 657
A12.2 CHHS and CHHL Processes 659

13 Langevin Noise **661**

A13.1 Properties of Langevin Noise Sources 661
A13.1.1 Correlation Functions and Spectral Densities 661
A13.1.2 Evaluation of Langevin Noise Correlation Strengths 664

A13.2 Specific Langevin Noise Correlations 665
A13.2.1 Photon Density and Carrier Density Langevin Noise Correlations 665
A13.2.2 Photon Density and Output Power Langevin Noise Correlations 666
A13.2.3 Photon Density and Phase Langevin Noise Correlations 667

A13.3 Evaluation of Noise Spectral Densities 669
A13.3.1 Photon Noise Spectral Density 669
A13.3.2 Output Power Noise Spectral Density 670
A13.3.3 Carrier Noise Spectral Density 671

References 672

Problems 672

14 Derivation Details for Perturbation Formulas **675**

Reading List 676

15 Multimode Interference **677**

A15.1 Multimode Interference-Based Couplers 677

A15.2 Guided-Mode Propagation Analysis 678
A15.2.1 General Interference 679
A15.2.2 Restricted Multimode Interference 681

A15.3 MMI Physical Properties 682
A15.3.1 Fabrication 682
A15.3.2 Imaging Quality 682
A15.3.3 Inherent Loss and Optical Bandwidth 682
A15.3.4 Polarization Dependence 683
A15.3.5 Reflection Properties 683

Reference 683

16 The Electro-Optic Effect **685**

References 692

Reading List 692

17 Solution of Finite Difference Problems **693**

A17.1 Matrix Formalism 693

A17.2 One-Dimensional Dielectric Slab Example 695

Reading List 696

Index **697**

Preface

Diode lasers and related photonic integrated circuits have become even more commercially important since the first edition of this book was published, in 1995. They are used in a wide variety of applications ranging from the readout sources in DVD and Blu-ray disk players, laser printers, mice and pointers, to the complex multiwavelength transmitters and receivers in optical fiber communication systems that carry hundreds of gigabits per second of information. New applications, such as solid-state lighting sources, or sources for high-spectral-efficiency telecommunications networks continue to emerge as the devices become more varied, reliable, manufacturable, and inexpensive.

In this edition, the GaN-based materials, which have become important in the UV/blue/green wavelength regions as well as for solid-state lighting, are included with equal emphasis to the GaAs and InP-based materials, which provide emission from the red to about 1 micron in wavelength on GaAs and over the 1.3–1.6 micron wavelength range on InP. Thus, the range of applications that can be addressed with a mastery contents of this edition is very broad.

This book has been written to be a resource for professors, graduate students, industry researchers, and design engineers dealing with the subject of diode lasers and related photonic integrated circuits for a range of applications. The depth of coverage is relatively advanced, but the initial chapters provide a working knowledge of semiconductor lasers before delving into much of the advanced material. Appendices are used both to provide a review of background material as well as some of the details of the more advanced topics. Thus, by appropriate use of the appendices, the text can support teaching the material at different academic levels, but it remains self-contained.

Significant new material has been added, to both improve on the original text, and to address important technology developments over the last decade. One of the key novel features is the addition of many worked examples throughout all the chapters to better illustrate how to apply the theory that is being covered. New homework problems have also been added to supplement the previous ones, some

of which are less complex than the previous problems, because many found them too difficult for beginning students or casual reference readers.

New topics that are being covered in this second edition are more introductory material related to benefits, applications and basics of laser diodes and photonic ICs; additional methods for analytic calculation of S parameters based on Mason's rule; expanded treatment of DFB and VCSEL lasers; additional material on quantum dots, gain and other material parameters for both GaN, InP and GaAs based active regions and devices; treatment of the mode-locked lasers and injection locking; total internal reflection mirrors and beam splitters; a new appendix and section on mutimode interference effects and devices; treatment of star couplers and photonic multiplexers, demultiplexers and routers, and their design; expanded treatment of losses in dielectric waveguides; treatment of light propagation in curved waveguides; significantly expanded treatment of tunable and widely tunable laser diodes; expanded treatment of externally modulated lasers, including Mach-Zehnder modulators and semiconductor optical amplifiers; additional material on waveguide photodiodes, optical transceivers and triplexers; and a full section on basics and PICs for coherent communications.

Also available online with the second edition will be a number of password-protected tools, such as BPM and S and T matrix computation code, DFB laser code, mode solving code, as well as color versions of all figures, all of which should be useful for instructors and students, as well as other readers.

The full text is intended for use at the graduate level, although a fairly comprehensive introductory course on diode lasers at an advanced undergraduate level could be based around the material in Chapters 1 through 3 together with Appendices 1 through 7.

It is assumed that the readers have been exposed to elementary quantum mechanics, solid-state physics, and electromagnetic theory at the undergraduate level. It is also recommended that they have had an introductory optoelectronics course. Appendices 1 and 3 review most of the necessary background in just about all of the required detail. Thus, it is possible to use the book with less prior educational background, provided these review appendices are covered with some care.

For use in a more advanced graduate class, it would not be necessary to cover the material in the first seven appendices. (Of course, it would still be there for reference, and the associated homework problems could still be assigned to ensure its understanding. Nevertheless, it is still recommended that Appendix 5, which covers the definitions of modal gain and loss, be reviewed because this is not well understood by the average worker in the field.) The coverage could then move efficiently through the first three chapters and into Chapters 4 and 5, which deal with the details of gain and laser dynamics in a first course. For more focus on the gain physics some of Appendices 8 through 12 could be included in the coverage. In any event, their inclusion provides for a very self-contained treatment of this important subject matter.

Chapters 6 and 7 deal more with the electromagnetic wave aspects of photonic ICs and diode lasers. This material is essential for understanding the more advanced PIC type of devices used in modern fiber-optic links and networks.

However, keeping this material to last allows the student to develop a fairly complete understanding of the operation of diode lasers without getting bogged down in the mathematical techniques necessary for the lateral waveguide analysis. Thus, a working understanding and appreciation of laser operation can be gained in only one course. Chapter 6 deals with perturbation, coupled-mode theory and modal excitation while Chapter 7 deals with dielectric waveguide analysis. Putting Chapter 6 first emphasizes the generality of this material. That is, one really does not need to know the details of the lateral mode profile to develop these powerful techniques. Using the coupled-mode results, gratings and DFB lasers are again investigated. Historically, these components were primarily analyzed with this theory. However, in this text grating-based DFB and DBR lasers are first analyzed in Chapter 3 using exact matrix multiplication techniques, from which approximate formulas identical to those derived with coupled mode theory result. The proliferation of computers and the advent of lasers using complex grating designs with many separate sections has led the authors to assert that the matrix multiplication technique should be the primary approach taught to students. The advent of the vertical-cavity laser also supports this approach. Nevertheless, it should be realized that coupled-mode theory is very important to reduce the description of the properties of complex waveguide geometries to simple analytic formulae, which are especially useful in design work. Chapter 7 also introduces some basic numerical techniques, which have become indispensable with the availability of powerful personal computers and efficient software for solving complex numerical algorithms. The finite-difference technique is introduced for optical waveguide analysis, while the beam-propagation method is discussed as a key tool for analyzing real PIC structures.

Chapter 8 pulls together most of the material in the first seven chapters by providing a comprehensive overview of the development of photonic integrated circuits, with a series of design examples of relatively complex photonic integrated circuits.

Unlike many books in this field, this book is written as an engineering text. The reader is first trained to be able to solve problems on real diode lasers, based on a phenomenological understanding, before going into the complex physical details such as the material gain process or mode-coupling in dielectric waveguides. This provides motivation for learning the underlying details as well as a toolbox of techniques to immediately apply each new advanced detail in solving real problems. Also, attention has been paid to accuracy and consistency. For example, a careful distinction between the internal quantum efficiency in LEDs and injection efficiency in lasers is made, and calculations of gain not only illustrate an analysis technique, but they actually agree with experimental data. Finally, by maintaining consistent notation throughout all of the chapters and appendices, a unique self-contained treatment of all of the included material emerges.

L. A. Coldren
S. W. Corzine
M. L. Mašanović

Acknowledgments

The original text grew out of lecture notes developed for a graduate-level course on diode lasers and guided-wave optics (Coldren) as well as a PhD dissertation written on vertical-cavity lasers with strained quantum-well active regions at UC-Santa Barbara (Corzine). This second edition benefited from the intensive photonic integration research conducted at UC Santa Barbara, from further course notes development by Professor Coldren and Dr. Mašanović, and from ten or so more generations of graduate students who took the class, asked important questions, and generally have helped come up with better examples and answers to homework problems. The supplemental materials on DFB and tunable lasers, MMIs, AWGRs, and injection locking were especially due to the significant efforts of Dr. Mašanović.

Outside the classroom, and generally independent of the course in question, interactions with other faculty and students at UC Santa Barbara have contributed greatly to our understanding of the subject on which the book expounds. Original contributors include Professors Kroemer, Dagli, Bowers, Yeh, and Suemune (now at Hokkaido University), and students include Drs. Ran-Hong Yan, Randy Geels, Jeff Scott, Bruce Young, Dubravko Babić, Zuon-Min Chuang, Vijay Jayaraman, and Radha Nagarajan.

In addition, we would like to acknowledge numerous people that we have interacted and collaborated with while this edition has been formulated, including Professors Blumenthal, Bowers, and Dagli as research colleagues, Dr. Pietro Binetti for his expertise on arrayed waveguide grating routers, as well as Abirami Sivanathan, John Parker and Erik Norberg, who have all participated in grading the course and proofreading of the manuscript. Finally, our thanks go to Brian Kerr, who has provided a large portion of the excellent new illustrations for the second edition, Peter Allen, who has designed the book cover and Michael Belt, who has helped greatly with the new index.

List of Fundamental Constants

$$\pi = 3.141\,592\,653\,59$$

$$c = 2.997\,924\,58 \times 10^{10}\ \text{cm/s}$$

$$\mu_0 = 4\pi \times 10^{-9}\ \text{H/cm}$$

$$\varepsilon_0 = 8.854\,18782 \times 10^{-14}\ \text{F/cm or C}^2/(\text{J} \cdot \text{cm}) (= 1/\mu_0 c^2)$$

$$q = 1.602\,189\,2 \times 10^{-19}\ \text{C or J/eV}$$

$$m_0 = 9.109\,534 \times 10^{-31}\ \text{kg}$$

$$h = 6.626\,176 \times 10^{-34}\ \text{J} \cdot \text{s}$$

$$\hbar = 1.054\,588\,7 \times 10^{-34}\ \text{J} \cdot \text{s}$$

$$k_\text{B} = 1.380\,662 \times 10^{-23}\ \text{J/K} = 8.617347 \times 10^{-5}\ \text{eV/K}$$

$$k_\text{B} T = 25.852\ 04\ \text{meV (at } T = 300\,\text{K)}$$

$$\lambda_\text{ph} \cdot E_\text{ph} = 1.239\,85\ \mu\text{m} \cdot \text{eV}$$

Source: *CRC Handbook of Chemistry and Physics*, 62nd Edition, 1981–1982.

CHAPTER ONE

Ingredients

1.1 INTRODUCTION

Diode lasers, like most other lasers, incorporate an optical gain medium in a resonant optical cavity. The design of both the gain medium and the resonant cavity are critical in modern lasers. A sample schematic of a laser cavity and its elements is shown in Fig. 1.1. In this case, an optional mode selection filter is also added to permit only one cavity mode to lase. The gain medium consists of a material that normally absorbs incident radiation over some wavelength range of interest. But, if it is *pumped* by inputting either electrical or optical energy, the electrons within the material can be excited to higher, nonequilibrium energy levels, so the incident radiation can be amplified rather than absorbed by stimulating the de-excitation of these electrons along with the generation of additional radiation. The resonant optical cavity supports a number of cavity standing waves, or modes. As illustrated in Figs. 1.1b and c, these occur where the cavity length is a multiple of a half wavelength. If the resulting gain is sufficient to overcome the losses of some resonant optical mode of the cavity, this mode is said to have reached *threshold*, and relatively coherent light will be emitted. The resonant cavity provides the necessary positive feedback for the radiation being amplified, so that a lasing oscillation can be established and sustained above threshold pumping levels. A typical diode laser light-pump current characteristic is shown in Fig. 1.1d. The *threshold* can be identified on an output light power vs. pump characteristic by a sharp knee, as illustrated in Fig. 1.1d.

For various applications, a single lasing mode inside a laser cavity is preferred. Different methods in cavity design can be used to favor the lasing of one mode

Diode Lasers and Photonic Integrated Circuits, Second Edition.
Larry A. Coldren, Scott W. Corzine, and Milan L. Mašanović.
© 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

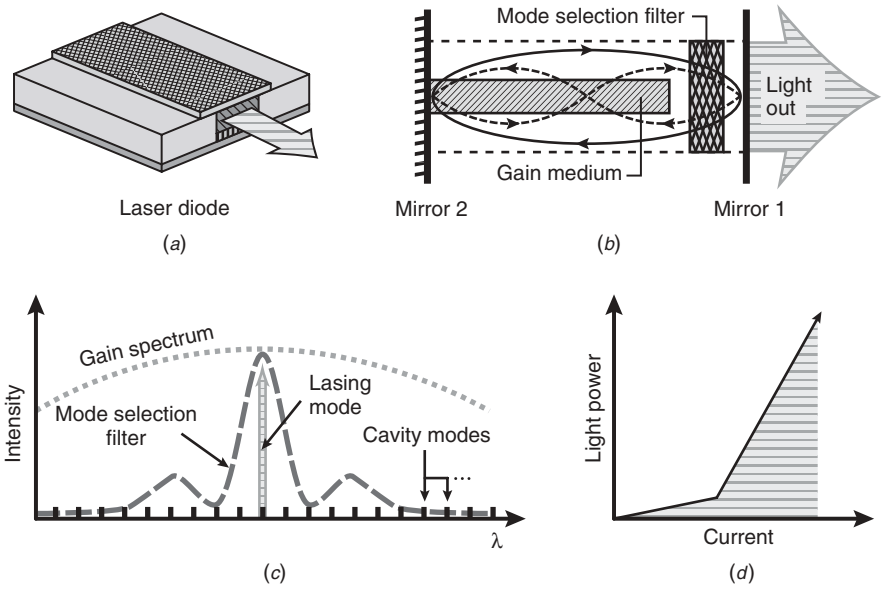


FIGURE 1.1: (a) A schematic of a simple laser diode. (b) Necessary ingredients for a single-frequency laser cavity—two mirrors, a gain medium, and a mode selection filter, which is required only for single wavelength λ operation. (c) Spectral characteristics of laser elements that get superimposed for single mode operation: cavity modes are given by $m \cdot \frac{\lambda}{2} = \bar{n}L$, where the mode number m is an integer, and $\bar{n}$ is the effective index of refraction (d) Typical light-current diode laser characteristic.

relative to others. The response of the optical mirrors can be tailored to support a single mode. Often, additional optical filtering elements will be incorporated inside the resonant cavity, to insure single mode operation of the laser. Fig. 1.1c shows the spectral response of the various elements of this cavity. This resonant optical cavity is defined by two broadband mirrors, with flat spectral responses, which define a number of cavity modes. An additional mode filtering element, with a defined bandpass optical transfer function, is included. The optical gain medium has a certain spectral response, which, in combination with the spectral response of the filter, will define which cavity mode will be singled out. As in any other oscillator, the output power level saturates at a level equal to the input minus any internal losses.

Since their discovery, lasers have been demonstrated in solid, liquid, gas and plasma materials. Today, the most important classes of lasers, besides the widespread *diode/(or semiconductor)* lasers are, *gas, dye, solid-state*, and *fiber* lasers, the latter really being fiber-optic versions of solid-state lasers. The helium–neon gas laser, the widely tunable flowing-dye laser, the Nd-doped YAG (yttrium–aluminum–garnet) solid-state and the Er or Yb-doped silica fiber lasers are four popular examples. Figure 1.2 shows commercial examples of Nd-YAG and dye lasers, an Er-doped fiber amplifier (EDFA), as well as a packaged diode

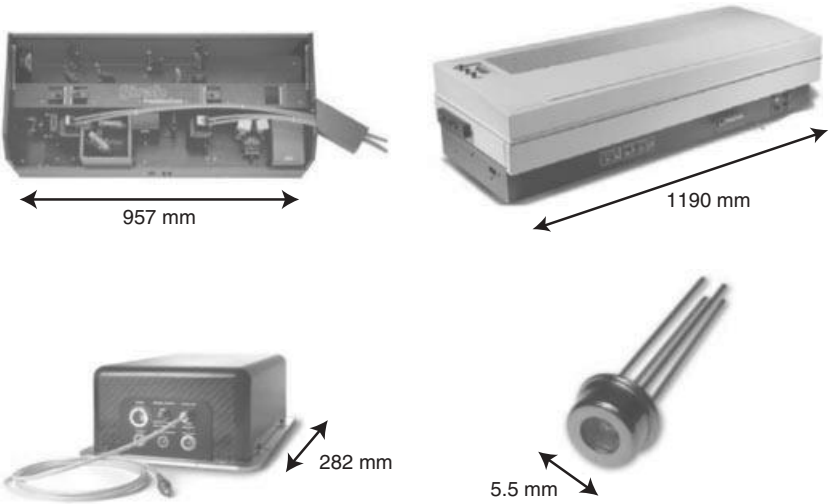


FIGURE 1.2: Examples of solid-state (upper left), dye (upper right), and fiber laser (bottom left) systems compared to a packaged diode laser chip (bottom right). To function, the diode laser also requires some drive electronics, and this increases its net size somewhat.

laser for comparison. The EDFA is used in fiber-optic systems to compensate losses, and with the addition of mirrors placed in the fiber, it can also become a laser. Diode lasers are distinguished from these other types primarily by their ability to be pumped directly by an electrical current. Generally, this results in a much more efficient operation. Overall power conversion efficiencies of ~50% are not uncommon for a diode laser, whereas efficiencies on the order of 1% are common for gas and solid-state lasers, which traditionally have been pumped by plasma excitation or an incoherent optical flashlamp source, respectively. However, in recent years diode laser pumps have been used for both bulk solid-state lasers as well as fiber lasers, and wall plug efficiencies better than 25% have been achieved. Efficiencies of some gas lasers can be somewhat higher than that of the He-Ne laser, such as in the case of the CO_2 gas laser, which has a typical efficiency of over 10%. Another type of gas laser, the so-called Excimer laser, uses transitions between highly excited atomic states to produce high-power ultraviolet emission, and these are used in the medical industry for a variety of surgical procedures as well as in the semiconductor industry for patterning very fine features. Dye lasers are almost always used in a research environment because of their relatively high maintenance requirements, and they are generally pumped by other high power bench-top lasers. Their appeal is that their output wavelength can be tuned by as much as 10% for a given dye and mirror set, and by changing these, wavelengths from the near IR through much of the visible can be provided from a single commercial product.

Because of their longer cavities gas, dye, solid-state and fiber lasers also tend to have more coherent outputs than simple semiconductor lasers. However, more

sophisticated single-frequency diode lasers can have comparable linewidths in the low megahertz range.

Another major attribute of diode lasers, their high reliability or useful lifetime, has led to their widespread use in important applications such as fiber-optic communications systems. Whereas the useful life of gas or flash-lamp-pumped solid-state lasers is typically measured in thousands of hours, that of carefully qualified diode lasers is measured in hundreds of years. Recent use of diode lasers to pump solid-state and fiber lasers may, however, provide the best advantages of both technologies, providing high reliability, improved efficiency, and low linewidth.

Net size is another striking difference between semiconductor and other lasers. Whereas gas, solid-state and fiber lasers are typically tens of centimeters in length, diode laser chips are generally about the size of a grain of salt, although the mounting and packaging hardware increases the useful component size to the order of a cubic centimeter or so. The diode lasers are mass-produced using wafer scale semiconductor processes, which makes them really inexpensive compared to all other types of lasers. The semiconductor origins of diode lasers allows for semiconductor integration techniques to be applied, and for multiple building blocks to be defined along the common waveguide, yielding functionally complex devices and opening a new field of photonic integrated circuits. Diode lasers with integrated optical amplifiers, modulators and similar other functions have been realized. In addition, monolithic widely tunable diode lasers and transmitters have been conceived and developed, in a footprint much smaller than that of external-cavity widely tunable lasers. Arrays of diode lasers and transmitters have been commercialized as well, for both optical pumping, and telecom purposes.

Diode lasers are used in many consumer products today. Examples are illustrated in Fig. 1.3. The most widely used diode lasers on the planet by far are those used in CD/DVD players, DVD ROM drives and optical mice. These diode lasers produce light beams in the red part of the visible spectrum at a wavelength of 0.65 μm. Recent improvements of the diode lasers emitting in the blue visible part of the spectrum have allowed for higher density DVD discs to be developed, resulting in the Blu-ray Disc technology, operating at 0.405 μm. Visible red diode lasers have replaced helium-neon lasers in supermarket checkout scanners and other bar code scanners. Laser printers are commonly used to produce high-resolution printouts, enabled by the high resolution determined by the wavelength of the diode laser used (780 nm or lower). Laser pointers, patient positioning devices in medicine utilize diode lasers emitting in the visible spectrum, both red and green.

In fiber-optic communication systems, diode lasers are primarily used as light sources in the optical links. For short reach links, a directly modulated diode laser is used as a transmitter. For longer reach links, diode lasers are used in conjunction with external modulators, which can be external to the diode laser chip, or integrated on the same chip. Complex diode laser-based photonic integrated circuits are currently deployed in a number of optical networks. In addition, Erbium doped fiber amplifiers, a key technology that is utilized for signal amplification in the existing fiber-optic networks, has in part been enabled by the development of high power, high reliability diode pump lasers.



FIGURE 1.3: Examples of the most common products that utilize diode lasers. (left) red laser in a DVD player shown in laptop computer; (center-top) blue laser in a Blu-ray Disc player; (center-bottom) red laser in a laser printer; (right-top) red laser in a bar-code scanner; (right-bottom) red (and sometimes green or blue) laser in a pointer.

There are many other areas where diode lasers are utilized. In medical applications, diode lasers are used in optical coherence tomography, an optical signal acquisition and processing method allowing extremely high quality, micrometer-resolution, three-dimensional images from within optical scattering media (e.g., biological tissue) to be obtained. In remote sensing, diode lasers are used in light detection and ranging (LIDAR) technology, new generation of optical radars that offer much improved resolution compared to the classical radio-frequency radars, due to light's much shorter wavelength. Other, similar applications are in range finding and military targeting.

In this chapter, we shall attempt to introduce some of the basic ingredients needed to understand semiconductor diode lasers. First, energy levels and bands in semiconductors are described starting from background given in Appendix 1. The interaction of light with these energy levels is next introduced. Then, the enhancement of this interaction by carrier and photon confinement using heterostructures is discussed. Materials useful for diode lasers and how epitaxial layers of such materials can be grown is briefly reviewed. The lateral patterning of these layers to provide lateral current, carrier, and photon confinement for practical lasers is introduced. Finally, examples of different diode lasers are given at the end of the chapter.

1.2 ENERGY LEVELS AND BANDS IN SOLIDS

To begin to understand how gain is accomplished in lasers, we must have some knowledge of the energy levels that electrons can occupy in the gain medium. The allowed energy levels are obtained by solving Schrdinger's equation using the appropriate electronic potentials. Appendix 1 gives a brief review of this important solid-state physics, as well as the derivation of some other functions that we shall need later. Figure 1.4 schematically illustrates the energy levels that might be associated with optically induced transitions in both an isolated atom and a semiconductor solid. Electron potential is plotted vertically.

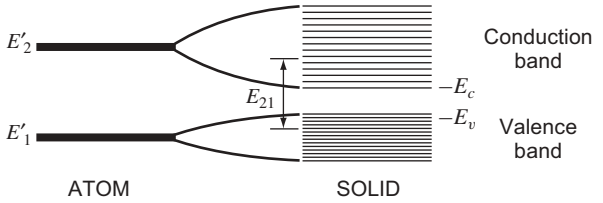


FIGURE 1.4: Illustration of how two discrete energy levels of an atom develop into bands of many levels in a crystal.

In gas and solid-state lasers, the energy levels of the active atomic species are only perturbed slightly by the surrounding gas or solid host atoms, and they remain effectively as sharp as the original levels in the isolated atom. For example, lasers operating at the 1.06 μm wavelength transition in Nd-doped YAG, use the $^4F_{3/2}$ level of the Nd atom for the upper laser state #2 and the $^4I_{11/2}$ level for the lower laser state #1. Because only these atomic levels are involved, emitted or absorbed photons need to have almost exactly the correct energy, $E_{21} = hc/1.06\ \mu m$.

On the other hand, in a covalently bonded solid like the semiconductor materials we use to make diode lasers, the uppermost energy levels of individual constituent atoms each broaden into bands of levels as the bonds are formed to make the solid. This phenomenon is illustrated in Fig. 1.4. The reason for the splitting can be realized most easily by first considering a single covalent bond. When two atoms are in close proximity, the outer valence electron of one atom can arrange itself into a low-energy *bonding* (symmetric) charge distribution concentrated between the two nuclei, or into a high-energy *antibonding* (antisymmetric) distribution devoid of charge between the two nuclei. In other words, the isolated energy level of the electron is now split into two levels due to the two ways the electron can arrange itself around the two atoms.[1] In a covalent bond, the electrons of the two atoms both occupy the lower energy bonding level (provided they have opposite spin), whereas the higher energy antibonding level remains empty.

If another atom is brought in line with the first two, a new charge distribution becomes possible that is neither completely bonding nor antibonding. Hence, a third energy level is formed between the two extremes. When N atoms are covalently bonded into a linear chain, N energy levels distributed between the lowest-energy bonding state and the highest-energy antibonding state appear, forming a band of energies. In our linear chain of atoms, spin degeneracy allows all N electrons to fall into the lower half of the energy band, leaving the upper half of the band empty. However in a three-dimensional crystal, the number of energy levels is more generally equated with the number of *unit cells*, not the number of atoms. In

[1]The energy level splitting is often incorrectly attributed to the Pauli exclusion principle, which forbids electrons from occupying the same energy state (and thus forces the split, as the argument goes). In actuality, the splitting is a fundamental phenomenon associated with solutions to the wave equation involving two coupled systems and applies equally to probability, electromagnetic, or any other kind of waves. It has nothing to do with the Pauli exclusion principle.

typical semiconductor crystals, there are two atoms per primitive unit cell. Thus, the first atom fills the lower half of the energy band (as with the linear chain), whereas the second atom fills the upper half, such that the energy band is entirely full.

The semiconductor *valence* band is formed by the multiple splitting of the highest occupied atomic energy level of the constituent atoms. In semiconductors, the valence band is by definition entirely filled with no external excitation at $T = 0$ K. Likewise, the next higher-lying atomic level splits apart into the *conduction* band, which is entirely empty in semiconductors without any excitation. When thermal or other energy is added to the system, electrons in the valence band may be excited into the conduction band analogous to how electrons in isolated atoms can be excited to the next higher energy level of the atom. In the solid then, this excitation creates holes (missing electrons) in the valence band as well as electrons in the conduction band, and both can contribute to conduction.

Although Fig. 1.4 suggests that many conduction—valence band state pairs may interact with photons of energy E_{21}, Appendix 1 shows that the imposition of momentum conservation in addition to energy conservation limits the interaction to a fairly limited set of state pairs for a given transition energy. This situation is illustrated on the electron energy versus k-vector ($E - k$) plot shown schematically in Fig. 1.5. (Note that momentum $\equiv \hbar\mathbf{k}$.) Because the momentum of the interacting photon is negligibly small, transitions between the conduction and valence band must have the same k-vector, and only vertical transitions are allowed on this diagram. This fact will be very important in the calculation of gain.

1.3 SPONTANEOUS AND STIMULATED TRANSITIONS: THE CREATION OF LIGHT

With a qualitative knowledge of the energy levels that exist in semiconductors, we can proceed to consider the electronic transitions that can exist and the interactions

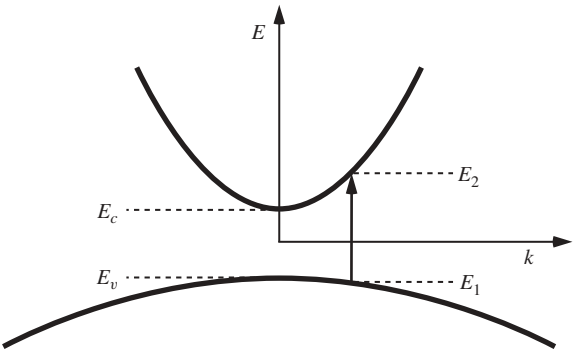


FIGURE 1.5: Electron energy vs. wave vector magnitude in a semiconductor showing a transition of an electron from a bound state in the valence band (E_1) to a free carrier state in the conduction band (E_2). The transition leaves a hole in the valence band. The lowest and highest energies in the conduction and valence bands are E_c and E_v, respectively.

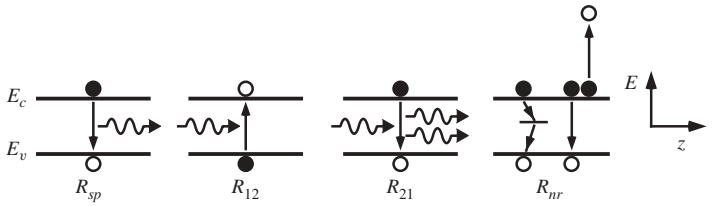


FIGURE 1.6: Electronic transitions between the conduction and valence bands. The first three represent radiative transitions in which the energy to free or bind an electron is supplied by or given to a photon. The fourth illustrates two nonradiative processes.

with lightwaves that are possible. Figure 1.6 illustrates the different kinds of electronic transitions that are important, emphasizing those that involve the absorption or emission of photons (lightwave quanta).

Although we are explicitly considering semiconductors, only a single level in both the conduction and valence bands is illustrated. As discussed earlier and in Appendix 1, momentum conservation selects only a limited number of such pairs of levels from these bands for a given transition energy. In fact, if it were not for a finite bandwidth of interaction owing to the finite state lifetime, a single pair of states would be entirely correct. In any event, the procedure to calculate gain and other effects will be to find the contribution from a single state pair and then integrate to include contributions from other pairs; thus, the consideration of only a single conduction–valence band state pair forms an entirely rigorous basis.

As illustrated, *four basic electronic recombination/generation (photon emission/absorption) mechanisms* must be considered separately:

1. Spontaneous recombination (photon emission)
2. Stimulated generation (photon absorption)
3. Stimulated recombination (coherent photon emission)
4. Nonradiative recombination

The open circles represent unfilled states (holes), and the solid circles represent filled states (electrons). Because electron and hole densities are highest near the bottom or top of the conduction or valence bands, respectively, most transitions of interest involve these carriers. Thus, photon energies tend to be only slightly larger than the bandgap (i.e., $E_{21} = h\nu \sim E_g$). The effects involving electrons in the conduction band are all enhanced by the addition of some pumping means to increase the electron density to above the equilibrium value there. Of course, the photon absorption can still take place even if some pumping has populated the conduction band somewhat.

The first case (R_{sp}) represents the case of an electron in the conduction band recombining spontaneously with a hole (missing electron) in the valence band

to generate a photon. Obviously, if a large number of such events should occur, relatively incoherent emission would result because the emission time and direction would be random, and the photons would not tend to contribute to a coherent radiation field. This is the primary mechanism within a light-emitting diode (LED), in which photon feedback is not provided. Because spontaneous recombination requires the presence of an electron–hole pair, the recombination rate tends to be proportional to the product of the density of electrons and holes, NP. In undoped active regions, charge neutrality requires that the hole and electron densities be equal. Thus, the spontaneous recombination rate becomes proportional to N^2.

The second illustration (R_{12}) outlines photon absorption, which stimulates the generation of an electron in the conduction band while leaving a hole in the valence band.

The third process (R_{21}) is exactly the same as the second, only the sign of the interaction is reversed. Here an incident photon perturbs the system, stimulating the recombination of an electron and hole and simultaneously generating a new photon. *Of course, this is the all-important positive gain mechanism that is necessary for lasers to operate.* Actually, it should be realized that the net combination of stimulated emission and absorption of photons, $(R_{21} - R_{12})$, will represent the net gain experienced by an incident radiation field. In an undoped active region, net stimulated recombination (photon emission) depends on the existence of photons in addition to a certain value of electron density to overcome the photon absorption. Thus, as we shall later show more explicitly, the net rate of stimulated recombination is proportional to the photon density, N_p, multiplied by $(N - N_{tr})$, where N_{tr} is a *transparency* value of electron density (i.e., where $R_{21} = R_{12}$).

Finally, the fourth schematic in Fig. 1.6 represents the several nonradiative ways in which a conduction band electron can recombine with a valence band hole without generating any useful photons. Instead, the energy is dissipated as heat in the semiconductor crystal lattice. Thus, this schematic represents the ways in which conduction band electrons can escape from usefully contributing to the gain, and as such these effects are to be avoided if possible. In practice, there are two general nonradiative mechanisms for carriers that are important. The first involves nonradiative recombination centers, such as point defects, surfaces, and interfaces, in the active region of the laser. To be effective, these do not require the simultaneous existence of electrons and holes or other particles. Thus, the recombination rate via this path tends to be directly proportional to the carrier density, N. The second mechanism is Auger recombination, in which the electron–hole recombination energy, E_{21}, is given to another electron or hole in the form of kinetic energy. Thus, again for undoped active regions in which the electron and hole densities are equal, Auger recombination tends to be proportional to N^3 because we must simultaneously have the recombining electron–hole pair and the third particle that receives the ionization energy. Appendix 2 gives techniques for calculating the carrier density from the density of electronic states and the probability that they are occupied, generally characterized by a Fermi function.

1.4 TRANSVERSE CONFINEMENT OF CARRIERS AND PHOTONS IN DIODE LASERS: THE DOUBLE HETEROSTRUCTURE

As discussed in the previous section, optical gain in a semiconductor can only be achieved through the process of stimulated recombination (R_{21}). Therefore, a constant flow of carriers must be provided to replenish the carriers that are being recombined and converted into photons in the process of providing gain. For this flow of carriers in the gain material to happen, the semiconductor must be pumped or excited with some external energy source. A major attribute of diode lasers is their ability to be pumped directly with an electrical current. Of course, the active material can also be excited by the carriers generated from absorbed light, and this process is important in characterizing semiconductor material before electrical contacts are made. However, we shall focus mainly on the more technologically important direct current injection technique in most of our analysis.

The carrier-confining effect of the double–heterostructure (DH) is one of the most important features of modern diode lasers. After many early efforts that used homojunctions or single heterostructures, the advent of the DH structure made the diode laser truly practical for the first time and led to two Nobel prize awards in physics in the year 2000. Figure 1.7 gives a schematic of a broad-area *pin* DH laser diode, along with transverse sketches of the energy gap, index of refraction, and resulting optical mode profile across the DH region. As illustrated, a thin slab of undoped active material is sandwiched between p- and n-type cladding layers, which have a higher conduction–valence band energy gap. Typical thicknesses of the active layer for this simple three-layer structure are $\sim 0.1-0.2$ μm. Because the bandgap of the cladding layers is larger, light generated in the active region will not have sufficient photon energy to be absorbed in them (i.e., $E_{21} = h\nu < E_{gcl}$).

For this DH structure, a transverse (x-direction) potential well is formed for electrons and holes that are being injected from the n- and p-type regions, respectively, under forward bias. As illustrated in part (b), they are captured and confined together, thereby increasing their probability of recombining with each other. In fact, unlike in most semiconductor diodes or transistors that are to be used in purely electronic circuits, it is desirable to have all the injected carriers recombine in the active region to form photons in a laser or LED. Thus, simple pn-junction theory, which assumes that all carriers entering the depletion region are swept through with negligible recombination, is totally inappropriate for diode lasers and LEDs. In fact, a better assumption for lasers and LEDs is that all carriers recombine in the i-region. Appendix 2 also discusses a possible "leakage current," which results from some of the carriers being thermionically emitted over the heterobarriers before they can recombine.

To form the necessary resonant cavity for optical feedback, simple cleaved facets can be used because the large index of refraction discontinuity at the semiconductor–air interface provides a reflection coefficient of ~30%. The lower bandgap active region also usually has a higher index of refraction, n, than the cladding, as outlined in Fig. 1.7c, so that a transverse dielectric optical waveguide is formed with its axis along the z-direction. The resulting transverse optical energy density

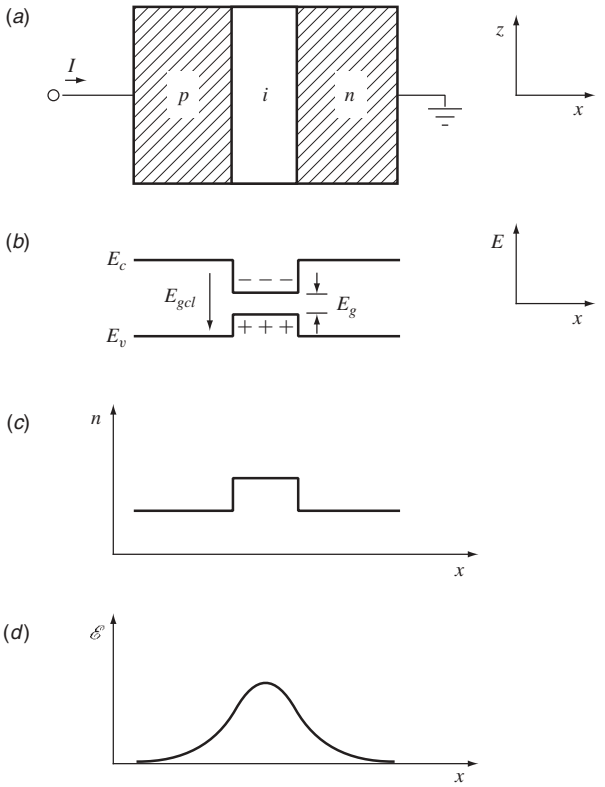


FIGURE 1.7: Aspects of the double-heterostructure diode laser: (a) a schematic of the material structure; (b) an energy diagram of the conduction and valence bands vs. transverse distance; (c) the refractive index profile; and (d) the electric field profile for a mode traveling in the z-direction.

profile (proportional to the photon density or the electric field magnitude squared, $|\mathscr{E}|^2$) is illustrated in Fig. 1.7d.

Thus, with the in-plane waveguide and perpendicular mirrors at the ends, a complete resonant cavity is formed. Output is provided at the facets, which only partially reflect. Later we shall consider more complex reflectors that can provide stronger feedback and wavelength filtering function, as illustrated in Fig 1.1. One should also realize that if the end facet reflections are suppressed by antireflection coatings, the device would then function as an LED. When we analyze lasers in the next chapter, the case of no feedback will also be considered.

The thickness of the active region in a DH plays an important role in its optical properties. If this thickness starts to get below $\sim$100 nm, quantum effects on optical properties must be taken into account, and this regime of operation will be referred to as the *quantum confined* regime. For dimensions larger than 100 nm, we can assume that we are working with a continuum of states, and this regime is called the *bulk* regime.

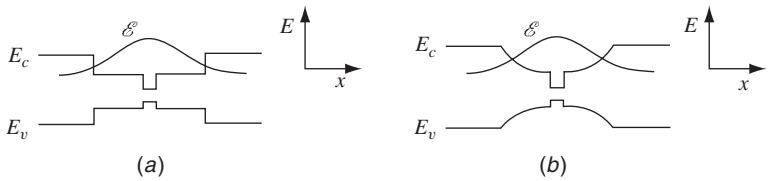


FIGURE 1.8: Transverse band structures for two different separate-confinement heterostructures (SCHs): (a) standard SCH; (b) graded-index SCH (GRINSCH). The electric field (photons) are confined by the outer step or graded heterostructure; the central quantum well confines the electrons.

It turns out that many modern diode lasers involve a little more complexity in their transverse carrier and photon confinement structure as compared to Fig. 1.7, but the fundamental concepts remain valid. For example, with in-plane lasers, where the light propagates parallel to the substrate surface, a common departure from Fig. 1.7 is to use a thinner *quantum-well* carrier-confining active region ($d \sim 10$ nm) and a surrounding intermediate bandgap *separate confinement* region to confine the photons. Figure 1.8 illustrates transverse bandgap profiles for such separate-confinement heterostructure, single quantum-well (SCH-SQW) lasers. The transverse optical energy density is also overlaid to show that the photons are confined primarily by the outer heterointerfaces and the carriers by the inner quantum well. Quantum-well active regions reduce diode laser threshold current and improve their efficiency and their thermal properties. The important concepts related to the quantum-well active regions are introduced in Appendix 1 and further discussed in detail in Chapter 4.

Example 1.1 An InP/InGaAsP double-heterostructure laser cross-section consists of a 320 nm tall InGaAsP separate confinement heterostructure waveguide region with the bandgap corresponding to 1.3 μm (1.3 Q), clad by InP on both sides.

Problem: (1) Determine the effective index of the fundamental transverse mode of this waveguide. (2) Determine the rate of decay of the normalized electric field U.

Solution: To solve this problem, we utilize the tools from the Appendix 3. Because this optical waveguide structure is symmetric, we can utilize the expression (A3.14) to solve for the effective index. Then, we can compute the wave vector component along x, k_x, and the decay constant γ using Equation (A3.7). From the problem statement, the refractive index of the InGaAsP region can be found in Table 1.1, $n_{\text{II}} = 3.4$.
For the cladding, the refractive index value at 1.55 μm is $n_{\text{I}} = n_{\text{III}} = 3.17$. From Eq. (A3.12), the normalized frequency, V, is given by

$$V = k_0 d (n_{\text{II}}^2 - n_{\text{III}}^2)^{\frac{1}{2}} = \frac{2\pi}{1.55\ \mu\text{m}}\ 0.32\ \mu\text{m}\ (3.4^2 - 3.17^2)^{\frac{1}{2}} = 1.594$$

Using Eq. (A3.14), we can compute the value of the normalized propagation parameter b,

$$b = 1 - \frac{\ln\left(1 + \frac{V^2}{2}\right) \cdot}{\frac{V^2}{2}} = 1 - \frac{\ln\left(1 + \frac{1.594^2}{2}\right)}{\frac{1.594^2}{2}} = 0.354$$

and the effective index value,

$$\bar{n} = (n_{\text{II}}^2 b + n_{\text{I}}^2(1-b))^{\frac{1}{2}} = (3.4^2 \cdot 0.354 + 3.17^2 \cdot (1 - 0.354))^{\frac{1}{2}} = 3.253$$

To determine the minimum thickness of the top p doped cladding, we can compute the decay constant γ using Eq. (A3.7), remembering that the propagation constant $\beta = k_0\bar{n}$,

$$\gamma = k_0(\bar{n}^2 - n_{\text{I}}^2)^{\frac{1}{2}} = \frac{2\pi}{1.55\ \mu\text{m}}\ (3.253^2 - 3.17^2)^{\frac{1}{2}} = 2.9591\ \mu\text{m}^{-1}$$

Thus, for a 1 μm thick cladding, the optical energy decays to exp $(-2.9591) = 0.0027$ at the top surface. We can observe that the rate of decay is strongly dependent on the refractive index difference between the waveguide and the cladding—for a larger difference, the field intensity outside the waveguide region decays faster. In a real laser, the active region would probably be defined by a set of quantum wells in the center of the InGaAsP double heterostructure region. This would complicate solving for the effective index, and this case will be treated in Chapter 6, when we talk about the perturbation theory.

1.5 SEMICONDUCTOR MATERIALS FOR DIODE LASERS

The successful fabrication of a diode laser relies very heavily on the properties of the materials involved. There is a very limited set of semiconductors that possess all the necessary properties to make a good laser. For the desired double heterostructures at least two compatible materials must be found, one for the cladding layers and another for the active region. In more complex geometries, such as the SCH mentioned earlier, three or four different bandgaps may be required within the same structure. The most fundamental requirement for these different materials is that they have the same crystal structure and nearly the same lattice constant, so that single-crystal, defect-free films of one can be *epitaxially* grown on the other. Defects generally become nonradiative recombination centers, which can steal many of the injected carriers that otherwise would provide gain and luminescence. In a later section we shall discuss some techniques for performing this epitaxial growth, but first we need to understand how to select materials that meet these fundamental boundary conditions.

Table 1.1 lists the lattice constants, bandgaps, effective masses, and indices of refraction for some common materials. (Subscripts on effective masses, C, HH,

TABLE 1.1: Material Parameters for III–V Compounds

III–V Compounds	a (Å)	E_g (eV) 0 K	E_g (eV) 300 K	m_C	m_{HH}	m_{LH}	m_{SH}	ε@dc	$n@E_g$	$n@(\lambda\ \mu m)$
GaAs	5.6533	1.519	1.424	0.067	0.38	0.09	0.15	13.2	3.62	3.52(0.98)
AlGaAs (0.2)	5.6548	1.769	1.673	0.084	0.39	0.1	0.16	12.5	3.64	3.46(0.87); 3.39(0.98)
AlAs*	5.660	2.228	2.153	0.19	0.48	0.2	0.29	10.06	3.2	2.98(0.87); 2.95(0.98)
InGaAs (0.2) comp. strained on GaAs	5.6533	1.296	1.215	0.059	0.37 0.078†	0.062 0.16†	0.11	13.6	3.6	
InP	5.8688	1.424	1.351	0.077	0.61	0.12	0.20	12.4	3.41	3.21(1.3); 3.17(1.55)
InGaAsP (1.3 μm)	5.8688	1.029	0.954	0.056	0.42	0.055	0.1	13.3	3.52	3.40(1.55)
InGaAsP (1.55 μm)	5.8688	0.874	0.800	0.045	0.37	0.044	0.08	13.75	3.55	
InGaAs (1.65 μm)	5.8688	0.818	0.748	0.046	0.36	0.041	0.07	13.9	3.56	
InAs	6.0583	0.418	0.359	0.027	0.34	0.027	0.05	15.15	3.52	
GaP*	5.4505	2.35	2.272	0.254	0.67	0.17	0.46	11.1	3.5	
AlP*	5.4635	2.505	2.41	0.21	0.51	0.21	0.3	9.8	2.97	
AlSb*	6.1355	1.696	1.63	0.33	0.47	0.16	0.24	12.0	3.5	
GaSb	6.0959	0.811	0.70	0.041	0.27	0.05	0.08	15.69	3.92	
InSb	6.4794	0.237	0.175	0.014	0.34	0.016	0.03	16.8	3.5	
GaN‡ (hexagonal)	$a = 3.189$ $c = 5.185$	3.50	3.39	0.20 (∥) 0.20 (⊥)				8.9	2.67	2.52(0.41)
AlN‡ (hexagonal)	$a = 3.112$ $c = 4.982$	6.28	6.20	0.32 (∥) 0.30 (⊥)				8.5	2.5	2.18(0.14)
InN‡ (hexagonal)	$a = 3.545$ $c = 5.703$	0.78	-	0.07 (∥) 0.07 (⊥)				15.3	3.15	2.93(0.82); 3.12(0.66)
$In_{0.12}Ga_{0.88}N$‡ (410 nm)										2.77(0.41)

*Indirect gap.
†In-plane masses.
‡For more details on hole mass parameters, see [5].

LH, and SH, denote values in the conduction, heavy-hole, light-hole, and split-off bands, respectively.)

Figure 1.9 plots the bandgap versus lattice constant for several families of III–V semiconductors. These III–V compounds (which consist of elements from columns III and V of the periodic table) have emerged as the materials of choice for lasers that emit in the 0.7–1.6 μm wavelength range. This range includes the important fiber-optic communication bands at 0.85, 1.31, and 1.55 μm, the pumping bands for fiber amplifiers at 1.48 and 0.98 μm, the window for pumping Nd-doped YAG at 0.81 μm, and the wavelength used for classic DVD disc players at 0.65 μm. Most of these materials have a *direct gap* in E–k space, which means that the minimum and maximum of the conduction and valence bands, respectively, fall at the same k-value, as illustrated in Fig. 1.5. This facilitates radiative transitions because momentum conservation is naturally satisfied by the annihilation of the equal and opposite momenta of the electron and hole. (The momentum of the photon is negligible.)

The lines on this diagram represent ternary compounds, which are alloys of the binaries labeled at their end-points. The dashed lines represent regions of indirect gap. The areas enclosed by lines between three or four binaries represent quaternaries, which obviously have enough degrees of freedom that the energy gap can

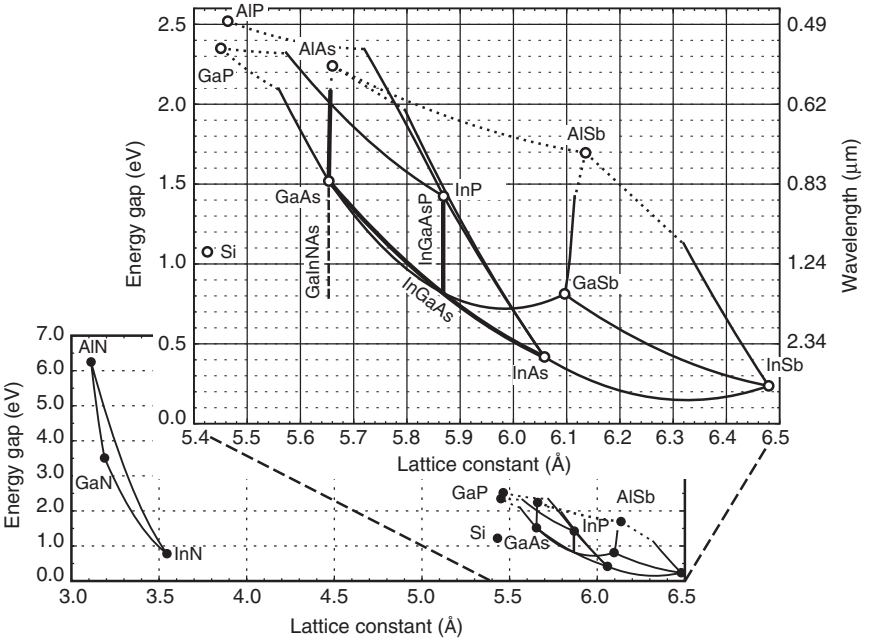


FIGURE 1.9: Energy gap vs. lattice constant of ternary compounds defined by curves that connect the illustrated binaries. The values in this plot were obtained from [2, 5] and are valid at $T = 0$ K. Details on how these values can be converted to room temperature values are given in the references cited.

be adjusted somewhat without changing the lattice constant. Thus, in general, a quaternary compound is required in a DH laser to allow the adjustment of the energy gap while maintaining lattice matching. Fortunately, there are some unique situations that allow the use of more simple ternaries. As can be seen, the AlGaAs ternary line is almost vertical. That is, the substitution of Al for Ga in GaAs does not change the lattice constant very much. Thus, if GaAs is used as the substrate, any alloy of $Al_xGa_{1-x}As$ can be grown, and it will naturally lattice match, so that no misfit dislocations or other defects should form. As suggested by the formula, the x-value determines the percentage of Al in the group III half of the III–V compound. The AlGaAs/GaAs system provides lasers in the 0.7–0.9 μm wavelength range. For DH structures in this system, about two-thirds of the band offset occurs in the conduction band. For shorter wavelengths into the red (e.g., 650 nm as used in DVDs), the AlInGaP/GaAs system is generally employed. In this case lattice matching requires a precise control of the ratios of Al:In:Ga in the quaternary regions.

The most popular system for long-distance fiber optics is the InGaAsP/InP system. Here the quaternary is specified by an x and y value (i.e., $In_{1-x}Ga_xAs_yP_{1-y}$). This is grown on InP to form layers of various energy gap corresponding to wavelengths in the 1.0–1.6 μm range, where silica fiber traditionally had minima in loss (1.55 μm) and dispersion (1.3 μm). Using InP as the substrate, a range of lattice-matched quaternaries extending from InP to the InGaAs ternary line can be accommodated, as indicated by the vertical line in Fig. 1.9. Fixing the quaternary lattice constant defines a relation between x and y. It has been found that choosing x equal to $\sim 0.47y$ results in approximate lattice matching to InP. The ternary endpoint is $In_{0.53}Ga_{0.47}As$. For DH structures in this system, only about 40% of the band offset occurs in the conduction band.

InGaAsP lasers and photonic integrated circuits (PICs) generally need to be operated at a constant temperature to maintain their performance. This is primarily due to the fact that with the increasing temperature, the electron leakage current from the quantum well increases. The main material parameter controlling the current leakage is the conduction band offset. Due to their much lighter effective mass, electrons require much tighter confinement with increasing temperature than holes. To improve the diode laser performance at high temperatures, particularly of interest for the fiber-optic metropolitan area network deployment, material engineering was successfully employed to increase the conduction band offset, through introduction of InGaAlAs material system. This material system enables quantum wells with the conduction band offset of $\Delta E_c = 0.7\Delta E_g$. Changing the barrier from $\Delta E_c = 0.4\Delta E_g$ to $\Delta E_c = 0.7\Delta E_g$ will lead to the reduction of the leakage current density from $J = 50$ A/cm^2 to $J = 1.5$ A/cm^2. Uncooled operation of lasers and integrated laser electroabsorption modulator PICs have been demonstrated at both 1310 nm and 1550 nm, and this remains an active area of research and deployment.

Lattice constants of ternary and quaternary compounds can be precisely calculated from Vegard's law, which gives a value equal to the weighted average of all the four possible constituent binaries. For example, in $In_{1-x}Ga_xAs_yP_{1-y}$, we

obtain

$$a(x,y) = xya_{\mathrm{GaAs}} + x(1-y)a_{\mathrm{GaP}} + (1-x)ya_{\mathrm{InAs}} + (1-x)(1-y)a_{\mathrm{InP}}. \quad (1.1)$$

Similarly, the lattice constants for other alloys can be calculated. For example, in the AlInGaP/GaAs case, we would be considering a linear superposition of the lattice constants of the binaries AlP, InP, and GaP to match that of GaAs. And for InGaAlAs/InP, it would be the InAs, GaAs, AlAs binary lattice constants superimposed to match that of InP. The following example illustrates the application of the Vegard's law to the crystal lattice.

Example 1.2 InP-based 1550 nm vertical cavity surface emitting lasers have been made with AlAsSb/AlGaAsSb multilayer mirrors.

Problem: Calculate the fraction of As in the AlAsSb mirror layers for lattice matching to InP.

Solution: To solve this problem, we will use the Vegard's law. The composition of any AlAsSb alloy can be specified by value x, where x is the percentage of As in the alloy $AlSb_{1-x}As_x$. From Table 1.1, the lattice constants of InP, AlAs and AlSb are $a_{\mathrm{InP}} = 5.8688$ Å, $a_{\mathrm{AlAs}} = 5.660$ Å, and $a_{\mathrm{AlSb}} = 6.1355$ Å, respectively. Using Vegard's law,

$$a_{\mathrm{InP}} = xa_{\mathrm{AlAs}} + (1-x)a_{\mathrm{AlSb}}.$$

Therefore, the fraction of As in the lattice matched mirror layer is

$$x = \frac{a_{\mathrm{InP}} - a_{\mathrm{AlSb}}}{a_{\mathrm{AlAs}} - a_{\mathrm{AlSb}}} = 0.56.$$

Other parameters, for example, bandgap, can also be interpolated in a similar fashion to Eq. (1.1), however a second-order *bowing* parameter must oftentimes be added to improve the fit. The ternary lines in Fig. 1.9 were obtained using the following modified version of Vegard's law,

$$E_{gABC}(x) = xE_{gAB} + (1-x)E_{gAC} - x(1-x)C_{ABC}, \quad (1.2)$$

where C_{ABC} is an empirical bowing parameter.

When interpolating, one must be careful if different bands come into play in the process. For example, in AlGaAs, the values for GaAs and $Al_{0.2}Ga_{0.8}As$ can be linearly extrapolated for direct gap AlGaAs up to $x \sim 0.45$, but at this point the indirect band minimum becomes the lowest, so the gap for higher x-values is then interpolated from this point using energy gap values that correspond to the first indirect band for both GaAs and AlAs. This extrapolation will be needed in some homework exercises. Here, we give a simple example of bandgap calculation in a ternary compound.

Example 1.3

Problem: A wurtzite structure GaInN quantum well contains 53% of Ga and 47% of N. The bowing parameter for the direct bandgap of this ternary is $C = 1.4$ eV. What is the bandgap of this quantum well?

Solution: To calculate the bandgap, we need to use Vegard's law, including the correction introduced by the bowing parameter. From Table 1.1, the direct bandgaps for GaN and InN are $E_{g1} = 3.510$ eV and $E_{g2} = 0.78$ eV, respectively. Using Eq. (1.2),

$$E_{gGaInN} = 0.53 \cdot E_{g1} + 0.47 \cdot E_{g2} - 0.53(0.47)1.4 \text{ eV} = 1.878 \text{ eV}.$$

In addition to the usual III–V compounds discussed earlier, Table 1.1 also lists some of the nitride compounds. These compounds had originally gained attention because of their successful use in demonstrating LEDs emitting at high energies in the visible spectrum. Whereas the InAlGaAsP based compounds are limited to emission in the red and near infrared regions, the nitrides have demonstrated blue and UV emission. GaN-based optoelectronic devices have achieved considerable progress since their first demonstration in 1996. Development of advanced epitaxial growth techniques, defect-reduced substrates, and sophisticated device design has resulted in high performance light-emitting diodes (LEDs) and laser diodes (LDs) with wide commercial presence. GaN-based LEDs have been particularly successful in solid-state lighting and display applications (traffic signal lights, automobile lights, flashlights), while laser diodes have emerged as critical components in the next generation of the high density DVD disk players (Blu-ray). Nitride components continue to generate considerable interest in such applications as projection displays, high resolution printing, and optical sensing, and thus remain an active area of research, development and commercialization.

Lattice matching is generally necessary to avoid defects that can destroy the proper operation of diode lasers. However, it is well known that a small lattice mismatch ($\Delta a/a \sim 1\%$) can be tolerated up to a certain thickness ($\sim$20 nm) without any defects. Thus, for a thin active region, one can move slightly left or right of the lattice matching condition illustrated in Fig. 1.9 or by Eq. (1.1). In this case, the lattice of the deposited film distorts to fit the substrate lattice in the plane, but it also must distort in the perpendicular direction to retain approximately the same unit cell volume it would have without distortion. Figure 1.10 shows a cross section of how unit cells might distort to accommodate a small lattice mismatch. After a critical thickness is exceeded, misfit defects are generated to relieve the integrated strain. However, up to this point, it turns out that such *strained layers* may have more desirable optoelectronic properties than their unstrained counterparts. In particular, due to their small dimensions of less than 10 nm, strained quantum wells can be created without introducing any undesired defects into the crystal. These structures will be analyzed in some detail in Chapter 4. In fact, it is fair to say that such strained-layer quantum wells, contained within separate-confinement

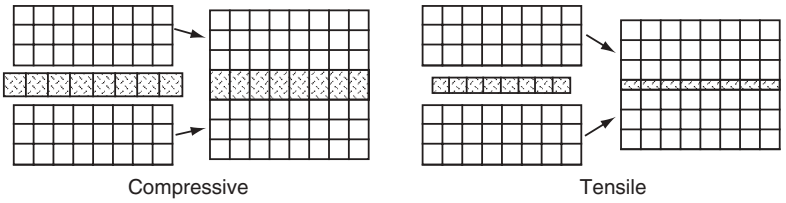


FIGURE 1.10: Schematic of sandwiching quantum wells with either a larger or smaller lattice constant to provide either compressive or tensile strain, respectively.

heterostructures as illustrated in Fig. 1.7, have become the most important form of active regions in modern diode lasers.

One of the key factors in determining the output wavelength of a quantum-well laser besides the material composition is the well width, or the so-called quantum-size effect. As illustrated in the following example, as a quantum-well is made more narrow, the lowest state energy is squeezed up from the well bottom, and the transition wavelength is made shorter. The barrier height plays an important role in this process as well because this limits the amount the energy can move away from the well bottom. These issues are covered in Appendix 1.

Example 1.4 An 80 Å wide quantum well composed of InGaAsP, lattice matched to InP, with the bandgap corresponding to 1.55 μm (1.55 Q), is surrounded by an InGaAsP barrier, lattice matched to InP, with the bandgap wavelength of 1.3 μm (1.3 Q).

Problem: Determine the energy and wavelength for photons generated in recombination between the ground states of the quantum well at room temperature.

Solution: In order to solve this problem, we utilize the tools from Appendix 1. We will determine the energy levels of this quantum well using Eq. (A1.14). First, we need to compute the energy of the ground state for this quantum well with infinitely high walls. Then, we need to determine the quantum numbers taking into account that this quantum well has finite walls, using Eq. (A1.17).

As mentioned in the problem statement, both the quantum well and the barrier are lattice matched to InP. From Table 1.1, $E_{\text{barrier}} = 0.954$ eV, and $E_{\text{well}} = 0.800$ eV. In this material system, only 40% of the band offset occurs in the conduction band. Therefore, the quantum well barrier height in the conduction band is given by $V_{0C} = 0.4\,(E_{\text{barrier}} - E_{\text{well}}) = 61.6$ meV and $V_{0V} = 0.6\,(E_{\text{barrier}} - E_{\text{well}}) = 92.4$ meV in the valence band.

From Eq. (A1.14), the ground state energy for a quantum well with infinite walls, E_{1c}^{∞}, is given by

$$E_{1c}^{\infty} = 3.76\,\frac{m_0}{m}\left(\frac{100\ \text{Å}}{l}\right)^2\ \text{meV} = 3.76\,\frac{1}{0.045}\left(\frac{100}{80}\right)^2\ \text{meV} = 130.55\ \text{meV}$$

where $m = m_c$ was taken from Table 1.1. Similarly, for the valence band, $E_{1v}^{\infty} = 15.88$ meV, with $m = m_{HH}$. Now, we can calculate n_{max} for both quantum wells using Eq. (A1.17), $n_{maxc} = \sqrt{\frac{V_{0c}}{E_{1c}^{\infty}}} = 0.69$, and $n_{maxv} = \sqrt{\frac{V_{0v}}{E_{1v}^{\infty}}} = 2.41$. The normalized variable n_{max}, when rounded up to the nearest integer, yields the largest number of bound states possible. Either by reading the chart in Fig. A1.4 or using Eq. (A1.18), we can calculate the lowest quantum numbers for both cases:

$$n_{1c} \approx \frac{2}{\pi} \arctan\left[n_{maxc}\ (1 + 0.6^{n_{maxc}+1})\right] = 0.49$$

$$n_{1v} \approx \frac{2}{\pi} \arctan\left[n_{maxv}\ (1 + 0.6^{n_{maxv}+1})\right] = 0.78$$

Thus, $E_{1c} = n_{1c}^2 E_{1c}^{\infty} = 31.35$ meV and $E_{1v} = n_{1v}^2 E_{1v}^{\infty} = 9.66$ meV. Finally, the photon energy is given by

$$E_{photon} = E_{well} + E_{1c} + E_{1v} = 841.01 \text{ meV}.$$

This energy corresponds to the wavelength

$$\lambda = \frac{1.23985 \text{ eV } \mu\text{m}}{0.84101 \text{ eV}} = 1.47 \ \mu\text{m}.$$

1.6 EPITAXIAL GROWTH TECHNOLOGY

To make the multilayer structures required for diode lasers, it is necessary to grow single-crystal lattice-matched layers with precisely controlled thicknesses over some suitable substrate. We have already discussed the issue of lattice matching and some of the materials involved. Here we briefly introduce several techniques to perform epitaxial growth of the desired thin layers.

We shall focus on the three most important techniques in use today: liquid-phase epitaxy (LPE), molecular beam epitaxy (MBE), and organometallic vapor-phase epitaxy (OMVPE). OMVPE is often also referred to as metal-organic chemical vapor deposition (MOCVD), although purists do not like the omission of the word *epitaxy*. As the names imply, the three techniques refer to growth either in liquid, vacuum, or a flowing gas, respectively. The growth under liquid or moderate pressure gas tends to be done near equilibrium conditions, so that the reaction can proceed in either the forward or reverse direction to add or remove material, whereas the MBE growth tends to be more of a physical deposition process. Thus, the near-equilibrium processes, LPE and MOCVD, tend to better provide for the removal of surface damage at the onset of growth, and they are known for providing higher quality interfaces generally important in devices. MBE on the other hand provides the ultimate in film uniformity and thickness control.

Figure 1.11 gives a cross section of a modern LPE system. In this system the substrate is placed in a recess in a graphite slider bar, which forms the bottom of

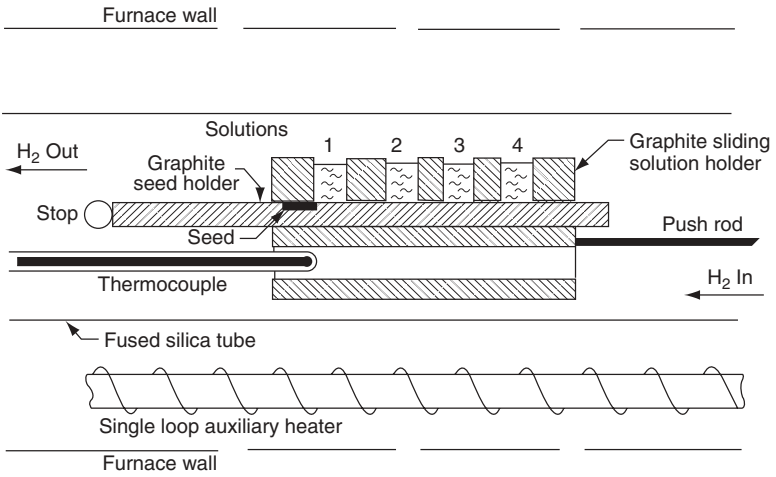


FIGURE 1.11: Schematic of a liquid-phase epitaxy (LPE) system [3]. (*Reprinted, by permission, from Applied Physics Letters.*)

a sequence of bins in a second graphite housing. The bins are filled with solutions from which a desired layer will grow as the substrate is slid beneath that bin. This entire assembly is positioned in a furnace, which is accurately controlled in temperature. There are several different techniques of controlling the temperature and the dwell time under each melt, but generally the solutions are successively brought to saturation by reducing the temperature very slowly as the substrate wafer is slid beneath alternate wells. In modern systems, the process of slider positioning and adjusting furnace temperature is done by computer control for reproducibility and efficiency. However, LPE is rapidly being replaced by MOCVD for the manufacture of most diode lasers.

The melts typically consist mostly of one of the group III metals with the other constituents dissolved in it. For InGaAsP growth, In metal constitutes most of the melt. For an $In_{0.53}Ga_{0.47}$ as film only about 2.5% of Ga and 6% As is added to the melt for growth at 650°C. For InP growth only about 0.8% of P is added. Needless to say, the dopants are added in much lesser amounts. Thus, LPE growth requires some very accurate scales for weighing out the constituents and an operator with a lot of patience.

Figure 1.12 shows a schematic of an MOCVD system. The substrate is positioned on a susceptor, which is heated typically by rf induction, or in some cases, by resistive heaters. The susceptor is placed into a reactor, which is designed to produce a laminar gas flow over the substrate surface. The gas carrier different growth species, and by precise control of the species concentration and flow rates, substrate temperature and pressure, highly precise growth is accomplished. Both low-pressure and atmospheric-pressure systems are being used. Whereas the atmospheric-pressure system uses the reactant gases more effectively, the layer uniformity and the time required to flush the reactor before beginning a new layer

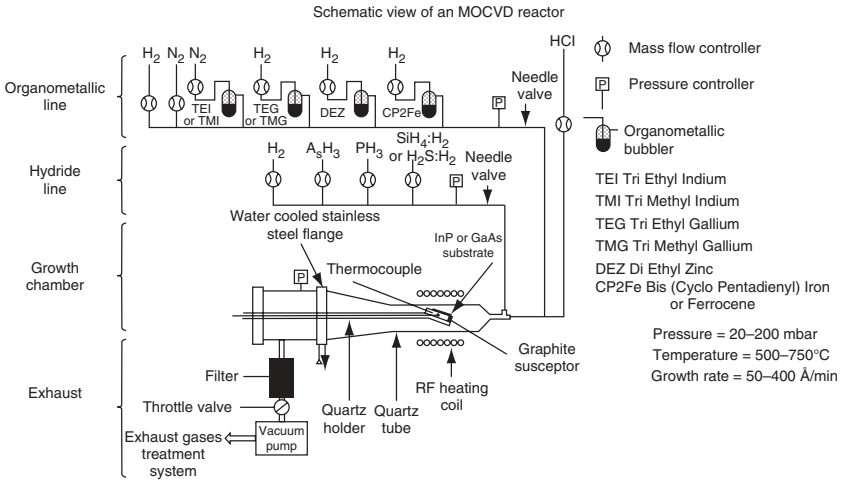


FIGURE 1.12: Schematic of a metal-organic chemical vapor deposition (MOCVD) system [1]. (*From GaInAsP Alloy Semiconductors, T. P. Pearsall, Ed., Copyright © John Wiley & Sons, Inc. Reprinted by permission of John Wiley & Sons, Inc.*)

is long. Low pressure is more popular where very abrupt interfaces between layers are desired, and this is very important for quantum-well structures. Typical growth temperature for InP based compounds is around 625°C.

As can be seen, a large part of the system is devoted to gas valving and manifolding to obtain the proper mixtures for insertion into the growth reactor chamber. The sources typically used for MOCVD consist of a combination of hydrides such as arsine (AsH_3), phosphine (PH_3), and organometallic liquids, which are used to saturate an H_2 carrier gas. Example organometallics are triethylindium and triethylgallium. Dopants can be derived from either other hydrides or liquid sources. For example, H_2S or triethyl-zinc can be used for *n*- or *p*-type dopants, respectively.

One of the key concerns with MOCVD is safety. The problems are primarily with the hydrides, which are very toxic. Thus, much of the cost of an MOCVD facility is associated with elaborate gas handling, monitoring, and emergency disposal techniques. Recently, there has been some industry effort to work with less toxic liquid sources for As and P (e.g., *t*-butyl-arsine and *t*-butyl-phosphine). Although still toxic, liquid sources give off only modest amounts of poisonous gases due to their vapor pressure, and such quantities could be accommodated by conventional fume hoods. Still, even though the hydrides have to be contained in high-pressure gas cylinders, which conceivably can fail and release large quantities of concentrated toxic gas in a short time, they remain the industry standard to this day.

Figure 1.14 shows a cross section of a solid-source MBE growth chamber. As illustrated, MBE is carried out under ultrahigh vacuum (UHV) conditions. Constituent beams of atoms are evaporated from effusion cells, and these condense on a heated substrate. Liquid nitrogen cryoshields line the inside of the system to

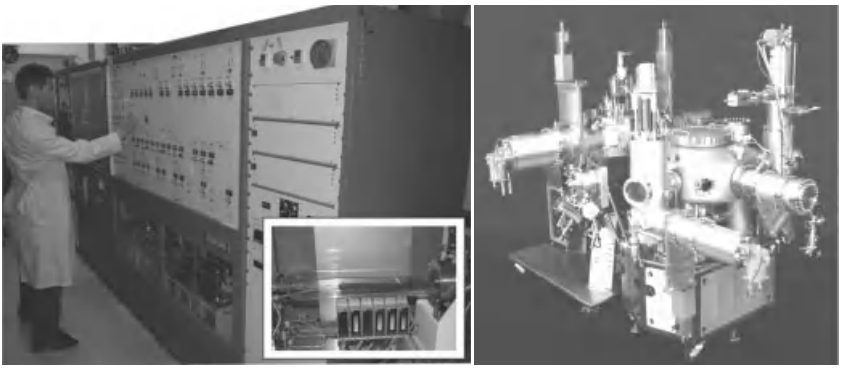


FIGURE 1.13: Photographs of (left) a single-wafer, horizontal flow MOCVD growth system at UC Santa Barbara, manufactured by Thomas Swan. The glass reactor, housing a graphite susceptor, is shown in the inset. Quartz heating lamps are visible underneath the reactor. (right) GEN20 MBE system, manufactured by Veeco, an ultra-flexible tool with a design configurable for III-V and emerging materials.

condense any stray gases. For stoichiometry control MBE makes use of the fact that the group V elements are much more volatile than the group III elements. Thus, if the substrate is sufficiently hot, the group V atoms will reevaporate unless there is a group III atom with which to form the compound. At the same time, the substrate must be sufficiently cool so that the group III atoms will stick. Therefore, the growth rate is determined by the group III flux, and the group V flux is typically set to several times that level. Typical growth temperatures for the AlGaAs system are in the 600–650°C range. However, because Al tends to oxidize easily, and such oxides create nonradiative recombination centers, AlGaAs lasers may be grown over 700°C.

One of the key features of MBE is that UHV surface analysis techniques can be applied to the substrate either before or during growth in the same chamber. One of the most useful tools is reflection high-energy electron diffraction (RHEED), which is an integral part of any viable MBE system. It is particularly useful in monitoring the growth rate *in situ* because the intensity of the RHEED pattern varies in intensity as successive monolayers are deposited.

Some hybrid forms of the last two techniques have also been developed (i.e., gas source MBE, metal-organic MBE (MOMBE), and chemical beam epitaxy (CBE) [7, 8]). These techniques are particularly interesting for the phosphorous-containing compounds, such as the important InGaAsP. Basically, gas-source MBE involves using the hydrides for the group V sources. Generally, these gases must be cracked by passing them through a hot cell prior to arriving at the substrate. MOMBE uses the metal-organics for the group III sources; again, some cracking is necessary. CBE is basically just ultralow-pressure MOCVD because both the group III and V sources are the same as in an MOCVD system. However, in the CBE case one still retains access to the UHV surface analysis techniques that have made MBE viable.

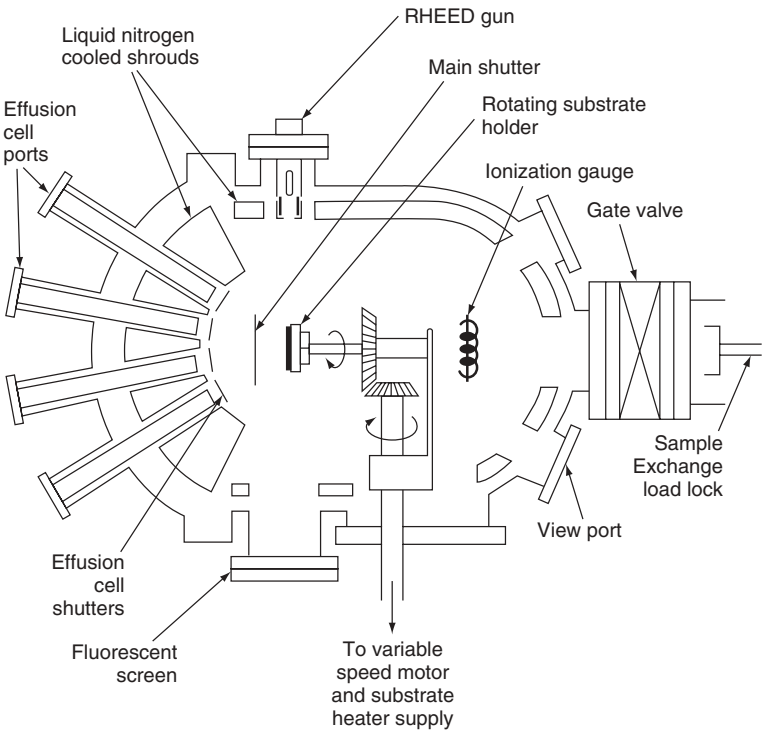


FIGURE 1.14: Schematic of a molecular beam epitaxy (MBE) system [4]. (*Reprinted, by permission, from Journal of Applied Physics.*)

Photographs of a research MOCVD and a research MBE system are shown in Fig. 1.13.

1.7 LATERAL CONFINEMENT OF CURRENT, CARRIERS, AND PHOTONS FOR PRACTICAL LASERS

Practical diode lasers come in two basic varieties: those with *in-plane* cavities and those with *vertical* cavities. The in-plane (or edge-emitting) types have been in existence since the late 1960s, whereas the vertical cavity (or vertical-cavity surface-emitting-laser–VCSEL for short) types have been viable only since about 1990. As mentioned earlier, feedback for the in-plane type can be accomplished with a simple cleaved-facet mirror; however, for vertical-cavity lasers a multilayer reflective stack must be grown below and above the active region for the necessary cavity mirrors. Figure 1.15 illustrates both types. Of course, many practical edge-emitters also use more complex mirrors as discussed in Chapter 3.

As suggested by this figure, practical lasers must emit light in a narrow beam, which implies that a *lateral* patterning of the active region is necessary. In the case

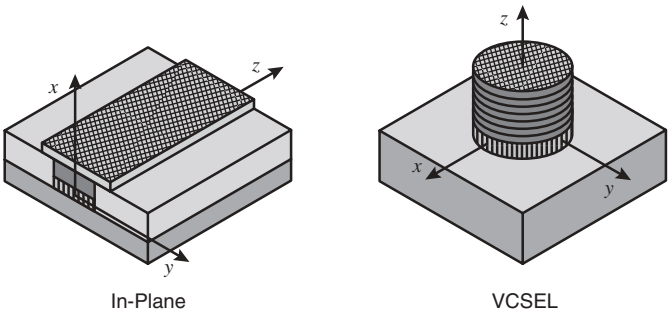


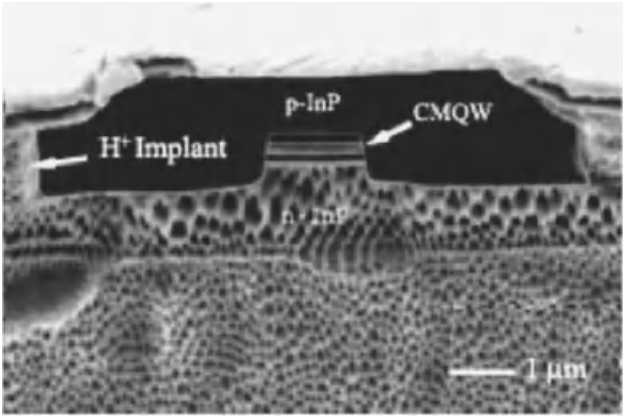
FIGURE 1.15: Schematic of in-plane and vertical-cavity surface-emitting lasers showing selected coordinate systems.

of the in-plane types, a stripe laser is formed that typically has lateral dimensions of a few microns. Similarly, the vertical-cavity types typically consist of a circular dot geometry with lateral dimensions of a few microns. This emitting aperture of a few microns facilitates coupling to optical fibers or other simple optics because it is sufficiently narrow to support only a single lateral mode of the resulting optical waveguide, but sufficiently wide to provide an emerging optical beam with a relatively small diffraction angle. However, in the case of the VCSEL there continues to be an important multimode-fiber optic market in which it is usually desired to use multimoded VCSELs that can fill the modal spectrum of the fibers.

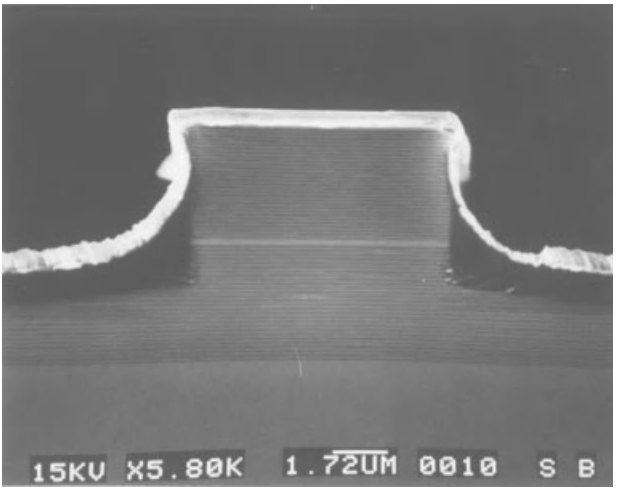
Figure 1.16 shows cross-sectional scanning electron micrographs (SEMs) of both the in-plane and vertical-cavity lasers. The reference coordinate systems, also introduced in Fig. 1.15, are somewhat different for these two generic types of diode lasers. The difference arises from our insistence on designating the optical propagation axis as the z-axis. We shall also refer to this direction as the *axial* direction. For both types the *lateral* y-direction is in the plane of the substrate. For in-plane lasers the vertical to the substrate is the *transverse* x-direction, as illustrated in Fig. 1.7, whereas for vertical-cavity lasers the x-direction lies in the plane and is deemed a second lateral direction.

Once we have decided that a lateral patterning of the active region is desirable for lateral carrier and photon confinement, we also must consider lateral *current* confinement. That is, once the active region is limited in lateral extent, we must ensure that all the current is injected into it rather than finding some unproductive shunt path. In fact, current confinement is the first and simplest step in moving from a broad-area laser to an in-plane stripe or vertical-cavity dot laser. For example, current can be channeled to some degree simply by limiting the contact area. However, in the best lasers current confinement is combined with techniques to laterally confine the carriers and photons in a single structure.

Lateral confinement of *current*, *carriers*, and *photons* has been accomplished in literally dozens of ways, and there are even more acronyms to describe all of these. For brevity's sake, we will focus on only a few generic types illustrated in Fig. 1.17. Here we illustrate schematic transverse-lateral (x-y) edge-emitter laser



(*a*)



(*b*)

FIGURE 1.16: Cross-sectional SEMs of (a) in-plane and (b) vertical-cavity semiconductor lasers.

cross sections along with calculated optical mode intensity profiles. The first two types only provide current confinement, the third adds a weak photon confinement, and the last four provide all three types of confinement. The first four can be created by patterned etching, deposition, and implantation techniques typically found in semiconductor fabrication facilities, but the last three require at least one semiconductor regrowth step in addition, which is more specialized and perhaps costly. The examples are explicit for in-plane lasers, but many are also applicable to vertical-cavity lasers as illustrated by the analogous schematics shown in Fig. 1.18.

Figure 1.17a illustrates a simple oxide stripe laser. This stripe laser is the simplest to make because the area of current injection is limited simply by limiting

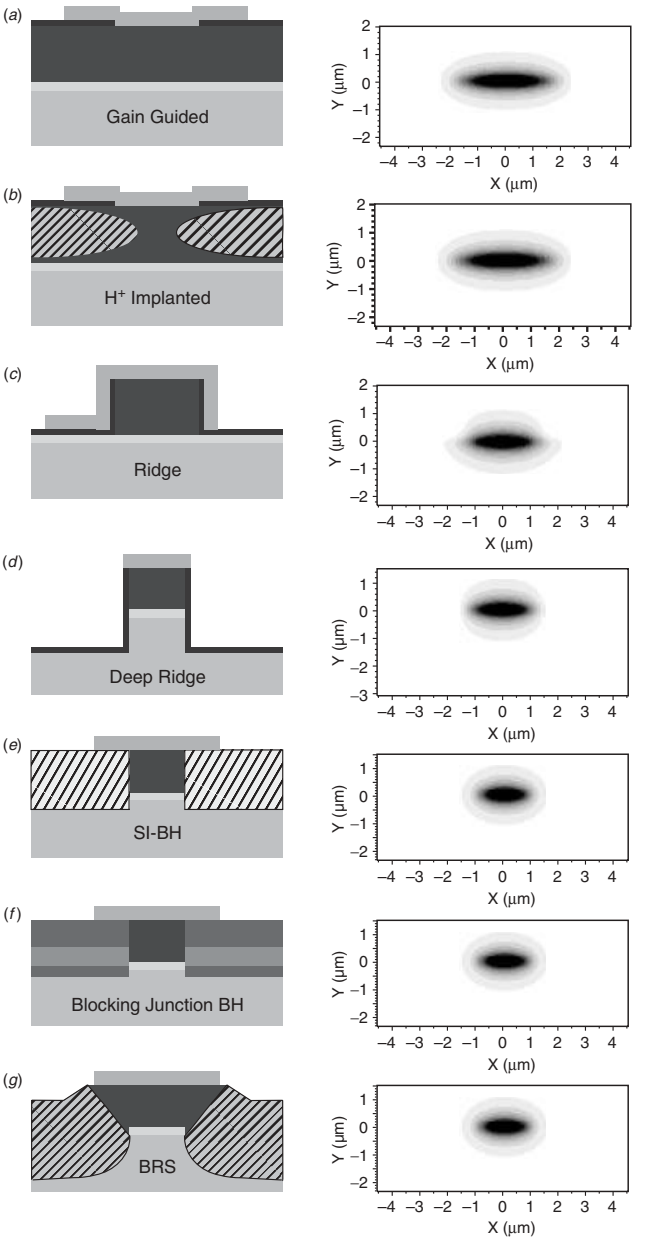


FIGURE 1.17: Lateral confinement structures for in-plane lasers: (a) oxide-stripe provides weak current confinement; (b) proton-implant provides improved current confinement; (c) ridge structure provides good current plus photon confinement; (d) deep-ridge gives strong current, carrier and photon confinement; (e) semi-insulating regrowth buried-heterostructure (SI-BH) provides good current, photon, and carrier confinement; (f) blocking-junction BH can provide current, photon, and carrier confinement; and (g) the buried-ridge-stripe BRS can also provide current, photon, and carrier confinement.

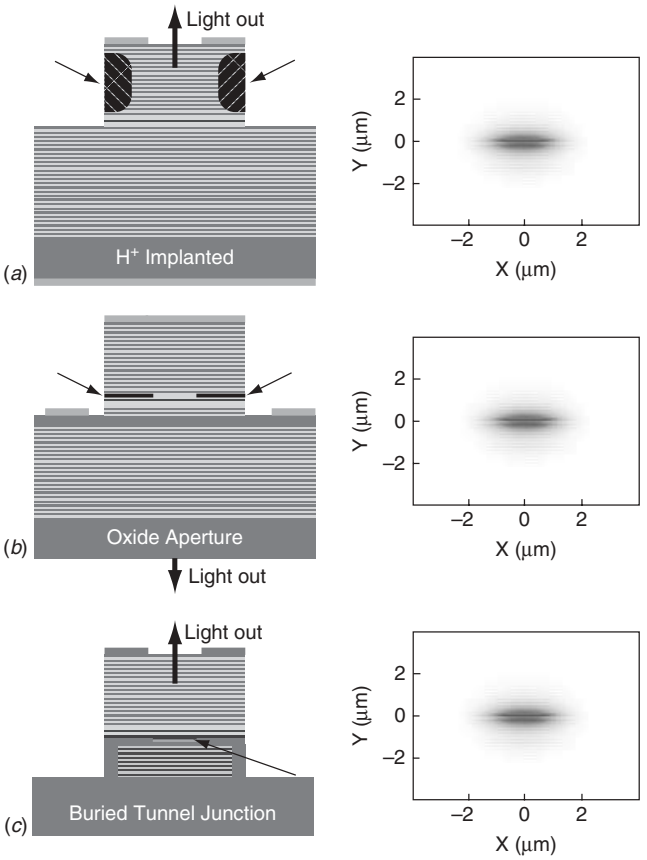


FIGURE 1.18: Schematics of VCSELs. (a) Proton-implanted, multimode with only current confinement; (b) dielectrically apertured with current and photon confinement; (c) mesa-confined with possible current and photon confinement.

the contact area. This laser has some current confinement, but no carrier or photon confinement. The proton-implanted configuration of Fig. 1.17b can provide current confinement by creating semi-insulating regions beneath the contact surface. It uses the fact that implanted hydrogen ions (protons) create damage and trap out the mobile charge, rendering the implanted material nearly insulating. Generally, the implant and resulting semi-insulating region is limited to a region between the contact and the active region, so that the contact can be wider while nonradiative recombination is avoided in the active region.

The configurations of Fig. 1.17a and b are described as *gain-guided* stripe lasers because the current is apertured, but there is no lateral index step or heterobarrier to provide a potential well for photons or carriers. Thus, carriers injected into the active region can diffuse laterally, decreasing the laser's efficiency, and there is also no lateral index change to guide photons along the axis of the cavity, so

optical losses tend to be high. Although these two configurations were of some commercial importance in the early days of diode lasers before the advent of viable etching or regrowth techniques, currently their use is limited for commercial edge-emitters. However, multimode VCSELs, which use a relatively large diameter >10 microns, are still produced using the proton-implant technology because this can be done at very low cost. The oxide stripe laser is still used in the lab to characterize material because the processing is so simple, and proton isolation is still used to limit current leakage in combination with other confinement techniques in practical edge-emitters, but the implant areas are kept away from the active region for reduced loss and improved reliability.

Figure 1.17c illustrates a surface *ridge* laser that combines current confinement with a weak photon confinement. The efficiency of current injection can be high, but because the processing involves etching down to just above the active region, carriers are still not confined laterally. They are free to diffuse laterally and recombine without contributing to the gain. The etching depth is adjusted to provide just enough effective lateral index change to provide a single lateral mode optical waveguide. Quantum-well-SCH transverse waveguide structures together with selective chemical etches can provide the desired lateral index profiles and separation from the active region to avoid surface-related carrier recombination. The ridge structure continues to be a popular choice, both because of its fabrication simplicity and its reasonable performance characteristics, especially if operated far above threshold where the lateral carrier diffusion effects are less important because they saturate at threshold. Because of the weak optical confinement, the surface ridge only allows relatively large radius waveguide bends. For sharper bends a stronger lateral index step is needed. Figure 1.17d shows the case of a deeply etched ridge that provides strong lateral confinement of photons, carriers, and current. It provides for much sharper waveguide bends. However, in this case it is very important that the sidewalls be formed very smoothly to limit optical losses, and some procedure must be used to limit the nonradiative recombination of carriers at the surface of the active region, which forms an inherent array of defects. Also, the deeply etched ridge will need to be much narrow than the surface ridge to support only a single lateral optical mode. Thus, it may be more difficult to couple light into and out of it because of the resulting large diffraction angle, and its thermal impedance will also be higher because of its smaller connection area to the substrate.

Figures 1.17e, f, & g illustrate examples of *buried-heterostructure* (BH) lasers, in which the active region is surrounded by higher-bandgap, lower-index-of-refraction semiconductor materials so that carriers and photons are confined laterally as well as transversely. Thus, a lateral (y) cut through the active region gives very analogous plots for the bandgap and electric field as the transverse (x) cut in Fig. 1.7. The three different cases confine the injected current in three different ways to prevent it from leaking around the active region. The first (e) uses the growth of a semi-insulating layer of semiconductor on either side of an etched ridge to confine the current to flow only into the active region; the second (f) uses two or more reverse biased junctions in the regrown regions; and the third

(g) follows an overgrowth of doped material above and beside the active region by an implant to prevent current flow along the sides. In this latter case, the lower turn-on voltage of the active-region diode also tends to funnel the current through this region even without the lateral implants.

There are still other ways of forming BH lasers that are not illustrated in Fig. 1.17, using techniques such as *quantum-well intermixing* (QWI). For QWI, only one epitaxial growth is performed, and then the active layer to the left and right of the desired active stripe is *modified to increase its bandgap after growth* by the diffusion of impurities or vacancies across the quantum-well active region. The diffusing species cause an intermixing of the original cladding and active quantum-well lattice atoms by requiring them to hop from site to site as the diffusing species move through. Thus, the absorption edge energy of the quantum wells is increased in these regions providing lateral carrier confinement to the central undiffused region.

Figure 1.18 shows schematic VCSEL cross sections together with calculated optical mode profiles for three types of lateral confinement structures. These differ from Fig. 1.17 because for the VCSEL, axial-lateral (z-y) cross sections are given, even though cross sections of the epitaxial wafer with its surface normal being the vertical axis are shown in both cases. Circular symmetry in the x-y plane of the wafer is assumed. Nearly all VCSELs use quantum wells placed at a standing wave maximum of the electric field, as will be discussed in later chapters and specifically Appendix 5. The first case (a) illustrates a proton-implanted VCSEL that uses the implant to generate a current confinement aperture much like the edge-emitter of Fig. 1.17b. As already mentioned, these are usually formed with relatively large diameters >10 μm to deliberately have multimode outputs. The second example (b) illustrates the dielectrically apertured VCSEL. The aperture is usually formed by an oxidation of a AlAs or a high Al content AlGaAs layer very close to the active region (sometimes referred to as "oxide-confined"), although it has also been formed by an etched air-gap. The low-index, insulating aperture provides lateral current and optical confinement. As the aperture is close to the active region, the current confinement tends to be quite good. The optical confinement results from the lensing action of the dielectric aperture that refocuses light resonating between the two planar mirror stacks. Blunt apertures are generally used, and some optical scattering loss results, but if tapered to be more lenslike, these apertures can act as near perfect lenses and scattering losses or internal diffraction loss can be effectively eliminated [6]. The third VCSEL example (c) also provides good optical and current confinement, but in this case the inclusion of a small thickness semiconductor disk above the active region provides the optical lensing element. It also yields current confinement if the disk provides a lower resistance path to the active region than the surrounding areas. One popular approach is to form a low resistance tunnel junction layer structure that is etched off everywhere except for this small disk. This configuration does require a regrowth of the mirror material above the disk confinement element, whereas the first two VCSEL geometries can be formed after a single epitaxial growth step.

1.8 PRACTICAL LASER EXAMPLES

Figures 1.19 and 1.20 give cross-sectional schematics and illustrative characteristics of a number of experimental in-plane and vertical-cavity lasers. The figure captions contain some of the relevant descriptive details. In the chapters to follow, the operating principles of such lasers will be detailed.

Figure 1.19 illustrates results from five edge-emitting in-plane lasers. The first (a) is a simple gain-guided stripe laser that is illustrative of designs used for high-power pump lasers. In this case the active region is a strained InGaAs quantum well in an AlGaAs SCH waveguide to provide emission in the 910–990 nm range. Narrow stripe lasers that can be coupled into a single mode optical fiber and have wavelengths 980 nm of this general composition are used to pump Er-doped

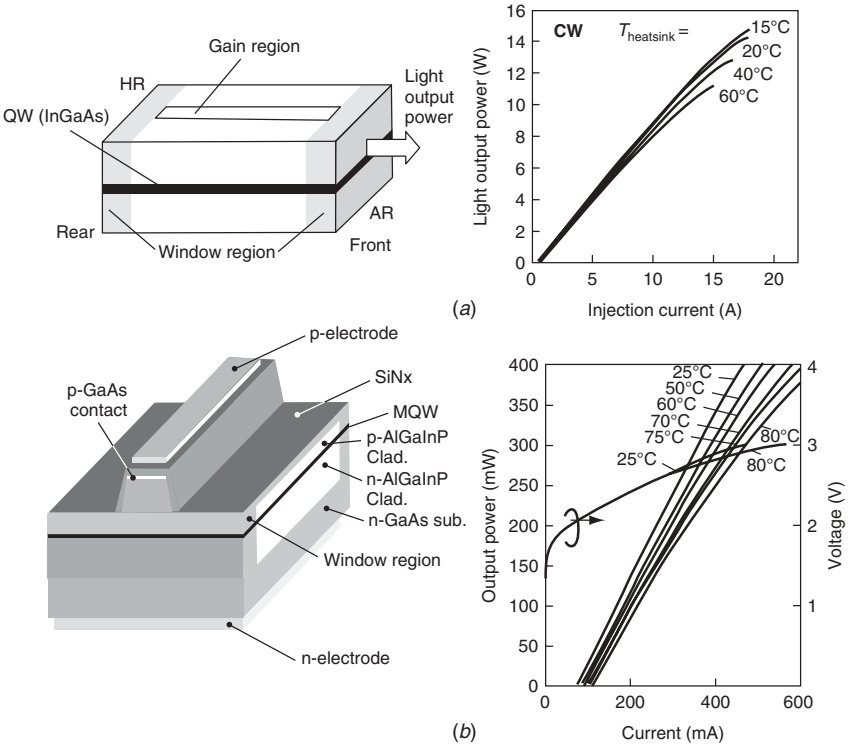


FIGURE 1.19: (a) schematic and *L-I* characteristic for 915 nm wavelength InGaAs/GaAs 100 μm wide broad-area laser using vacancy-induced QWI at the facets to avoid COD [7] (© IEEE 2007); (b) schematic and *L-I* characteristic for 650 nm wavelength AlGaInP/GaAs ridge laser using impurity-induced QWI at facets [8] (© IEEE 2007); (c) schematic and *L-I* characteristic of 404 nm wavelength InGaN/GaN ridge laser using nonpolar design [9] (© JJAP 2007); (d) schematic and *L-I* characteristic for 1300 nm wavelength InGaAlAs/InP MQW-DFB laser [10] (© IEEE 2006); (e) schematic and *L-I* characteristic of 1550 nm wavelength InGaAsP/InP SIBH-MQW-DFB laser [11] (© IEEE 2007).

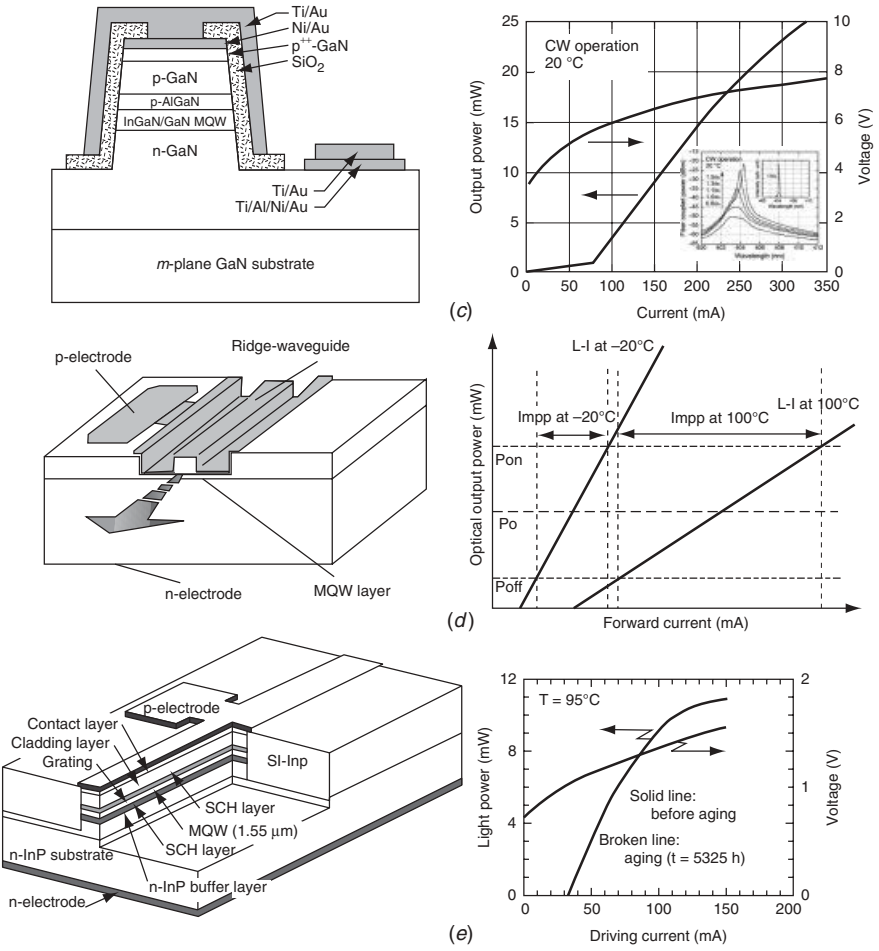


FIGURE 1.19: *(Continued)*

fiber amplifiers that are very important in optical fiber communication systems. Other important diode pumps include ones for pumping Nd-doped YAG, and these typically are composed of AlInGaP with emission in the 808 nm range. These are also formed on GaAs substrates.

An important limitation on the output power of such pump lasers is a catastrophic optical damage (COD) of the cleaved mirror facets due to nonradiative recombination of carriers at surface related defects and the resulting optical absorption and run-away heating. In the case of Fig. 1.19a, the COD is addressed by creating a nonabsorbing region near the facet by locally increasing the bandgap. The bandgap increase is accomplished by diffusing vacancies across the quantum well in this region to intermix the SCH and QW materials—what we have termed QWI above. The vacancy source results from the absorption of Ga

ions by a special film deposited on the surface in this region. CW output powers of >8 W were obtained with input currents of 10 A up to temperatures of 60°C for 100-μm-wide stripes.

Other approaches have been employed to limit COD. QWI can also be achieved by diffusing impurities across the quantum-well active regions as well as vacancies. Some companies have also been able to develop surface passivation treatments and thin-film coatings that appear to limit COD and other forms of facet degradation during aging and high-power operation.

Figure 1.19b gives an example of a relatively high power AlInGaP/GaAs surface-ridge waveguide laser that emits at 650 nm. With the powers shown, it can be used for high speed DVD writing as well as reading. For this application the power out as well as the output beam shape need to be free from kinks and glitches over a wide range of input currents. Diffraction angles parallel and perpendicular to the substrate that differ at most by a factor of 2 are also desired. These are accomplished by using a relatively narrow ridge width that ensures a single lateral optical mode as well as a relatively circular output beam.

QWI was used to form a transparent window region near the facets to prevent COD in this device as well. In this case the QWI was accomplished by diffusing Zn from a thin film of ZnO deposited on the surface only within 10 μm of the cleaved facet. The facets were also coated with low reflectivity (3%) and high reflectivity (95%) coatings on the output and back facets, respectively.

Similar lasers, although with lower power requirements, are used in most DVD players. Similar designs are also used in a wide variety of other "red laser" applications that range from supermarket scanners, to handheld pointers, to indicator lines on carpenter's tools.

Figure 1.19c shows results from an experimental GaN-based ridge wave guide blue laser. In this case the device is a nonpolar InGaN/GaN design, which does not have an AlGaN cladding. Unlike most of the early GaN laser work that used c-plane (0001) structures, this one is grown on m-plane $(1\bar{1}00)$ that avoids polarization effects as well as increased problems with parasitic wave guides and threading dislocations. This device emits at 404 nm; it has a threshold current density of 6.8 kA/cm^2 and a threshold voltage of 5.6 V. Although significant improvements in output power and efficiency are expected with this design, it still will serve as a representative of an important class of lasers based on GaN and its related alloys: AlGaInN.

The AlGaInN system can provide direct bandgaps from the infared (InN) to the UV (AlN), but good lasers have only been made in the blue-green to the near UV range. Nevertheless, this is a very important range for a wide variety of applications. Blue and near UV diode lasers are of great interest for solid-state lighting, high density optical data storage, projection displays, optical sensing, and medical applications. "Blu-ray" Disc is an important commercial product at this writing, but projection displays and solid-state lighting may have the largest long-term impact.

Figures 1.19d and e illustrate fiber-optic communication sources, both based on the InP materials system. The first (d) is an InGaAlAs/InP multiple-quantum-well

(MQW) 1300-nm wavelength ridge-waveguide laser that uses a grating along its length to select a single longitudinal output mode. As discussed in Chapter 3, this distributed grating provides a distributed feedback or (DFB) mirror, so it is called a DFB laser. The 1300-nm wavelength is the point of lowest dispersion in optical fiber, and it is used to send high data rates over modest distances. Because it is not used in long-haul networks, this wavelength tends to be limited to metropolitan, local area, or on-campus networks.

As discussed previously, the InAlGaAs material system has most of the quantum-well bandgap offset in the conduction band, so it is better at confining the low-effective-mass electrons than the InGaAsP/InP system. In this example, this material is selected to enable the lasers to operate to higher temperatures so that thermoelectric coolers are not needed, and much lower cost, lower power packages can be made. By making the cavity short (200 μm) the relaxation resonance was also increased to >20 GHz even at 100°C. The laser is designed to operate with direct current modulation at a data rate of 10.7 Gb/s. The laser is never turned completely off, as indicted in Fig. 1.19d, but the on/off ratio is kept to ~7 dB.

Figure 1.19e illustrates a semi-insulating regrowth buried-heterostructure (SIBH) laser formed with multiple quantum wells in the InGaAsP/InP system for operation at 1550 nm, which is the wavelength of minimum loss in optical fiber. It also is a DFB laser to ensure single-mode operation. Long-haul communication systems desire single-mode sources at 1550 nm to maximize the distance possible between amplifiers or repeater nodes. Because it is not the natural dispersion minimum of optical fiber, this needs to be managed in some other manner. It is also the wavelength region used for wavelength division multiplexed (WDM) multichannel communication. Thus, 1550 nm also tends to be used in metro and local area networks as these extend toward the edge.

Figure 1.20a illustrates results from a proton-implanted, gain-guided vertical-cavity laser emitting at 850 nm. The active region uses GaAs quantum-wells, and the epitaxial mirrors are composed of alternate quarter-wave layers of AlGaAs having a low (~15%) and high (~85%) Al composition, respectively. About two dozen periods are needed to provide >99.5% reflectivity as required to reach threshold in practical environments given the very small gain length (typically, the thickness of three quantum wells ~8 nm each) in these devices. The proton implant is generally concentrated midway between the top contact and the active region within the top, p-type mirror stack. It is usually slightly larger in diameter than the emission window in the top contact. This positioning ensures a low overall resistance path from the contact as well as a good aperturing of the current without damaging the carrier lifetime in the active region. This latter point emphasizes that a proton implant is *not* a carrier confinement approach, but rather just the opposite—it yields near zero carrier lifetime, so allowing the proton implant to extend into the lateral extremes of the active region would drain away carriers faster than simply allowing them to laterally diffuse away.

The GaAs-based device of Fig. 1.20a was the earliest type of VCSEL to be qualified for use in telecommunication systems and widely used with multimode fiber in data communication (datacom) links. As multimode fiber has core diameters

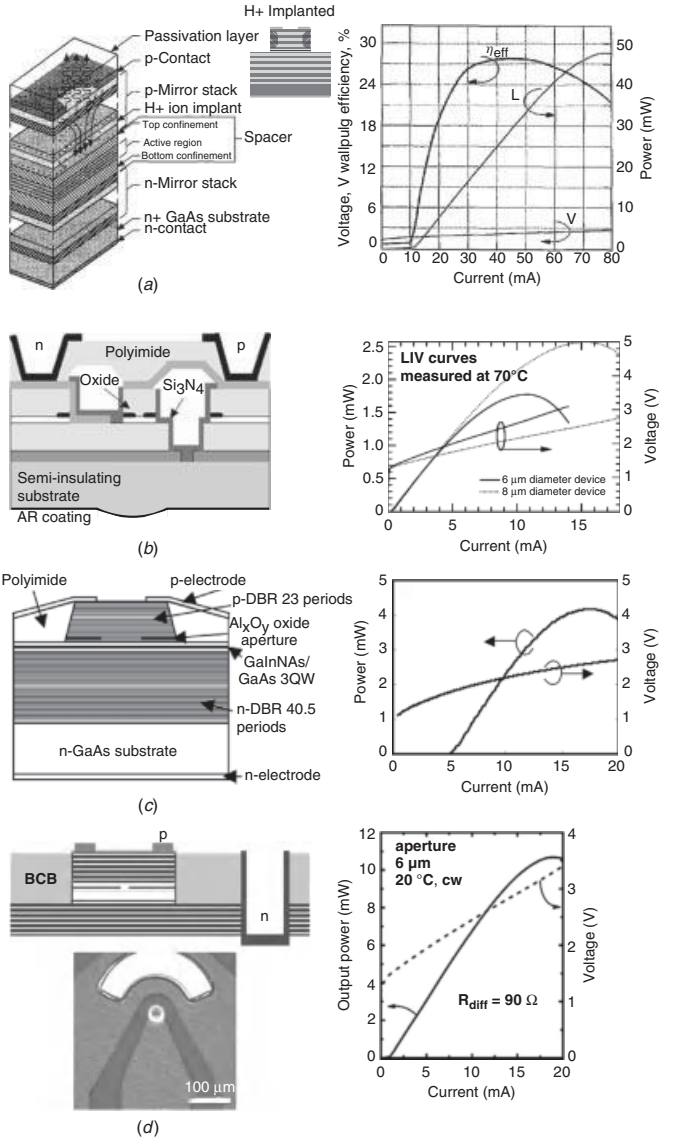


FIGURE 1.20: Practical VCSEL examples: (a) schematic and *L-I* characteristic for proton-implanted 850 nm wavelength AlGaAs/GaAs laser with 15 μm contact window and 20 μm diameter implant [12] (© IEE 1995); (b) schematic, *L-I* and temperature characteristics of 985 nm InGaAs/GaAs oxide-apertured VCSEL with an integrated backside microlens [13] (© IEEE 2007); (c) schematic and *L-I* characteristic of GaInNAs/GaAs VCSEL emitting at 1262 nm [14] (© IEEE 2007); (d) schematic, *L-I,* and temperature characteristics of 980 nm InGaAs/GaAs quantum-dot VCSEL [15] (© IEEE 2007); (e) schematic and L-I characteristics for both 1300 and 1550 nm InGaAlAs/InP VCSELs incorporating one dielectric mirror and a tunnel junction confinement structure [19] (© IEEE 2005).

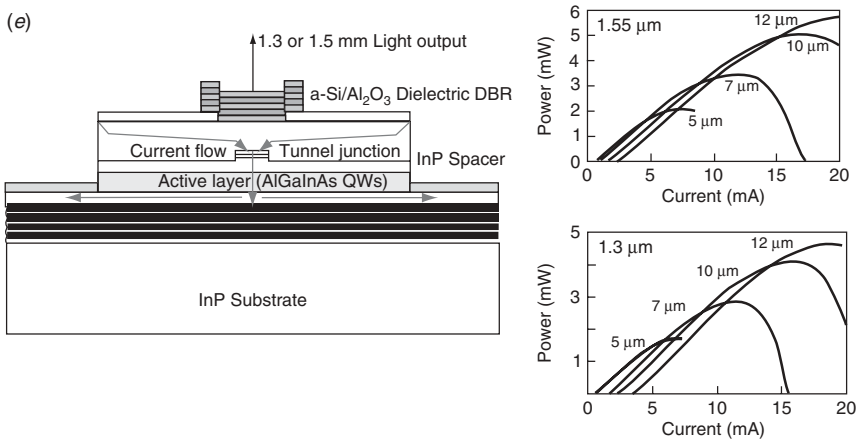


FIGURE 1.20: *(Continued)*

~50–60 μm, VCSEL diameters of ~20 μm with many lateral lasing modes are typically used. "Gigabit Ethernet" and "Fiber-channel" are examples of IEEE link standards that incorporate such VCSEL sources. Data rates in the gigabit per second range and distances up to a few hundred meters are typically involved. Linear arrays of such VCSELs together with fiber ribbon cable have also found applications in parallel data links, which can multiply the overall link capacity by 10 or more times.

A key attribute of the datacom links incorporating such multimode VCSELs and fiber is their low cost, given the relatively high level of performance. Unlike edge-emitters, VCSELs can be manufactured and tested at wafer scale—no cleaves or edges are necessary to evaluate them. Because of the large multimode fiber core and the small, circular diffraction angle of the VCSEL emission, passive alignment techniques are possible in the packaging of the VCSEL. Optical connectors can also be simple. All of this enables low cost production.

Figure 1.20b shows a somewhat more complex VCSEL design that employs an oxide aperture for lateral current and photon confinement. Illustrated is a bottom-emitting device with an InGaAs MQW active region emitting at 980 nm, GaAs/AlGaAs mirror stacks, and an integrated microlens. As shown, trenches must be etched down to the AlAs oxidation layer just above the active to perform the lateral oxidation after the epitaxial growth of the semiconductor layers. As illustrated, oxide-apertured devices can easily be formed with smaller diameters that support single-mode operation and very low threshold currents. The use of the InGaAs strained quantum wells provides an emission at which the GaAs substrate is transparent, and this enables bottom emission and flip-chip bonding for good heat sinking and simple packaging, including the option of integrated, prealigned microlenses on the backside of the substrate. The InGaAs also has better gain properties for more efficient and higher modulation bandwidth operation. However,

for many datacom applications the standards have been adopted for 850 nm, so many of these advantages go unrealized in systems.

The temperature characteristics shown in Fig. 1.20b also show another interesting feature designed into most modern VCSELs. Because the cavity is so short and the multilayer mirror reflectivity band is narrow, only a single axial mode can lase, and this tunes with temperature at the rate of ~0.1 nm/°C due to the change in the index of refraction. Also, the gain has a maximum value near the energy separation between the lowest quantum-well levels in the conduction and valence bands, but this tunes in temperature at the rate of ~0.5 nm/°C due to the change in bandgap energy. Thus, there is a differential tuning of the gain and mode wavelength, or explicitly, the peak gain will tune across and past the mode wavelength at a rate of ~0.4 nm/°C. So the trick is to deliberately misalign the peak gain at room temperature and as the temperature rises, enable the gain to move into alignment with and then past the mode. Of course, this results in a wide region of relative insensitivity to temperature.

Figure 1.20c illustrates another oxide-confined GaAs-based VCSEL, but in this case the MQW-active region is composed of GaInNAs, or a similar composition to Fig. 1.20b with a few percent of nitrogen added. Somewhat counterintuitively, the result is a significant increase in the emission wavelength toward 1300–1262 nm in this case. Normally, one might expect that the addition of a smaller atom to an alloy would reduce the wavelength, as in the case of GaN, for example. The difference here is that the N is so different in atomic size that it doesn't fit into the GaAs lattice in a normal bonding configuration. So, in small percentages, it creates a local deviation in the lattice potential, much like an impurity or a defect might, but of course it has the right valence, so it is neither. But, it does have the effect of pulling the conduction band down a bit resulting in the lower transition energy. Another complementary effect is that the smaller size of the N adds tensile strain, partially compensating the compressive strain of the In, and thus enabling an increase in the In composition, and this also tends to increase the emission wavelength.

The key reason for pursuing efforts with this material has been the ability to make GaAs-based VCSELs with well-developed high index contrast AlGaAs/GaAs mirror stacks and AlO_x oxide confinement, but with wavelengths ~1300 nm for longer-reach datacom applications. However, there have also been problems associated with reproducibility of the wavelength and gain properties of the GaInNAs quantum wells. A critical annealing step generally has been found to be necessary to bring out the optimal properties of the material, and transfer to a manufacturing environment has not always been successful. Although good insensitivity to temperature variations has been reported, this appears to be due to a large nonradiative component that is temperature insensitive. Thus, the efficiency has generally not been comparable to other GaAs-based VCSELs, or even as good as the best InP-based long wavelength VCSELs as discussed later. Nevertheless, the GaInNAs approach may still be the most cost effective if high performance is not needed.

Figure 1.20d gives an example of a GaAs-based VCSEL with a quantum-dot active region. Although many efforts with quantum-dot actives on GaAs have

aimed to push the wavelength out to the 1300-nm range using increased strain InAs dots, this example emits at 980 nm. It uses a growth procedure that avoids a lateral interconnecting "wetting layer" common to most other quantum dot devices that have shown relatively slow modulation responses. Tilted indium-rich columns result from a multisubmonolayer growth procedure. In this case direct modulation to 20 Gb/s at 85°C was demonstrated for the 6-μm diameter device characterized in the figure. In this case, over 10 mW of power is emitted at room temperature, and the threshold current is again engineered to decrease with temperature due to a deliberate gain-mode peak offset as discussed earlier.

Figure 1.20e introduces an InP-based VCSEL technology that incorporates an etched-tunnel-junction mesa as the current and photon confinement element. Results for both 1300- and 1550-nm emission are shown. InP-based VCSELs generally suffer from at least three problems relative to their GaAs-based counterparts. First, there are no well-developed lattice-matched mirrors that have high index contrast and high thermal conductivity; second, there is no simple well-developed oxide aperturing technology; third, the active materials tend to have somewhat lower gain per unit current and to be more sensitive to temperature. The configuration of Fig. 1.16e addresses most of these issues. However, a number of other alternative fabrication technologies have also been explored over the years. These include (1) wafer bonding GaAs-based mirrors onto an InP-base active region [16]; (2) use of a very thick InGaAsP bottom (output) mirror stack together with a lattice mismatched (metamorphic) GaAs/AlAs mirror stack for a top mirror [17]; and (3) use of lattice-matched high-index contrast GaAlAsSb mirror stacks in an all-epitaxial InP-based structure [18]. After some development, none have been manufactured in any volume. Of course, long-wavelength GaAs-based VCSELs have also been researched using both GaInNAs and quantum-dots, as already mentioned.

In the case of Fig. 1.20e, three key layer structures are grown in sequence. First the bottom semiconductor mirror stack, bottom n-type contact layer, active region, a thin p-type layer, and the $p^+ - n^+$ tunnel junction are formed on the InP substrate. The tunnel junction, which enables low-resistance n-type contacts to the top as well as the bottom of the diode, is then etched away except where the optical mode is to exist. In the second semiconductor growth, the n-type InP spacer/top contact layer is grown over the tunnel junction mesa and the rest of the p-type top surface. Note that no current will flow across the reverse biased n−p junction except where the tunnel junction exists. Ohmic contacts are formed to the InP spacer. Finally, the top amorphous-Si/Al_2O_3 dielectric mirror stack is grown. Because the tunnel junction also adds optical length to the center of the cavity, it also acts as a lensing element to confine to optical mode to this region. The thermal impedance can be kept relatively low in this design if heat sinking is added to the top InP layer because it has relatively good thermal conductivity. This can be addressed by designing the device for bottom emission and flip-chipping it onto a heat sink. There has been similar work that has emphasized this approach [20].

The applications for these long wavelength VCSELs are in single-mode fiber data links with data rates ∼10 Gb/s (e.g., 10 Gigabit Ethernet or 10 GbE) and distances extending to tens of kilometers. Parallel data links for applications like

100 GbE are also being developed. Competition with simple edge-emitting solutions such at those in Fig. 1.19d and 1.19e are key issues. The low-cost manufacturability of VCSELs remains one of the possible advantages, but as the VCSELs become a little more complex this becomes a weaker argument.

REFERENCES

1. Hirtz JP, Razeghi M, Bonnet M, Duchemin JP. Ga.sub.0.47 In.sub.0.53 As/InP double heterostructures grown by low-pressure metal-organic vapour-phase epitaxy. In: Pearsall TP, editor. *GaInAsP alloy semiconductors*. New York: Wiley; 1982. P 61–83.
2. Vurgaftman I, Meyer JR, Ram-Mohan LR. Band parameters for III–V compound semiconductors and their alloys. *J Appl Phys* 2001;**89**:5815.
3. Hayachi I, Panish MB, Foy PW, and Sumski S. *Appl. Phys. Lett.*, 1970;**109**:17.
4. Chaub KY, and Cho AY. *J. Appl. Phys.*, 1982;**53**:441.
5. Vurgaftman I, Meyer Jr. Band parameters for nitrogen-containing semiconductors. *J Appl Phys* 2003;**94**:3675.
6. Coldren LA, Thibeault BJ, Hegblom ER, Thompson GB, Scott JW. Dielectric apertures as intracavity lenses in vertical-cavity lasers. *Appl Phys Lett* 1996;**68**:313.
7. Taniguchi H, Ishii H, Minato R, Ohki Y, Namegaya T, Kasukawa A. 25-W 915-nm lasers with window structure fabricated by impurity-free vacancy disordering (IFVD). *IEEE Journal of Selected Topics in Quantum Electronics*. 2007;**13**(5):1176–1179.
8. Sumitomo H, Kajiyama S, Oguri J, Sakashita T, Yamamoto T, Nakao K, Domoto S, Ueda M, Amano H, Satoyoshi H, Kita T, Izumi, S. Uniform and high-power characteristics of AlGaInP-based laser diodes with 4-inch-wafer process technology. *IEEE Journal of Selected Topics in Quantum Electronics* 2007;**13**(5):1170–1175.
9. Farrell RM, Feezell DF, Schmidt MC, Haeger DA, Kelchner KM, Iso K, Yamada H, Saito M, Fujito K, Cohen DA, Spec JS, DenBaars SP, Nakamura S. Continuous-wave operation of AlGaN-cladding-free nonpolar *m*-plane InGaN/GaN laser diodes. *Jpn J Appl Phys* 2007;**46**:L761–L763.
10. Singh H, Motoda K, Mukaikubo M, Okamoto K, Washino R, Sakuma Y, Uchida K, Yamamoto H, Uomi K, Nakahara K, Aoki M. Proceedings of Optical Fiber Communication Conference and National Fiber Optic Engineers Conference; 2006 Mar 5–10.
11. Takeshita T, Tadokoro T, Iga R, Tohmori Y, Yamamoto M, Sugo M. Highly reliable 1.5-μm DFB laser with high-mesa SIBH structure. *IEEE Transactions on Electron Devices* 2007;**54**(3):456–461.
12. Morgan FA, Hibbs-Brenner MK, Walterson RA, Lehman JA, Marta TM, Bounnak S, Kelweit EL, Akinwande R, Nohava JC. Producible GaAs-based MOVPE-grown vertical-cavity top-surface emitting lasers with record performance. *Electronics Letters* 1995;**31**:462.
13. Lin C-K, Tandon A, Djordjev K, Corzine SW, Tan MRT. High-speed 985 nm bottom-emitting VCSEL arrays for chip-to-chip parallel optical intercommects. *IEEE Journal of Selected Topics in Quantum Electronics* 2007;**13**(5):1332–1339.
14. Nishida T, Takaya M, Kakinuma S, Kaneko T. *IEEE Journal of Selected Topics in Quantum Electronics* 2005;**11**(5):958–961.
15. Hopfer F, Mutig A, Fiol G, Kuntz M, Shchukin BA, Haisler FA, Warming T, Stock E, Mikhrin SS, Krestnikov IL. 20 Gb/s 85°C error-free operation of VCSELs based on

submonolayer deposition of quantum dots. *IEEE Journal of Selected Topics in Quantum Electronics* 2007;**13**(5):1302–1308.

16. Okuno YL, Gan K-G, Chou J-F, Chiu Y-J, Wang CS, Wu S, Geske J, Bjorlin ES, Bowers JE. *IEEE Journal of Selected Topics in Quantum Electronics* 2005;**11**(5): 1006–1014.
17. Starck C, Plais A, Derouin E, Pinquier A, Gaborit F, Fortin C, Goldstein L, Boucart J, Salet P, Carpentier D. Indium phosphide and related materials. 1998 International Conference on. 1998 May 11–15. P369–372.
18. Nakagawa S, Hall, E, Almuneau G, Kim JK, Buell DA, Kroemer H, Coldren LA. 1.55-μm InP-lattice-matched VCSELs with AlGaAsSb-AlAsSb DBRs. *IEEE Journal of Selected Topics in Quantum Electronics* 2001;**7**(2):224–230.
19. Nishiyama N, Caneau C, Hall B, Guryanov G, Hu MH, Liu XS, Li M-J, Bhat R, Zah CE. Long-wavelength vertical-cavity surface-emitting lasers on InP with lattice matched AlGaInAs– InP DBR grown by MOCVD. *IEEE Journal of Selected Topics in Quantum Electronics* 2005;**11**(5):990–998.
20. Amann M-C, Hofmann W. InP-based long-wavelength VCSELs and VCSEL arrays. *IEEE Journal of Selected Topics in Quantum Electronics* 2009;**15**(3):861–868.

READING LIST

P. Bhattacharya, *Semiconductor Optoelectronic Devices*,Chs. 1 and 2, Prentice Hall, Englwood Cliffs, NJ (1994).

B.E.A. Saleh and M.C. Teich, *Fundamentals of Photonics*, Chs. 15 and 16, Wiley, New York (1991).

A. Yariv, *Optical Electronics*, 4th ed., Ch. 15, Saunders College Publishing, Philadelphia (1991).

G. P. Agrawal and N. K. Dutta, *Semiconductor Lasers*, 2d ed., Chs. 4 and 5, Van Nostrand Reinhold, New York (1993).

PROBLEMS

These problems draw on material from Appendices 1 through 3.

1. Define the necessary elements of a laser cavity.

2. A laser cavity is formed by two ideal mirrors with flat frequency response over the 1.4–1.6 μm wavelength range, positioned 1 mm apart, by an ideal bandpass filter centered at 1.55 μm and 4 nm wide, and by a gain region with a parabolic function gain function $g = -7.5 \cdot 10^{12}\ \text{cm}^{-3}(\lambda - 1.55\ \mu\text{m})^2 + 10^4\ \text{cm}^{-1}$. For this simple cavity, referring to Fig. 1.1, determine the lasing wavelength of this laser. Assume that most of the photons are propagating through free space.

3. List three advantages and three disadvantages diode lasers have relative to gas or solid-state lasers.

4. What are the most common applications of diode lasers? Can you think of any other applications not mentioned in the text?

5. What is the difference between the energy levels in the solid state and diode lasers?

6. List and explain all the basic electronic recombination/generation mechanisms. Which one is required for lasers to operate?

7. Describe the main ways of nonradiative recommbination in semiconductor lasers.

8. Explain how a double heterostructure works. Who are the Nobel prize winners for this invention?

9. Discuss the differences in carrier transport and recombination in direct-bandgap double heterostructure PN junctions relative to indirect-bandgap homojunction PN junctions?

10. A blue diode laser cross section consists of a set of five 80-nm-wide InGaN quantum wells with 12% In and 88% Ga, surrounded by six 80-nm-wide GaN barriers, and clad by P and N type GaN.

(a) Determine the effective index of the fundamental transverse mode of this waveguide. To do this, simplify the structure by calculating an average optical index of refraction for the quantum well/barrier region.

(b) Determine the rate of decay of the normalized transverse electric field U.

11. Give three differences between using a bulk active region and a quantum-well active region in a diode laser.

12. What are the fundamental requirements for the different materials involved in forming a complex separate-confinement heterostructure?

13. Why are III-V materials better than Si for LEDs and lasers?

14. What type of lasers is the GaAs/AlGaAs material system used for? How about InP/InGaAsP?

15. What is band offset, and why is it important? How much of the band offset occurs in the conduction band for (1) GaAs based lasers (2) InP based lasers?

16. An electron is trapped in a one-dimensional potential well 5 nm wide and 100 meV deep.

(a) How many bound energy states exist?

(b) What are the energy levels of the first three measured relative to the well bottom?

(c) If the well energy depth were doubled, how many states would be confined? (Assume the free electron mass.)

17. Repeat Problem 16 for a 10-nm-wide GaAs well and AlGaAs barriers.

18. Ten potential wells that each have two bound states are brought together so that their wavefunctions overlap slightly. How many bound energy states exist in this system?

19. How can the temperature range of InGaAsP lasers be extended?

20. The Blu-ray DVD disc has a much higher capacity than the original DVD disc. Explain how this was accomplished. How would you create an even higher capacity optical disc?

21. List two major manufacturers of the MOCVD growth systems.

22. Redo Example 1.2 for an InP-based VCSEL using AlInGaAs/AlInAs DBR mirror stacks. That is, determine the percentage of Al and In in the AlInAs layers to lattice match to InP.

23. Plot the minimum bandgap versus lattice constant for InAlSb. The bowing parameters for the direct, and the first indirect valley are 0.43 eV and 0 respectively. At which composition and lattice constant does the indirect bandgap equal direct bandgap?

24. A very long one-dimensional chain consists of atoms covalently bonded together with a resulting center-to-center spacing of 0.3 nm. The band structure of this system can be determined from the overlap of the individual atomic wavefunctions. The coupling energy given by Eq. (A1.21) for a particular atomic energy level, E_a, is 0.2 eV.

(a) Calculate the band structure over the first two Brillouin zones.

(b) Calculate the electron effective mass at the band extrema.

25. A light source emits a uniform intensity in the wavelength range $0.4 - 2.0$ μm. A polished wafer of GaAs with antireflection coatings on both surfaces is placed between the source and an optical spectrum analyzer.

(a) Sketch the wavelength spectrum received.

(b) Which processes in Fig. 1.6 are significant in forming this spectrum?

26. The light source in Problem 1.25 is replaced by a GaAs laser emitting at 850 nm, and it is found that 99.5% of the incident light is absorbed in the GaAs wafer. Now an Ar-ion laser emitting at 488 nm is trained to the same place on the wafer.

(a) As the power of the Ar-ion laser is increased to 3 W, the absorption of the GaAs laser beam is reduced to 50%. Assuming the heat is conducted away, explain what might be happening.

(b) The power in the Ar-ion laser is further increased, and it is found that at about 10 W the GaAs laser beam passes through the wafer unattenuated. Again, neglecting heating effects, explain why it requires 10 W rather than ~6 W to reach transparency.

27. For good carrier confinement it has been found that the quasi-Fermi levels should remain at least 5 kT below the top of a quantum well at operating temperature. In a particular GaAs quantum-well SCH laser, the operating active region temperature is found to be 125°C. If the quantum well is 80 Å wide, how much Al should be in the separate confinement region to provide the desired 5 kT margin in the conduction band at a carrier density of 4×10^{18} cm^{-3}?

28. (a) Plot the carrier density vs. the quasi-Fermi level for the conduction band in bulk GaAs and InGaAsP (1.3 μm) at 300 K. Cover the carrier density range from 1×10^{17} cm^{-3} to 1×10^{19} cm^{-3}, and use a logarithmic scale for the carrier density axis.

(b) With this result answer Problem 9 for a AlGaAs/GaAs bulk DH structure.

29. Calculate the density of states vs. energy for a "quantum wire" potential well in which two dimensions are relatively small. That is, assume a large dimension ($\gg$10 nm) in the z-direction and quasi-continuous state energies only for k_z.

30. Calculate the density of states vs. momentum for a quantum well.

31. Derive Eq. (A3.3).

32. Photons are transversely confined in a simple three-layer waveguide in a DH laser consisting of an InGaAsP active region 0.2 μm thick sandwiched between InP cladding layers. The bandgap wavelength of the active region is 1.3 μm.

(a) How many transverse TE modes can exist in this slab waveguide?

(b) Plot the transverse electric field for the lowest-order TE mode.

(c) What is the energy density 0.5 μm above the active-cladding interface relative to the peak value in the active region?

(d) What is the effective index of the guided mode?

(e) What is the transverse confinement factor?

33. Suppose the DH laser of Problem 1.32 is now used to form a BH laser with an active region 2 μm wide and InP lateral cladding regions, as in Fig. 1.17.

(a) What is the effective index for the fundamental two-dimensionally guided mode?

(b) How many lateral modes are possible?

(c) What is the lateral confinement factor for the fundamental mode?

34. VCSELs have been formed by etching 5 μm square pillars through the entire laser structure, creating rather large index discontinuities at the lateral surfaces. Assuming the axial propagation constant, β, is fixed at the same value for all resonant modes, and that the lowest-order mode has a wavelength of 1.0 μm, plot the mode spectrum including the first six lateral modes.

35. Derive Eq. (A3.14) and verify Eq. (A3.15).

36. It has been proposed that if the lateral dimensions of VCSELs or in-plane lasers become sufficiently small, the density of states for electrons and holes can be modified by the lateral size effect. In VCSEL material with an 80 Å thick GaAs quantum-well active region and high barriers, devices of various lateral widths are formed. How narrow must the device be before the lateral size effect shifts the lowest state energy up by 10 meV (about $\frac{1}{2}$ kT at room temperature)? Neglect any indirect surface-state pinning effect.

CHAPTER TWO

A Phenomenological Approach to Diode Lasers

2.1 INTRODUCTION

Now that we have a basic understanding of what diode lasers are, what is involved in their fabrication and operation, and what characteristics can generally be expected, we are perhaps in a position to delve into the mechanics of how an injected current actually results in an optical output. In this chapter we attempt to develop an engineering toolbox of diode laser properties based largely on phenomenological arguments. In the course of this development, we make heavy reference to several appendices for a review of some of the underlying physics.

The chapter begins by developing a rate equation model for the flow of charge into double-heterostructure active regions and its subsequent recombination. Some of this electron–hole recombination generates photons by spontaneous emission. This incoherent light is important in LEDs, and a section is devoted to deriving the relevant equations governing LED operation.

Sections 2.4 through 2.6 provide a systematic derivation of the dc light-current characteristics of diode lasers. First, the rate equation for photon generation and loss in a laser cavity is developed. This shows that only a small portion of the spontaneously generated light contributes to the lasing mode. Most of it comes from the stimulated recombination of carriers. All of the carriers that are stimulated to recombine by light in a certain mode contribute more photons to that same mode. Thus, the stimulated carrier recombination/photon generation process is a *gain*

Diode Lasers and Photonic Integrated Circuits, Second Edition.
Larry A. Coldren, Scott W. Corzine, and Milan L. Mašanović.
© 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

process. The threshold gain for lasing is studied next, and it is found to be the gain necessary to compensate for cavity losses. The current required to reach this gain is called the threshold current, and it is shown to be the current necessary to supply carriers for the unproductive nonradiative and spontaneous recombination processes, which clamp at their threshold value as more current is applied. Above threshold, all additional injected carriers recombining in the active region are shown to contribute to photons in the lasing mode. A fraction escape through the mirrors; others are absorbed by optical losses in the cavity.

The next section deals with the modulation of lasers. Here for the first time we solve the rate equations for a modulated current. Under small-signal modulation, the rate equations for carriers and photons are found to be analogous to the differential equations that describe the current and voltage in an *RLC* circuit. Thus, the optical modulation response is found to have a resonance and to fall off rapidly above this frequency.

Finally, this chapter reviews techniques for characterizing real lasers. These techniques can be used to extract the important device parameters used in the theoretical derivations. They also provide practical terminal parameters that are useful in the design of optoelectronic circuits.

2.2 CARRIER GENERATION AND RECOMBINATION IN ACTIVE REGIONS

In Chapter 1, when we considered the current injected into the terminals of a diode laser or LED, we suggested that it was desirable to have all the current contribute to electrons and holes, which recombine in the active region. However, in practice only a fraction, η_i, of the injected current, I, does contribute to such carriers. In Fig. 2.1 we again illustrate the process of carrier injection into a double-heterostructure active region using a somewhat more accurate sketch of the energy gap versus depth into the substrate.

Because the definitions of the active region and the injection efficiency, η_i, are so critical to further analysis, we highlight them here for easy reference.

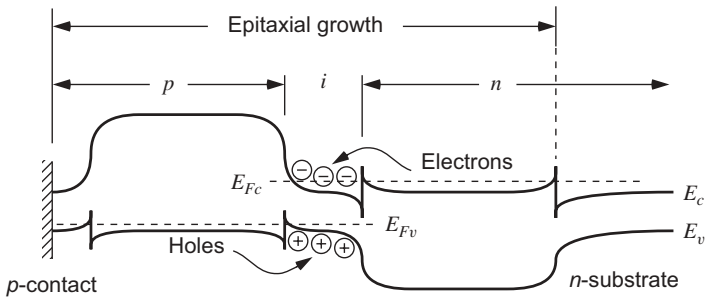


FIGURE 2.1: Band diagram of forward biased double-heterostructure diode.

Active region: the region where recombining carriers contribute to useful gain and photon emission.

The active region is usually the lowest bandgap region within the depletion region of a *pin* diode for efficient injection. However, it occasionally is convenient to include some of the surrounding intermediate bandgap regions. Also in this definition, *useful* is the operative word. There may be photon emission and even gain at some undesired wavelength elsewhere in the device.

Injection efficiency, η_i: *the fraction of terminal current that generates carriers in the active region.* It is important to realize that this definition includes *all* the carriers that are injected into the active region, not just carriers that recombine radiatively at the desired transition energy. This definition is oftentimes misstated in the literature.

We also will specifically analyze active regions that are undoped or lightly doped, so that under high injection levels relevant to LEDs and lasers, charge neutrality dictates that the electron density equals the hole density (i.e., $N = P$ in the active region). Thus, we can greatly simplify our analysis by specifically tracking only the electron density, N.

The carrier density in the active region is governed by a dynamic process. In fact, we can compare the process of establishing a certain steady-state carrier density in the active region to that of establishing a certain water level in a reservoir that is being simultaneously filled and drained. This is shown schematically in Fig. 2.2. As we proceed, the various filling (generation) and drain (recombination) terms illustrated will be defined. The current leakage illustrated in Fig. 2.2 contributes to reducing η_i and is created by possible shunt paths around the active region. The carrier leakage, R_l, is due to carriers "splashing" out of the active region (by thermionic emission or by lateral diffusion if no lateral confinement exists) before recombining. Thus, this leakage contributes to a loss of carriers in the active region that could otherwise be used to generate light.

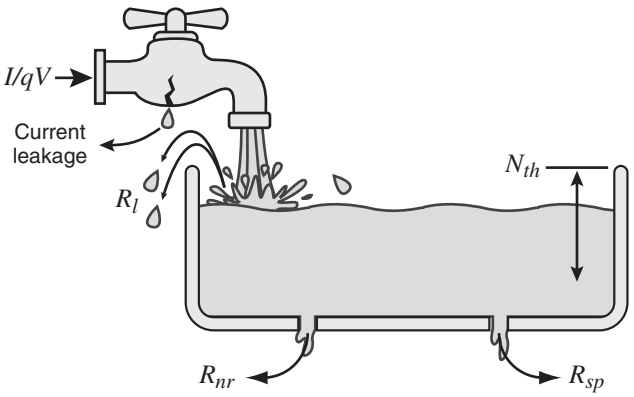


FIGURE 2.2: Reservoir with continuous supply and leakage as an analog to a double heterostructure active region with current injection for carrier generation and radiative and nonradiative recombination (LED or laser below threshold).

For the DH active region, the injected current provides a generation term, and various radiative and nonradiative recombination processes as well as carrier leakage provide recombination terms. Thus, we can write the rate equation,

$$\frac{dN}{dt} = G_{gen} - R_{rec}, \tag{2.1}$$

where G_{gen} is the rate of injected electrons and R_{rec} is the rate of recombining electrons per unit volume in the active region. Because there are $\eta_i I/q$ electrons per second being injected into the active region,

$$G_{gen} = \frac{\eta_i I}{qV}, \tag{2.2}$$

where V is the volume of the active region. For example, if a current of $I =$ 20 mA is flowing into the laser's terminals, a fraction $\eta_i = 80\%$ of the carriers are injected into the active region, and if the active volume is 100 μm^3, then $G_{gen} = 10^{27}$ electrons/s-cm^3. Or, 10^{18} cm^{-3} electrons are injected in 1 ns.

The recombination process is a bit more complicated because several mechanisms must be considered. As introduced in Fig. 1.6, there is a spontaneous recombination rate, R_{sp}, and a nonradiative recombination rate, R_{nr}. And as depicted in Fig. 2.2, a carrier leakage rate, R_l, must sometimes be included if the transverse and/or lateral potential barriers are not sufficiently high (see Appendix 2 for a discussion of R_l). Finally, under the right conditions, a net stimulated recombination, R_{st}, including both stimulated absorption and emission, is important. Thus, we can write

$$R_{rec} = R_{sp} + R_{nr} + R_l + R_{st}. \tag{2.3}$$

The first three terms on the right refer to the natural or unstimulated carrier decay processes. The fourth one, R_{st}, requires the presence of photons.

It is common to describe the natural decay processes by a *carrier lifetime*, τ. In the absence of photons or a generation term, the rate equation for carrier decay is just, $dN/dt = -N/\tau$, where $N/\tau \equiv R_{sp} + R_{nr} + R_l$, by comparison to Eq. (2.3). This rate equation defines τ, and we can also think of its definition as the decay constant in the solution to this rate equation for carrier decay,

$$N(t) = N(0)e^{-t/\tau}.$$

As mentioned in Chapter 1, this natural decay rate can be expressed in a power series of the carrier density, N, because each of the terms depends on the existence of carriers. Thus, we can rewrite Eq. (2.3) in several ways.

$$R_{rec} = R_{sp} + R_{nr} + R_l + R_{st}, \tag{2.3a}$$

$$R_{rec} = \frac{N}{\tau} + R_{st}, \tag{2.3b}$$

$$R_{rec} = BN^2 + (AN + CN^3) + R_{st}, \tag{2.3c}$$

Whereas the grouping suggests in Eq. (2.3c), it has been found that $R_{sp} \sim BN^2$ and $R_{nr} + R_l \sim (AN + CN^3)$. The coefficient B is called the *bimolecular recombination coefficient*, and it has a magnitude, $B \sim 10^{-10}$ cm^3/s for most AlGaAs and InGaAsP alloys of interest. We also note that the carrier lifetime, τ, is not independent of N in most circumstances.

Thus, so far we can write our carrier rate equation in several equivalent ways. We shall deal with R_{st} a little later, but using Eq. (2.3b), our carrier rate equation for the *active region* may be expressed as

$$\frac{dN}{dt} = \frac{\eta_i I}{qV} - \frac{N}{\tau} - R_{st}. \tag{2.4}$$

In the absence of a large photon density, such as in a laser well below threshold or in most LEDs, it can be shown that R_{st} can be neglected. Figure 2.2 illustrates each of these terms in our reservoir analogy, explicitly showing "leaks" R_{sp}, R_{nr}, and R_l for N/τ.

2.3 SPONTANEOUS PHOTON GENERATION AND LEDS

Before proceeding to the consideration of lasers, where R_{st} will become a dominant term above threshold, let us first try to gain some understanding of the situation where the photon density is relatively low, such as in an LED where no feedback is present to provide for the buildup of a large photon density. This case is actually similar to a laser below threshold, in which the gain is insufficient to compensate for cavity losses, and generated photons do not receive net amplification.

The spontaneous photon generation rate per unit volume is exactly equal to the spontaneous electron recombination rate, R_{sp}, because by definition every time an electron–hole pair recombines radiatively, a photon is generated. (Again, N equals the density of electron–hole pairs as well as electrons for relatively light doping). Under *steady-state* conditions ($dN/dt = 0$), the generation rate equals the recombination rate (i.e., from Eqs. (2.2) and (2.3)), with $R_{st} \approx 0$,

$$\frac{\eta_i I}{qV} = R_{sp} + R_{nr} + R_l. \tag{2.5}$$

The spontaneously generated optical power, P_{sp}, is obtained by multiplying the number of photons generated per unit time per unit volume, R_{sp}, by the energy per photon, $h\nu$, and the volume of the active region, V. We could solve Eq. (2.5) for R_{sp}, but because the exact dependence of $R_{nr} + R_l$ on I is unknown, this leads only to a parametric equation. The conventional approach is to bury this problem by defining a *radiative efficiency*, η_r, where

$$\eta_r = \frac{R_{sp}}{R_{sp} + R_{nr} + R_l}. \tag{2.6}$$

We must not forget that η_r usually depends on carrier density somewhat. Then, from Eqs. (2.5) and (2.6),

$$P_{sp} = h\nu V R_{sp} = \eta_i \eta_r \frac{h\nu}{q} I. \tag{2.7}$$

The product of $\eta_i \eta_r$ is sometimes referred to as the LED internal efficiency. However, we shall *not* use this definition here because it can lead to serious confusion when we move on to lasers. As we shall see, only η_i appears in the laser output power, and we have called it alone the internal efficiency.

If we are interested in how much power the LED emits into some receiving aperture, P_{LED}, we must further multiply P_{sp} by the *net collection efficiency*, η_c, experienced in transmitting photons out of the semiconductor and into this aperture. Historically, this value was relatively low (<10%) for most LEDs, because the light is emitted in all directions, and much of it is totally reflected at the semiconductor–air interface. However, with recent developments of LEDs for lighting applications, novel structures and techniques have been developed to improve the extraction efficiency, and efficiencies in the range of 70% have been achieved and reported. Several LED cross sectional schematics are shown in Fig. 2.3.

In a conventional planar surface structure, as shown in Fig. 2.3a, much of the light is reflected back toward the active region rather than being coupled out of the semiconductor chip. This can lead to the regeneration of new carriers by the reabsorption of this light and "photon recycling," which can increase LED efficiency.

The simplest way to increase light extraction efficiency is to roughen the semiconductor air interface, which will drastically reduce the amount of total internal reflection back into the LED. This technique has become a common practice with thin film LEDs. With the addition of external encapsulation, 80% efficiencies have been reported [2].

Another commercially viable technique is use of shaped transparent substrates, which allow that most generated light satisfy criteria for escape. A cross section of one such structure, a truncated-inverted pyramid, is shown in Fig. 2.3b.

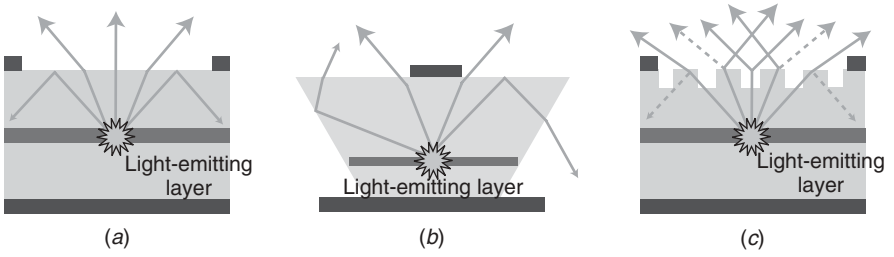


FIGURE 2.3: LED cross-sectional schematics for (a) a standard LED where only a small portion of the generated light reaches a desired detector, (b) shaped transparent substrate LED, and (c) diffraction grating LED.

Figure 2.3c shows the latest approach, where the light propagation is controlled and 2.3c extraction increased through use of multidirectional diffraction effect. Introduction of this surface roughness reduces the amount of total internal reflection, and has allowed for well controlled, unencapsulated extraction efficiency of 73% [3].

Going back to our LED analysis, the product of the three efficiencies (fraction of carriers injected into the active region, fraction of these recombining radiatively, and the fraction of those usefully coupled out) gives the *external LED quantum efficiency*, η_{ex}. That is,

$$P_{LED} = \eta_c \eta_i \eta_r \frac{h\nu}{q} I = \eta_{ex} \frac{h\nu}{q} I. \tag{2.8}$$

Thus, ignoring the slight dependence of η_{ex} on I, we see that the power coupled from an LED is directly proportional to the drive current. The external LED quantum efficiency, η_{ex}, is the number of photons coupled to the receiving aperture per electron flowing into the LED.

The frequency response of the LED can also be derived from the carrier rate in Eq. (2.4), with $R_{st} \approx 0$. We shall use the theorem that the Fourier transform of the impulse response in the time domain gives the frequency response. An impulse of current is simply a quantity of charge, which will establish an initial condition of $N(t = 0^+) = N_i$. For $t > 0$, the rate equation can be written as

$$\frac{dN}{dt} = -\frac{N}{\tau} = -AN - BN^2 - CN^3. \tag{2.9}$$

With the polynomial expansion of the recombination rate, we are reminded that the carrier lifetime, τ, is generally a function of the carrier density. If it were independent of N, the solution would be a simple exponential decay, and the frequency response would be analogous to that of a simple RC circuit in which the 3 dB cutoff frequency, $\omega_c = 1/\tau$. For τ to be constant: (1) the cubic term must be negligible and (2) either the linear term, AN, must dominate (not good, because this represents nonradiative recombination) or the active region must be heavily doped, such that the BN^2 term, which really equals BNP, can be written as $(BP_d)N$. That is, the p-type doping level, P_d, must be greater than the injection level, N, so that $P_d + P \approx P_d$. Under these conditions, then, the time response is just a simple exponential decay given by Eq. (2.4) and the frequency response is a Lorentzian function,

$$N(\omega) = \frac{N(0)}{1 + j\omega\tau}, \tag{2.10}$$

which drops to 0.707 $N(0)$ at $\omega\tau = 1$. For $R_{sp} \approx (BP_d)N$, the power out, P_{LED}, which is proportional to R_{sp}, will also have the same frequency response. The other cases are left as exercises for the reader, but it should be clear that the cutoff frequency will be reduced if the carrier lifetime is increased.

2.4 PHOTON GENERATION AND LOSS IN LASER CAVITIES

For the diode laser, we must now further investigate the nature of the net stimulated recombination rate, R_{st}, in generating photons as well as the effect of the resonant cavity in storing photons. In analogy with Section 2.2, we wish to construct a rate equation for the *photon density*, N_p, which includes the photon generation and loss terms. We shall use the subscript p to indicate that variables are referring to photons.

A main difference between the laser and LED, discussed in Section 2.3, is that we only consider light emission into a single mode of the resonant cavity in the laser. Because there are typically thousands of possible optical modes in a diode laser cavity, only a small fraction of R_{sp} contributes to the photon generation rate for a particular mode. Appendix 4 discusses the possible optical modes of a resonant cavity using some of the results of Appendix 3. Note that the number of effective modes in a small vertical-cavity laser can be much fewer, typically dozens rather than thousands.

The main photon generation term above threshold (the regime of interest in lasers) is R_{st}. Every time an electron–hole pair is stimulated to recombine, another photon is generated. However, as indicated in Fig. 2.4, because the cavity volume occupied by photons, V_p, is usually larger than the active region volume occupied by electrons, V, the photon *density* generation rate will be $[V/V_p]R_{st}$ not just R_{st}.

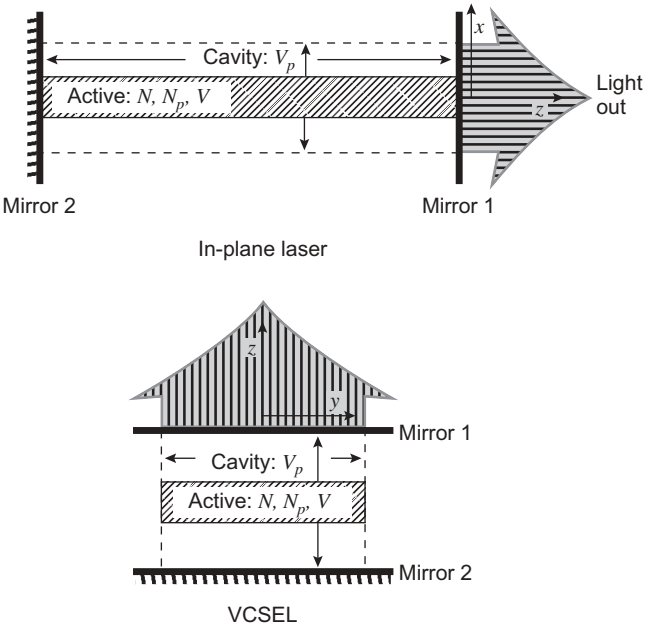


FIGURE 2.4: Schematics of in-plane and vertical-cavity lasers illustrating the active (cross-hatched) and cavity (within dashed lines) volumes as well as the coordinate systems.

This electron–photon overlap factor, V/V_p, is generally referred to as the *confinement factor*, Γ. Sometimes it is convenient to introduce an effective thickness, width, and length that contains the photons, d_{eff}, w_{eff}, and L, respectively. That is, $V_p = d_{eff} w_{eff} L$. Then, if the active region has dimensions, d, w, and L_a, the confinement factor can be expressed as, $\Gamma = \Gamma_x \Gamma_y \Gamma_z$, where $\Gamma_x = d/d_{eff}$, $\Gamma_y = w/w_{eff}$, and $\Gamma_z = L_a/L$. Appendix 5 puts the derivation of Γ on a more rigorous foundation, pointing out that Γ_z is subject to an enhancement factor for $L_a \lesssim \lambda$. Photon loss occurs within the cavity due to optical absorption and scattering out of the mode, and it also occurs at the output coupling mirror, where a portion of the resonant mode is usefully coupled to some output medium. These losses will be quantified in the next section, but for now we can characterize the net loss by a *photon* (*or cavity*) *lifetime*, τ_p, analogous to how we handled electron losses. A first version of the photon density rate equation for a single optical mode in the *active region* takes the form:

$$\frac{dN_p}{dt} = \Gamma R_{st} + \Gamma \beta_{sp} R_{sp} - \frac{N_p}{\tau_p}, \tag{2.11}$$

where β_{sp} is the *spontaneous emission factor*, defined as the percentage of the total spontaneous emission coupled into the lasing mode. As indicated in Appendix 4, β_{sp} is just the reciprocal of the number of available optical modes in the bandwidth of the spontaneous emission for uniform coupling to all modes. As also indicated by Eq. (2.11), in the absence of generation terms, the photons decay exponentially with a decay constant of τ_p. Again, this is really the definition of τ_p. That is, for no sources the photon density decays as,

$$N_p(t) = N_p(0) e^{-t/\tau_p}. \tag{2.12}$$

Equations (2.4) and (2.11) are two coupled equations that can be solved for the steady-state and dynamic responses of a diode laser. However, in their present form, several terms still need to be written explicitly in terms of N and N_p before such solutions are possible. First, we shall consider R_{st}.

R_{st} represents the photon-stimulated net electron–hole recombination that generates more photons. This is a *gain* process for photons. As illustrated in Fig. 1.6 and discussed more fully in Appendix 6, the net effect of the upward and downward electronic transitions, corresponding to stimulated absorption and emission of photons, respectively, are included. In Fig. 2.5 we show the growth of a photon density from an incoming value of N_p to an exiting value of $N_p + \Delta N_p$ as it passes through a small length, Δz, of active region. Without loss of generality, but for simplicity, we assume full overlap between the active region and the photon field (i.e., $\Gamma = 1$). As shown, we can also describe this growth in terms of an incremental *gain per unit length*, g, by

$$N_p + \Delta N_p = N_p e^{g \Delta z}. \tag{2.13}$$

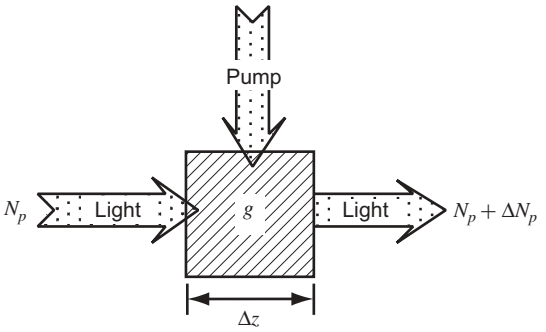


FIGURE 2.5: Definition of gain in terms of the increase in photon number across a small segment of gain material.

If Δz is sufficiently small, $\exp(g\Delta z) \approx (1 + g\Delta z)$. Also, using the fact that $\Delta z = v_g \Delta t$, where v_g is the group velocity, we find that, $\Delta N_p = N_p g v_g \Delta t$. That is, the generation term for dN_p/dt is given by

$$\left(\frac{dN_p}{dt}\right)_{gen} = R_{st} = \frac{\Delta N_p}{\Delta t} = v_g g N_p. \tag{2.14}$$

Put another way, we simply define the incremental gain, g, as the exponential growth rate with distance in the photon density (neglecting any losses) due to stimulated recombination; that is,

$$N_p(z) = N_p(0)e^{gz}. \tag{2.15}$$

Then, substituting, $z = v_g t$, and taking the time derivative we again get the same result as Eq. (2.14), again assuming $\Gamma = 1$ for simplicity.

Thus, we can now rewrite the carrier and photon density rate equations for values in the active region,

$$\frac{dN}{dt} = \frac{\eta_i I}{qV} - \frac{N}{\tau} - v_g g N_p, \tag{2.16}$$

$$\frac{dN_p}{dt} = \Gamma v_g g N_p + \Gamma \beta_{sp} R_{sp} - \frac{N_p}{\tau_p}. \tag{2.17}$$

Of course, we still have not made all the substitutions necessary to directly solve the two equations simultaneously. In Appendix 6 it is suggested that the gain can be expressed as a linear function of carrier density about some bias point, at least under small-signal conditions.

However, a logarithmic function fits the gain better over a wider range of N, as we shall introduce in Section 2.6 and discuss in detail in Chapter 4. Of course, we also know that N/τ can be replaced by the polynomial $AN + BN^2 + CN^3$,

where the terms estimate defect, spontaneous (R_{sp}), and Auger recombination, respectively. We also have the definitions of the constants η_i, Γ, and β_{sp}, and in the next section, we will derive the photon lifetime, τ_p, in terms of the physical cavity losses. Nevertheless, we shall leave the rate equations in the general form of Eqs. (2.16) and (2.17) for future reference.

2.5 THRESHOLD OR STEADY-STATE GAIN IN LASERS

In Section 2.4, we characterized the cavity loss by a phenomenological photon decay constant or lifetime, τ_p, and one can express the gain necessary to overcome losses, and thus reach threshold, from Eq. (2.17) by assuming steady-state conditions (i.e., $dN_p/dt = 0$), and solving for this steady-state or *threshold gain*, g_{th}, where the generation term just equals the recombination term for photons. If we assume that only a small fraction of the spontaneous emission is coupled into the mode (i.e., β_{sp} is quite small), then the second term can be neglected, and we are left with the solution,

$$\Gamma g_{th} = \frac{1}{v_g \tau_p}. \tag{2.18}$$

The product, Γg_{th}, is referred to as the *threshold modal gain* because it now represents the net gain required for the mode as a whole, and it is the mode as a whole that experiences the cavity loss. Now in this section, we wish to explicitly express τ_p in terms of the individual losses associated with optical propagation along the cavity and the cavity mirrors. We shall also attempt to be a little more rigorous with respect to the confinement factor, Γ, by introducing modal averages.

As shown in Appendix 3 and discussed in Chapter 1, the optical energy of a modern diode laser propagates in a dielectric waveguide mode, which is confined both transversely and laterally as defined by a normalized transverse electric field profile, $U(x, y)$. In the axial direction this mode propagates as $\exp(-j\tilde{\beta}z)$, where $\tilde{\beta}$ is the complex propagation constant, which includes any loss or gain. Thus, the time- and space-varying electric field can be written as

$$\mathscr{E} = \hat{\mathbf{e}}_y E_0 U(x, y) e^{j(\omega t - \tilde{\beta} z)}, \tag{2.19}$$

where $\hat{\mathbf{e}}_y$ is the unit vector indicating TE polarization and E_0 is the magnitude of the field. The complex propagation constant, $\tilde{\beta}$, includes the incremental *transverse modal gain*, $\langle g \rangle_{xy}$ and *internal modal loss*, $\langle \alpha_i \rangle_{xy}$. That is,

$$\tilde{\beta} = \beta + j\beta_i = \beta + \frac{j}{2}(\langle g \rangle_{xy} - \langle \alpha_i \rangle_{xy}), \tag{2.20}$$

where the real part of $\tilde{\beta}$; $\beta = 2\pi \bar{n}/\lambda$, and $\bar{n}$ is an effective index of refraction for the mode, also defined in Appendix 3. As shown in Appendix 5, the transverse *modal*

gain, $\langle g \rangle_{xy}$, and loss, $\langle \alpha_i \rangle_{xy}$, are found from weighted averages of the gain and loss, respectively, across the mode shape, $U(x, y)$. Both are related to power; thus, the factor of $\frac{1}{2}$ in this equation for the amplitude propagation coefficient. From Appendix 5, we can let $\langle g \rangle_{xy} = \Gamma_{xy} g$, where Γ_{xy} is the transverse confinement factor, if $g(x, y)$ is constant across the active region and zero elsewhere, so that it can be removed from the integral. This is generally valid for in-plane lasers, but not as good for VCSELs. Also, for notational convenience, we shall let $\langle \alpha_i \rangle_{xy} = \alpha_i$. It is perhaps worth reiteratering that the gain, g, can have both positive and negative values as a function of pumping level, and it is due to the stimulated band-to-band carrier transitions involved in R_{st}; α_i on the other hand, encompasses all other "passive" scattering and absorptive losses the optical mode might experience that are not included in R_{st}, and it has only a second-order dependence on carrier density due to "free-carrier absorption" as the material becomes more conductive.

As illustrated in Fig. 2.6, most laser cavities can be divided into two general sections: an active section of length L_a and a passive section of length L_p. Also, g and α_i will clearly be different in these two sections. In the passive section, by definition $g = 0$, and α_i can be given a second subscript to designate its location. The propagating mode is reflected by end mirrors, which have amplitude reflection coefficients of r_1 and r_2, respectively, to provide a resonant cavity. The amount transmitted is potentially useful output.

For a mode of the laser to reach threshold, the gain in the active section must be increased to the point where all the propagation and mirror losses are compensated, so that the electric field exactly replicates itself after one round trip in the cavity. Equivalently, we can unravel the round trip to lie along the z-axis and require that $\mathscr{E}(z = 2L) = \mathscr{E}(z = 0)$, provided we insert the mode reflection coefficients at $z = 0$ and $z = L$. As a consequence of inserting these boundaries into Eq. (2.19), we obtain

$$r_1 r_2 e^{-2j\tilde{\beta}_{ath} L_a} e^{-2j\tilde{\beta}_{pth} L_p} = 1. \tag{2.21}$$

The subscript *th* denotes that this characteristic equation only defines the threshold value of $\tilde{\beta}$. (In Chapter 3 we shall take a more basic approach to obtain this same

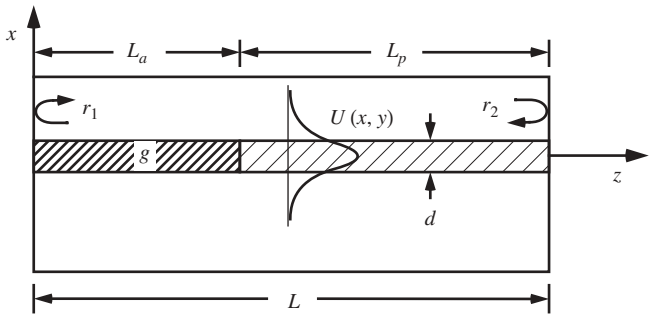


FIGURE 2.6: Generic laser cavity cross-section showing active and passive sections (no impedance discontinuity assumed) and the guided-mode profile.

characteristic equation). Using Eq. (2.20), we can break the complex Eq. (2.21) into two equations for its magnitude and phase. For the magnitude,

$$r_1 r_2 e^{(\Gamma_{xy} g_{th} - \alpha_{ia}) L_a} e^{-\alpha_{ip} L_p} = 1, \tag{2.22}$$

where we have chosen reference planes to make the mirror reflectivities real. Solving for $\Gamma_{xy} g_{th} L_a$ we obtain

$$\Gamma_{xy} g_{th} L_a = \alpha_{ia} L_a + \alpha_{ip} L_p + \ln\left(\frac{1}{R}\right), \tag{2.23}$$

where the mean mirror intensity reflection coefficient, $R = r_1 r_2$. For cleaved facet lasers based upon GaAs or InP, $R \sim 0.32$. Dividing Eq. (2.23) by the total cavity length, L, we realize that $\Gamma_{xy} L_a / L \approx \Gamma_{xy} \Gamma_z = \Gamma$ (exact for $L_a \gg \lambda$), and define the average internal loss as

$$\frac{(\alpha_{ia} L_a + \alpha_{ip} L_p)}{L} = \langle \alpha_i \rangle. \tag{2.24}$$

We then have

$$\langle g_{th} \rangle = \Gamma g_{th} = \langle \alpha_i \rangle + \frac{1}{L} \ln\left(\frac{1}{R}\right). \tag{2.25}$$

For convenience, the mirror loss term is sometimes abbreviated as α_m,

$$\alpha_m \equiv \frac{1}{L} \ln\left(\frac{1}{R}\right). \tag{2.26}$$

Finally, from Eqs. (2.18) and (2.25), the photon decay lifetime is given by,

$$\frac{1}{\tau_p} = \frac{1}{\tau_i} + \frac{1}{\tau_m} = v_g (\langle \alpha_i \rangle + \alpha_m). \tag{2.27}$$

Thus, we can also write

$$\Gamma g_{th} = \langle \alpha_i \rangle + \alpha_m = \frac{1}{v_g \tau_p}. \tag{2.28}$$

As noted in Appendix 5, if the averaging is initially done over the whole volume, the three-dimensional modal gain and loss used in Eqs. (2.25) and (2.28) are obtained directly. However, this obscures the physics of the recirculating mode in the cavity, so we have chosen to show the longitudinal weighting separately here. As suggested earlier, the general $\langle g \rangle_{th}$ form is always valid, but the Γg_{th} form only should be used for in-plane lasers, where the gain can be assumed to be constant over the active volume. The limitation that $L_a \gg \lambda$ for $L_a / L = \Gamma_z$ listed earlier is also discussed in Appendix 5. As explained there, the axial averaging

of gain and loss must also use a weighted average over the axial standing wave pattern in the general case. In fact, for very short active regions ($L_a \ll \lambda$), such as in many vertical-cavity lasers, it is possible for $\Gamma_z \approx 2L_a/L$, if the active segment is placed at the peak of the electric-field standing wave (see Appendix 5).

It is most important to realize that Eqs. (2.25) and (2.28) give only the cavity loss parameters necessary to calculate the threshold gain. That is, the calculation is really a calculation of net cavity loss, and not a fundamental calculation of gain at all. The equations have *nothing to do with the stimulated emission physics*, which determines what the gain is for a given injection current. This physics is briefly summarized in Appendix 6, and it will be the primary subject of Chapter 4. So when g_{th} appears in various equations in the the future, one should immediately "see" net cavity loss, which just happens to be the value the gain needed to approach lasing threshold.

For the phase part of Eq. (2.21), $\exp(2j\beta_{tha}L_a)\exp(2j\beta_{thp}L_p) = 1$, requires that $\beta_{tha}L_a + \beta_{thp}L_p = m\pi$, which gives a condition on the modal wavelength,

$$\lambda_{th} = \frac{2}{m}[\bar{n}_a L_a + \bar{n}_p L_p], \tag{2.29}$$

where m is the longitudinal mode number. It should also be realized that $\bar{n}$ varies with wavelength ($\partial\bar{n}/\partial\lambda$, dispersion), and it generally is also dependent on the carrier density ($\partial\bar{n}/\partial N$, plasma loading). Thus, when making computations, these dependences must be included. That is, to determine $\bar{n}$ at a wavelength $\lambda = \lambda_0 + \Delta\lambda$ and a carrier density, $N = N_0 + \Delta N$, we use

$$\bar{n}(\lambda, N) = \bar{n}(\lambda_0, N_0) + \frac{\partial\bar{n}}{\partial\lambda}\Delta\lambda + \frac{\partial\bar{n}}{\partial N}\Delta N. \tag{2.30}$$

Typically, $\partial\bar{n}/\partial\lambda \sim -1\ \mu\text{m}^{-1}$, and $\partial\bar{n}/\partial N \approx \Gamma_{xy}\partial n_A/\partial N \sim -\Gamma_{xy}10^{-20}\ \text{cm}^3$, where n_A is the index in the active region. Rearranging Eq. (2.29) we take a total derivative of both sides, assuming that ΔN is constant,

$$d(m\lambda) = d(2[\bar{n}_a L_a + \bar{n}_p L_p]) => \lambda \cdot dm + m \cdot d\lambda = 2\left(\frac{\partial\bar{n}_a}{\partial\lambda}L_a + \frac{\partial\bar{n}_p}{\partial\lambda}L_p\right)d\lambda. \tag{2.31}$$

Further manipulating this expression, noting that $dm = \pm 1$, and using Eq. (2.29) to eliminate m, we can find the wavelength separation between two modes, m and $m + 1$, to be

$$d\lambda = \frac{\lambda^2}{2(\bar{n}_{ga}L_a + \bar{n}_{gp}L_p)}, \tag{2.32}$$

where the group effective index for the jth section $\bar{n}_{gj} = \bar{n}_j - \lambda(\partial\bar{n}/\partial\lambda) = \overline{n}_j + \omega(\partial\overline{n}/\partial\omega)$. The group index in semiconductors is typically 20–30% larger than the index of refraction, depending on the specific wavelength relative to the band edge. From experiments, the values of $\bar{n}_g$ for the active sections of various in-plane

TABLE 2.1: Estimated Values of Group Indices for Different Wavelength Ranges, to be used for Problems in this Textbook

λ (nm)	$\bar{n}_g$
400	3
850–980	4.5
1300	4
1550	3.8

lasers have been estimated. A set of estimated group index values, for different laser operating wavelength ranges, is given in Table 2.1.

Finally, it is important to note that the steady-state gain in a laser operating *above* threshold must also equal its threshold value as given by Eq. (2.25). That is, *in a laser cavity*,

$$g(I > I_{th}) = g_{th}. \qquad \text{(steady state)} \tag{2.33}$$

If the gain were higher than g_{th}, then the field amplitude would continue to increase without bound, and this clearly cannot exist in the steady state. Furthermore, because the gain is monotonically related to the carrier density, this implies that the carrier density must also *clamp* at its threshold value. That is,

$$N(I > I_{th}) = N_{th}. \qquad \text{(steady state)} \tag{2.34}$$

In fact, what happens when the current is increased to a value above threshold is that the carrier density and gain initially (on the order of a nanosecond) increase to values above their threshold levels, and the photon density grows. But then, the stimulated recombination term R_{st} also increases, reducing the carrier density and gain until a new steady-state dynamic balance is struck where Eqs. (2.33) and (2.34) are again satisfied. Put another way, the stimulated recombination term in Eq. (2.15) uses up all additional carrier injection above threshold. In terms of our reservoir analogy depicted in Fig. 2.2, the water level has reached the spillway and any further increase in input goes over the spillway without increasing the water depth. Of course, the spillway represents simulated recombination. Fig. 2.7 shows the analogy in this case.

Figure 2.8 summarizes this carrier clamping effect in a laser cavity. The physics of the g versus N curve never changes. The feedback effect causes the carrier density to clamp, in order to keep the gain at its threshold value.

Example 2.1 An active-passive cleaved laser chip, operating at 1.55 μm, consists of a multiple quantum-well InP/InGaAsP active section, whose length is 500 μm, and internal loss $\alpha_{ia} = 15\ \text{cm}^{-1}$, and a passive section, whose length is 300 μm

and internal loss $\alpha_{ip} = 10\ \text{cm}^{-1}$. The active region contains 4–3 nm wide strained InGaAs quantum wells.

Problem: (1) Determine the threshold modal gain of this laser. (2) Determine the cavity mode spacing for this laser.

Solution: To calculate the threshold modal gain, we will use the expression from Eq. (2.25). For this, we need to determine the average loss of the cavity, $\langle \alpha_i \rangle$, and the mirror loss, α_m. From Eq. (2.24),

$$\langle \alpha_i \rangle = \frac{(\alpha_{ia} L_a + \alpha_{ip} L_p)}{L_a + L_p} = \frac{15 \cdot 10^{-4} \cdot 500 + 10 \cdot 10^{-4} \cdot 300}{(500 + 300) 10^{-4}\ \text{cm}} = 13.125\ \text{cm}^{-1}.$$

From Eq. (2.26), using the fact that we are dealing with an InP based laser with cleaved facets, where $R \sim 0.32$,

$$\alpha_m = \frac{1}{L_a + L_p} \ln\left(\frac{1}{R}\right) = \frac{1}{800 \cdot 10^{-4}\ \text{cm}} \ln\left(\frac{1}{0.32}\right) = 14.243\ \text{cm}^{-1}$$

Then, the threshold modal gain is given by Eq. (2.25),

$$\Gamma g_{th} = \langle \alpha_i \rangle + \alpha_m = 27.368\ \text{cm}^{-1}$$

Cavity mode spacing can be determined from Eq. (2.32), using the fact that $n_{ga} = n_{gp} = 3.8$ for this InGaAsP based laser, and the lasing wavelength of 1.55 μm,

$$d\lambda = \frac{\lambda^2}{2(\bar{n}_{ga} L_a + \bar{n}_{gp} L_p)} = \frac{1.55^2}{2(3.8 \cdot 500 + 3.8 \cdot 300)}\ \mu\text{m} = 0.395\ \text{nm}.$$

2.6 THRESHOLD CURRENT AND POWER OUT VERSUS CURRENT

2.6.1 Basic P–I Characteristics

Although the rate Eqs. (2.16) and (2.17) are valid both above and below threshold, and they can be solved in parametric form, we shall piece together a below-threshold LED characteristic with an above-threshold laser characteristic to construct the power out versus current in for a diode laser with useful analytic expressions. The LED part is already largely complete with Eq. (2.8). Thus, we shall here concentrate on the above-threshold laser part. The first step is to use the below-threshold steady-state carrier rate equation, Eq. (2.5) *almost* at threshold. That is,

$$\frac{\eta_i I_{th}}{qV} = (R_{sp} + R_{nr} + R_l)_{th} = \frac{N_{th}}{\tau}. \tag{2.35}$$

Then, recognizing that $(R_{sp} + R_{nr} + R_l) = AN + BN^2 + CN^3$ depends monotonically on N, we observe from Eq. (2.34) that above threshold $(R_{sp} + R_{nr} + R_l)$

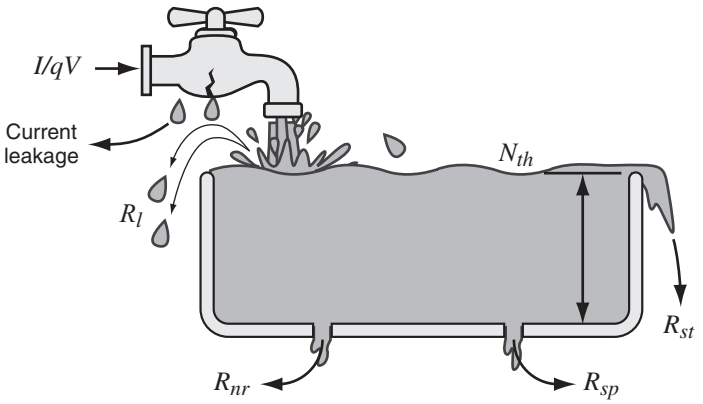


FIGURE 2.7: Reservoir analogy above threshold where water level has risen to the spillway so that an increased input results in an increased output (R_{st}) but no increase in carrier density (water level). The flows R_{nr} and R_{sp} do not change above threshold.

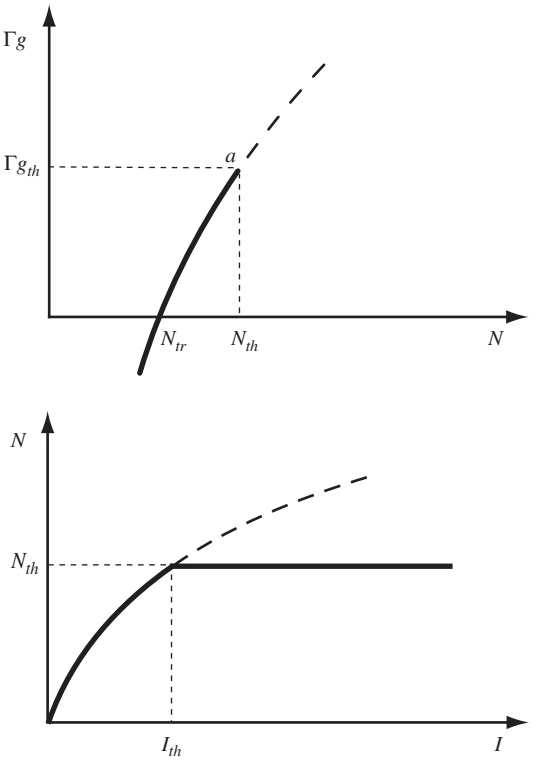


FIGURE 2.8: Gain versus carrier density and carrier density versus input current. The carrier density clamps at threshold causing the gain to clamp also.

will also clamp at its threshold value, given by Eq. (2.35). Thus, we can substitute Eq. (2.35) into the carrier rate equation, Eq. (2.16), to obtain a new above-threshold carrier rate equation,

$$\frac{dN}{dt} = \eta_i \frac{(I - I_{th})}{qV} - v_g g N_p, \quad (I > I_{th}) \tag{2.36}$$

where we have assumed η_i is not a function of current above threshold. From Eq. (2.36) we can now calculate a steady-state photon density above threshold where $g = g_{th}$. That is,

$$N_p = \frac{\eta_i (I - I_{th})}{q v_g g_{th} V}. \quad \text{(steady state)} \tag{2.37}$$

Now with some relatively straightforward substitutions, we can calculate the power out, because it must be proportional to N_p. To obtain the power out, we first construct the *stored optical energy in the cavity*, E_{os}, by multiplying the photon density, N_p, by the energy per photon, $h\nu$, and the cavity volume, V_p. That is, $E_{os} = N_p h\nu V_p$. Then, we multiply this by the *energy loss rate through the mirrors*, $v_g \alpha_m = 1/\tau_m$, to get the optical power output from the mirrors,

$$P_0 = v_g \alpha_m N_p h\nu V_p. \tag{2.38}$$

Substituting from Eqs. (2.37) and (2.28), and using $\Gamma = V/V_p$, in Eq. (2.38),

$$P_0 = \eta_i \left(\frac{\alpha_m}{\langle \alpha_i \rangle + \alpha_m} \right) \frac{h\nu}{q} (I - I_{th}). \quad (I > I_{th}) \tag{2.39}$$

Now, by defining

$$\eta_d = \frac{\eta_i \alpha_m}{\langle \alpha_i \rangle + \alpha_m}, \tag{2.40}$$

we can simplify Eq. (2.39) to be

$$P_0 = \eta_d \frac{h\nu}{q} (I - I_{th}). \quad (I > I_{th}) \tag{2.41}$$

Equation (2.41) represents the total power out of both mirrors. If the mirrors have equal reflectivity, then exactly half will be emitted out of each. If one is totally reflecting, then all will be emitted out the other. On the other hand, if the mirrors have partial but unequal reflectivity, the fraction emitted from each is a nontrivial function, which we shall derive in Chapter 3. Equation (2.41) also shows that the power out above threshold is a linear function of the current above threshold. This is true regardless of our assumptions about the form of the gain–current relationship or the nature of the nonradiative recombination mechanisms. The assumptions necessary for this P–I linearity are that the gain–current relationship,

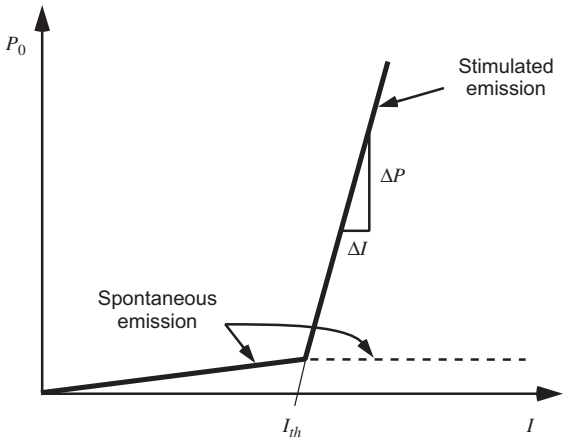


FIGURE 2.9: Illustration of output power versus current for a diode laser. Below threshold only spontaneous emission is important; above threshold the stimulated emission power increases while the spontaneous emission is clamped at its threshold value.

the internal efficiency, the confinement factor, and the cavity losses remain constant. As shown in Appendix 5, by confinement factor, we really mean that the *modal gain* must remain constant.

To determine what we should call η_d, we can compare the calculated result of Eq. (2.41) to a measurement. Postulating that it might be related to a quantum efficiency, we calculate a differential quantum efficiency, defined as the number of photons out per electron in from a measured P–I characteristic. As shown in Fig. 2.9, the differential quantum efficiency would be found by measuring the slope $[\Delta P_0/\Delta I]$ in watts/amp above threshold (including output from both ends) and then multiplying this number by $[q/h\nu]$ in Coulombs/joule to get an empirical number of photons per electron equal to $[\Delta P_0/\Delta I][q/h\nu]$. Now, if we take the derivative with respect to current of Eq. (2.41), and solve for η_d, we get the same result. This shows that η_d is indeed *the differential quantum efficiency*. To repeat then,

$$\eta_d = \left[\frac{q}{h\nu}\right]\frac{dP_0}{dI}. \quad (I > I_{th}) \tag{2.42}$$

The region in Fig. 2.9 below threshold $(I < I_{th})$ can be approximated by neglecting the stimulated emission term in Eq. (2.17) and solving for N_p, again under steady-state conditions. In this case we find that

$$N_p = \Gamma\beta_{sp}R_{sp}\tau_p. \quad (I < I_{th}) \tag{2.43}$$

Using Eqs. (2.43), (2.25), (2.6), and (2.5) in Eq. (2.38), we get the spontaneous emission into the laser mode as

$$P_0(I < I_{th}) = \eta_r\eta_i\left(\frac{\alpha_m}{\langle\alpha_i\rangle + \alpha_m}\right)\frac{h\nu}{q}\beta_{sp}I. \tag{2.44}$$

Comparing this to the LED expression of Eq. (2.8) shows that $\eta_c = \alpha_m \beta_{sp}/(\langle \alpha_i \rangle + \alpha_m)$ as might have been expected.

At threshold the spontaneous emission clamps as the carrier density clamps because R_{sp} depends on N. Thus, as the current is increased above threshold, the spontaneous emission *noise* remains constant at the value of Eq. (2.44) with $I = I_{th}$, while the coherent stimulated emission power grows according to Eq. (2.39). As we shall find in Chapter 5, this results in a gradual reduction in the linewidth of the output wavelength as the power is increased.

2.6.2 Gain Models and Their Use in Designing Lasers

Equation (2.41) gives the output power in terms of the additional current applied above threshold. The proportionality factors are constants involving the cavity losses, the lasing wavelength, and the internal efficiency. To design lasers for minimum current at a given output power, we also need an analytic expression for the threshold current. So far, we have introduced the threshold modal gain dependence in terms of the cavity losses, Eq. (2.28), and we have suggested that the gain can be related to the carrier density by a logarithmic (Chapter 4) relationship. The threshold current is also related to the threshold carrier density via the recombination rates, which can be expressed as a polynomial in N (e.g., Eq. (2.35)).

In Chapter 4 it will be shown that the gain versus carrier density can be well approximated by a simple three-parameter logarithmic formula.

$$g = g'_{0N} \ln \frac{N + N_s}{N_{tr} + N_s}. \tag{2.45}$$

In this approximation, g'_{0N} is an empirical gain coefficient, N_{tr} is the transparency carrier density, and N_s is a shift to force the natural logarithm to be finite at $N = 0$ such that the gain equals the unpumped absorption due to the band-to-band transitions. However, if we restrict our attention to positive gains, $g \geq 0$, Eq. (2.45) can be further approximated by the simple two-parameter expression,

$$g = g_{0N} \ln \frac{N}{N_{tr}}, \quad (g \geq 0) \tag{2.46}$$

provided that we use a new gain coefficient g_{0N}. In this case the differential gain, $\partial g/\partial N \equiv a$, which will show up in a number of important relationships, is given by

$$\frac{\partial g}{\partial N} \equiv a = \frac{g_{0N}}{N}. \tag{2.47}$$

Of course, in a laser above threshold, $N = N_{th}$. Generally, N_{tr} and $\partial g/\partial N$ will be quite different for bulk, quantum-well, and strained-layer quantum-well active regions. The value of this parameter is the basis for many of the arguments for and against certain of these structures. Figure 2.10 illustrates schematically the modal gain versus carrier density with some of the relevant parameters labeled. The value

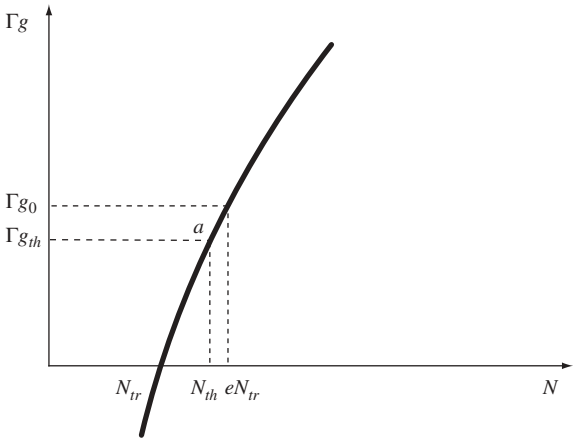


FIGURE 2.10: Schematic illustration of modal gain versus injected carrier density with values labeled from the two-parameter logarithmic fit of Eq. (2.46).

for differential gain a is determined by the slope of the tangent to the gain curve at the threshold point, Fig. 2.10. As with any other system, local gain–carrier relationship can be linearized at point a.

Fitting Eq. (2.46) to numerical gain plots to be found in Chapter 4, a strained 80 Å InGaAs/GaAs quantum well yields $g_{0N} \sim 2100\ \text{cm}^{-1}$ and $N_{tr} \sim 1.8 \times 10^{18}\ \text{cm}^{-3}$; and an 80 Å GaAs quantum well gives $g_{0N} \sim 2400\ \text{cm}^{-1}$ and $N_{tr} \sim 2.6 \times 10^{18}\ \text{cm}^{-3}$. For InP substrate cases, a strained 30 Å InGaAs/InP well gives $g_{0N} \sim 4000\ \text{cm}^{-1}$ and $N_{tr} \sim 3.3 \times 10^{18}\ \text{cm}^{-3}$; and an unstrained 60 Å InGaAs quantum well gives $g_{0N} \sim 1800\ \text{cm}^{-1}$ and $N_{tr} \sim 2.2 \times 10^{18}\ \text{cm}^{-3}$.

Now we can combine Eqs. (2.28) and (2.46) to get the threshold carrier density,

$$N_{th} = N_{tr} e^{g_{th}/g_{0N}} = N_{tr} e^{(\langle \alpha_i \rangle + \alpha_m)/\Gamma g_{0N}}. \tag{2.48}$$

Using the polynomial fit for the recombination rates in Eq. (2.35), and recognizing that for the best laser material the recombination at threshold is dominated by spontaneous recombination, we have, $I_{th} \cong BN_{th}^2 qV/\eta_i$. Thus,

$$I_{th} \cong \frac{qVBN_{tr}^2}{\eta_i} e^{2(\langle \alpha_i \rangle + \alpha_m)/\Gamma g_{0N}}, \tag{2.49}$$

where for most III–Vs of interest the bimolecular recombination coefficient, $B \sim 10^{-10}\ \text{cm}^3/\text{s}$.

Equations (2.41) and (2.49) can now be used for a closed-form expression of output power versus applied current. However, because we are usually trying to minimize the current needed for a given required power from one mirror, P_{01}, we solve for I.

$$I \cong \frac{qP_{01}(\langle \alpha_i \rangle + \alpha_m)}{F_1 \eta_i h\nu \alpha_m} + \frac{qVBN_{tr}^2}{\eta_i} e^{2(\langle \alpha_i \rangle + \alpha_m)/\Gamma g_{0N}}, \tag{2.50}$$

where the first term is the additional current required above threshold to obtain power P_{01} from Eq. (2.41), and the second term is the threshold current, or Eq. (2.49). The factor F_1 is the fraction of the total output power coming out of mirror 1. An exact analytic formula for it will be derived in Chapter 3, but clearly, it is one-half for equal mirrors and unity if mirror 2 has a reflectivity of one.

Equations (2.49) and (2.50) give reasonable accuracy in simple analytic expressions, which correctly show that it is always desirable to reduce the transparency value and increase the differential gain of the active material. Both points argue in favor of using quantum-well, especially strained-layer quantum-well, active regions. Relative to the cavity design, the equations also indicate that it is desirable to reduce the cavity loss ($\langle\alpha_i\rangle + \alpha_m$) and volume, V, subject to retaining a reasonably large confinement factor, Γ. Thus, the merits of using vertical-cavity surface emitters or short-cavity in-plane lasers with coated facets are also suggested.

Figure 2.11 plots Eq. (2.50) for a typical set of assumed parameters to quantify the relationship between mirror reflectivity and cavity volume for a specific set of parameters. Note that for any given values of internal loss and power out there is a trough in required drive current that slopes monotonically downward as R approaches unity and L tends to zero. In practice, this $R = 1, L = 0$ minimum cannot be approached too closely because high current densities lead to device heating, and the deleterious effects of heating on the device parameters are not included here. In fact, the trade-off between the desire to minimize device volume and the need to maintain reasonable current densities and thermal impedances generally forces laser designs to be considerably larger than the plots in Fig. 2.11 tend to suggest. For higher powers out or internal losses, the bottom of the current trough moves to smaller L_s and R_s. Figure 2.11b isolates two specific lengths from part (a) and adds other power levels to illustrate this point. Further discussion of optimum laser design is left until Chapter 8 and Appendix 17. Due to the exponential dependence on g_{th}/g_{0N} in Eqs. (2.49) and (2.50), it may be beneficial to use more than one quantum well to increase Γ in a quantum-well laser. This dependence is a result of the saturation of the gain as the carrier density is increased to nearly fill the lowest set of states, as discussed in Appendix 6 and Chapter 4. Thus, by distributing the carriers over N_w wells, the gain within each well is reduced by less than N_w times, but the modal gain is still multiplied by nearly N_w times this value. For such multiple quantum-well (MQW) lasers Eqs. (2.49) and (2.50) are still valid but one must be sure to multiply the single-well confinement factor, Γ_1, and volume, V_1, by the number of wells, N_w. That is, for an MQW laser, from Eq. (2.49) or the second term in Eq. (2.50), one can explicitly write

$$I_{th_{MQW}} \cong \frac{qN_wV_1BN_{tr}^2}{\eta_i} e^{2(\langle\alpha_i\rangle+\alpha_m)/N_w\Gamma_1 g_{0N}}. \tag{2.51}$$

Here, we have assumed a separate confinement waveguide, so that the optical mode does not change significantly as more wells are added. Also, the number of wells is limited to the number that can be placed near the maximum of the optical mode. The optimum number of wells is the number that minimizes Eq. (2.51), neglecting

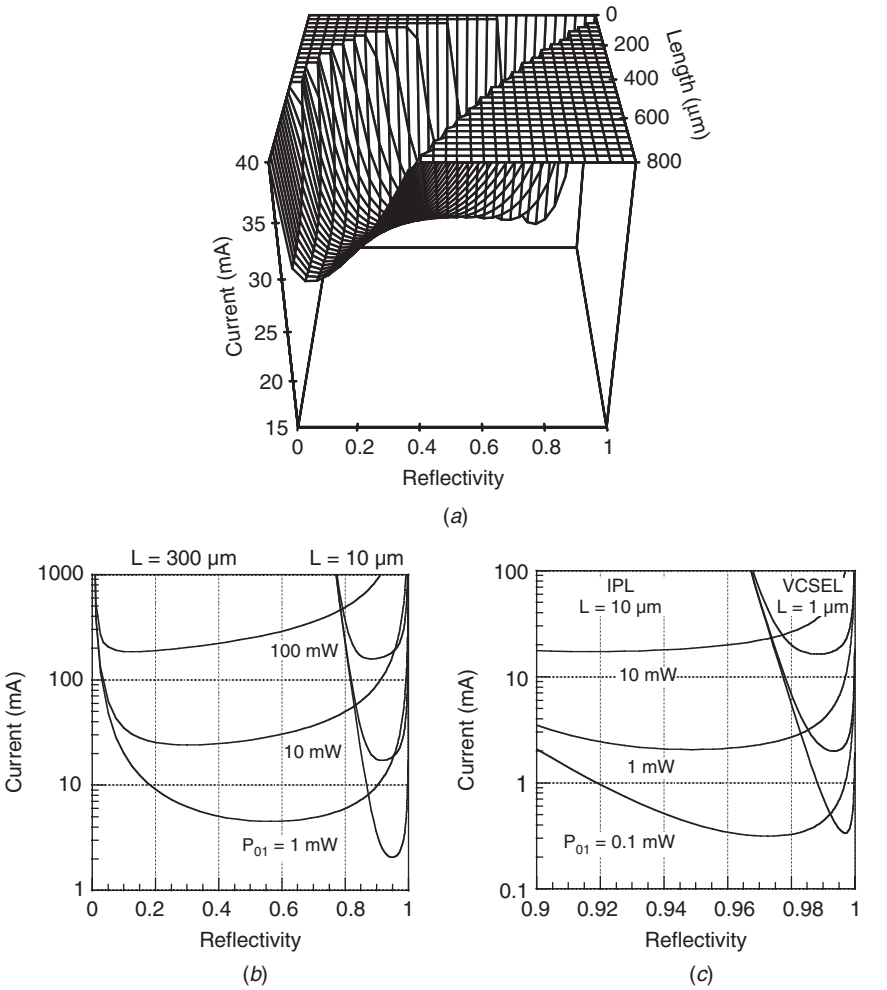


FIGURE 2.11: (a) Three-dimensional plot of current required for 10 mW of power out for lasers of variable length, L, and mean mirror reflectivity, R. Other parameters are: $w = 2.0$ μm; $d = 10$ μm; $\langle \alpha_i \rangle = 20$ cm^{-1}; $\Gamma g_0 = 50$ cm^{-1}; $F_1 = 0.5$; $N_{tr} = 2 \times 10^{18}$ cm^{-3}; $\eta_i = 1$; $\Gamma_{xy} = 0.033$; $\Gamma_z = 1$. (b) I versus R for two lengths from (a) with two additional power levels shown. (c) Comparison of small in-plane laser and VCSEL. $\Gamma_{xy} = 1$; $\Gamma_z = 0.06$ for VCSEL. Active area ($A = 20$ μm^2) is the same in both. The VCSEL uses three quantum wells rather than one, tripling the active volume. This reduces the value of the optimum R for the VCSEL and broadens the current minimum. However, the VCSEL minimum current is nearly the same as the in-plane laser here, because three wells are used.

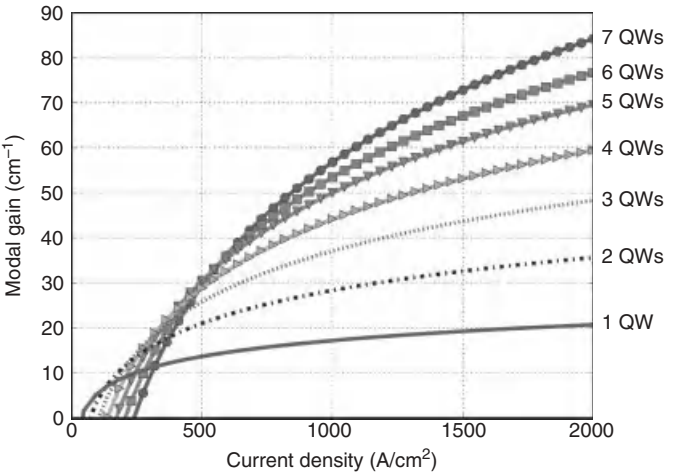


FIGURE 2.12: Example plots of modal gain versus current density for increasing numbers of quantum wells in an SCH-MQW laser. For the calculations, we assumed a quantum well active material used in Examples 2.2 and 2.4, and that the active region was offset relative to a 300 nm thick 1.4Q waveguide. Confinement factor Γ was computed using numerical mode solving.

nonradiative recombination. With the increased confinement factor we also see that higher powers can be obtained efficiently without moving too far up the gain curve. Perhaps a more instructive way of visualizing the optimal design of MQW lasers is by way of the graphical example in Fig. 2.12, which gives a family of modal gain versus current density plots as the number of quantum wells is increased. From this for a given threshold modal gain (net modal loss), one can determine the minimum current (and number of quantum wells) needed to reach threshold. Sometimes one might also wish to have a large differential gain, a, or have some margin for operation at higher temperatures, so moving to the next higher curve than this minimum might be more desirable.

Example 2.2 The active material from Example 2.1 is used to fabricate a 500-μm-long and 50-μm-wide broad area laser. The material gain for the quantum wells can be approximated using Eq. (2.46), with $g_{0N} = 1207.29$ cm^{-1} and $N_{tr} = 1.2284 \cdot 10^{18}$ cm^{-3}. The transverse confinement factor Γ_1 is 1% per well, and $\eta_i = 80\%$.

Problem: (1) Calculate the threshold modal gain for this laser. (2) Determine the threshold current for this laser. (3) Determine the threshold current for a laser of the same dimensions, but with 7 quantum wells, assuming that the internal modal loss α_{ia} remains the same.

Solution: To calculate the threshold current, we first need to calculate the threshold carrier density using the Eq. (2.48). Because we are dealing with a long wavelength InP/InGaAsP material system, both spontaneous, Eq. (2.51), and

nonradiative, Eq. (2.52) threshold current components need to be included. To compute the threshold carrier density, we need to determine the threshold modal gain of this all–active laser,

$$\Gamma g_{th} = \langle \alpha_{ia} \rangle + \frac{1}{L} \ln\left(\frac{1}{R}\right) = 37.79 \text{ cm}^{-1}.$$

The transverse confinement factor is determined by the number of quantum wells,

$$\Gamma = N_w \cdot \Gamma_1 = 0.04.$$

Using Eq. (2.48), threshold carrier density is

$$N_{th} = N_{tr} e^{\frac{g_{th}}{g_{0N}}} = N_{tr} e^{\frac{37.79}{0.04 \cdot 1207.29}} = 2.6865 \cdot 10^{18} \text{ cm}^{-3}.$$

Now we can compute the threshold current, using the volume of the active region $V = L \cdot W \cdot N_w \cdot d = 500 \cdot 50 \cdot 4 \cdot 0.003 \cdot 10^{-12}$ cm^{-3},

$$I_{th} = \frac{qVN_{th}}{\eta_i \tau} \cong \frac{qV}{\eta_i}(BN_{th}^2 + CN_{th}^3) = 47.94 \text{ mA}$$

where $B = 0.3 \cdot 10^{-10}$ cm^3/s and $C = 3 \cdot 10^{-29}$ cm^6/s. If we now have a laser with 7 quantum wells, the confinement factor will be changed,

$$\Gamma_2 = N_w \cdot \Gamma_1 = 0.07.$$

leading to a changed threshold gain $g_{th2} = 539.84$ cm^{-1}, threshold carrier density $N_{th2} = 1.9211 \cdot 10^{18}$ cm^{-3}, volume $V = 500 \cdot 50 \cdot 7 \cdot 0.003 \cdot 10^{-12}$ cm^{-3}, and the threshold current of

$$I_{th} = \frac{qV}{\eta_i}(BN_{th}^2 + CN_{th}^3) = 34.00 \text{ mA}.$$

In the plots of Fig. 2.12 it is assumed that as more identical quantum wells are added within the separate-confinement-heterostructure (SCH) region, the injected current is distributed equally across the wells; thus, for example, it requires twice as much current density to reach transparency for two wells as for one. For only a few wells it can also be assumed, as mentioned earlier, that the confinement factor will increase linearly as the number of wells increase. In short, to a good approximation, successive curves can be generated by simply multiplying both coordinate values by N_w, although in Fig. 2.12 an actual numerical calculation was performed.

If nonradiative recombination is important at threshold, an additional nonradiative threshold current component must be added as outlined earlier. For the long-wavelength InGaAsP/InP materials, nonradiative recombination is known to be very important. In fact, were it not for such recombination, the threshold current densities of lasers using such materials would be lower than those using GaAs

quantum wells, as indicated by the gain parameters listed after Eq. (2.46). If such higher-order nonradiative carrier recombination is important at threshold, one must add another component to the threshold current due to the CN_{th}^3 term in the recombination rate. Then, Eqs. (2.49) and (2.50) should be increased by

$$I_{nrth} = \frac{qVCN_{tr}^3}{\eta_i} e^{3(\langle \alpha_i \rangle + \alpha_m)/\Gamma g_0}, \tag{2.52}$$

where for 1.3 μm InGaAsP material, the Auger coefficient, $C \sim 3 \times 10^{-29}$ cm^6/s, and for 1.55 μm material it is about two or three times larger. The cubic dependence on N_{th} places more importance on reducing the threshold carrier density in this material system. In fact, this additive Auger term dominates Eq. (2.49) for carrier densities above $N_{th} \sim 3 \times 10^{18}$ or 1.5×10^{18} cm^{-3} at 1.3 and 1.55 μm, respectively. This fact focuses more attention on reducing cavity losses, $(\langle \alpha_i \rangle + \alpha_m)$, and maintaining a large confinement factor, Γ. With the use of strained-layer InGaAs/InGaAsP or InGaAs/InGaAlAs quantum wells on InP, a considerable improvement is possible because all the parameters affecting N_{th} move in the right direction. In fact, the Auger coefficient, C, may also be reduced due to the splitting of the valence bands.

2.7 RELAXATION RESONANCE AND FREQUENCY RESPONSE

Chapter 5 will discuss dynamic effects in some detail. Here, we wish to use Eqs. (2.16) and (2.17) to briefly outline the calculation of relaxation resonance frequency and its relationship to laser modulation bandwidth. As shown in Chapter 5, because of gain compression with increasing photon density and possible transport effects, the calculations are a bit oversimplified, particularly with respect to quantum-well structures. However, these simple equations do seem to work well for standard DH structures, and the method of attack for calculating resonance frequency is also instructive for the more complex calculations to follow in Chapter 5.

Consider the application of an above-threshold DC current, I_0, superimposed with a small AC current, I_1, to a diode laser. Then, under steady-state conditions the laser's carrier density and photon density would respond similarly, with some possible harmonics of the drive frequency, ω, that we shall ignore. Using complex frequency domain notation,

$$I = I_0 + I_1 e^{j\omega t}, \tag{2.47a}$$

$$N = N_0 + N_1 e^{j\omega t}, \tag{2.47b}$$

$$N_p = N_{p0} + N_{p1} e^{j\omega t}. \tag{2.47c}$$

Before applying these to Eqs. (2.16) and (2.17), we first rewrite the rate equations using $g = a\Delta N$ for the gain. This is valid because small-signal conditions are assumed, and the gain can be well approximated by a linear

deviation from its steady-state value, g_{th}, over some small range of carrier densities, $\triangle N$, *provided the local slope,* $a = \partial g/\partial N$, at the bias point N_0 is used. We also assume the DC current is sufficiently far above threshold that the spontaneous emission can be neglected. That is,

$$\frac{dN}{dt} = \frac{\eta_i I}{qV} - \frac{N}{\tau} - v_g(g_{th} + a\triangle N)N_p, \tag{2.48}$$

$$\frac{dN_p}{dt} = \Gamma v_g(g_{th} + a\triangle N)N_p - \frac{N_p}{\tau_p}. \tag{2.49}$$

Now, after plugging in Eq. (2.47) for I, N, and N_p, we note that $\triangle N \equiv N_1 \exp(j\omega t)$ and recognize that the DC components satisfy the steady-state versions of Eqs. (2.48) and (2.49) (i.e., with $d/dt \to 0$); and they can be grouped together and set to zero. Next, we recognize that the steady-state gains, g_{th}, can be replaced by $[\Gamma v_g \tau_p]^{-1}$ according to Eq. (2.28). Finally, we delete the second-harmonic terms that involve $e^{j2\omega t}$ and divide out an $e^{j\omega t}$ common factor. Then,

$$j\omega N_1 = \frac{\eta_i I_1}{qV} - \frac{N_1}{\tau} - \frac{N_{p1}}{\Gamma\tau_p} - v_g a N_1 N_{p0}, \tag{2.50}$$

$$j\omega N_{p1} = \Gamma v_g a N_1 N_{p0}. \tag{2.51}$$

With the preceding manipulations we have generated frequency domain equations that can easily be solved for the transfer function $N_{p1}(\omega)/I_1(\omega)$. First, solving for N_1 in Eq. (2.51), we have $N_1 = j\omega N_{p1}/\Gamma v_g a N_{p0}$. Then eliminating N_1 from Eq. (2.50) and using $P_{ac} = v_g \alpha_m N_{p1} h\nu V_p$, we obtain

$$\frac{P_{ac}(\omega)}{I_1(\omega)} = \frac{\eta_i h\nu}{q} \frac{v_g \alpha_m (v_g a N_{p0})}{v_g a N_{p0}/\tau_p - \omega^2 + j\omega[v_g a N_{p0} + 1/\tau]}. \tag{2.52}$$

Now we observe that the first term in the denominator is the square of a natural resonance frequency, so we define,

$$\omega_R^2 \equiv \frac{v_g a N_{p0}}{\tau_p}. \tag{2.53}$$

Using Eqs. (2.28) and (2.40), the transfer function can be written in a more normalized form:

$$\frac{P_{ac}(\omega)}{I_1(\omega)} = \frac{\eta_d h\nu/q}{1 - (\omega/\omega_R)^2 + j(\omega/\omega_R)[\omega_R \tau_p + 1/\omega_R \tau]}. \tag{2.54}$$

For sufficiently low modulation frequencies, the denominator reduces to one and Eq. (2.54) reduces to the AC equivalent of Eq. (2.41). For higher modulation frequencies, the $1 - (\omega/\omega_R)^2$ term in the denominator creates a strong resonance in the response. Figure 2.13 illustrates the frequency dependence for a wide range

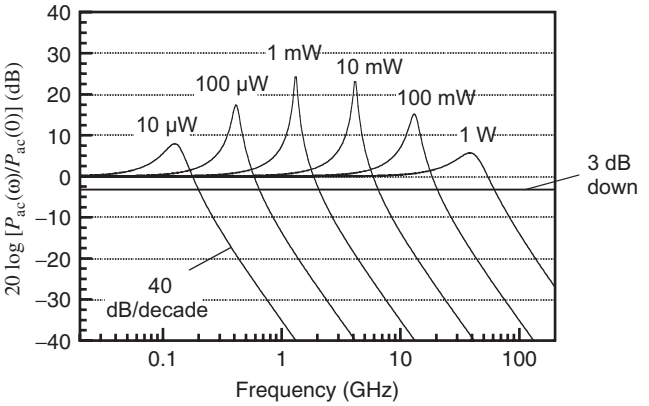


FIGURE 2.13: Frequency response of an idealized diode laser for several different output powers. The active region is characterized by: $h\nu = 1.5$ eV, $a = 5 \times 10^{-16}$ cm^2, $\tau = 3 \times 10^{-9}$ s, $\eta_i = 86.7\%$, and $v_g = 3 \times 10^{10}/4$ cm/s. The laser cavity is characterized by: $\tau_p = 2 \times 10^{-12}$ s (with $\alpha_m = 60$ cm^{-1} and $\alpha_i = 5$ cm^{-1}), $\eta_d = 80\%$, and $V_p = 5$ μm $\times$ 0.25 μm $\times$ 200 μm. The 20 log $[P_{ac}(\omega)/P_{ac}(0)]$ is used because photodetection generates an electrical *current* in direct proportion to the optical *power*. Thus, for a power ratio in the electrical circuit, this current must be squared.

of output powers. Note that the resonance is damped at low and high output powers. This occurs because the imaginary damping term in Eq. (2.54) depends on both ω_R and $1/\omega_R$. In Chapter 5, we will find that inclusion of gain compression and transport effects creates significantly more damping than predicted here. In fact, on real laser devices the resonance is typically limited to 5–10 dB (as opposed to the peak ~25 dB suggested in Fig. 2.13).

Beyond the strong resonance, the transfer characteristics degrade significantly. Thus, effective modulation of the output power can only be achieved over a modulation bandwidth of $\sim\omega_R$. When the damping is small, the electrical 3 dB down frequency (i.e., the frequency that reduces the received *electrical* power to one-half its DC value) is given by $\omega_{3\text{ dB}} = \sqrt{1+\sqrt{2}}\omega_R$. Expanding Eq. (2.53) using Eqs. (2.28), (2.38), and (2.40), we can express this result in terms of the output power:

$$f_{3\text{ dB}} \approx \frac{1.55}{2\pi}\left[\frac{\Gamma v_g a}{h\nu V}\frac{\eta_i}{\eta_d}\right]^{1/2}\sqrt{P_0}. \quad \text{(small damping)} \tag{2.55}$$

The modulation bandwidth of the laser can be steadily enhanced by increasing the output power. However, increased damping of the resonance at high powers, thermal limitations, and high power mirror facet damage set practical limits on the maximum average operating power we can use.

Because thermal limits are usually associated with the drive current, it is also convenient to express ω_R in terms of current. Using Eq. (2.37) for N_{p0}, with g_{th}

given by Eq. (2.28), Eq. (2.53) becomes

$$\omega_R = \left[\frac{\Gamma v_g a}{qV}\eta_i (I - I_{th})\right]^{1/2}. \tag{2.56}$$

In this form we observe that it is desirable to enhance the differential gain, minimize the volume of the mode ($\Gamma/V = 1/V_p$), and maximize the current relative to threshold for maximum bandwidth. If we want to keep the overall drive current low, then we should also try to minimize the threshold current, perhaps by increasing the facet reflectivity. If, however, we are more concerned about keeping the photon density low (for example, to reduce the risk of facet damage), then from Eq. (2.52) we should try to decrease the cavity lifetime instead, perhaps by *decreasing* the facet reflectivity. Thus, the optimum cavity design for a high-speed laser depends on what constraints we place on the device operation. In Chapter 5 we will find that at very high powers, the maximum bandwidth actually becomes independent of ω_R and is more fundamentally related to the damping factor (the K-factor), which is affected by gain compression and transport effects.

Example 2.3 A cleaved facet, active 3-μm-wide and 500-μm-long ridge laser is created from the laser structure from Example 2.1. This ridge laser is biased 30 mA above threshold and directly modulated by applying a small signal sine wave current to its active section.

Problem: Determine the resonance peak frequency $f_R = \frac{\omega_R}{2\pi}$ of this laser, assuming injection efficiency η_i of 80%, and assuming no change in the internal losses.

Solution: To calculate the resonance frequency, we will use the Eq. (2.56). Therefore, we need to compute the differential gain a at threshold, since the carrier density is clamped at the threshold carrier density. Thus, we first need to determine the threshold carrier density. Because the internal losses and the laser length are unchanged, the threshold modal gain for this laser is the same as that for the laser in Example 2.2, $\Gamma g_{th} = 37.79\ \text{cm}^{-1}$, leading to the same threshold carrier density as calculated in Example 2.2,

$$N_{th} = N_{tr} e^{\frac{g_{th}}{g_{0N}}} = N_{tr} e^{\frac{37.79}{0.04 \cdot 1207.29}} = 2.6865 \cdot 10^{18}\ \text{cm}^{-3}.$$

Differential gain a can be computed using expression Eq. (2.46),

$$a = \frac{\partial g}{\partial N}|_{N=N_{th}} = \frac{g_{0N}}{N_{th}} = \frac{1207.29\ \text{cm}^{-1}}{2.6865 \cdot 10^{18}\ \text{cm}^{-3}} = 4.49 \cdot 10^{-16}\ \text{cm}^2.$$

The resonance frequency is given by $f_R = \frac{\omega_R}{2\pi}$, where ω_R is defined in Eq. (2.56),

$$f_R = \frac{1}{2\pi}\left[\frac{\Gamma v_g a}{qV}\eta_i (I - I_{th})\right]^{1/2} = \left[\frac{\Gamma_1 v_g a}{qV_1}\eta_i (I - I_{th})\right]^{1/2}.$$

V_1 is the volume of a single quantum well, $V_1 = 3 \cdot 500 \cdot 0.003 \cdot 10^{-12}$ cm^{-3} = $4.5 \cdot 10^{-12}$ cm^{-3}, and the group velocity for InGaAsP laser is $v_g = \frac{c}{3.8} = 0.7894 \cdot 10^{10}$ cm/s. Therefore, the resonance frequency is

$$f_R = \frac{1}{2\pi}\left[\frac{0.01 \cdot 0.7894 \cdot 10^{10} \cdot 4.49 \cdot 10^{-16}\ \text{cm}^2}{1.6 \cdot 10^{-19} \cdot 4.5 \cdot 10^{-12}} 0.8(0.3)\frac{1}{s^2}\right]^{1/2} = 5.47\ \text{GHz}.$$

To gain a somewhat more intuitive understanding of the natural resonant behavior that results between the carrier-photon coupling described by the coupled rate equations, we consider only the third term on the right side of Eq. (2.50), for the small-signal carrier density, N_1, together with Eq. (2.51) for the small-signal photon density, N_{p1}. This is approximately valid, because the carrier density is almost clamped, $N_{p1} \gg N_1$, so the other terms in Eq. (2.50) are much smaller. If we view the left-hand sides of these two equations as time derivatives, then we observe from Eq. (2.51) that as N_1 increases and becomes positive, N_{p1} increases in time due to increased gain in the laser. However, from the third term in Eq. (2.50), once N_{p1} becomes positive, it serves to decrease N_1 through increased stimulated emission. As N_1 decreases and becomes negative, N_{p1} begins to fall, and once it becomes negative, it again produces an increase in N_1. At this point, the cycle repeats itself. This phenomenon produces a natural resonance in the laser cavity which shows up as a ringing in the output power of the laser in response to sudden changes in the input current. The natural frequency of oscillation associated with this mutual dependence between N_1 and N_{p1} can be found by multiplying Eqs. (2.50) and (2.51) together, again ignoring all but the third term on the right-hand side of the first equation, we again obtain ω_R, given by Eq. (2.53), and commonly referred to as the *relaxation resonance frequency* (where *relaxation* refers to an attempt by the photons and carriers to relax to their steady-state values). It is directly proportional to the square root of the differential gain and average photon density in the cavity (output power) and inversely proportional to the square root of the photon lifetime in the cavity. The relaxation resonance of the laser cavity is much like the natural oscillation of an *LC* circuit. However, the additional terms present in Eq. (2.50) lead to more of an *RLC* circuit behavior, dampening the resonant response.

2.8 CHARACTERIZING REAL DIODE LASERS

In this section we wish to review some of the common measurements that are made on diode lasers. We shall emphasize those that can be used to extract internal parameters that we have used in the rest of this chapter. More complex characterization techniques will be delayed until after the discussion of dynamic effects and the introduction of more complex cavity geometries.

2.8.1 Internal Parameters for In-Plane Lasers: $\langle\alpha_i\rangle$, η_i, and g versus J

Perhaps the most fundamental characteristic of a diode laser is the P–I characteristic as has been illustrated in Fig. 2.9. From a measured P–I characteristic one can immediately determine the experimental threshold current, I_{th}, from the intercept of the above-threshold curve with the abscissa. The differential quantum efficiency, η_d, can be calculated from Eq. (2.41), provided the wavelength is known. Usually the mean mirror reflectivity, $R = r_1 r_2$, can be calculated with good accuracy, and the length, L, can be measured. Thus, the mirror loss, $\alpha_m \equiv (1/L)\ln(1/R)$ can be calculated. However, the net internal optical loss $\langle\alpha_i\rangle$ and quantum efficiency, η_i, cannot be determined from a single device.

To determine these important internal parameters, one commonly uses two or more lasers of different length fabricated from the same material with identical mirrors. This is relatively straightforward for in-plane lasers because the length can be varied at the final cleaving step. From Eq. (2.40) it can be seen that by measuring the differential efficiency of two such lasers, one is left with two equations containing two unknowns, $\langle\alpha_i\rangle$ and η_i. That is,

$$\eta_d = \frac{\eta_i \ln\left(\frac{1}{R}\right)}{L\langle\alpha_i\rangle + \ln\left(\frac{1}{R}\right)},$$

and

$$\eta_d' = \frac{\eta_i \ln\left(\frac{1}{R}\right)}{L'\langle\alpha_i\rangle + \ln\left(\frac{1}{R}\right)}, \tag{2.57}$$

where L and L' are the lengths of the two different lasers. Solving, we find

$$\langle\alpha_i\rangle = \frac{\eta_d' - \eta_d}{L\eta_d - L'\eta_d'} \ln\left(\frac{1}{R}\right),$$

and

$$\eta_i = \eta_d \eta_d' \frac{L - L'}{L\eta_d - L'\eta_d'}. \tag{2.58}$$

If indeed one can make two identical lasers except for their lengths, then Eqs. (2.58) will give the desired internal parameters. However, experimental data usually have some uncertainty which limits the utility of these expressions. For more reliability, it is generally better to plot a number of data points on a graph and determine the unknowns by fitting a curve to the data. In the present case it is most convenient to plot the reciprocal of the measured differential efficiencies versus L. Then a straight line through the data has a slope and intercept from which $\langle\alpha_i\rangle$ and η_i can be determined. More specifically,

$$\frac{1}{\eta_d} = \frac{\langle\alpha_i\rangle}{\eta_i \ln(1/R)} L + \frac{1}{\eta_i}. \tag{2.59}$$

Thus, the intercept gives η_i, and this can be used in the slope to get $\langle\alpha_i\rangle$.

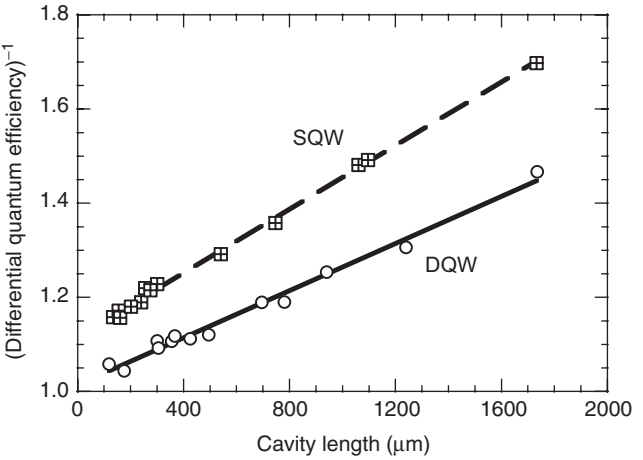


FIGURE 2.14: Plot of experimental reciprocal external differential efficiencies versus laser cavity length for 50 μm wide $In_{0.2}Ga_{0.8}As$ GRINSCH quantum well lasers. For both single-quantum-well (SQW) and double-quantum-well (DQW) cases, the $In_{0.2}Ga_{0.8}As$ well(s) was 80 Å wide. For the DQW a 12 nm GaAs separation barrier was used, and in both cases 40 nm of GaAs was used on each side of the well(s). On each side of this active region, the barrier stepped to $Al_{0.2}Ga_{0.8}As$ for 8 nm and then tapered to $Al_{0.8}Ga_{0.2}As$ over 80 nm to form the graded-index GRINSCH structure [1]. From these data the SQW had $\alpha_i = 3.2$ cm^{-1} and $\eta_i = 89.6\%$; and the DQW had $\alpha_i = 2.6$ cm^{-1} and $\eta_i = 98.6\%$.

Figure 2.14 shows such a plot for some data taken from broad-area in-plane InGaAs/GaAs quantum-well lasers. Single (SQW) and double (DQW) quantum-well cases are included [1].

For shorter cavity lengths, the data in Fig. 2.14 will fall above the line indicated. This data was ignored when determining the line fit because it represents a region where higher-order effects result in an incomplete clamping of the carrier density above threshold. The result is an apparent decrease in η_i (see Appendix 2 for details). If one assumes that the net internal loss does not change in this process, it is possible to estimate the decrease in η_i by repeated use of Eq. (2.59) or 2.35 for this high-gain points.

In the process of taking data, one can also generate a table of threshold current densities in the active region, $J_{th} = (\eta_i I_{th}/wL)$ versus L. These are usually taken from broad-area devices so that lateral current and carrier leakage can be neglected. From Eq. (2.25) we also see that corresponding threshold modal gains Γg_{th} can be calculated for each length once the internal loss is found. Thus, it is possible to construct the modal gain versus current density characteristic for the laser from these threshold values. This characteristic can usually be fitted well to an exponential, two parameter (J_{tr} and g_0) curve,

$$J = J_{tr} e^{\frac{g}{g_0}}. \tag{2.60}$$

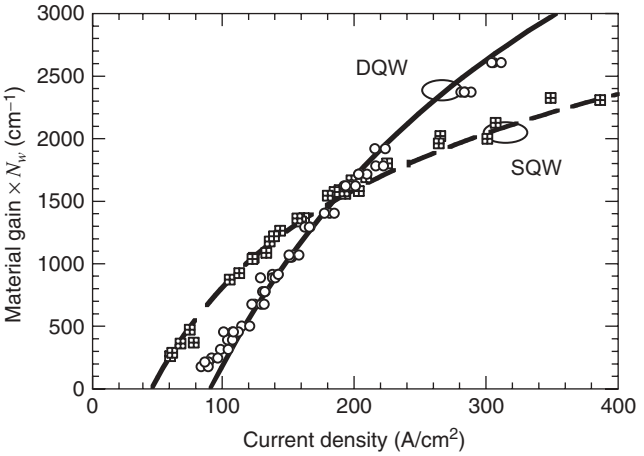


FIGURE 2.15: Experimentally determined gain versus current density for an InGaAs/GaAs quantum well laser described in Fig. 2.14 [1]. For the ordinate the modal gain Γg is divided by the confinement factor for one well Γ_w to give the material gain g times the number of wells, N_w. The solid curves are from a calculation based on the theory to be developed in Chapter 4.

Because the confinement factor, Γ, can usually be calculated as discussed in Appendix 5, one can ultimately determine the basic material gain versus current density characteristic for the active material. Figure 2.15 gives the result for the example in Fig. 2.14.

Example 2.4 To characterize the active material from Example 2.1, a 50-μm-wide and 500-μm-long broad area laser is cleaved from the wafer. Its pulsed threshold current is 47.94 mA and differential efficiency from both facets 48.24%. This laser is then re-cleaved into two 250-μm-long lasers, having pulsed threshold currents of 59.44 mA and differential efficiencies of 60.19%.

Problem: (1) What is the injection efficiency η_i? (2) What is the average internal modal loss? (3) Determine J_{tr} and g_0 in the gain vs current density characteristic for each quantum well, assuming $J = J_{tr} e^{\frac{g}{g_0}}$.

Solution: To calculate the injection efficiency η_i and the internal modal loss $\langle \alpha_i \rangle$, we will use the Eq. (2.58). Once the modal loss is known, and knowing the threshold current densities in the active region, we can construct the modal gain Γg versus current J density curve. Because the confinement factor is known, we can determine the basic material gain g versus current density J characteristic for this active material using the Eq. (2.60). From Eq. (2.58),

$$\langle \alpha_i \rangle = \frac{\eta_d' - \eta_d}{L\eta_d - L'\eta_d'} \ln\left(\frac{1}{R}\right) = \frac{(0.6019 - 0.4824)\ln\left(\frac{1}{0.32}\right)}{0.05 \cdot 0.4824 - 0.025 \cdot 0.6019}) = 15 \text{ cm}^{-1}$$

and

$$\eta_i = \eta_d \eta_d' \frac{L - L'}{L\eta_d - L'\eta_d'} = 0.6019 \cdot 0.4824 \frac{0.025 \text{ cm}^{-1}}{0.0091} = 0.80.$$

For the gain versus current density characteristic, we utilize two data points from two different laser lengths:

$$J_{th1} = \frac{\eta_i I_{th1}}{w \cdot L_1} = \frac{0.8 \cdot 47.94 \text{ mA}}{500 \cdot 50 \cdot 10^{-8} \text{ cm}^2} = 153.41 \text{ A/cm}^2$$

$$J_{th2} = \frac{\eta_i I_{th2}}{w \cdot L_2} = \frac{0.8 \cdot 59.44 \text{ mA}}{250 \cdot 50 \cdot 10^{-8} \text{ cm}^2} = 380.42 \text{ A/cm}^2.$$

Threshold modal gain Γg_{th1} for the 500-μm cavity was calculated in Example 2.2 to be 37.79 cm^{-1}. Similarly, for the 250-μm-long cavity, the threshold modal gain is $\Gamma g_{th1} = 60.58 \text{ cm}^{-1}$, where $\Gamma = 0.04$, as discussed in Example 2.2. Finally,

$$g_0 = \frac{g_{th1} - g_{th2}}{ln \frac{J_{th1}}{J_{th2}}} = \frac{944.71 - 1514.43 \text{ cm}^{-1}}{ln \frac{153.41}{380.42}} = 627 \text{ cm}^{-1}$$

$$J_{tr} = \frac{J_{th1}}{\exp \frac{g_{th1}}{g_0}} = \frac{153.41 \text{ A/cm}^2}{\exp \frac{944.71}{627}} = 34 \text{ A/cm}^2.$$

2.8.2 Internal Parameters for VCSELs: η_i and g versus J, $\langle \alpha_i \rangle$, and α_m

In vertical-cavity lasers the preceding procedure is a little difficult to carry out because the cavity length is set by the crystal growth. Multiple growths may result in other changes in the material besides the cavity length. Therefore, it has been proposed that the desired information can be determined by making in-plane cleaved lasers fabricated from the vertical-cavity laser material. However, a somewhat different approach is followed. Clearly, the internal loss determined for the in-plane laser will not be the same as for the VCSEL because the optical mode travels through a different cross section of materials. Nevertheless, if the electrical pumping current follows the same path and the threshold current densities covered in the in-plane diagnostic lasers includes the VCSEL values, the measured internal quantum efficiency should be the same.

The most valuable piece of information provided by the diagnostic lasers is the gain versus current density characteristic. Combining this characteristic and the internal quantum efficiency from the in-plane lasers together with the measured threshold current density and differential quantum efficiency from the VCSEL, we now have enough information to unambiguously determine the VCSEL internal loss and mirror loss (and thus, reflectivity). That is, Eqs. (2.40) and (2.25) can be solved for $\langle \alpha_i \rangle$ and α_m because $\Gamma g_{th}, \eta_d$, and η_i are known. The results are

$$\alpha_m = \Gamma g_{th} \frac{\eta_d}{\eta_i} \tag{2.61}$$

and

$$\langle \alpha_i \rangle = \Gamma g_{th} \left[1 - \frac{\eta_d}{\eta_i} \right] \tag{2.62}$$

As before, the confinement factors for both the in-plane diagnostic lasers and the VCSELs must be calculated.

2.8.3 Efficiency and Heat Flow

Just as the differential efficiency is important in determining the electrical to optical modulation efficiency, the overall net power conversion efficiency is also important in determining the achievable optical power out as well as the circuit heating and system power requirements. This so-called wall-plug efficiency is simply the optical power out relative to the electrical power in, $\eta = P_0/P_{in}$. The optical power out is given by Eq. (2.39), and the electrical power in is the product of the drive current and the total voltage across the diode's terminals. We can express this as

$$P_{in} = I^2 R_s + I V_d + I V_s, \tag{2.63}$$

where R_s is the series resistance, V_s is a current-independent series voltage, and V_d is the ideal diode voltage, which is equal to the quasi-Fermi level separation. This voltage is clamped at its threshold value above threshold.

The power dissipated in the laser is

$$P_D = P_{in} - P_0 = P_{in}[1 - \eta], \tag{2.64}$$

and the temperature rise is

$$\Delta T = P_D Z_T, \tag{2.65}$$

where Z_T is the thermal impedance. Analytic expressions for Z_T, which are approximately valid for several practical cases of interest, exist. Three are illustrated in Fig. 2.16. For a heat sink plane positioned much closer than the lateral dimensions

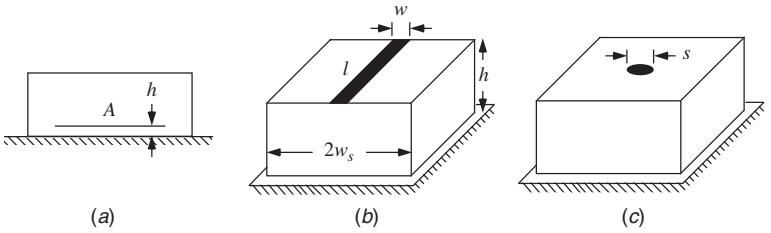


FIGURE 2.16: Schematics of heat flow geometries relevant to lasers: (a) planar or one-dimensional flow for a heat sink relatively near the heat source, (b) a line source on a thick substrate, and (c) a disk source on a half-space.

of the regions generating the heat (Fig. 2.16a), a one-dimensional heat flow can be assumed. In this case,

$$Z_T = \frac{h}{\xi A}, \quad \text{(1-D flow)} \tag{2.66}$$

where ξ is the thermal conductivity of the material separating the source of area A a distance h from the ideal sink. For GaAs and AlAs, $\xi \sim 0.45$ and 0.9 W/cm-$^\circ$C. For $Al_xGa_{1-x}As$, alloy scattering reduces ξ to a minimum of 0.11 W/cm-$^\circ$C at $x \approx 0.5$. For InP and InGaAsP, the values are $\xi \sim 0.68$ and 0.06 W/cm-$^\circ$C.

For a linear stripe heat source of length l and width w on a thick substrate (thickness h), which is somewhat wider (width 2 w_s) than this thickness (Fig. 2.16b), a quasi two-dimensional heat flow results. Then,

$$Z_T \approx \frac{\ln(4h/w)}{\pi \xi l}. \quad \text{(line : } w \ll h < w_s\text{)} \tag{2.67}$$

A narrow stripe in-plane laser mounted active region up on a relatively thick substrate approximates this case. For a disk heat source of diameter s on a half-space (Fig. 2.16c), a three-dimensional flow into the half-space can be assumed. Then,

$$Z_T = \frac{1}{2\xi s}. \quad \text{(small disk)} \tag{2.68}$$

This is approximately valid for a small-diameter VCSEL mounted on the top side of a relatively thick substrate.

2.8.4 Temperature Dependence of Drive Current

The required drive current for a given power out of a laser is given by Eq. (2.50), in which the first and second terms give the needed current above threshold and the threshold current, respectively. In this equation it is assumed that the recombination below threshold is dominated by spontaneous emission events. If significant nonradiative recombination exists, an additional threshold term, such as Eq. (2.52), must be added. For both in-plane and vertical cavity lasers these expressions are functions of temperature. Generally, more current is required both for threshold and the increment above threshold as the temperature is increased, and we can estimate the nature of this dependence by exploring the temperature dependence of each of the factors in the terms of Eqs. (2.50) and (2.52).

However, for VCSELs as well as single axial-mode in-plane lasers, the situation is complicated by the integrated mode selection filter (e.g., Bragg mirrors), which can force the lasing mode to be well off the wavelength where the gain is a maximum. Thus, such lasers can be designed to have anomolous temperature behavior because the wavelength of the cavity mode and the gain peak shift at different rates versus temperature. In fact, by deliberately misaligning the mode

from the gain peak at room temperature, it is even possible to make the threshold go down with increasing temperature as the gain moves into alignment with the mode [4]. In this section, we will not consider these relative mode–gain alignment issues. Rather, we will assume that a spectrum of modes exist, as in a simple in-plane laser, so that lasing always can occur at the gain peak. Thus, we again can focus only on the temperature dependence of the various factors in Eqs. (2.50) and (2.52).

For the threshold current in Eq. (2.49), three factors generally have a significant temperature dependence: N_{tr}, g_0, and $\langle \alpha_i \rangle$. From the gain calculations of Chapter 4, it may be shown that over some range of temperatures, $N_{tr} \propto T$, $g_0 \propto 1/T$, and $\langle \alpha_i \rangle \propto T$. The transparency carrier density is increased and the gain parameter is reduced because injected carriers spread over a wider range in energy with higher temperatures. The increased internal loss results from the required higher carrier densities for threshold. From Eq. (2.49), we conclude that both the gain and the internal loss variations result in an exponential temperature dependence of the threshold current, whereas the linear dependence of the transparency carrier density is not significant over small temperature ranges. Additional threshold components such as Eq. (2.52) will introduce further temperature dependencies. For example, in Chapter 4 and Appendix 2 it is shown that $C \propto \exp(\gamma_C T)$ and $R_l \propto \exp(\gamma_l T)$. Thus, Auger recombination and carrier leakage both contribute additional exponential increases in the threshold current. These observations suggest that the threshold current can be approximately modeled by

$$I_{th} = I_0 e^{T/T_0}, \tag{2.69}$$

where T_0 is some overall characteristic temperature, and both temperatures are given in degrees Kelvin, K. Note that small values of T_0 indicate a larger dependence on temperature (since $dI_{th}/dT = I_{th}/T_0$). It should also be noted that any minor temperature dependence of other parameters can easily fit into this model over some limited temperature range. For example, the internal efficiency can decrease at higher temperatures due to increased leakage currents and/or higher-order effects discussed in Appendix 2. This decrease in η_i will show up as a reduction of T_0 over a limited temperature range, regardless of the exact dependence of η_i on temperature.

For good near-infrared (~850 nm) GaAs/AlGaAs DH lasers, observed values of T_0 tend to be greater than 120 K near room temperature. For quantum-well GaAs/AlGaAs the values are somewhat higher ($\sim 150 - 180$ K), and for strained-layer InGaAs/AlGaAs quantum wells, $T_0 \geq 200$ K have been observed. For $1.3 - 1.55$ μm InGaAsP/InP DH and quantum-well lasers the characteristic temperature is generally quite a bit lower as expected. Measured values tend to fall in the 50–70 K range, due to Auger recombination as well as possible carrier leakage and intervalence band absorption effects. Thus, the threshold tends to change significantly between room temperature and 100°C, usually resulting in relatively poor performance at the higher temperatures, and generally requiring the use of thermoelectric coolers. Shorter wavelength (600–800 nm) AlGaAs/GaAs

and AlInGaP/GaAs lasers also tend to have a smaller T_0 than the near-infrared variety, presumably due to increased carrier leakage.

The above-threshold current required to obtain a desired output power is also temperature dependent, although the dependence is usually smaller than for the threshold current. This dependence results from a reduction in the differential quantum efficiency. As suggested by the constituent factors in the first term in Eq. (2.50), an increase in $\langle \alpha_i \rangle$ as well as a drop in η_i are usually the cause of the increase in $I - I_{th}$. In analogy with Eq. (2.69), we can write

$$I - I_{th} = I_{p0} e^{T/T_\eta}, \tag{2.70}$$

where T_η is the characteristic temperature for the above-threshold current increment. T_η is generally two or three times larger than T_0, as might be expected from the above discussion. That is, T_0 includes several effects in addition to those in T_η.

In practice one is typically more interested in the relative changes in threshold current and differential efficiency as the temperature varies rather than the absolute values as given in Eqs (2.69) and (2.70), so more useful expressions tend to be the ratios of currents and differential efficiencies at two different temperatures, T_1 and T_2:

$$\frac{I_{th1}}{I_{th2}} = e^{(T_1 - T_2)/T_0} \tag{2.71}$$

$$\frac{\eta_{d1}}{\eta_{d2}} = e^{-(T_1 - T_2)/T_\eta}. \tag{2.72}$$

Example 2.5 Another batch of lasers similar to those from Example 2.3 is made, but this time, the laser contacts exhibit large series resistance, and thus lead to significant amount of heating under CW operation. Consider a 250-μm-long, 3-μm-wide all-active ridge laser, which can be modeled by a 50 Ω series resistance and an ideal diode with voltage $V_d = 0.88$V. The InP substrate is 100 μm thick and 500 μm wide, and it is bonded to a good heat sink. The characteristic temperature for threshold current is $T_0 = 25$ K and that for differential efficiency is $T_\eta = 110$ K. Pulsed threshold current is 15 mA, and differential efficiency out of both ends 48.24%.

Problem: (1) What is the thermal impedance (2) What is the new CW threshold current (3) At a bias of 50 mA, what is the power out and the temperature rise of the active region?

Solution: Thermal impedance can be calculated using Eq. (2.67),

$$Z_T = \frac{\ln(4h/w)}{\pi \xi l} = \frac{\ln \frac{4 \cdot 100}{3}}{\pi \cdot 0.6 \cdot 25010^{-4}} = 103.78^\circ \text{ C/W}$$

To find the new threshold current, we will need to iterate, given that the temperature increase depends on the new threshold current, which in turn is determined by

the temperature increase. At threshold, the output power from the laser can be neglected; therefore, the dissipated power is equal to the input power,

$$P_D = P_{in} = I_{th}^2 \cdot R_s + I \cdot V_d,$$

where V_d is the ideal diode voltage and is approximately 0.88 V for InGaAsP/InP. Assuming that the threshold current increase due to heating is 1 mA, $I_{th} = 16$ mA, the dissipated power and temperature increase are

$$P_D = (0.016)^2 \cdot 50 + (0.016)(0.88) \text{ mW} = 26.88 \text{ mW}$$
$$\Delta T = P_d Z_T = 0.02688 \cdot 103.78^\circ\text{C} = 2.79^\circ\text{C}.$$

To check, we plug in the value for ΔT to calculate the threshold current based on known $T_0 = 25$ K,

$$I'_{th}(\Delta T = 2.79^\circ\text{C}) = 15 \text{ mA} \cdot \exp \frac{2.79}{25} = 16.77 \text{ mA}.$$

Thus, we conclude that we have underestimated the heating effects, and we use I'_{th} to calculate the dissipated power and repeat the process. After a couple of iterations, we end up with the final value for I_{th},

$$I_{th} = 16.9 \text{ mA}.$$

To calculate the output power for the bias current of $I = 50$ mA, we do the following: assuming that the output power is negligible, we calculate the dissipated power, the temperature increase, and then the increase in the threshold temperature and the decrease in the differential efficiency. At that point, we can compute the output power. To iterate, we reduce the dissipated power by the output power value and repeat the process. After a couple of steps, the process converges.

$$P_D \approxeq P_{in} = (0.05)^2 \cdot 50 + (0.05)(0.88) \text{ mW} = 169.00 \text{ mW}$$
$$\Delta T = P_D Z_T = 0.169 \cdot 103.78^\circ\text{C} = 17.54^\circ\text{C}$$

The threshold current and the differential efficiency with this much temperature increase are given by

$$I'_{th} = 15 \text{ mA} \cdot \exp \frac{17.54}{25} = 30.25 \text{ mA}$$
$$\eta'_d = 0.4824 \cdot \exp \frac{-17.54}{110} = 0.4113.$$

The output power is given by

$$P_o = \frac{h\nu}{q} \eta'_d (I - I_{th}{}') = 0.8 \cdot 0.4113(50 - 30.25) \text{ mW} = 6.49 \text{ mW}.$$

Reducing the dissipated power by P_o yields $\Delta T = 16.87^{\circ}\text{C}$ and $P_o = 6.80$ mW. Repeating the process yields the final values of

$$\Delta T = 16.83^{\circ}\text{C}$$

$$P_o = 6.82 \text{ mW}.$$

The L-I characteristic of this laser is illustrated in Fig. 2.17.

2.8.5 Derivative Analysis

Real diode lasers do not always have perfectly linear $P-I$ characteristics above threshold, and they have parasitic series resistance as well as a possible series voltage as outlined in Eq. (2.63) above. Derivatives of the $P-I$ and $V-I$ characteristics can be useful in sorting out these nonidealities. The dP/dI characteristic in an ideal laser would only provide a good measure of the threshold current and a slope to determine η_d above threshold. However, actual $P-I$ characteristics can have kinks, and they tend to be nonlinear. The kinks can indicate a switching between lateral or axial modes or an additional parasitic mirror in the device. These are obviously emphasized in a derivative curve. Premature saturation of the output power may indicate the existence of current leakage paths that "turn-on" at higher current levels or excessive heating of the gain material. The derivative curve gives a good quantitative measure of these symptoms. Figure 2.18(a) gives example plots of $P-I$ and dP/dI for an in-plane laser.

In addition to the $V-I$ characteristic, it is common to plot IdV/dI versus I. This latter characteristic gives a sensitive measure of the series resistance, and it

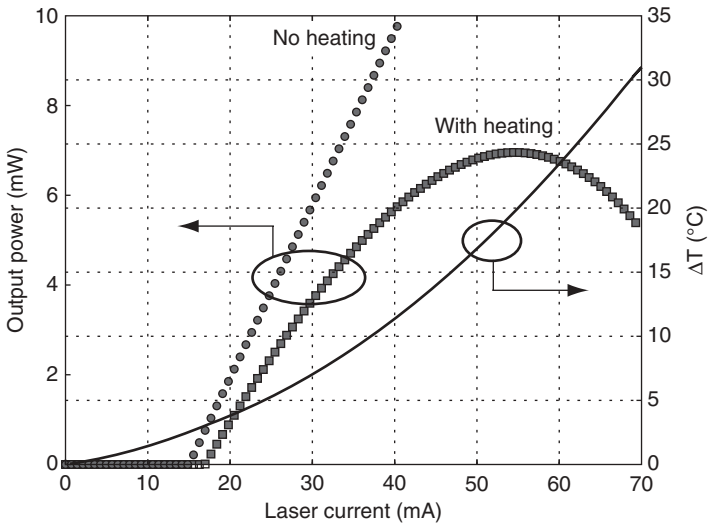


FIGURE 2.17: Illustration of Example 2.5—laser heating as a function of bias current, L-I curve with thermal roll off (temperature effects taken into account) and ideal L-I curve (for no heating) are shown.

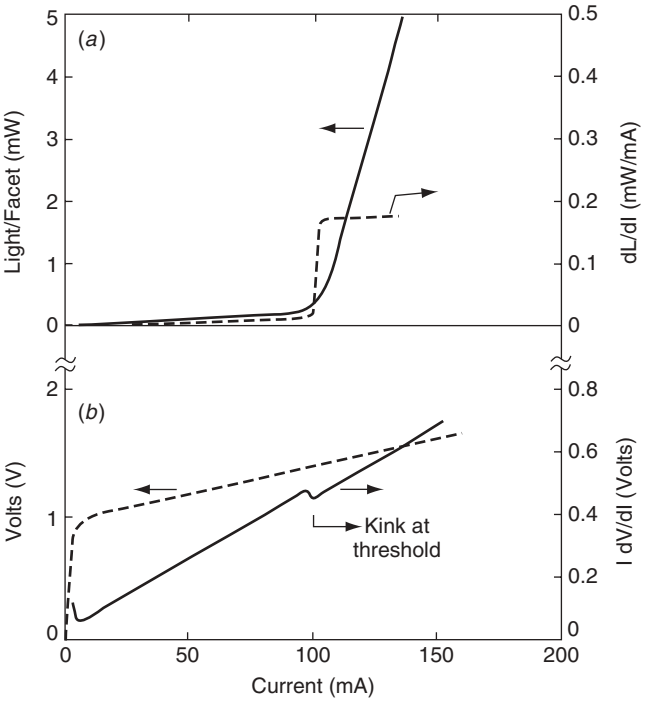


FIGURE 2.18: (a) Plots of optical output power, P, and its derivative, $dP/dI \equiv dL/dI$, versus drive current, I, for a stripe-geometry gain-guided InGaAsP laser. (b) Plots of the terminal voltage, V, and IdV/dI versus current for the same device [5]. (*Reproduced, by permission, from Semiconductor Lasers*.)

is particularly useful in identifying shunt current paths. Because the voltage across the junction clamps at threshold with the carrier density, a kink in the curve occurs at that point. Figure 2.18b shows example plots of both V and IdV/dI versus I. The information contained in the plot can be derived by considering an equivalent circuit with a parasitic resistance in series with an ideal heterojunction diode. The diode $V - I$ is described by

$$I = I_0[e^{qV_d/nkT} - 1]. \tag{2.73}$$

Taking the derivative of the terminal voltage, $V = V_d + IR$, and solving for dV_d/dI from Eq. (2.73), we obtain for $I \gg I_0$ but below threshold,

$$I\frac{dV}{dI} = \frac{nkT}{q} + IR. \tag{2.74}$$

Above-threshold V_d is constant, so

$$I\frac{dV}{dI} = IR. \tag{2.75}$$

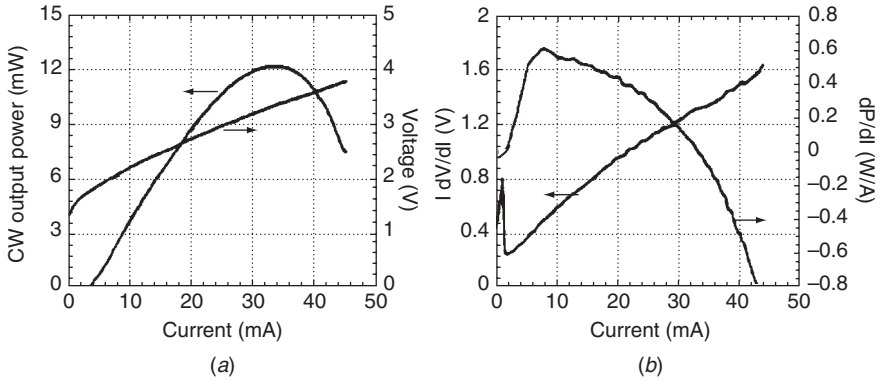


FIGURE 2.19: (a) Output power and terminal voltage for a 20 μm diameter VCSEL with three InGaAs strained quantum wells and AlAs/GaAs DBR mirrors. (b) Derivative curves for the same device [6].

Thus, we see that the slope above and below threshold should be R, but there is a positive offset of nkT/q below threshold, which provides a kink of this magnitude at threshold. Now if a shunt resistance is added to the equivalent circuit, it turns out that an additional term must be added to Eq. (2.74). This provides a peak in the $I(dV/dI)$ characteristic below threshold. For common DH structures the diode ideality factor $n \sim 2$.

Figure 2.19 gives plots analogous to Fig. 2.18 for an InGaAs/GaAs VCSEL. Here, significant local heating causes the $P-I$ curve to roll over at relatively low powers. This results in a negative dP/dI beyond this point. In addition, significant series resistance makes it difficult to discern the nkT/q kink in the IdV/dI characteristic at threshold. Thus, the derivative analysis is not always very effective for VCSELs.

REFERENCES

1. Hu SY, Young DB, Corzine SW, Gossary AC, Coldren LA. High-efficiency and low-threshold InGaAs/AlGaAs quantum-well lasers. *J Appl Phys* 1994;**76**: 3932–3934.
2. Krames MR, Shchekin OB, Mueller-Mach R, Mueller GO, Ling Zhou, Harbers G, Craford MG, Philips Lumileds Lighting Co. Status and future of high-power light-emitting diodes for solid-state lighting. *J Display Tech* 2007;**3**(2): 160–175.
3. Oder TN, Kim KH, Lin JY, Jiang HX. III-nitride blue and untraviolet photonic crystal light emitting diodes. *Appl Phys. Lett.* 2004;**84**: 466.
4. Young DB, Scott JW, Peters FH, Peters MG, Majewski ML, Thibeault BJ, Corzine SW, Coldren LA. Enhanced performance of offset-gain high-barrier vertical-cavity surface-emitting lasers. *IEEE Journal of Quantum Electronics* 1993;**29**(6): 2013–2022.
5. Agrawal GP, Dutta NK. *Semiconductor lasers*. 2nd ed. New York: Van Nostrand Reinhold; 1993.
6. Peters MG. [dissertation]. Santa Barbara, CA: University of California, Santa Barbara; 1995.

READING LIST

Agrawal GP, Dutta NK. *Semiconductor lasers*. 2nd ed. New York: Van Nostrand Reinhold; 1993.

Ebeling KJ. *Integrated opto-electronics*. Berlin: Springer-Verlag; 1993.

Verdeyen JT. *Laser electronics*. 2nd ed. Upper Saddle River, NJ: Prentice Hall; 1989.

Yariv A. *Optical electronics*. 4th ed. Philadelphia: Saunders College Publishing; 1991.

PROBLEMS

These problems draw on material from Appendices 4 through 6.

1. In a diode laser, the terminal current is I, the current bypassing the active region is I_b, the current due to carriers leaking out of the active region before they recombine is I_l, the current contributing to nonradiative recombination in the active region is I_{nr}, the current contributing to spontaneous emission in the active region is I_{sp}, the currents contributing to spontaneous emission and nonradiative recombination outside the active region are I'_{sp} and I'_{nr}, respectively, and the current contributing to stimulated emission in the active region is I_{st}.
   (a) What is the injection efficiency?
   (b) If the measured external differential efficiency above threshold is η_d, what is the ratio of the mirror loss to the total cavity loss?
   (c) For below-threshold operation, what is the radiative efficiency?

2. A reservoir of area A is filled at a rate of R_f (in $\text{ft}^3/\text{min.}$) and simultaneously drained from two pipes that have flow rates that depend on the height of water, h. The drain rates are, $R_{d1} = C_1 h$ and $R_{d2} = C_2 h^2$, respectively.
   (a) Write a rate equation for the water height.
   (b) What is the steady-state water height?
   (c) If $A = 100\ \text{ft}^2$, $R_f = 10\ \text{ft}^3/\text{min}$, and $C_1 \approx 0$, what is C_2 for a steady-state depth of 5 ft?

3. What is the approximate intrinsic cutoff frequency of an LED with a p-type active region doping of $6.3 \times 10^{18}\ \text{cm}^{-3}$?

4. The relative increase in photons in passing through a piece of GaAs is found to be $[1/N_p][dN_p/dt] = 10^{13}\ \text{s}^{-1}$. What is the material gain in cm^{-1}?

5. A 1.3 μm wavelength InGaAsP/InP diode laser cavity is found to have an optical loss rate of $4 \times 10^{12}/\text{s}$.
   (a) What is the photon lifetime?
   (b) What is the threshold modal gain?

6. In a cleaved-facet 1.55 μm InGaAsP/InP multiple quantum-well laser 400 μm in length, it is known that the injection efficiency and losses are 80% and $10\ \text{cm}^{-1}$, respectively.

(a) What is the threshold modal gain?
(b) What is the differential efficiency?
(c) What is the axial mode spacing?

7. A cleaved-facet, DH GaAs laser has an active layer thickness of 0.1 μm, a length of 300 μm, and a threshold current density of 1 kA/cm^2. Assume unity injection efficiency, an internal loss of 10 cm^{-1}, a confinement factor of 0.1, and only radiative recombination.

(a) What is the threshold carrier density in the active region?
(b) What is the power out of one cleaved facet per micrometer of width at a current density of 2 kA/cm^2?
(c) What are the photon and carrier densities at 2 kA/cm^2?

8. In the device of Problem 2.7, gain transparency ($g = 0$) is found to occur at 0.5 kA/cm^2 and the transverse confinement factor is 0.15. What is the relaxation resonance frequency at 2 kA/cm^2?

9. Two broad-area DH 1.3 μm InGaAsP/InP lasers are cleaved from the same material. One is 200 μm long and the other is 400 μm long. The threshold current densities are found to be 3 kA/cm^2 and 2 kA/cm^2, respectively, and the differential efficiencies including both ends are measured to be 60% and 50%, respectively.

(a) What are the injection quantum efficiency and internal loss for this material?
(b) For a ±1% error in each of the measured differential efficiencies, what are the errors in the calculated internal loss and quantum efficiency?

10. For the material of Problem 2.9, the relaxation resonance frequency for the 200-μm laser biased at twice threshold is found to be 3 GHz. What is the resonance frequency for the 400-μm device also biased at twice threshold?

11. A VCSEL is formed with multilayer AlGaAs mirrors and a 3-quantum-well GaAs active region. Current is injected through the mirrors. At a terminal current density of 1 kA/cm^2 the active region provides 1% of one-pass gain for the propagating axial mode. The injection efficiency is assumed to be 80%, and the average internal loss is 25 cm^{-1}. The effective cavity length is 1.5 μm.

(a) What mean mirror reflectivity is necessary for the device to reach threshold at 1 kA/cm^2?
(b) For this case, plot the output power density versus terminal current density.
(c) If we assume the gain is linear with carrier density, and that only spontaneous recombination is important below threshold, plot the threshold current density versus mean mirror reflectivity for $0.98 < R < 1.0$. On the opposite axis label the differential efficiency at each 0.005 reflectivity increment.

12. With the VCSEL material of Problem 2.11, etched square mesas are now formed measuring s on each side. Assuming a spontaneous bandwidth of 30 nm, an axial confinement factor of 2 L_a/L, lateral confinement factors of unity, and that the approximations of Appendix 4 are valid, plot the spontaneous emission factor versus s.

13. Again with the VCSEL material of Problem 2.11, square mesas are formed measuring s on each side by etching down to the active region, as illustrated in Fig. 1.15. The GaAs substrate may be assumed to be thick and wide. If we assume a threshold current density of 1 kA/cm^2 independent of area, a series voltage of 1 V, and a series resistance that is inversely proportional to device area, $R_s = 20$ kΩ-μm^2/s^2.

(a) Plot the temperature at the base of the mesa (active region location) versus s for a current of twice threshold. Cover $1 < s < 20$ μm.

(b) Assuming a differential efficiency of 50%, plot the power out and required current versus s for the conditions of (a).

14. An MQW-SCH InGaAsP/InP laser wafer with emission wavelength near 1.55 μm has been grown and characterized with broad area chips. Alternate active and passive regions are included. The active region contains 6–7 nm thick unstrained quaternary quantum wells (with lowest conduction and valence band quantum state energies spaced by 0.8 eV) separated by 5–8 nm quaternary barriers (having photoluminescense emission at 1.3 μm), all centered within a 1.3 μm-Q SCH waveguide, which all measures 350 nm in total thickness; the passive region does not have quantum wells, but is all 1.3 μm-Q material, again 350 nm thick. We can neglect optical reflections between the active and passive waveguides. A calculation indicates that the transverse confinement factor in the active is $\Gamma = 0.08$. Measurements on three different broad-area, cleaved-facet chips show that the modal losses in the active and passive sections are 15 cm^{-1} and 5 cm^{-1}, respectively, and that the injection efficiency to the active is 70%. A material gain curve is also derived for the quantum wells. Fitting to an expression of the form $g = g_{0J} \ln(J/J_{tr})$, it is found that $g_{0J} = 600$ cm^{-1}, and $J_{tr} = 100$ A/cm^2 per well. It is also found that the recombination rate for nominal threshold gains is composed of a mix of radiative and Auger recombination, such that the carrier density varies approximately as $J^{0.4}$.

We create three cleaved facet, 3-μm-wide ridge lasers from this material (lateral current and carrier spreading as well as any additional scattering losses can be neglected): the first is an active–passive laser with active and passive sections each 500 μm long; the second is an all active laser 500 μm in length; the third is another all active laser 250 μm in length. Measured small-signal frequency response of the 500–500 μm active–passive laser gives a resonance peak at 5 GHz, for a bias current 30 mA above threshold. Neglecting heating effects:

(a) What are the differential efficiencies (considering the emission from both ends) of each of the three lasers?

(b) What are the threshold currents of each of the three lasers?

(c) What is the differential gain, dg/dN, for this laser above threshold?

(d) What is the approximate threshold carrier density?

(e) What is the transparency carrier density? (Assume a logarithmic g versus N dependence)

(f) What are the expected relaxation resonance frequencies for the 500-μm and 250-μm all-active devices at 30 mA above their threshold currents?

15. We now make another batch of similar lasers from the same material as in Problem 2.15, but this time we mess up the contacting procedure such that the lasers have a relatively large series resistance, and thus, a significant amount of heating under continuous wave operation. Consider the 250-μm-long, 3-μm-wide, all-active device, which can be modeled by a 50 Ω series resistance and an ideal diode with an ideality factor of 3. The InP substrate is 90 μm thick and 500 μm wide; it is bonded to a good heat sink. The characteristic temperature for threshold current is $T_0 = 55$ K, and the characteristic temperatur for differential efficiency is $T_\eta = 110$ K.

(a) What is the thermal impedance of this laser?

(b) What is the new CW threshold current?

(c) At a bias of 50 mA, what is the power out and the temperature rise at the active region?

(d) Plot the output power and temperature rise versus CW current up to 100 mA. Compare to the pulsed output power on the same plot.

16. In a 1.55 μm InGaAsP/InP BH laser, the active region is 0.2 μm thick, 3 μm wide and 300 μm long. The injection efficiency is 70%. In addition, there is a 400-μm-long passive waveguide channel with the same lateral and transverse dimensions butted to the end of the active region. The transverse and lateral confinement factors are 0.2 and 0.8, respectively. Cleaved mirrors form a 700-μm-long cavity, and other internal reflections can be neglected. The material losses are 80 cm^{-1}, 20 cm^{-1}, and 5 cm^{-1} in the active, passive, and cladding regions, respectively. The gain versus carrier density characteristic for the active material is linear with a transparency carrier density of 2×10^{18} cm^{-3} and a differential gain of 5×10^{-16} cm^2. Assume a spontaneous emission bandwidth of 100 nm. At transparency the Auger recombination rate equals the spontaneous recombination rate, and other nonradiative terms can be neglected.

(a) Plot the $P-I$ characteristic, labeling the threshold current, the spontaneous emission power into the mode at threshold, and the differential efficiency above threshold.

(b) Plot the small-signal frequency response for a bias current of twice threshold.

17. Using Eq. (A6.25) calculate the gain 50 meV above the band edge in GaAs as a function of $(f_2 - f_1)$. (Assume $\tau_{sp}^{21} = 0.3$ ns, and consider only the heavy-hole band.)

CHAPTER THREE

Mirrors and Resonators for Diode Lasers

3.1 INTRODUCTION

The analysis in Chapter 2 formally assumed a simple two-mirror Fabry–Perot cavity for the axial photon confinement and some waveguide for the lateral and the transverse confinement. However, modern diode lasers use a variety of cavity structures, and the axial dimension is of great importance for their design. Therefore, in this chapter, we shall focus on axial confinement, which is generally provided by a slightly more complex mirror structure for feedback to form the desired resonant cavity. Together with the material from Chapter 1 and Appendix 3 on the transverse and lateral guiding mechanisms and structures, this chapter will form a full tool set for understanding of modern diode lasers. To illustrate the importance of the axial confinement, it is worth mentioning that in gas and solid-state lasers, the entire cavity may be defined by the axial mirrors because in many cases no lateral guiding structure is employed. By using proper axial cavity design, which includes some mode-selective filtering characteristics, we can produce single-mode as well as tunable single-mode diode lasers that are of great practical interest.

In this chapter, we first develop a scattering matrix formalism so the various structures can be analyzed rigorously and easily. The use of the associated transmission matrices reduces the analysis of axial structures with numerous impedance discontinuities to a mathematical exercise in matrix multiplication.

Diode Lasers and Photonic Integrated Circuits, Second Edition.
Larry A. Coldren, Scott W. Corzine, and Milan L. Mašanović.
© 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

Based on the scattering matrix formalism, we then successively explore several different axial geometries for diode lasers. Three- and four-mirror cavities are treated first. The mathematical condensation of the additional cavities and mirrors into a single effective complex mirror allows many of the basic formulae for Fabry–Perot lasers to be used. Because the mirror loss varies with wavelength, the loss of a single axial mode can be less than others. Thus, lasing in a single axial mode is possible. Also, if the phase in various laser sections can be changed independently, a tuning of the laser's wavelength can result.

Subsequently, we introduce the concept of Bragg grating mirrors. These mirrors are of interest for both in-plane and vertical-cavity lasers and form the basis for distributed Bragg reflector lasers. They consist of a series of relatively small impedance discontinuities along the axial propagation direction, phased so that the reflections add constructively at some (Bragg) frequency. Thus, the grating mirrors can provide a high level of net reflection when only small impedance differences exist, and because these gratings may be many wavelengths in length, the desired phasing can only occur over a narrow band of wavelengths to provide single-frequency operation. If substituted for a discrete mirror of a diode laser, a distributed Bragg reflector (DBR) laser results. In such cases the net complex grating reflection can replace the discrete mirror reflectivity in the Fabry–Perot formulae of Chapter 2. Multiple section DBR lasers have emerged as the most practical for wide wavelength tunability and have been successfully commercialized and deployed in modern fiber-optic networks. In the final part of this section, we deal with the spectral purity of axial mode selection that is possible with the aforementioned compound cavity lasers. The mode suppression ratio (MSR) is the ratio of the power out of the primary mode to the next largest mode. Experiments have shown that an MSR of at least 30 dB is necessary for single-frequency system applications.

Finally, in the last section, we consider gratings with gain. Lasers made with such gratings are called distributed feedback (DFB) lasers. They can be simpler to fabricate than DBRs because no transitions from active to passive regions are necessary, but their analysis is a little more complex. Fortunately, the transmission matrix formalism still works for complex propagation constants, so the threshold gains and wavelengths can still be obtained numerically. Because of their relative fabrication simplicity, DFB lasers have emerged as commercially the best choice for single-frequency operation. DFB lasers achieve high mode suppression ratios, and we investigate the influence of different DFB design parameters on their MSR. Wavelength tuning in DFBs is possible, but the range is limited as compared to the DBR.

3.2 SCATTERING THEORY

When working with complex laser cavities it is convenient to work with normalized amplitudes, a_j, which have a magnitude equal to the square root of the power flow and a phase equal to a selected observable such as the electric field. If we choose to reference the phase to the electric field, which we have written as

$$\mathscr{E}(x,y,z,t) = \hat{\mathbf{e}} E_0 U(x,y) e^{j(\omega t - \tilde{\beta} z)}, \tag{3.1}$$

we can define

$$a_j = \frac{E_0}{\sqrt{2\eta_j}} e^{-j\tilde{\beta}z}, \tag{3.2}$$

where $\eta_j = 377\ \Omega/\bar{n}_j$ is the mode impedance (the ratio of the transverse electric to transverse magnetic field magnitudes of the mode). Thus, provided $\int |U|^2\, dx\, dy = 1$, we see that $a_j a_j^* = P_j^+$, the power flowing in the positive z-direction in the mode.

At some waveguide reference plane, in general, there are incident and reflected powers. We characterize the incident waves (or inputs) by a normalized amplitude, a_j, and the reflected waves (or outputs) by a normalized amplitude, b_j, where the j's refer to the reference plane or port in question. Thus, at port j, the net power flowing into the port is

$$P_j = a_j a_j^* - b_j b_j^*. \tag{3.3}$$

Note that the impedance can be different at each port, but the definitions are unchanged. This is one of the most important features of the normalized amplitudes. Figure 3.1 shows a multiport scattering junction with inputs and outputs at each port for reference.

If the outputs can be linearly related to the inputs, a matrix formalism can be developed to express the outputs as a weighted combination of the inputs:

$$b_i = \sum_j S_{ij} a_j, \tag{3.4}$$

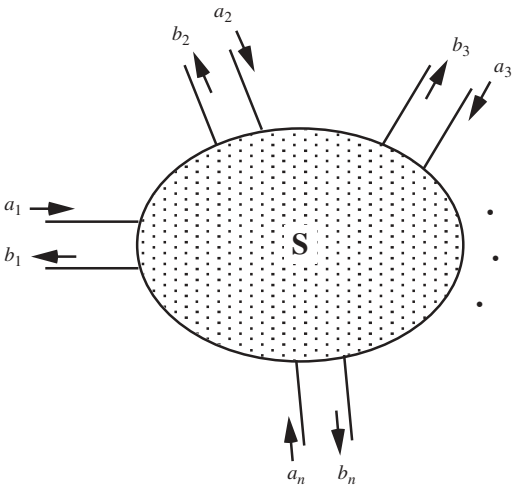


FIGURE 3.1: Generic scattering junction illustrating the inputs, a_j, and outputs, b_j, for the various ports.

where the S_{ij} are called the *scattering coefficients*. Note from Eq. (3.4) that to determine a particular S_{ij}, all inputs except a_j must be set to zero. That is,

$$S_{ij} = \left.\frac{b_i}{a_j}\right|_{a_k=0, k\neq j}. \tag{3.5}$$

This is equivalent to terminating all ports in their characteristic impedance to prevent reflections back into the network.

More generally, $\mathbf{b} = \mathbf{Sa}$, where $\mathbf{a}$ and $\mathbf{b}$ are column vectors and $\mathbf{S}$ is a matrix. For example, for a two-port scattering junction, such as a partially transmissive mirror, we have

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}. \tag{3.6}$$

The scattering coefficients are particularly useful because they have direct physical significance. All represent the ratio of a normalized output amplitude to a normalized input amplitude. The diagonal elements of the matrix are the respective complex amplitude reflection coefficients. For example in the two-port, we have previously referred to S_{11} and S_{22} as simply r_1 and r_2. The power reflection coefficients in this case are $|S_{11}|^2$ and $|S_{22}|^2$, respectively. The off-diagonal terms represent the complex (amplitude and phase) output at one port due to the input at another. Thus, they are really *transfer functions*. In all cases the magnitude squared of a scattering coefficient, $|S_{ij}|^2$, represents the fraction of power appearing at the port i due to the power entering port j.

Scattering matrices may have a number of interesting properties if the networks they describe satisfy certain criteria. For example, the scattering matrix of a linear reciprocal system is symmetric. That is for a two-port, $S_{12} = S_{21}$. For a lossless two-port, power conservation yields $|S_{11}|^2 + |S_{21}|^2 = 1$, and $|S_{22}|^2 + |S_{12}|^2 = 1$. Also, the scattering matrix of a lossless system is unitary.

Another important matrix that relates the normalized amplitudes is the *transmission matrix*. The transmission matrix expresses the inputs and outputs at a given port in terms of those at the others. In the case of a two-port, it is used most often to cascade networks together because simple matrix multiplication can be used (as we will see in a moment). Referring to Fig. 3.2a, the transmission matrix of a two-port is defined as

$$\begin{bmatrix} A_1 \\ B_1 \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \begin{bmatrix} A_2 \\ B_2 \end{bmatrix}, \tag{3.7}$$

where instead of using the input and output amplitudes, a_i and b_i, defined earlier, we have chosen to denote right or forward-going waves as A_i, and left or backward-going waves as B_i. From Fig. 3.2a, the correspondence between the **T**-matrix and **S**-matrix amplitudes is as follows: $A_1 = a_1, B_1 = b_1, A_2 = b_2$, and $B_2 = a_2$. This change of notation is convenient when cascading two-port networks in a serial

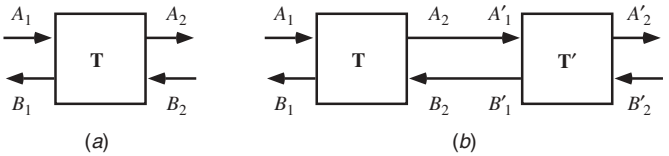


FIGURE 3.2: (a) Single two-port network. (b) Two networks cascaded together.

chain. For example, in Fig. 3.2b, port #1 of the second is connected to port #2 of the first. By equating $A_2 = A_1'$ and $B_2 = B_1'$, we can relate the fields on the left side of the overall structure, A_1 and B_1, to the fields on the right side, A_2' and B_2', as follows:

$$\begin{bmatrix} A_1 \\ B_1 \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \begin{bmatrix} A_2 \\ B_2 \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \begin{bmatrix} T_{11}' & T_{12}' \\ T_{21}' & T_{22}' \end{bmatrix} \begin{bmatrix} A_2' \\ B_2' \end{bmatrix}. \tag{3.8}$$

This process can be continued to obtain the net transmission matrix of arbitrarily complex multisection waveguide devices. The **T**-matrix can be obtained directly from the **S**-matrix using the following:

$$\begin{aligned} T_{11} &= \frac{1}{S_{21}}, \qquad T_{12} = -\frac{S_{22}}{S_{21}}, \\ T_{21} &= \frac{S_{11}}{S_{21}}, \qquad T_{22} = -\frac{S_{11}S_{22} - S_{12}S_{21}}{S_{21}}. \end{aligned} \tag{3.9}$$

Definitions and relations between the **S**- and **T**-matrices are summarized in Table 3.1.

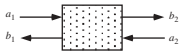
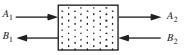
As mentioned earlier, various network properties allow us to specify relationships *between* the matrix coefficients, allowing us to reduce the total number of independent parameters. For example, the *normalized* fields of a system that satisfy Maxwell's equations with scalar ε and μ are known to obey reciprocity, which simply put, means that the scattering matrix is equal to its transpose, or that $S_{12} = S_{21}$.

In addition to being reciprocal (a property of any linear network), the network might also be lossless. If this is the case, then other simplifying relations can be derived. For example, it can be shown that $T_{12} = T_{21}^*$ and $T_{22} = T_{11}^*$, which simplifies the determination of the **T**-matrix significantly. Table 3.2 summarizes these relations for various network properties.

3.3 S AND T MATRICES FOR SOME COMMON ELEMENTS

The utility of the **S**- and **T**-matrices should become clearer as we consider a few common "scattering junctions." As we shall see, these form the basis of many more complex waveguide networks encountered in diode lasers. The further development of these matrices for more complex photonic integrated circuits, such as ones including directional couplers, will be left to a later chapter.

TABLE 3.1: Relations Between Scattering and Transmission Matrices

Scattering Matrix	Transmission Matrix
Definition	*Definition*
	
$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$	$\begin{bmatrix} A_1 \\ B_1 \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \begin{bmatrix} A_2 \\ B_2 \end{bmatrix}$
$b_1 = S_{11}a_1 + S_{12}a_2$	$A_1 = T_{11}A_2 + T_{12}B_2$
$b_2 = S_{21}a_1 + S_{22}a_2$	$B_1 = T_{21}A_2 + T_{22}B_2$
Relation to r and t	*Relation to r and t*
$r_{12} = \left.\frac{b_1}{a_1}\right\|_{a_2=0} = S_{11}$	$r_{12} = \left.\frac{B_1}{A_1}\right\|_{B_2=0} = \frac{T_{21}}{T_{11}}$
$t_{12} = \left.\frac{b_2}{a_1}\right\|_{a_2=0} = S_{21}$	$t_{12} = \left.\frac{A_2}{A_1}\right\|_{B_2=0} = \frac{1}{T_{11}}$
$r_{21} = \left.\frac{b_2}{a_2}\right\|_{a_1=0} = S_{22}$	$r_{21} = \left.\frac{A_2}{B_2}\right\|_{A_1=0} = -\frac{T_{12}}{T_{11}}$
$t_{21} = \left.\frac{b_1}{a_2}\right\|_{a_1=0} = S_{12}$	$t_{21} = \left.\frac{B_1}{B_2}\right\|_{A_1=0} = \frac{\det \mathbf{T}}{T_{11}}$
$\mathbf{S} = \begin{bmatrix} r_{12} & t_{21} \\ t_{12} & r_{21} \end{bmatrix}$	$\mathbf{T} = \frac{1}{t_{12}} \begin{bmatrix} 1 & -r_{21} \\ r_{12} & t_{12}t_{21} - r_{12}r_{21} \end{bmatrix}$
$\det \mathbf{S} = S_{11}S_{22} - S_{12}S_{21} = r_{12}r_{21} - t_{12}t_{21}$	$\det \mathbf{T} = T_{11}T_{22} - T_{12}T_{21} = t_{21}/t_{12}$
Relation to **T***-Matrix*	*Relation to* **S***-Matrix*
$\mathbf{S} = \frac{1}{T_{11}} \begin{bmatrix} T_{21} & \det \mathbf{T} \\ 1 & -T_{12} \end{bmatrix}$	$\mathbf{T} = \frac{1}{S_{21}} \begin{bmatrix} 1 & -S_{22} \\ S_{11} & -\det \mathbf{S} \end{bmatrix}$

3.3.1 The Dielectric Interface

Figure 3.3 illustrates the normalized amplitudes at a dielectric interface. The media are characterized by indices of refraction n_1 and n_2.

In any such problem, one has the freedom to select reference planes for each port. The phase of the scattering coefficients will clearly depend on the location of such planes. Some feel that it is important to make such an asymmetric problem have a symmetric scattering matrix by carefully selecting the reference planes, so that **S** will satisfy certain mathematical niceties. However, we believe that tends to create confusion, and it certainly obscures the physics of the problem. Thus, we shall always attempt to select "natural" reference planes at physical boundaries.

In the present case we select both reference planes at the physical interface between the two dielectrics, so that the scattering junction has zero length. Using

TABLE 3.2: Network Properties and Their Consequences on the Matrix Coefficients

Reciprocal Network (*valid for normalized fields with and without loss*)

$$\mathbf{S}_t = \mathbf{S} \rightarrow \begin{matrix} S_{12} = S_{21} \\ \det \mathbf{T} = 1 \end{matrix}$$

$$\mathbf{S} = \begin{bmatrix} S_{11} & S_{21} \\ S_{21} & S_{22} \end{bmatrix} = \frac{1}{T_{11}} \begin{bmatrix} T_{21} & 1 \\ 1 & -T_{12} \end{bmatrix}$$

$$\mathbf{T} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & (T_{12}T_{21} + 1)/T_{11} \end{bmatrix} = \frac{1}{S_{21}} \begin{bmatrix} 1 & -S_{22} \\ S_{11} & S_{21}^2 - S_{11}S_{22} \end{bmatrix}$$

Lossless Reciprocal Network

$$|S_{11}|^2 + |S_{21}|^2 = 1 \qquad |T_{21}|^2 + 1 = |T_{11}|^2$$

$$\mathbf{S}_t^*\mathbf{S} = 1 \rightarrow |S_{12}|^2 + |S_{22}|^2 = 1 \rightarrow 1 + |T_{12}|^2 = |T_{11}|^2$$

$$S_{11}^* S_{12} + S_{21}^* S_{22} = 0 \qquad T_{21}^* - T_{12} = 0$$

$$\mathbf{S} = \begin{bmatrix} S_{11} & S_{21} \\ S_{21} & -S_{11}^*(S_{21}/S_{21}^*) \end{bmatrix} = \frac{1}{T_{11}} \begin{bmatrix} T_{21} & 1 \\ 1 & -T_{21}^* \end{bmatrix}$$

$$\mathbf{T} = \begin{bmatrix} T_{11} & T_{21}^* \\ T_{21} & T_{11}^* \end{bmatrix} = \begin{bmatrix} 1/S_{21} & S_{11}^*/S_{21}^* \\ S_{11}/S_{21} & 1/S_{21}^* \end{bmatrix}$$

Lossless Reciprocal Network with r and t Phase Shifts of 0 *or* π

$$\begin{matrix} S_{22} = -S_{11} \\ S_{11} = S_{11}^* \\ S_{21} = S_{21}^* \end{matrix} \rightarrow \det \mathbf{S} = -1$$

$$T_{22} = T_{11}, \quad T_{12} = T_{21}$$

$$\mathbf{S} = \begin{bmatrix} S_{11} & S_{21} \\ S_{21} & -S_{11} \end{bmatrix} = \frac{1}{T_{11}} \begin{bmatrix} T_{21} & 1 \\ 1 & -T_{21} \end{bmatrix}$$

$$\mathbf{T} = \begin{bmatrix} T_{11} & T_{21} \\ T_{21} & T_{11} \end{bmatrix} = \frac{1}{S_{21}} \begin{bmatrix} 1 & S_{11} \\ S_{11} & 1 \end{bmatrix}$$

Eq. (3.5), this leads to

$$S_{11} = \left.\frac{b_1}{a_1}\right|_{a_2=0} = -r_1 = \frac{n_1 - n_2}{n_1 + n_2}, \tag{3.10}$$

where for the last equality, we have assumed normally incident plane waves. This is approximately true for weakly guided dielectric waveguide modes, but n_1 and n_2 should be replaced by the effective indices of the modes, $\bar{n}_1$ and $\bar{n}_2$. Note that we have chosen to label the reflection from the n_1 side $-r_1$ rather than $+r_1$ in Fig. 3.3. The reason is for compatibility with other calculations in this book. Also, if $n_2 > n_1, r_1$ would be a positive real number.

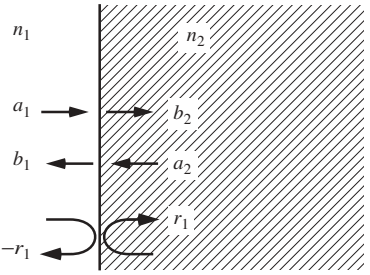


FIGURE 3.3: Interface between two dielectrics illustrating reference planes at interface for both ports.

Similarly, for the second port

$$S_{22} = \left.\frac{b_2}{a_2}\right|_{a_1=0} = r_2 = -(-r_1), \tag{3.11}$$

and

$$S_{12} = S_{21} = t = \sqrt{1 - r_1^2}, \tag{3.12}$$

where we have used power conservation for Eq. (3.12). This is clearly valid for plane waves because there can be no loss in zero length, but for waveguide modes, power conservation also implies that the transverse mode profiles are equal. (If the waveguide modes do not match, there can be scattering loss. This will be taken up in a later chapter.) For the normally incident plane wave case, $t = 2(n_1 n_2)^{1/2}/(n_1 + n_2)$.

Thus, the complete scattering matrix for the dielectric interface can be written as

$$\mathbf{S} = \begin{bmatrix} -r_1 & t \\ t & r_1 \end{bmatrix}, \tag{3.13}$$

where again, the sign of r_1 follows the convention in Fig. 3.3. The corresponding **T**-matrix is

$$\mathbf{T} = \frac{1}{t}\begin{bmatrix} 1 & -r_1 \\ -r_1 & 1 \end{bmatrix}. \tag{3.14}$$

3.3.2 Transmission Line with No Discontinuities

Figure 3.4 shows a network that is a length of waveguide, L, in which there are no discontinuities. In fact, this network consists only of two reference planes on a waveguide. Thus, the problem is to find how to relate variables from one reference plane to another where there are no scattering junctions in between.

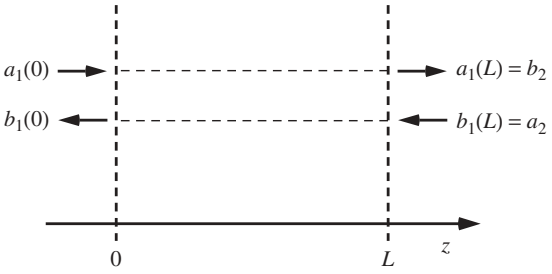


FIGURE 3.4: Transmission line section of length L.

In such cases it is common to express the normalized amplitudes as a function of distance as defined by Eq. (3.2). Thus, if the origin is put at port #1, then $b_2 = a_1(L)$ and $a_2 = b_1(L)$, as indicated in Fig. 3.4. Because there is no coupling between the waves propagating in the forward and backward directions, $S_{11} = S_{22} = 0$. From Eq. (3.2), we have $a_1(z) = a_1(0)e^{-j\tilde{\beta}z}$ and $b_1(z) = b_1(0)e^{j\tilde{\beta}z}$, for the forward and backward modes. Thus,

$$b_2 = a_1(L) = a_1(0)e^{-j\tilde{\beta}L} = a_1 e^{-j\tilde{\beta}L}, \tag{3.15}$$

and

$$a_2 = b_1(L) = b_1(0)e^{j\tilde{\beta}L} = b_1 e^{j\tilde{\beta}L}.$$

Forming S_{12} and S_{21}, we find that

$$S_{12} = S_{21} = e^{-j\tilde{\beta}L}. \tag{3.16}$$

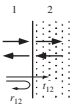
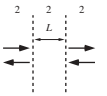
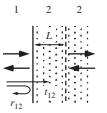
This is a very important result because it verifies that a propagation delay is the same for both forward and backward waves. In other words, the waveguide modes do not know how you have chosen the coordinate system. Mode propagation a distance L in any direction results in a phase shift of $-\beta L$ and a growth rate $\beta_i L$ for that mode, assuming that $\tilde{\beta} = \beta + j\beta_i$ as in Eq. (2.20). The scattering matrix is summarized as

$$\mathbf{S} = \begin{bmatrix} 0 & e^{-j\tilde{\beta}L} \\ e^{-j\tilde{\beta}L} & 0 \end{bmatrix}. \tag{3.17}$$

The corresponding **T**-matrix found using Eq. (3.9) is given by

$$\mathbf{T} = \begin{bmatrix} e^{j\tilde{\beta}L} & 0 \\ 0 & e^{-j\tilde{\beta}L} \end{bmatrix}. \tag{3.18}$$

TABLE 3.3: Summary of S- and T-matrices for Simple "Building-Block" Components

Scattering Matrix	Structure	Transmission Matrix
$\begin{bmatrix} r_{12} & t_{12} \\ t_{12} & -r_{12} \end{bmatrix}$		$\frac{1}{t_{12}}\begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix}$
$r_{21} = -r_{12}\ t_{21} = t_{12}$		$r_{12}^2 + t_{12}^2 = 1$
$\begin{bmatrix} 0 & e^{-j\phi} \\ e^{-j\phi} & 0 \end{bmatrix}$		$\begin{bmatrix} e^{j\phi} & 0 \\ 0 & e^{-j\phi} \end{bmatrix}$
$\phi = \tilde{\beta}_2 L$		
$\begin{bmatrix} r_{12} & t_{12}e^{-j\phi} \\ t_{12}e^{-j\phi} & -r_{12}e^{-j2\phi} \end{bmatrix}$		$\frac{1}{t_{12}}\begin{bmatrix} e^{j\phi} & r_{12}e^{-j\phi} \\ r_{12}e^{j\phi} & e^{-j\phi} \end{bmatrix}$
		$r_{12}^2 + t_{12}^2 = 1$

The basic component matrices are summarized in Table 3.3. For the **T**-matrix, lossless mirrors have been assumed in the first and third cases. To include the possibility of a lossy mirror, T_{22} for these two cases should be multiplied by $r_{12}^2 + t_{12}^2$. An advantage of the **T**-matrices is that more complicated structures can be constructed simply by matrix multiplying together the basic components shown in the table. For example, the third matrix in the list is easily constructed by matrix multiplying the first by the second.

3.3.3 Dielectric Segment and the Fabry–Perot Etalon

Figure 3.5 shows a dielectric block of length L and index n_2. To the left is a region of index, n_1, and to the right is a region of index n_3. We shall use r_1 and t_1 for the left interface and r_2 and t_2 for the right interface viewed from the central medium. (Thus, choosing the reference planes at the physical interfaces and assuming that $n_2 > n_1, r_1$ would be a positive real number.) Such a structure is known as a Fabry–Perot etalon.

As indicated in the diagram, we can look at this problem as three scattering networks cascaded—two dielectric interfaces and one transmission line. Thus, the problem could be solved by multiplying the **T**-matrices. This is to be shown in one of the problems at the end of the chapter. Here, however, we would like to demonstrate how to solve the system of normalized amplitudes to obtain the **S**-matrix directly. As we shall see as this chapter proceeds, once we have the **S**- and **T**-matrices of such a dielectric segment, most multisection diode laser problems can be solved by their repetitive use.

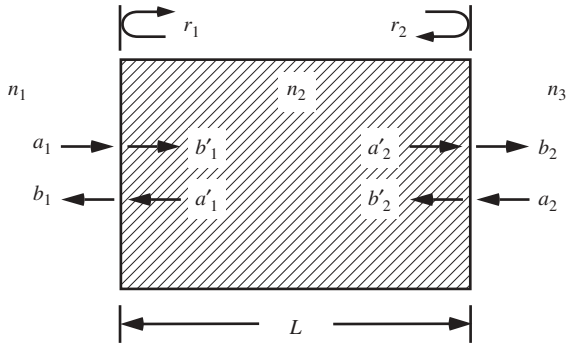


FIGURE 3.5: Dielectric block of length L.

Referring to Fig. 3.5 where primes are used for internal variables, and initially assuming that $a_2 = 0$, we can write the following relationships for the outputs, b_j:

$$\begin{aligned} b_1 &= -a_1 r_1 + a'_1 t_1, \\ b'_1 &= a_1 t_1 + a'_1 r_1, \\ b_2 &= a'_2 t_2, \\ b'_2 &= a'_2 r_2. \end{aligned}$$

Also, we can express

$$\begin{aligned} a'_1 &= b'_2 e^{-j\tilde{\beta}L}, \\ a'_2 &= b'_1 e^{-j\tilde{\beta}L}. \end{aligned} \tag{3.19}$$

Solving this system of equations for $S_{11} = b_1/a_1$ and $S_{21} = b_2/a_1$ (because $a_2 = 0$), we obtain

$$S_{11} = -r_1 + \frac{t_1^2 r_2 e^{-2j\tilde{\beta}L}}{1 - r_1 r_2 e^{-2j\tilde{\beta}L}}, \tag{3.20}$$

$$S_{21} = \frac{t_1 t_2 e^{-j\tilde{\beta}L}}{1 - r_1 r_2 e^{-2j\tilde{\beta}L}}. \tag{3.21}$$

Similarly, with $a_1 = 0$,

$$S_{22} = -r_2 + \frac{t_2^2 r_1 e^{-2j\tilde{\beta}L}}{1 - r_1 r_2 e^{-2j\tilde{\beta}L}}, \tag{3.22}$$

$$S_{12} = S_{21}. \tag{3.23}$$

The common factor of $1 - r_1 r_2 e^{-2j\tilde{\beta}L}$ gives rise to the characteristic resonances and antiresonances associated with Fabry–Perot etalons.

The corresponding **T**-matrix for the Fabry–Perot etalon can be obtained in a number of ways: (1) we can use the **S**-matrix coefficients and apply Eq. (3.9); (2) we can multiply the third and first matrices in Table 3.3, using the appropriate refractive indices; or (3) we can solve for the relevant ratios, $T_{11} = a_1/b_2$ and $T_{21} = b_1/b_2$, (with $a_2 = 0$); $T_{12} = -T_{11}b_2/a_2$ and $T_{22} = b_1/a_2 - T_{21}b_2/a_2$, (with $a_1 = 0$). With $a_2 = 0$,

$$T_{11} = \frac{1}{t_1 t_2}[e^{j\tilde{\beta}L} - r_1 r_2 e^{-j\tilde{\beta}L}], \tag{3.24}$$

$$T_{21} = -\frac{1}{t_1 t_2}[r_1 e^{j\tilde{\beta}L} - r_2 e^{-j\tilde{\beta}L}]. \tag{3.25}$$

Similarly, with $a_1 = 0$,

$$T_{12} = -\frac{1}{t_1 t_2}[r_1 e^{-j\tilde{\beta}L} - r_2 e^{j\tilde{\beta}L}]. \tag{3.26}$$

$$T_{22} = \frac{1}{t_1 t_2}[e^{-j\tilde{\beta}L} - r_1 r_2 e^{j\tilde{\beta}L}], \tag{3.27}$$

For the latter three T-parameters we have assumed lossless interfaces, $t_1^2 = 1 - r_1^2$ and $t_2^2 = 1 - r_2^2$, to simplify the expressions. For the S-parameters, Eqs. (3.20) to (3.23), the reflection and transmission at each mirror has been left in general form. Thus, these apply to any Fabry–Perot etalon of length L. For example, the dielectric interfaces could include loss, or they could be coated to enhance the reflectivities, r_1 and r_2, to values greater than that indicated by Eq. (3.10).

The absolute squares of S_{11} and S_{21} give the amount of power reflected by and transmitted through the Fabry–Perot etalon as a function of the wavelength of the incident field. Using Eqs. (3.20) and (3.21), and assuming $t_i^2 = 1 - r_i^2$, we obtain

$$|S_{11}|^2 = \frac{(r_1 - r_2 e^{2\beta_i L})^2 + 4R\sin^2\beta L}{(1-R)^2 + 4R\sin^2\beta L} \rightarrow \frac{4R\sin^2\beta L}{(1-R)^2 + 4R\sin^2\beta L},$$

$$|S_{21}|^2 = \frac{(1-r_1^2)(1-r_2^2)e^{2\beta_i L}}{(1-R)^2 + 4R\sin^2\beta L} \rightarrow \frac{(1-R)^2}{(1-R)^2 + 4R\sin^2\beta L},$$

where $R = r_1 r_2 e^{2\beta_i L}$ and $\tilde{\beta} = \beta + j\beta_i$. The arrows indicate the special case of a symmetric ($r_1 = r_2$), lossless ($\beta_i = 0$) Fabry–Perot cavity. Note that without loss, $|S_{11}|^2 + |S_{21}|^2 = 1$, as required by power conservation. Figure 3.6 plots both the magnitude and phase of S_{11} and S_{21} vs. $2\beta L$ using three different reflectivities. The periodic maxima in the transmission spectrum of S_{21} go to unity for this symmetric, lossless case, and become extremely sharp at high values of r. These maxima occur at the axial resonances or *modes* of the Fabry–Perot cavity where $e^{-2j\beta L} = |r_1 r_2|/r_1 r_2 = 1$. By power conservation, the minima of the net reflection,

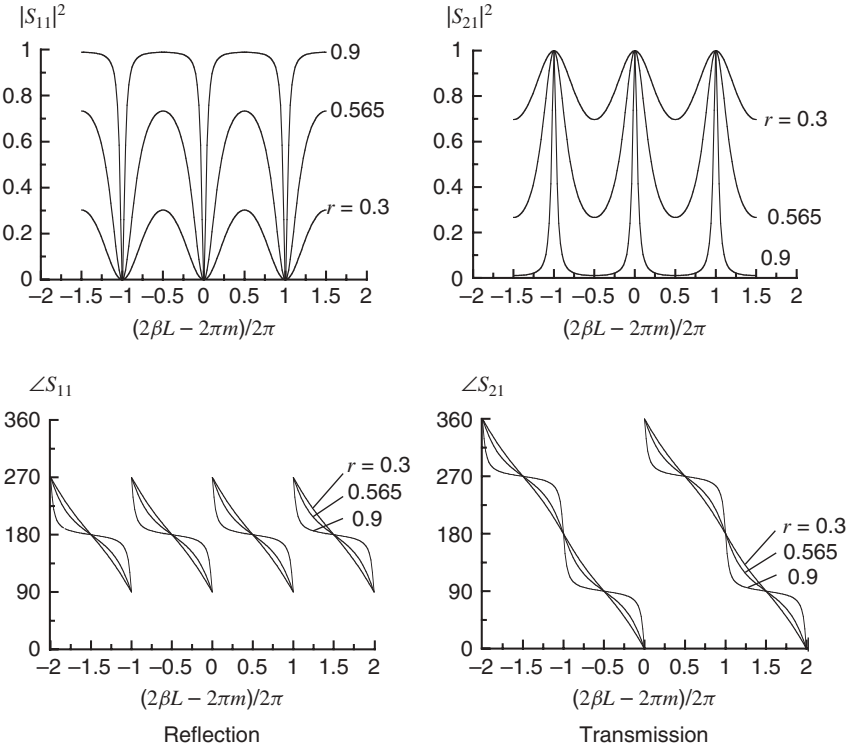


FIGURE 3.6: Magnitude and phase of the reflection (S_{11}) and transmission (S_{21}) coefficients of a Fabry–Perot etalon as a function of the cavity round-trip phase relative to an integer multiple of 2π (in the lower-right plot, the phase is relative to an *even* multiple of 2π). Three mirror reflectivities are plotted assuming $r = r_1 = r_2$ and zero loss (a cleaved-facet cavity corresponds to $r = 0.565$).

S_{11}, go to zero at the resonances in this case. If the cavity has loss, the maxima of S_{21} cannot reach unity. Of course, if the cavity has gain as in a laser, the maxima can be larger than unity. In the lossy case, it is also worth noting that S_{11} can still be adjusted to zero by reducing the reflectivity of the input mirror to be lower than the output mirror such that $r_1 = r_2 e^{-\alpha L}$. This is equivalent to setting the first term equal to the second in Eq. (3.20) on resonance. This asymmetric Fabry–Perot is useful in efficient optical modulators and detectors.

Example 3.1 A dielectric transmission line with an air interface is shown in Fig. 3.7.

Problem: Write the **T**-matrix parameters for this two port.

Solution: To illustrate the usefulness of **T**-matrix formalism, we will solve this problem by multiplying the **T**-matrices corresponding to the building blocks of

this structure. For the dielectric transmission line, the **T**-matrix is given by

$$\mathbf{T_1} = \begin{bmatrix} e^{j\phi_1} & 0 \\ 0 & e^{-j\phi_1} \end{bmatrix},$$

where $\phi_1 = \beta_1 L_1$. For the dielectric/air interface, we have

$$\mathbf{T_2} = \frac{1}{t_{12}} \begin{bmatrix} 1 & -r_{12} \\ -r_{12} & 1 \end{bmatrix},$$

where $-r_{12} = \frac{n_1 - n_2}{n_1 + n_2}$, and $r_{12}^2 + t_{12}^2 = 1$. Finally, the air segment can be described by the same matrix as the one for the dielectric segment, $\mathbf{T_3} = \mathbf{T_1}$, except that $\phi_2 = \beta_1 L_1$ in this case.

To get the full **T**-matrix of the system, we need to multiply through the matrices corresponding to the individual segments,

$$\mathbf{T} = \mathbf{T_1} \cdot \mathbf{T_2} \cdot \mathbf{T_3} = \begin{bmatrix} e^{j\phi_1} & 0 \\ 0 & e^{-j\phi_1} \end{bmatrix} \cdot \frac{1}{t_{12}} \begin{bmatrix} 1 & -r_{12} \\ -r_{12} & 1 \end{bmatrix} \cdot \begin{bmatrix} e^{j\phi_2} & 0 \\ 0 & e^{-j\phi_2} \end{bmatrix}$$

$$= \frac{1}{t_{12}} \begin{bmatrix} e^{j(\phi_1+\phi_2)} & -r_{12}e^{j(\phi_1-\phi_2)} \\ -r_{12}e^{j(\phi_2-\phi_1)} & e^{-j(\phi_1+\phi_2)} \end{bmatrix}.$$

The **T**-matrix can be converted to an **S**-matrix using the relationships between corresponding elements described in Table 3.2. This exercise will be useful when we start dealing with periodic grating sections and DFB lasers later in this Chapter.

3.3.4 S-Parameter Computation Using Mason's Rule

While **S**- and **T**-matrices work really well for performing the numerical analysis of complex dielectric structures, for cases where an analytic solution is prefered, in place of matrix multiplication, it is generally easier to apply a technique from control systems engineering, Mason's rule [2], which allows for relatively simple reduction of signal-flow graphs to single transfer functions relating the output of a

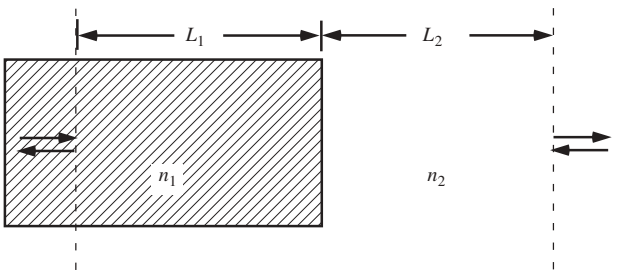


FIGURE 3.7: Dielectric transmission line with air interface, used in Example 3.1.

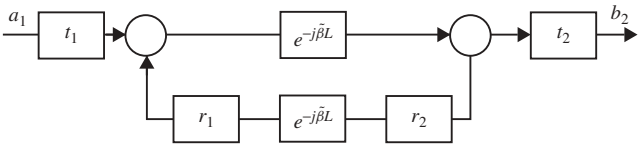


FIGURE 3.8: The signal flow graph for a Fabry–Perot segment.

system to its input. In particular, looking at the **S**-matrix formalism, Mason's rule will allow us to directly obtain the expressions for each of the S-parameters.

Mason's rule states that the transfer function of a system is given by

$$H = \frac{\sum_k T_k \Delta_k}{\Delta}, \tag{3.28}$$

where k is the number of forward paths for the signal, T_k is the kth forward path gain, $\Delta = 1\text{-}\sum$ loop gains $+ \sum$ nontouching loop gains taken two at a time $- \sum$ nontouching loop gains taken three at a time $+ \sum$ nontouching loop gains taken four at a time …, and Δ_k is formed by eliminating from Δ those loop gains that touch the kth forward path.

To illustrate the use and effectivness of Mason's rule, we will apply it to the example of a Fabry–Perot segment that was just analyzed in the previous section. The signal flow graph for S_{21} of a Fabry–Perot segment is shown in Figure 3.8. In this simple case, we only have one closed loop—therefore,

$$\Delta = 1 - r_1 r_2 e^{-2j\tilde{\beta}L}.$$

We also have only one forward path, therefore,

$$T_1 = t_1 t_2 e^{-j\tilde{\beta}L}, \quad \text{and} \quad \Delta_1 = 1.$$

Therefore,

$$H = S_{21} = \frac{t_1 t_2 e^{-j\tilde{\beta}L}}{1 - r_1 r_2 e^{-2j\tilde{\beta}L}},$$

which is the same result we've obtained previously, Eq. (3.21). Although this result is relatively straightforward, the true utility of Mason's rule is when the system becomes somewhat more complex, as we shall encounter in future examples, in Chapters 6 and 8.

3.3.5 Fabry–Perot Laser

In Chapter 2 we analyzed the use of a Fabry–Perot cavity to form a diode laser. Based on an intuitive argument, we labeled threshold as the point where the net round-trip gain equaled the net round-trip loss for an axial mode. We now can

argue this point more rigorously using the S-parameters. For lasing we must have coherent light being emitted for no optical inputs. Thus, the threshold for lasing must correspond to a pole of the S-parameters, which give the ratio of outputs over inputs. Referring to Eqs. (3.20) to (3.23), we observe that all the S-parameters have a factor $[1 - r_1 r_2 e^{-2j\tilde{\beta}L}]$ in the denominator of one term. Setting this factor to zero gives the needed pole. This definition of threshold is therefore equivalent to Eq. (2.21).

For lossless mirrors the differential efficiency is given by Eq. (2.40), and the power out is given by Eq. (2.41). However, we have not solved for the relative power out of each end of the laser. We now can determine the ratio of the powers out of each end, P_{01}/P_{02}, because it equals $|b_1/b_2|^2$. Looking at Fig. 3.5 with $a_1 = a_2 = 0$, we can see that $b_1 = a_1' t_1, b_2 = a_2' t_2$, and $a_1' = a_2' r_2 e^{-j\tilde{\beta}L}$. For a mode above threshold, $\text{Im}\{\tilde{\beta}\} = (\Gamma_{xy} g_{th} - \alpha_i)/2$. Thus,

$$\frac{P_{01}}{P_{02}} = \left|\frac{b_1}{b_2}\right|^2 = \frac{t_1^2}{t_2^2} r_2^2 e^{(\Gamma_{xy} g_{th} - \alpha_i)L}. \tag{3.29}$$

But, $\exp(\Gamma_{xy} g_{th} - \alpha_i)L = 1/(r_1 r_2)$ from Eq. (2.22). Therefore,

$$\frac{P_{01}}{P_{02}} = \frac{t_1^2 r_2}{t_2^2 r_1}. \tag{3.30}$$

The more general result for more complex cavities can be obtained directly from either the **T**-matrix or **S**-matrix coefficients using $P_{01}/P_{02} = |T_{21}|^2 = |S_{11}/S_{21}|^2$ evaluated at threshold.

Now if the mirrors are *lossy*, such that $r_j^2 + t_j^2 \neq 1$, we are more interested in the fraction, F_1, of *power delivered* from end 1, P_{01}, relative to the *total coupled out of the cavity* by the mirrors, P_m. This is because it is F_1 that must multiply the differential quantum efficiency and power expressions of Chapter 2, which assumed that all mirror loss was delivered to the outside. Referring again to Fig. 3.5 with $a_1 = a_2 = 0$, we construct the desired ratio, $F_1 = P_{01}/P_m = |b_1|^2/[|a_1'|^2(1 - r_1^2) + |a_2'|^2(1 - r_2^2)]$, which gives

$$F_1 = \frac{t_1^2}{(1 - r_1^2) + \frac{r_1}{r_2}(1 - r_2^2)}. \tag{3.31}$$

Therefore, using Eq. (2.40), the differential quantum efficiencies for light delivered out of end 1 and end 2 of the general Fabry–Perot laser are

$$\eta_{d1} = F_1 \eta_i \frac{\alpha_m}{\langle \alpha_i \rangle + \alpha_m},$$

and

$$\eta_{d2} = F_2 \eta_i \frac{\alpha_m}{\langle \alpha_i \rangle + \alpha_m}, \tag{3.32}$$

where F_1 is given by Eq. (3.31) and F_2 is found by switching subscripts in F_1. From the above, and Eq. (2.41) the power out of the jth port is

$$P_{0j} = \eta_{dj} \frac{h\nu}{q} (I - I_{th}). \tag{3.33}$$

For a symmetrical laser cavity with $r_1 = r_2 = r$ and $t_1 = t_2 = t$, the fractions reduce to

$$F_1 = F_2 = \frac{1}{2} \frac{t^2}{1 - r^2}. \tag{3.34}$$

Therefore, any loss occurring at the mirrors (other than what is coupled as useful output) is taken into account through F_1 and F_2.

3.4 THREE- AND FOUR-MIRROR LASER CAVITIES

Many modern diode laser configurations have at least one additional discontinuity within their cavities. In Fig. 2.6 and the associated derivations, we neglected any reflection from the interface between the active and passive sections. Here we shall discuss first one and then two discontinuities within the laser. The additional discontinuities may be within the semiconductor material or one may be at the semiconductor—air interface in an external cavity configuration.

3.4.1 Three-Mirror Lasers

Figure 3.9 gives a schematic of a three-mirror laser. The transmission coefficient across the interface, t_2, includes any scattering or coupling loss, so $r_2^2 + t_2^2 \neq 1$, in general. Generally, our analysis will hold for active–active as well as active–passive devices, but the results presented are only meaningful if the first cavity (labeled active) is the *dominant* cavity, which provides most of the gain.

Also shown in Fig. 3.9 is an equivalent two-mirror cavity, which replaces the passive section by an *effective mirror* with reflectivity, r_{eff}. This substitution is valid for steady-state analyses, but it will not necessarily properly model the compound cavity for dynamic operation. The value of r_{eff} was derived earlier, because it is really S_{11} for the passive section as viewed from the active section.

Using the reference planes and mirror reflectivities defined in Fig. 3.9 and Eq. (3.20), we find that

$$r_{eff} = r_2 + \frac{t_2^2 r_3 e^{-2j\tilde{\beta}_p L_p}}{1 + r_2 r_3 e^{-2j\tilde{\beta}_p L_p}}. \tag{3.35}$$

The complete characteristic equation for a three-mirror laser may be constructed by replacing r_2 by r_{eff} in Eq. (2.21) (or by solving for the poles of an active-cavity S-parameter with this substitution). However, this leads to a fairly complex equation.

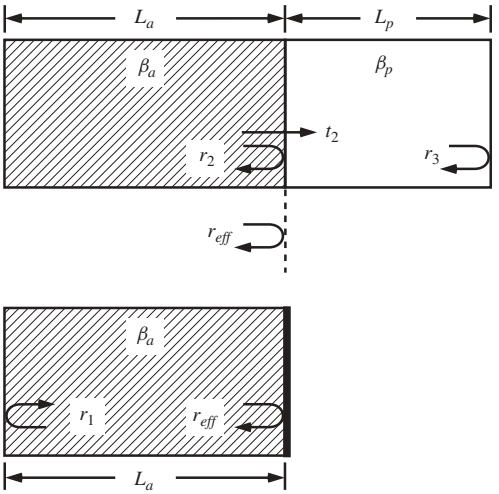


FIGURE 3.9: External cavity laser and equivalent cavity with effective mirror to model the external section.

We can obtain most of the information we want by using the equivalent cavity in Fig. 3.9 and carrying r_{eff} along in the threshold calculation. This results in only a slight modification to the threshold gain expression of Eq. (2.25). That is, the threshold gain of our three-mirror laser can be written as

$$\Gamma g_{th} = \langle \alpha_i \rangle_a + \frac{1}{L_a} \ln \left[\frac{1}{r_1 |r_{eff}|} \right], \tag{3.36}$$

where Γ and $\langle \alpha_i \rangle_a$ average over the active section of the cavity only (any losses encountered in the passive section are contained in r_{eff}). To complete the model, we need to specify the threshold condition for the round-trip phase. With $r_{eff} = |r_{eff}| e^{j\phi_{eff}}$ and r_1 positive and real, the round-trip phase must satisfy $e^{-2j\beta_a L_a} e^{j\phi_{eff}} = 1$, which translates into $2\beta_a L_a - \phi_{eff} = 2\pi m$. Taking derivatives of all variables dependent on frequency, we obtain

$$d\beta_a L_a - \tfrac{1}{2} d\phi_{eff} = \pi \, dm. \tag{3.37}$$

The spacing between adjacent modes is found by setting $dm = 1$ and solving for $d\beta_a$:

$$d\beta_a = \frac{\pi}{L_a - \frac{1}{2} d\phi_{eff}/d\beta_a}. \tag{3.38}$$

In this expression, the cavity length that defines the mode spacing includes the active section length plus an additional factor dependent on how the effective mirror phase changes with frequency, motivating us to consider the second quantity as

an effective length. However, because ϕ_{eff} explicitly depends on β_p, not β_a, we choose to define the effective length as

$$L_{eff} = -\frac{1}{2}\frac{d\phi_{eff}}{d\beta_p}. \tag{3.39}$$

The mode spacing can be defined in terms of wavelength, $d\beta_a = -d\lambda(2\pi/\lambda^2)\bar{n}_{ga}$; or in terms of frequency, $d\beta_a = d\nu(2\pi/c)\bar{n}_{ga}$, where $\bar{n}_g = \bar{n} - \lambda\partial\bar{n}/\partial\lambda$. Furthermore, it follows that $d\beta_p/d\beta_a = \bar{n}_{gp}/\bar{n}_{ga}$. Using these expressions in Eq. (3.38), the mode spacing in either wavelength or frequency becomes

$$d\lambda = \frac{\lambda^2}{2(\bar{n}_{ga}L_a + \bar{n}_{gp}L_{eff})} \quad \text{or} \quad d\nu = \frac{c}{2(\bar{n}_{ga}L_a + \bar{n}_{gp}L_{eff})}. \tag{3.40}$$

If no reflection exists at the active–passive interface ($r_2 = 0$) and r_3 is positive and real, then $\phi_{eff} = -2\beta_p L_p$, and $L_{eff} = L_p$, reducing Eq. (3.40) to Eq. (2.29) given earlier. For the more general Fabry–Perot etalon, the slope of the phase will be dependent on whether we are near a resonance or an antiresonance of the etalon (as shown earlier in Fig. 3.6). Thus, L_{eff} can be larger or smaller than L_p. However, if the phase varies rapidly and nonlinearly within the range of one mode spacing (for example, near the Fabry–Perot resonances ($\beta L = m\pi$) in Fig. 3.6), then Eq. (3.40) will most likely not be very accurate because this derivation assumes that ϕ_{eff} varies linearly over at least one mode spacing.

The differential quantum efficiency and power out of end 1 are given by Eqs. (3.32) and (3.33), respectively, using Eq. (3.31). The second mirror reflectivity, r_2, in these single-section laser expressions should be replaced by r_{eff} wherever it shows up, and the mirror loss, α_m, is given by the second term in Eq. (3.36).

Figure 3.6 gives plots of the magnitude of r_{eff} in the special case where $r_2 = -r_3$ and loss can be neglected. As can be seen, the magnitude of the reflectivity of this mirror can vary significantly, and this will provide a filtering effect on the cavity modes. As indicated by Eq. (3.36), the modes with the lowest loss or highest mirror reflectivity will tend to lase first. Thus, such a second section or etalon can be used to filter out unwanted modes. However, a point often confused is that the maxima of r_{eff} always occur at the antiresonances of the etalon. Thus, in the three-mirror configuration, having a high-Q external cavity actually leads to worse mode selectivity because the maxima become very flat in this case. In fact, there is an optimum value of net external cavity loss that provides the largest curvature at the maxima of r_{eff}. Figure 3.6 is actually not a very practical case because there is usually some loss both in traversing the passive cavity and in coupling back into the active section, and generally $r_2 \neq r_3$. Thus, the minima do not tend to be as deep, and the maxima have more shape. This is one case in nature where loss seems to help.

External cavities are not very useful for axial mode selection in the VCSEL case because their short cavities together with the finite gain bandwidth usually provides single-axial mode operation. Here lateral modes are the larger problem.

However, for in-plane lasers $\geq$100 μm in length, several axial modes will exist near the gain maximum, and the loss modulation caused by the external cavity is useful for axial mode selection.

The length of the external cavity (etalon) determines how modes will be selected. Ideally, the periodic loss modulation will combine with the gain roll-off to provide a single net gain maximum where the nearest axial mode will be selected. Figure 3.10 illustrates the three cases where the relative length, L_p/L_a is (a) $\ll$1, (*b*) $\sim$1 (but $\neq$1), and (c) $\gg$1.

Here we show the variations of α_m, from Eq. (3.36), and a generic net gain curve, $(\Gamma_{xy} g - \alpha_i)$ vs. wavelength along with indications of the mode locations. (Note that the maxima in r_{eff} correspond to minima in α_m.) The active cavity length and the width of the gain peak are held constant for all three cases. When the gain reaches the loss at some point, the mode at that wavelength reaches threshold. Also, we know that once one mode reaches threshold, the gain becomes clamped, and if the loss margin (the loss minus the gain) is large enough, the other modes are suppressed.

If the external cavity is somewhat shorter than the active section, as in case (a), the modes of the active cavity will be more closely spaced than the minima in α_m. In this case a single loss minimum can effectively select a single axial mode of the active cavity, if α_m varies enough. That is, the period of the loss modulation cannot be so large that the minimum of α_m is as wide as the gain peak, or no additional filtering will be provided. On the other hand, it must be large enough so that the next minima of α_m fall sufficiently far off the gain peak. (If too close, these secondary minima may select unwanted repeat modes.)

If the lengths are comparable as in Fig. 3.10b, the resonances of both cavities are spaced by about the same amount, and the active cavity modes will slowly slide across the minima of α_m providing an action similar to a vernier scale. Again, relatively good mode suppression is possible if the beat period is not too large or too small. In the third case, Fig. 3.10c, good mode suppression is generally not possible unless the external cavity mirror itself is a filter. In fact, a grating mirror is sometimes used to provide for single-frequency operation of the laser with a long external cavity.

One of the uses of a two-section (three-mirror) laser is to provide a tunable single-frequency source. The repeat modes can make this somewhat problematic, but the tuning mechanism is still worth reviewing. If the passive section shown in Fig. 3.9 is formed of electro-optic material, it would be possible to change its index and the round-trip phase, $-\beta_p L_p$, by applying an electric field across the material. Then, according to Eq. (3.35), the r_{eff} and α_m characteristics would tune in wavelength. This would cause a successive selection of different axial modes, as can be envisioned from Fig. 3.10. Moreover, because the phase of r_{eff} also varies across each period, some continuous tuning of the cavity mode is possible prior to the shift to a new mode. Perhaps one of the key reasons to have the tunability is to provide an active mechanism to optimally align the loss minimum with an axial mode for best spurious mode suppression.

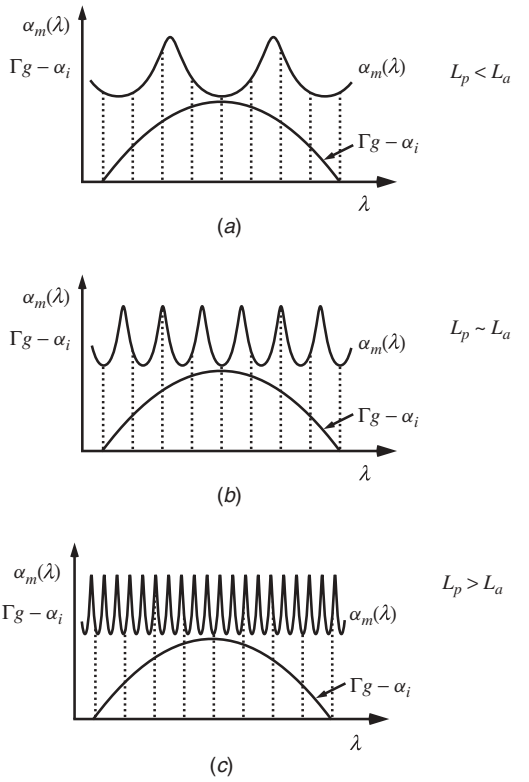


FIGURE 3.10: Schematic illustration of net propagation gain, $\Gamma_{xy} g - \alpha_i$, and the net mirror loss, α_m, as a function of wavelength for external cavities with length (a) shorter, (b) about the same, and (c) longer than the active section.

3.4.2 Four-Mirror Lasers

If the r_{eff} characteristic could be flipped over so that its maxima (α_m minima) also occurred at the resonances of the external cavity, the filtering action would be much better. As stated earlier this is not possible in a three-mirror cavity, but with a *four-mirror cavity* it is. However, the relative positions of the reflectors must provide the desired phasing. The best known example of a four-mirror laser is the coupled-cavity laser, which incorporates a narrow space to separate *two active sections*. This same effect can also be achieved by using coatings on the facet of the active section in an external-cavity laser.

To model the four-mirror laser, we replace the interface between the active and passive sections in Fig. 3.9 by another section, which can be fully described by another complex scattering matrix. (Again, this analysis also holds for the active–active situation, in which the highest gain cavity is referred to as the *active* cavity.) Physically, this interface section might be another dielectric region with

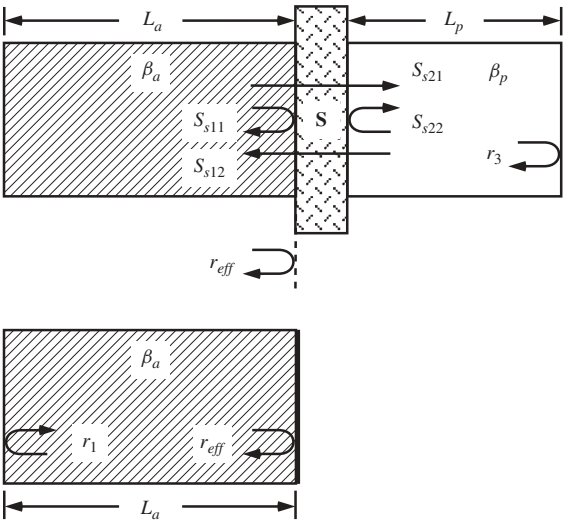


FIGURE 3.11: Generic four-mirror or three-section laser and equivalent mirror representation.

an index different from either the active or passive sections. Figure 3.11 shows the model and a possible implementation.

For this case we only need to modify the expression for r_{eff} slightly. Replacing $r_2, -r_2, t_2$, and t_2 with the more general terms, $S_{s11}, S_{s22}, S_{s21}$, and S_{s12}, respectively, in Eq. (3.35), we obtain

$$r_{eff} = S_{s11} + \frac{S_{s21}S_{s12}r_3 e^{-2j\tilde{\beta}_p L_p}}{1 - S_{s22}r_3 e^{-2j\tilde{\beta}_p L_p}}, \tag{3.41}$$

or rearranging,

$$r_{eff} = S_{s11}\left[1 + \frac{\sigma R'}{1 - R'}\right], \tag{3.42}$$

where $R' = S_{s22}r_3 e^{-2j\tilde{\beta}_p L_p}$, and $\sigma = (S_{s21}S_{s12}/S_{s11}S_{s22})$.

Equation (3.42) shows that for resonance R' is real and positive. Thus, for the second term to add to the first for a maximum in r_{eff}, the ratio σ must also be real and positive. As can be verified by reference to Eqs. (3.20) to (3.23), this occurs when the space is a multiple of a half-wavelength wide, if its index is either lower or higher than both end cavities (e.g., a simple air gap between two semiconductors). Other situations will be left as exercises for the reader, but it should be clear that the phase of this *gap factor*, σ, will determine whether the maxima of r_{eff} will occur at passive cavity resonances ($\angle\sigma = 0$), or antiresonances ($\angle\sigma = \pi$), as in Fig. 3.6. In the case of the three-mirror cavity, the phase of the gap factor is always π.

For the optimum case of $\angle\sigma = 0$ the minima in α_m are sharpest and the mode selection best. Also, if the index can be adjusted in the second cavity, the resulting tunability is somewhat improved. In particular, because the phase of r_{eff} varies most rapidly near the resonance of the second cavity, this optimum case leads to better continuous tunability.

3.5 GRATINGS

3.5.1 Introduction

Many important diode lasers use gratings or distributed Bragg reflectors (DBRs) for one or both cavity mirrors. With in-plane lasers the reason is to use their frequency selectivity for single-axial mode operation, and with vertical-cavity lasers the reason is to obtain a very high value of reflectivity. Gratings consist of a periodic array of index (sometimes gain) variations. At the Bragg frequency, the period of the grating is half of the average optical wavelength in the medium. Significant reflections can also occur at harmonics of this frequency. In the vertical-cavity case, quarter-wavelength-thick layers of two different index materials are alternated during growth. In the in-plane case, corrugations are typically etched on the surface of the waveguide, and these are refilled with a different index material during a second growth. Figure 3.12 illustrates the two cases. As indicated, the in-plane case usually has many more grating periods than the VCSEL.

In either case, the concept of the grating is that many small reflections can add up to a large net reflection. At the Bragg frequency the reflections from each discontinuity add up exactly in phase. For the rectangular gratings shown, there

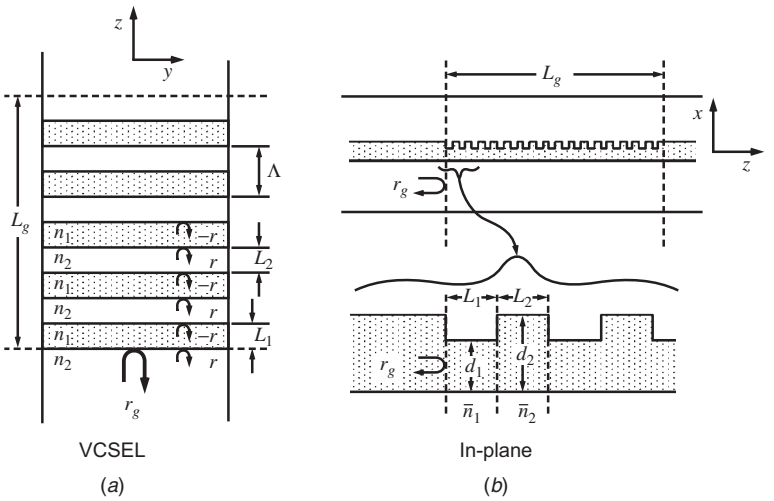


FIGURE 3.12: (a) Schematic of a DBR mirror for a vertical-cavity laser. (b) Schematic of a DBR mirror for an in-plane laser.

are two discontinuities per period, each of reflectivity r. Thus, the net reflection from m grating periods is $r_g \sim 2mr$ when the net reflection is weak, so that each discontinuity sees nearly the same incident field. For a significant reflection the field will fall off into the grating, and the problem becomes more difficult. As the frequency is deviated from the Bragg condition, the reflections from discontinuities further into the grating return with progressively larger phase mismatch. This causes a roll-off in the net reflection, which occurs more rapidly in longer gratings.

Example 3.2

Problem: For a grating from Figure 3.12 (a), calculate the contributions to the reflected electric field at the Bragg wavelength λ_0, from the first two interfaces. Assume that there are no losses in the grating, that the reflection amplitude $r \ll 1$, and that $L_1 = \lambda_0/4\bar{n}_1$ and $L_2 = \lambda_0/4\bar{n}_2$.

Solution: Assuming that electric field E is incident to the grating, the electric field contribution from the first two interfaces is given by the Eq. (3.20), modified for the different reflection sign convention used in Figure 3.12,

$$E_R = E\left(r + \frac{t^2(-re^{-2j\tilde{\beta}L})}{1 - r^2e^{-2j\tilde{\beta}L}}\right) \approx E\left(r - t^2re^{-2j(2\pi\bar{n}_1/\lambda_0)\cdot(\lambda_0/4\bar{n}_1)}\right)$$
$$= E\left(r - t^2re^{j\pi}\right) = E\left(r + t^2r\right).$$

The first term in the expression represents the reflection from the first interface, and the second term is the contribution from the second interface. We can conclude that the phase term of the round trip accross the half period of the grating will cancel out the π phase change due to the reflection from a higher index material. For the subsequent interface, the reflection sign will be positive again, and so will the phase component because the new round-trip phase component will be equal to $e^{j2\pi}$. Therefore, as expected, all the reflections from each interface will add in phase, and the final reflection from the grating will be a sum of interface reflections.

Because the dielectric interfaces extend uniformly across the mode in the vertical-cavity case, the results from Section 3.3.3 apply directly. That is, we can obtain the net grating reflectivity, r_g, exactly by cascading the **T**-matrices, Eqs. (3.24) to (3.27). For the in-plane case, we must assume some effective reflectivity at each discontinuity to use these results. This is indicated by the inset in Fig. 3.12b. For the rectangular gratings shown, this effective modal reflectivity at each discontinuity can be estimated by using the effective indices in each waveguide segment in Eq. (3.10). That is, if $\bar{n}_1$ and $\bar{n}_2$ are the effective indices in waveguide segments 1 and 2 of width d_1 and d_2, respectively, the reflectivity in going from segment 2 to segment 1 is approximately

$$\bar{r} \approx \frac{\bar{n}_2 - \bar{n}_1}{\bar{n}_2 + \bar{n}_1}. \tag{3.43}$$

The effective indices can be calculated following the procedure outlined in Appendix 3. The approximation is necessary because the transverse grating mode will be different from (actually something in between) the modes in uniform waveguides of width d_1 or d_2. The approximation is best for small impedance discontinuities. The reflectivity in going from segment 1 to segment 2 is $-\bar{r}$ by symmetry. As we shall discuss more in Chapter 6, for grating profiles other than rectangular, the effective reflectivity at each segment interface can be found by multiplying Eq. (3.43) by the relative Fourier coefficients for the harmonic of interest.

3.5.2 Transmission Matrix Theory of Gratings

Figure 3.13 shows how the periodic gratings of Fig. 3.12 can be represented using **T**-matrices. If the output fields are known, the input fields (and hence the reflectivity, $r_g = B_m/A_m$) can be determined by matrix multiplying the individual components of the grating, starting from the output and proceeding to the input. Now, for the simple uniform gratings depicted in Fig. 3.12, in which only two indices are involved, the matrix multiplication to obtain r_g can be simplified by realizing that each period is the same. (However, we may need to use a different **T**-matrix for the output segment.)

To determine the **T**-matrix for a single period, we must matrix multiply *four* simple **T**-matrices together. Starting at the reference plane and moving in the positive z-direction in Fig. 3.12, we encounter (1) a 2–1 dielectric interface, (2) a propagation delay of length L_1, (3) a 1–2 dielectric interface, and (4) a propagation delay of length L_2. At this point another 2–1 interface is encountered, marking the beginning the next period. The results of multiplying the first three of these matrices together has already been derived in Eqs. (3.24) to (3.27). To apply those equations to Fig. 3.12 we need to change notation slightly by setting $r_1 = r_2 \equiv -r, t_1 = t_2 \equiv t$ and $L = L_1$. Now, the associative property is maintained under matrix multiplication implying that $\mathbf{T_1T_2T_3T_4} = (\mathbf{T_1T_2T_3})\mathbf{T_4}$. Thus, by multiplying the composite of three matrices by the fourth propagation delay matrix defined in Eq. (3.18), we obtain the single period **T**-matrix.

Alternatively, we can group the **T**-matrices as follows: $(\mathbf{T_1T_2})(\mathbf{T_3T_4})$, and identify each group as being equivalent to the third structure listed in Table 3.3.

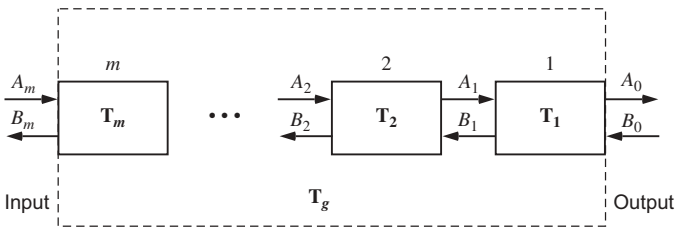


FIGURE 3.13: Cascaded scattering junctions characterized by transmission matrices. The net transmission matrix of the cascade is $\mathbf{T}_g$.

Multiplying two such matrices (using the appropriate refractive indices) also gives us the single period **T**-matrix. Either way, the general forms, and corresponding forms for a lossless DBR at the Bragg frequency (right of arrow), become

$$\begin{aligned} T_{11} &= \frac{1}{t^2}[e^{j\phi+} - r^2 e^{-j\phi-}] \rightarrow -\frac{1+r^2}{t^2}, \\ T_{21} &= \frac{r}{t^2}[e^{j\phi+} - e^{-j\phi-}] \rightarrow -\frac{2r}{t^2}, \\ T_{12} &= \frac{r}{t^2}[e^{-j\phi+} - e^{j\phi-}] \rightarrow -\frac{2r}{t^2}, \\ T_{22} &= \frac{1}{t^2}[e^{-j\phi+} - r^2 e^{j\phi-}] \rightarrow -\frac{1+r^2}{t^2}, \end{aligned} \tag{3.44}$$

where $\phi_\pm \equiv \tilde{\beta}_1 L_1 \pm \tilde{\beta}_2 L_2$, which becomes either π or 0 at the Bragg condition. For convenience, we define an *average* complex propagation constant of one grating period: $\tilde{\beta}\Lambda \equiv \tilde{\beta}_1 L_1 + \tilde{\beta}_2 L_2$, where $\tilde{\beta} = \beta - j\alpha/2$. Setting $\Lambda = L_1 + L_2, L_1 = \lambda_0/4n_1$, and $L_2 = \lambda_0/4n_2$ (where λ_0 is the Bragg wavelength), we find that

$$\beta = \frac{\beta_1/n_1 + \beta_2/n_2}{1/n_1 + 1/n_2} \tag{3.45}$$

$$\alpha = \frac{\alpha_1/n_1 + \alpha_2/n_2}{1/n_1 + 1/n_2} \tag{3.46}$$

In the first line, β is defined as the average propagation constant of the grating, $\beta = 2\pi/\lambda_0 \bar{n}$. In the second line, α is the average loss of the grating.

At the Bragg frequency, the phase delay of each layer is $\beta_1 L_1 = \beta_2 L_2 = \pi/2$, and $\beta_0 \Lambda = \beta_1 L_1 + \beta_2 L_2 = \pi$. Defining a *detuning parameter*,

$$\delta \equiv \beta - \beta_0, \tag{3.47}$$

the phase terms with no loss become simply

$$\phi_+ = \pi + \delta\Lambda, \quad \phi_- = 0. \tag{3.48}$$

The detuning parameter can alternatively be expressed as $\delta\Lambda = \pi(\nu - \nu_0)/\nu_0$. The arrows in Eq. (3.44) indicate this no loss case at the Bragg frequency ($\delta \rightarrow 0$).

For a cascade of m such matrices for the m grating segments of length Λ, we have

$$\mathbf{T}_g = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix}^m, \tag{3.49}$$

which can be simplified using a mathematical identity derived in Appendix 7. The only restriction for the form used here is that the system must be *reciprocal* such that $T_{11}T_{22} - T_{12}T_{21} = 1$. In other words, transmission through the dielectric stack

must be equivalent for light incident from either side of the stack. For all **T**-matrices we have discussed so far, this condition is satisfied even with loss or gain in the layers. Using the additional subscript g for the T-parameters of the entire grating, and assuming reciprocity, we can write

$$\begin{aligned} T_{g11} &= (1 + jm_{eff}\Delta)\cosh m\xi, \\ T_{g21} &= \frac{T_{21}}{T_{11}} m_{eff}(1 + j\Delta)\cosh m\xi, \\ T_{g12} &= \frac{T_{12}}{T_{22}} m_{eff}(1 - j\Delta)\cosh m\xi, \\ T_{g22} &= (1 - jm_{eff}\Delta)\cosh m\xi, \end{aligned} \tag{3.50}$$

where

$$\begin{aligned} \pm\xi &= \ln\{\tfrac{1}{2}[T_{11} + T_{22}] \pm \sqrt{\tfrac{1}{4}[T_{11} + T_{22}]^2 - 1}\}, \\ \Delta &= j\frac{T_{22} - T_{11}}{T_{22} + T_{11}}, \\ m_{eff} &= \frac{\tanh m\xi}{\tanh \xi}. \end{aligned} \tag{3.51}$$

The equation $\pm\xi = \ln\{\pm\}$ implies that the negative root $\{-\} = 1/\{+\}$ or $\{-\}\{+\} = 1$, which is easily verified. As will be discussed later, the three parameters, ξ, Δ, and m_{eff}, all have a physical significance and are important in the analysis of dielectric stacks.

The first parameter is the *discrete* propagation constant, ξ. Its value is very dependent on the wavelength of the incident light and is in general complex. It represents a discrete propagation constant because the fields (A_j, B_j) shown in Fig. 3.13 are multiplied by $e^{\pm\xi}$ upon passing to the next period. (As shown more exactly in Appendix 7, the fields when decomposed into the two eigenvectors of the matrix are multiplied by the eigenvalues $e^{+\xi}$ and $e^{-\xi}$ upon passing to the next period.) For example, if ξ is purely imaginary at some wavelength, the incident field will only encounter a phase shift of $m\xi$ upon passing through m periods, suffering no attentuation and hence, providing perfect transmission through the stack. Wavelength regimens for which this occurs are referred to as *passbands* of the dielectric stack. If ξ is purely real at some wavelength, the field will be attenuated by $e^{\pm m\xi}$ as it propagates through m periods, which can lead to very low transmission and hence, high reflection. These wavelength regimens are referred to as *stopbands* of the dielectric stack. Using Eq. (3.44) at the Bragg frequency, it is shown below that $\xi \approx j\pi + 2r$, revealing that the field experiences a π phase shift through each period and is attenuated by e^{-2r}. If r is high enough, and enough periods are used, the attentuation can be very high, leading to extremely low transmission and extremely high reflection of the field.

The second parameter is defined as the generalized detuning parameter, Δ, and is a measure of how far away we are from the Bragg condition. For small reflectivities, using Eq. (3.44) together with Eq. (3.48), we find $\Delta \approx \tan \delta\Lambda \approx \delta\Lambda$. At the Bragg condition, $T_{11} = T_{22}$ in a lossless dielectric stack, and $\Delta = 0$ exactly, from its definition in Eq. (3.51).

The third parameter defines the effective number of periods, m_{eff}, seen by the incident field. For very weak attenuation ($\mathrm{Re}\{\xi\} \ll 1$), and a small number of periods, the tanh functions reduce to their arguments and $m_{eff} = m$. For large attenuation, m_{eff} as a function of m saturates at a value of $m_{eff} = 1/\tanh \xi$, which when multiplied by Λ, determines the *penetration depth* of the field into the dielectric stack to be discussed later.

Example 3.3

Problem: Determine the values for the discrete propagation constant ξ, the generalized detuning parameter Δ, and the effective number of periods m_{eff} in a special case of when the wavelength satisfies the Bragg condition, and there is no loss or gain.

Solution: From Eq. (3.44), we observe that at the Bragg frequency, $T_{11} = T_{22}$ and $T_{21} = T_{12}$. The reciprocity condition for this case reduces to $T_{11}^2 - T_{21}^2 = 1$. Using these relations in the first equation of (3.49), we obtain $e^{\pm\xi} = T_{11} \pm T_{21}$, which allows us to set $\tanh \xi = T_{21}/T_{11}$. The three parameters defined in Eq. (3.51) then reduce to

$$\begin{aligned} \xi &= j\pi + \ln(-T_{11} - T_{21}), \\ \Delta &= 0, \\ m_{eff} &= \frac{T_{11}}{T_{21}} \tanh[m \ln(-T_{11} - T_{21})]. \end{aligned} \tag{3.52}$$

In the first equality, the identity: $\ln(-1) = j\pi$, was used to introduce a minus sign into the argument of the ln function. Inserting Eq. (3.44), we have

$$\begin{aligned} \mathrm{Re}\{\xi\} &= \ln\left(\frac{1+r}{1-r}\right) \approx 2r, \\ m_{eff} &= \frac{1+r^2}{2r} \tanh\left[m \ln\left(\frac{1+r}{1-r}\right)\right] \approx \frac{1}{2r} \tanh(2mr), \end{aligned} \tag{3.53}$$

where the latter relations are obtained by neglecting second- and higher-order terms in r (in the ln function expansion, the second-order terms actually cancel making it an excellent approximation). Thus, at the Bragg wavelength, the attenuating portion of the discrete propagation constant is roughly equal to the sum of the reflectivities encountered within one period, as one might intuitively expect. The effective number of periods is also inversely related to $2r$. In other words, as r increases, fewer and fewer periods are effectively seen by the field, which also agrees with intuition. Note that as m increases toward infinity, m_{eff} saturates at $1/2r$.

The reflectivity of the overall stack in the general case is given by $r_g = S_{g11} = T_{g21}/T_{g11}$. With Eq. (3.50), the reflectivity becomes

$$r_g = \frac{T_{21}}{T_{11}} m_{eff} \frac{1 + j\Delta}{1 + j m_{eff}\Delta}. \tag{3.54}$$

At the Bragg frequency, with no loss or gain, the second term disappears using Eq. (3.52), and the reflectivity reduces to

$$r_g = \tanh\left[m \ln\left(\frac{1+r}{1-r}\right)\right] \approx \tanh(2mr) \equiv r_g', \qquad (\Delta = 0, \alpha_g = 0, g = 0) \tag{3.55}$$

where the approximation is good to second order in r, and r_g' is defined as the lossless value of r_g. Expressing r in terms of the refractive indices, the reflectivity at the Bragg frequency, for a grating with no loss or gain, can also be written as

$$r_g = r_g' = \frac{1 - (n_1/n_2)^{2m}}{1 + (n_1/n_2)^{2m}}. \qquad (\Delta = 0, \alpha_g = 0, g = 0) \tag{3.56}$$

It is shown in Appendix 7 that Eq. (3.56) can more generally be applied to dielectric stacks with different values of index (for example, at the input or output of the stack) by replacing $(n_1/n_2)^{2m}$ with the product of the low-to-high index ratios of every interface in the stack.

In addition, it will be shown in the next section how the lossless grating reflection value r_g' can be used to model gratings with some loss.

Figure 3.14 shows example plots of the magnitude and phase of r_g from Eq. (3.54) for several values of $2mr$. For low reflection magnitudes it approaches a $\sin(\delta L_g)/(\delta L_g)$ function, whereas for high reflection magnitudes, the top flattens out (saturating at a value of unity) and the stopband broadens. In the phase spectrum, increasing $2mr$ has the effect of suppressing the phase slope over the range of the stopband. We also see that the phase jumps by π every time the reflectivity passes through a null. A phasor diagram of the reflectivity would reveal that as δL_g approaches a zero crossing, the reflectivity phasor aligns itself with the negative real axis, shrinks to zero, and then increases again, pointing along the positive real axis. For the low reflection magnitudes, the full width at half maximum of the reflection peak can be obtained by setting $\sin(\delta L_g)/(\delta L_g) = \frac{1}{2}$ because as $\delta \to 0, \sin(0)/0 \to 1$. Solving for δL_g which corresponds to the full width at half maximum, we have that $\delta L_g = 1.88$, yielding the spectral width of the peak at half maximum as

$$\Delta\lambda = \frac{\lambda^2}{2\bar{n}_g L_g} \frac{3.76}{\pi}. \tag{3.57}$$

In the low reflection limit the peak net reflection should intuitively approach $2mr$, since there are two discontinuities per grating period, and multiple reflections should be negligible. Also from Fourier transform theory, the net reflection from

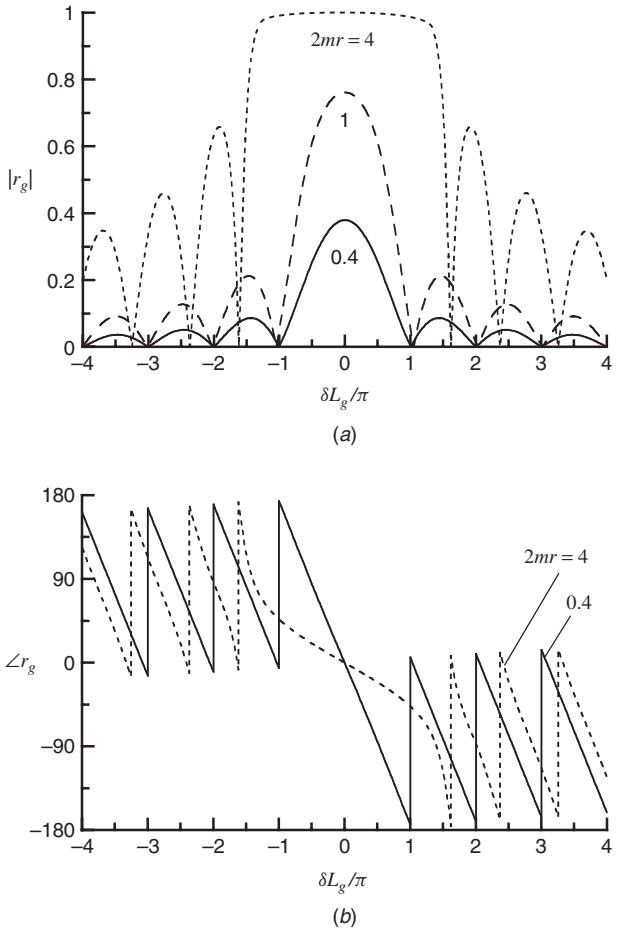


FIGURE 3.14: (a) Magnitude and (b) phase of the grating reflection coefficient versus the normalized frequency deviation from the Bragg condition for different values of the reflection parameter, $\kappa L_g \equiv 2mr$. For small r and large m, the reflection spectrum plotted above is only dependent on the product $2mr$. However, for larger values of r and/or smaller values of m, there is some dependence on the individual values of r and m. For example, $r = 0.01$ and $m = 100$ would be roughly equivalent to $r = 0.1$ and $m = 10$. However, some changes in the spectrum would occur with $r = 0.2$ and $m = 5$, even though the $2mr$ product is equivalent in all three cases. The above plots use $r = 0.1, 0.025, 0.01$ with $m = 20$ for all cases. In the phase plot, the $2mr = 1$ case is very similar to the 0.4 case and hence is not plotted. Here $\delta \equiv \beta - \beta_0$ where β is the average propagation constant of the grating.

m elements, equally spaced by half a wavelength, and each causing a reflection of $2r$, should produce the following spectral response:

$$|r_g| \approx 2mr \frac{\sin(\delta L_g)}{\delta L_g}, \qquad (mr < 0.2) \tag{3.58}$$

where δ is the deviation of the average propagation constant from the Bragg frequency. The qualifier reminds us that Eq. (3.58) is only valid in the weak reflection limit. Practical lasers usually require mirror power reflectivities of greater than 15% where multiple reflections cannot be ignored. Thus, Eq. (3.58) is of limited utility for diode laser work unless a relatively low reflectivity mirror is desired.

Historically, researchers working with long grating reflectors in in-plane lasers have chosen to use a different dimensionless parameter to quantify the net grating reflection rather than $2mr$, which is the reflection per grating segment, $2r$, times the number of segments, m. The parameter of choice is the reflection per unit length, κ, times the grating length, L_g. Thus, for the square wave grating, we see that the *coupling constant*, κ, is given by

$$\kappa L_g \equiv 2mr = \frac{m\Delta\bar{n}}{\bar{n}} = \frac{L_g}{\Lambda}\left(\frac{\Delta\bar{n}}{\bar{n}}\right), \tag{3.59}$$

where $\Delta\bar{n} \equiv |\bar{n}_2 - \bar{n}_1|$ and $\bar{n} \equiv (\bar{n}_2 + \bar{n}_1)/2$. For small index differences, $\Lambda = (\lambda_0/4)(1/\bar{n}_1 + 1/\bar{n}_2) \approx \lambda_0/2\bar{n}$ which gives $\kappa = 2\Delta\bar{n}/\lambda_0$. Later in Chapter 6 when we introduce coupled mode theory, we shall see that this form also naturally results (however, for the sinusoidal gratings analyzed there, κ is reduced by $\pi/4$). Therefore, the approximation for the grating reflection in Eq. (3.55) can also be written as, $r_g \approx \tanh(\kappa L_g)$. However, for very short gratings, such as in VCSELs, our original $2mr$ form seems more natural.

3.5.3 Effective Mirror Model for Gratings

From Fig. 3.14 we note that the phase varies relatively linearly near the reflection maximum. Such a reflection can be well approximated by a discrete mirror reflection, $|r_g'|$, but placed a distance L_{eff} away as shown in Fig. 3.15. For a lossless grating, $|r_g'|$ is equal to the magnitude of the grating's reflection $|r_g|$.

From Eq. (3.15) and the associated discussion in Section 3.3.2, we know that the incident and reflected wave amplitudes each experience a phase shift of $-\beta L_{eff}$ in traversing the distance to the effective mirror and back. Thus, knowing that the reflection phase is zero at the Bragg frequency, we can express r_g as

$$r_g \approx |r_g'| e^{-2j(\beta-\beta_0)L_{eff}}. \qquad (|\delta L_g| \ll \pi) \tag{3.60}$$

A small propagation loss can be approximately calculated using effective mirror approximation by replacing β by its complex form $\tilde{\beta}$, in Eq. (3.60). This results in

$$|r_g| = |r_g'| e^{-\alpha_i L_{eff}}. \tag{3.61}$$

This perturbation technique is not valid for significant levels of loss or gain, because if such levels exist, the rate of decay of energy into the grating will be significantly affected. In this case, r_g should be recalculated using the transmission matrix method incorporating complex propagation constants, $\tilde{\beta}$, throughout. But, for

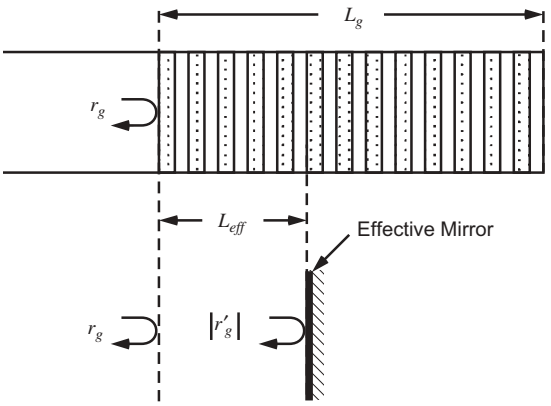


FIGURE 3.15: Definition of an effective mirror for a grating reflector.

many practical cases of interest we find that it is possible to model the grating near its reflection maximum by using the effective mirror concept, for which we need only know (1) the lossless reflection magnitude at the Bragg frequency, $|r_g|_{\max}$; (2) the effective mirror location, given by L_{eff}; and (3) the propagation loss over the entire grating length, given by α_i in the grating.[1] Expanding the true DBR reflection phase in a Taylor series about the Bragg frequency: $j\phi \approx j\phi_0 + j(\beta - \beta_0)(\partial\phi/\partial\beta) + \cdots$, and equating the linear $(\beta - \beta_0)$ coefficient with the exponent in Eq. (3.60), we find that the effective length is given by

$$L_{eff} = -\frac{1}{2}\frac{\partial\phi}{\partial\beta}, \tag{3.62}$$

which is the same result we found earlier in Eq. (3.39) by different means. From Fig. 3.14b, it is clear that as $2mr$ increases, L_{eff} decreases over the range of the stopband. Using Eqs. (3.54) and (3.44), it can be shown (after a bit of math) that if third- and higher-order terms in r are neglected, the effective length becomes

$$L_{eff} = \frac{1}{2\kappa}\tanh(\kappa L_g), \qquad (|\delta L_g| \ll \pi) \tag{3.63}$$

[1]If the propagation loss is not distributed evenly throughout the grating (as is often the case with VCSEL mirrors), we must use an effective propagation loss:

$$\alpha_{i,eff} = \frac{1}{L_{eff}}\int_0^{L_g} \alpha_i(z)_e^{-z/L_{eff}}(1 \pm \cos 2\beta z)\, dz.$$

The upper (lower) sign in the standing wave term is for a DBR with 0 (π) reflection phase at the Bragg wavelength. For a constant loss, $\alpha_i(z) = \alpha_{i0}$, and the equation reduces to $\alpha_{i,eff} = \alpha_{i0}$ for $L_g \gg L_{eff} \gg \lambda/4\pi n$.

where $\kappa L_g \equiv 2mr$ as given by Eq. (3.59); the exact expression for L_{eff} is given in Appendix 7. For weakly reflecting gratings ($\tanh \kappa L_g \rightarrow \kappa L_g$), the effective mirror plane is at the center of the grating, and for strongly reflecting gratings ($\tanh \kappa L_g \rightarrow 1$), $L_{eff} \rightarrow \Lambda/(4r) \approx \lambda_0/4\Delta\bar{n}$.

Although L_{eff} was defined to give the proper mirror phase and, thus, can be used to locate cavity modes, it also gives the approximate optical energy penetration depth into the grating mirror. As mentioned earlier, m_{eff} gives the effective number of periods seen by the incident field. The optical power is the square of the field; thus its exponential decay constant is twice that of the electric field, and hence the optical power penetrates half as far into the mirror. Therefore, the energy penetration depth is given by $L_{pen} = \Lambda m_{eff}/2$ (see Appendix 7 for more details). Substituting Eq. (3.53) for m_{eff} and setting $2r = \kappa\Lambda$, this definition reduces to the one given in Eq. (3.63) and hence $L_{pen} \approx L_{eff}$. Therefore, the total energy stored in the mirror is approximately equal to the energy density at its input multiplied by L_{eff}.

In the general case, $r_g = T_{g21}/T_{g11} = B_m/A_m|_{(B_0=0)}$ (or in terms of the **S**-parameters, $r_g = S_{g11} = b_m/a_m|_{(a_0=0)}$), which can be calculated numerically by performing the operations indicated in Fig. 3.13. The numerical procedure proceeds from the output of the grating backwards, after first assuming some value for A_0, such as unity. The intermediate A_j's and B_j's are evaluated by matrix multiplication moving to the left toward the beginning of the grating. Using this technique each segment in principle could be different, and the loss or gain can be naturally included in the T_{ij}'s, by using the appropriate complex propagation constants in Eqs. (3.24) to (3.27).

3.6 LASERS BASED ON DBR MIRRORS

3.6.1 Introduction

A distributed Bragg reflector (DBR) laser can be formed by replacing one or both of the discrete laser mirrors with a passive grating reflector. Figure 3.16 shows schematics of both in-plane and VCSEL configurations with one or two grating mirrors. By definition, the grating reflectors are formed along a passive waveguide section, so one of the issues is how to make the transition between the active and passive waveguides without introducing an unwanted discontinuity. This is of little concern in the VCSEL case because the axial direction is the growth direction, and switching materials is always done several times during growth. Thus, forming the mirrors only requires growing more uniform layers. However, for coherence along the axial direction, these layers must be very accurately controlled in thickness.

In the in-plane laser case making a DBR laser is relatively complex because a lot of structure must be created along the surface of the wafer. This generally includes a joint between the active and passive regions as well as grating patterning and regrowth. For this reason in-plane DBR lasers are only formed when their unique properties are required. Besides the single-frequency property provided by

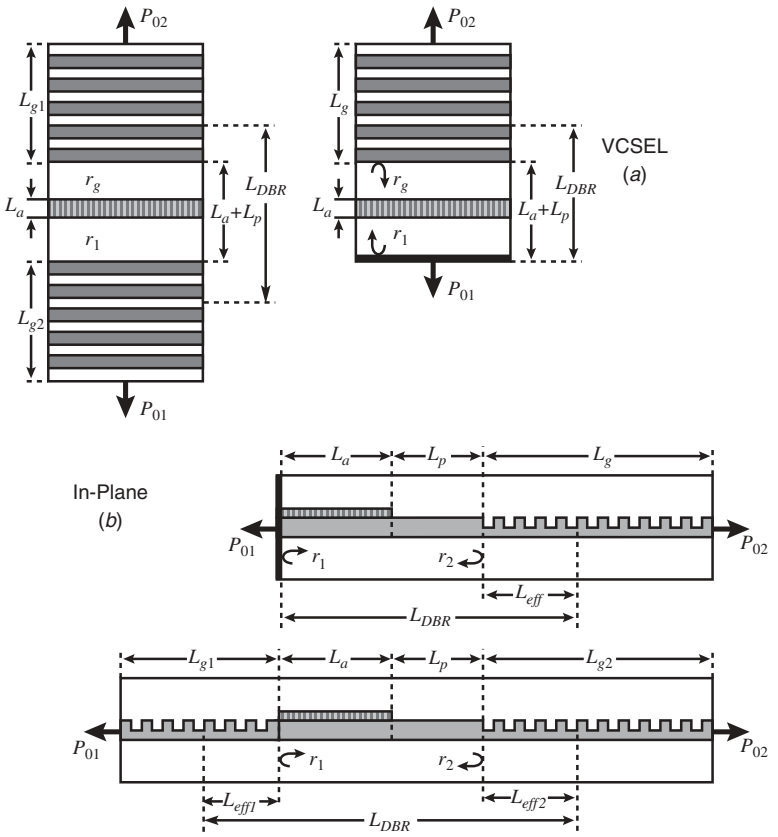


FIGURE 3.16: (a) Vertical-cavity surface-emitting laser schematic illustrating various lengths and reference planes. (b) In-plane laser schematic illustrating various lengths and reference planes.

the frequency-selective grating mirrors, these attributes can include wide tunability, if the effective index is varied electro-optically in the several sections by separate electrodes.

3.6.2 Threshold Gain and Power Out

The threshold gain of a DBR laser is the same as we have already calculated elsewhere, but we must interpret the parameters properly and consistently. This interpretation, however, can vary depending on how we choose to model the DBR. If we treat the grating reflector as a separate element characterized by reflection and transmission scattering parameters, Case (a), we logically choose the cavity length to be $L_a + L_p$. However, if we use the effective mirror model outlined in Fig. 3.16, Case (b), we would choose the cavity length to be L_{DBR} and include the grating

passive losses in the penetration depth, L_{eff}, rather than the reflection coefficient. Fortunately, these approaches lead to the same result near the Bragg frequency, provided the losses are not too large. In practice we find that the approximate Case (b) is most useful for design where a minimum of computation is desired. Of course, Case (a) is more exact for detailed analysis.

With reference to the single grating mirror configuration in Fig. 3.16, the threshold gain given by Eq. (2.25) becomes

Case (a)

$$\Gamma g_{th} = \langle \alpha_i \rangle + \frac{1}{L_a + L_p} \ln\left[\frac{1}{r_1 |r_g|}\right], \tag{3.64}$$

where Γ and $\langle \alpha_i \rangle$ average over the active and passive sections of the cavity only (any losses encountered in the DBR section are contained in $|r_g|$, the magnitude of the DBR reflectivity given by Eq. (3.54)). Using the effective mirror model,

Case (b)

$$\Gamma g_{th} = \langle \alpha_i \rangle + \frac{1}{L_{DBR}} \ln\left[\frac{1}{r_1 |r_g'|}\right], \tag{3.65}$$

where Γ and $\langle \alpha_i \rangle$ average over the entire effective cavity length, $L_{DBR} = L_a + L_p + L_{eff}$, and losses encountered in the DBR section are treated as propagation losses in L_{eff}. The prime on r_g denotes the use of its lossless value. In either case, the mode spacing is determined using Eq. (3.40) with the optical length set equal to $\bar{n}_{ga} L_a + \bar{n}_{gp} L_p + \bar{n}_{gDBR} L_{eff}$.

For VCSELs, the confinement factor Γ must generally be calculated by Eq. (5.10). Even for uniform gain within the active region Eq. (5.14) must be used for Γ_z. If two grating mirrors are used in a DBR laser, r_1 in Eq. (3.64) must be replaced by $|r_{g1}|$ for the other grating reflection, and the mode spacing needs to include the effective lengths of both DBRs.

Because the distributed mirrors are lossy in general, we must use Eqs. (3.32) and (3.33) for the differential quantum efficiency and the power out, respectively. However, we need to use the proper value for α_m, and we must replace r_2 and t_2 by the relevant grating S-parameters. Again, we have different expressions to be consistent with the two different models considered. That is, with a single Bragg mirror at end 2 and a discrete mirror at end 1 of the laser, Eq. (3.64) gives

Case (a)

$$\alpha_m = \frac{1}{L_a + L_p} \ln\left[\frac{1}{r_1 |r_g|}\right], \tag{3.66}$$

and for the effective mirror model,

Case (b)

$$\alpha_m = \frac{1}{L_{DBR}} \ln\left[\frac{1}{r_1 |r_g'|}\right].$$

For the fractional power out of each end in Case (a), we plug the grating S-parameters into Eq. (3.31). Setting $r_2 = S_{g11} \equiv r_g$ and $t_2 = S_{g21} \equiv t_g$, we obtain

Case (a)

$$F_1 = \frac{t_1^2}{(1 - r_1^2) + \frac{r_1}{|r_g|}(1 - |r_g|^2)},$$

and

$$F_2 = \frac{|t_g|^2}{(1 - |r_g|^2) + \frac{|r_g|}{r_1}(1 - r_1^2)}. \tag{3.67}$$

For the effective mirror model,

Case (b)

$$F_1 = \frac{t_1^2}{(1 - r_1^2) + \frac{r_1}{|r_g'|}(1 - |r_g'|^2)},$$

and

$$F_2 = \frac{|t_g'|^2}{(1 - |r_g'|^2) + \frac{|r_g'|}{r_1}(1 - r_1^2)}.$$

In this case, r_g is replaced by its lossless value r_g', and $|t_g|^2$ is replaced by the power transmission exclusively through the effective mirror, $|t_g'|^2$, which is found by setting $e^{-\alpha_i L_{eff}}|t_g'|^2 = |S_{g21}|^2$.

For both Case (a) and (b), it is useful to determine $|S_{g21}|^2$ for a lossy DBR. It can be shown using Eqs. (3.44) to (3.51) for a small uniform loss, α_i, at the Bragg condition that $\Delta \approx -j\alpha_i \Lambda/2$ and

$$|S_{g21}|^2 = 1/|T_{g11}|^2 \approx (1 - |r_g'|^2)e^{-2\alpha_i L_{eff}}. \tag{3.68}$$

Hence, we reach the unintuitive conclusion that the transmitted power is $\propto e^{-2\alpha_i L_{eff}}$ just like the reflected power. As a result, the transmission through the effective mirror becomes

$$|t_g'|^2 = (1 - |r_g'|^2)e^{-\alpha_i L_{eff}}, \tag{3.69}$$

which unfortunately does not correspond to the lossless transmission through the DBR (as one might have hoped).

With Eq. (3.68), we can show that the ratio of powers out of the two ends, F_1/F_2, is preserved in Case (a) and (b). Furthermore, we can show that F_1 and F_2 are both larger in Case (b). This increase in F_1 and F_2 compensates for the smaller mirror loss of Case (b), such that η_d is also approximately preserved in

Case (a) and (b) (the effective mirror model does tend to overestimate η_d, but not significantly for $\alpha_i L_{eff} < 0.1$, or for larger losses if the grating reflectivity is high).

For a second general lossy mirror at end 1, for example, another grating, its S-parameters should also be used for r_1 and t_1 in Eqs (3.66) and (3.67). In Case (b), the same additional substitutions are made. Finally, as mentioned earlier, the power out of each end can be obtained from Eq. (3.33).

3.6.3 Mode Selection in DBR-Based Lasers

Figure 3.17 illustrates two plots similar to those shown in Fig. 3.10 for the DBR case. One is for cavity lengths common to in-plane lasers, the other is for VCSELs.

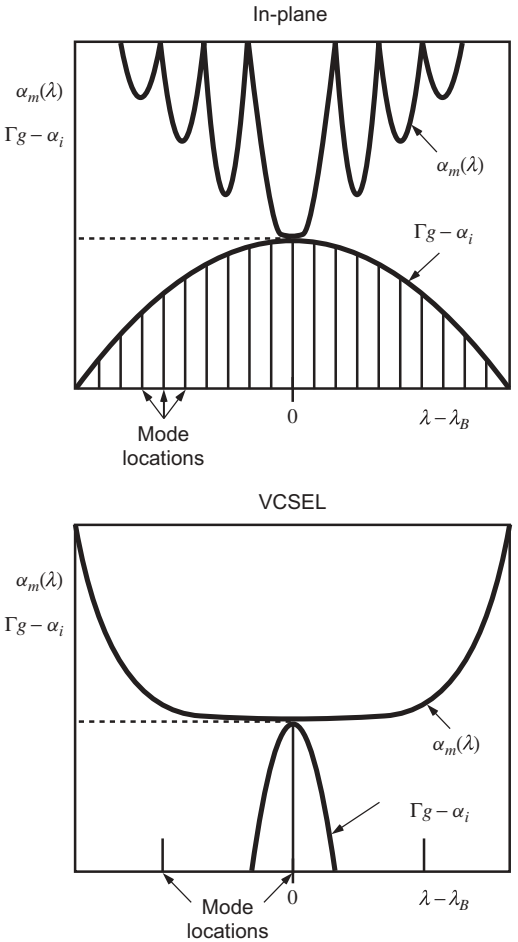


FIGURE 3.17: Schematic illustration of how a single axial mode is selected in an in-plane or vertical-cavity DBR laser. The VCSEL wavelength axis covers a five times larger range (i.e., the gain curve has the same width in both plots).

As compared to the coupled-cavity cases, the key difference with grating mirrors is that there is only a single loss minimum.

For in-plane lasers, the relatively rapid roll-up in α_m leads to good loss margins at the adjacent axial mode wavelength. As indicated in Fig. 3.17, the net loss margin at the adjacent axial mode is the sum of the roll-off in the net modal gain, $\Delta\Gamma g$, and the roll-up of the mirror loss, $\Delta\alpha_m$. In a later section, we shall show that this net loss can be used to derive an expression for the steady-state mode suppression ratio (MSR), which predicts how much unwanted modes are suppressed.

In the VCSEL case, both the mode spacing and the width of the loss minimum are much larger because the lengths of the cavity and grating, respectively, are much shorter. Thus, if we scale the wavelength axis to make them look comparable to the in-plane case, the primary effect is to make the gain look much more narrow relative to the mode spacing and mirror loss width. In fact, the roll-off in gain tends to be more important than the roll-up in loss for the VCSEL. That is, the primary cause of single-axial mode operation is just the short cavity.

3.6.4 VCSEL Design

As indicated in Chapter 1 both mirrors are DBRs in a typical VCSEL, unlike the simple schematic of Fig. 3.16. Also the cavity might include some different bandgap layers for current and optical confinement as well as possible intracavity electrical connections. In addition, the DBR mirrors may be more complex than a simple stack of two different index layers as suggested by Fig. 3.12a. Some layers may have a different index or thickness, and the interfaces may have a graded composition (and thus index) to facilitate better current conduction. In this section we shall include some of these aspects in an effort to focus on the design of real VCSEL structures. Of course, as in the rest of this chapter, the focus will remain on the axial cavity design.

As for other lasers the power out of one side of a VCSEL is given by Eq. (3.33). But for VCSELs it is sometimes useful to multiply the numerator and denominator of the conventional expression for differential efficiency by the DBR effective cavity length, L_{DBR}, to obtain for side j,

$$\eta_{dj} = F_j \eta_i \frac{T_m}{A_i + T_m}, \tag{3.70}$$

where $T_m = \ln(1/R)$ and $A_i = \langle\alpha_i\rangle L_{DBR}$. Because the gain length is small in VCSELs, the mean mirror reflectance R must be high, typically >99%, as was shown in the design curves of Fig. 2.11c. Therefore, T_m is a good approximation to the mean mirror power transmission in the absence of losses, and A_i is a very good approximation to the one-pass power losses, including those of the mirrors. That is, Case (b) is already being implied. Similarly, the one-pass threshold modal gain can be expressed as,

$$G_{th} = \Gamma g_{th} L_{DBR} = A_i + T_m. \tag{3.71}$$

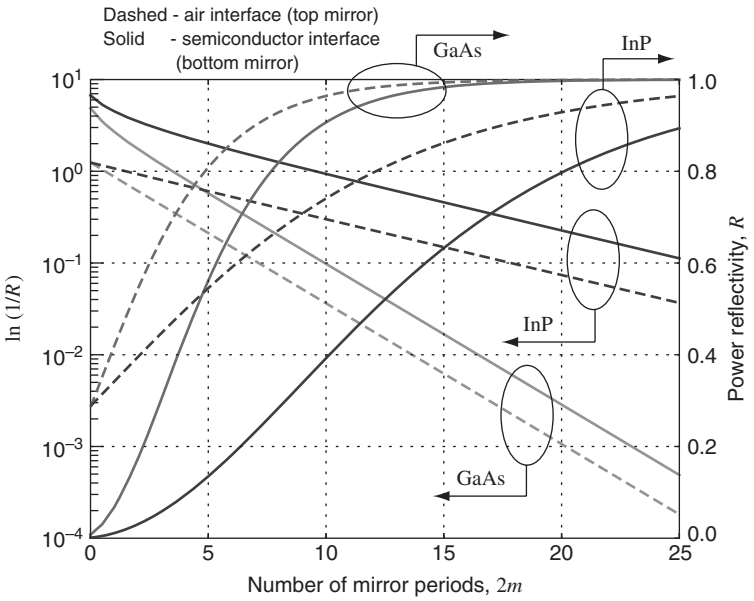


FIGURE 3.18: Plots of mirror reflectance, R, and $T_m = ln(1/R)$ for GaAs/AlGaAs mirrors ($n = 3.52$ and $n = 2.95$, respectively) and InP/InGaAsP lattice matched mirrors ($n = 3.17$ and $n = 3.40$, respectively).

Figure 3.18 gives plots of T_m vs. the number of mirror periods for some typical VCSEL mirror materials. As indicated, the AlGaAs system can provide a power reflectance >0.999 or $T_m < 10^{-3}$ with fewer than two dozen periods, but for the quaternary materials lattice-matched to InP, many more periods are necessary, and this is generally not seen as feasible. So, in this system at least one of the mirrors is usually formed by nonepitaxial mirrors. The exception to this has been the AlGaAsSb/InP system, which can provide nearly the same index discontinuity as in AlGaAs/GaAs [1], but it has not been very popular because it is difficult to grow by MOCVD.

In cases where the mirrors may contain a number of different layers, we can use the results of Appendix 7, Eq. (A7.43), to calculate the mirror amplitude reflectivity at the Bragg wavelength. It is repeated here for convenience,

$$|r_g| = \frac{1-b}{1+b},$$

where

$$b = \prod_{0}^{2m} \left[\frac{n_{1i}}{n_{2i}}\right], \tag{3.72}$$

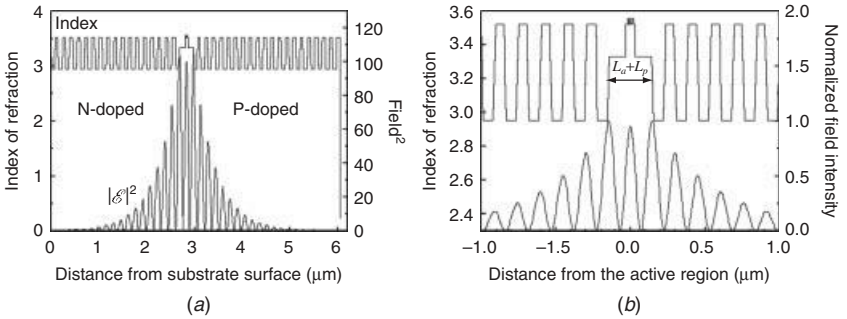


FIGURE 3.19: Examples of a VCSEL cavity showing the $|\mathscr{E}|^2$-field standing wave pattern together with the index of refraction of the various layers within the cavity: (left) entire cavity (right) expanded view of region near active. Lateral definition provided by proton implantation, as shown in Figure 1.18 (a). 3–8 nm thick $In_{0.2}Ga_{0.8}As/GaAs$ quantum wells with a gain maximum at 980 nm, a cavity spacer of $Al_{0.2}GaAs$ and uniform DBRs beyond composed of $Al_{0.9}Ga_{0.1}As/GaAs$ are assumed.

and n_{1i} is the low index and n_{2i} is the high index of any layer pair. Also, nonabrupt interfaces can be included, if we express the low and high index factors as deviations from the average of the minimum and maximum index levels and then multiply this deviation by the appropriate change in Fourier coefficient for that new grating period shape. That is, in Eq. (3.72) use

$$n_{1i} = \bar{n}_i - \chi_i \Delta n_i/2,$$
$$n_{2i} = \bar{n}_i + \chi_i \Delta n_i/2, \tag{3.73}$$

where $\bar{n}_i$ is the average of, and Δn_i is difference between, the maximum and minimum index values, and $\chi_i = 1$ for abrupt interfaces, $\pi/4$ for sinusoidal gratings, $2/\pi$ for triangular sawtooth gratings, and generally, $(\sin x/x)$ for trapezoidal gratings, with $x = 0$ for abrupt varying to x = $\pi/2$ for sawtooth [3].

An extremely important factor in the success of VCSELs is the enhancement in the confinement factor that results from placing the gain layers only at the maxima of the cavity's $|\mathscr{E}|^2$-field standing wave. As derived in Appendix 5, Eq. (A5.14), this can provide up to a factor of 2 enhancement over a simple fill-factor, and it is this enhancement factor that enables modern VCSELs to function as well as they do. In fact, it is easy to show that without this factor, most commercial VCSELs would not even reach threshold. Thus, from Appendix 5, we explicitly write the confinement factor as

$$\Gamma = \Gamma_{xy} \frac{L_a}{L_{DBR}} \left[1 + \cos 2\beta z_s \frac{\sin \beta L_a}{\beta L_a} \right], \tag{3.74}$$

where Γ_{xy} is the lateral confinement factor, L_a/L_{DBR} is the axial fill-factor, and the bracket is the enhancement factor in which z_s is the shift between the active layer center and the standing-wave peak.

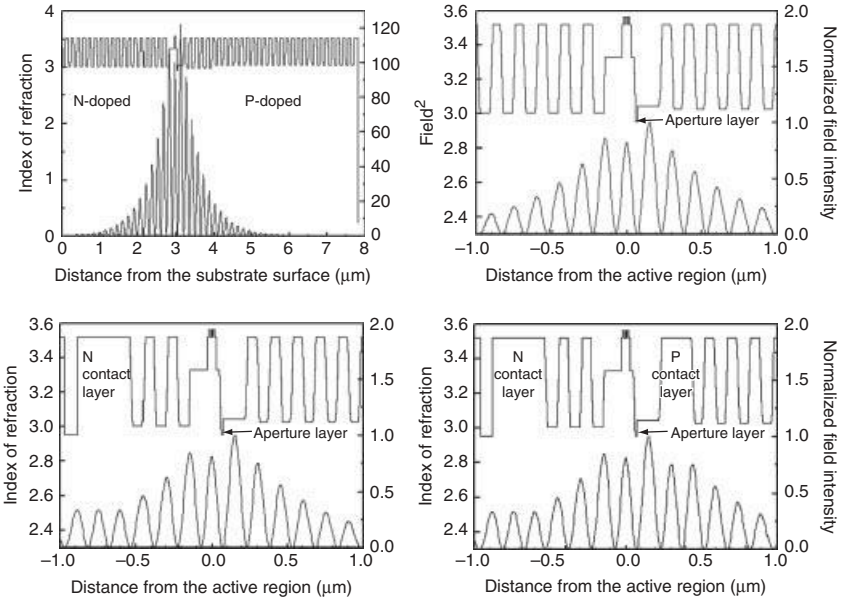


FIGURE 3.20: Examples of VCSEL cavities designed for oxide apertures, showing $|\mathscr{E}|^2$-field standing wave patterns and refractive indices of various layers within the cavity: (top) Full cavity and cavity expanded view of device designed for top and bottom contacts. P-doped mirror contrast is reduced by reducing Al content and doping is increased in upper portion to reduce series resistance. (bottom left) Expanded view of n intra-cavity contacted structure—typically grown on semi-insulating substrate for low parasitic capacitance high-speed operation. (bottom right) Expanded view of dual intra-cavity contacted structure that enables simple two layer mirror design that avoids complex bandgap engineering as illustrated in Figure 3.21.

Figures 3.19 and 3.20 illustrate a number of possible VCSEL cavity designs. Figure 3.19 represents a basic one-wavelength cavity design typical for proton implanted devices. In Fig. 3.20, the layer structures for three different oxide-apertured VCSEL designs are illustrated—one with both top and bottom contacts, one with a single intracavity contact layer, and one with two intra-cavity contracts. The oxide aperture is used for current and photon confinement, as discussed in Chapter 1. Again, the most important feature of a VCSEL design is the placement of the quantum-well gain layers on a peak of the $|\mathscr{E}|^2$-field standing-wave, as seen in Figures 3.19 and 3.20. The cavity spacing, $L_a + L_p$, between common mirror reference planes is indicated in Fig. 3.19. This when added to the effective penetration length into both mirrors, $L_{eff\,1} + L_{eff\,2}$, gives L_{DBR}. As can be seen the addition of only a single layer can make a significant difference in the total effective cavity length, L_{DBR}, and thus, also in quantities like the VCSEL confinement factor or relaxation resonance frequency.

Example 3.4 A 980-nm bottom-emitting VCSEL needs to be designed. The average internal losses and the injection efficiency are 20 cm^{-1} and 0.9, respectively. The mirrors are composed of quarter wave stacks of AlAs/GaAs. The top mirror needs to have a power reflection of 99.6% including air interface, and the bottom (output) mirror needs to have a power reflection of 98.4%. The cavity between the DBR mirrors has to be one optical wavelength measured in the material, and it consists of 3 InGaAs quantum wells, each 8 nm thick, separated by 2–8 nm thick GaAs barriers, and clad on each side by another 8-nm-thick layer of GaAs. The rest of the cavity consists of an AlGaAs spacer with 20% of Al. The mirrors begin with an AlAs quarter-wave layer next to the cavity. The quantum wells are centered between the mirrors for best confinement factor.

Problem: (a) Determine the thickness of the two AlGaAs spacer layers (b) Determine the number of the AlAs/GaAs periods in the top and bottom DBR mirrors (c) What is the differential efficiency measured out of the bottom (output) DBR mirror?

Solution: A schematic of this type of cavity is shown in Fig. 3.19. From the statement of the problem, we know that the cavity length is equal to one optical wavelength in the material, $\Sigma n_i l_i = \lambda$. From Table 1.1, we have $n_{\text{AlAs}} = 2.95$, $n_{\text{GaAs}} = 3.52$, $n_{\text{AlGaAs(0.2)}} = 3.39$ and $n_{\text{InGaAs}} = 3.60$. Therefore,

$$3(3.60)8\text{ nm} + 4(3.52)8\text{ nm} + 2(3.39)s = 980\text{ nm} => s = 115.2\text{ nm}$$

Both mirrors start with an AlGaAs/AlAs interface. For the bottom mirror, the mirror ends with the semiconductor substrate interface. The number of mirror periods, m, relates to the number of reflection pairs because there are two reflections per each grating period, from two interfaces. Because, per Fig. 3.19, the periodic mirror section begins with an AlAs/GaAs layer and ends with an AlAs/GaAs interface, there has to be an odd number of half periods in the mirror, and m will be a half integer. Using Eq. (3.72), we have

$$r_{gb} = \frac{1 - \left(\frac{n_1}{n_2}\right)^{2m} \cdot \frac{n_1}{n_3}}{1 + \left(\frac{n_1}{n_2}\right)^{2m} \cdot \frac{n_1}{n_3}} => \left(\frac{n_1}{n_2}\right)^{2m} = \frac{1 - r_{gb}}{(1 + r_{gb}) \cdot \frac{n_1}{n_3}},$$

where $n_1 = n_{\text{AlAs}}$, $n_2 = n_{\text{GaAs}}$ and $n_3 = n_{\text{AlGaAs(0.2)}}$, and the n_1/n_3 term is due to the first reflection interface. From here, we can calculate, m, as

$$m = \frac{\ln\left(\frac{1-r_{gb}}{(1+r_{gb})\cdot\frac{n_1}{n_3}}\right)}{2\ln\frac{n1}{n2}} = \frac{\ln\left(\frac{1-\sqrt{0.984}}{(1+\sqrt{0.984})\cdot 0.870}\right)}{2\ln\left(\frac{2.95}{3.52}\right)} = 15.21$$

The closest half integer is $m = 15.5$. With that selection, we need to recalculate the total bottom mirror reflectance,

$$r_{gb} = \frac{1 - \left(\frac{2.95}{3.52}\right)^{31} \cdot \frac{2.95}{3.39}}{1 + \left(\frac{2.95}{3.52}\right)^{31} \cdot \frac{2.95}{3.39}} = 0.993.$$

For the top mirror, we apply the same procedure, except that this time we have an air interface at the end. The parity of the number of half periods of the top mirror will depend on which layer is the last mirror layer grown. If we assume that the top mirror ends on a GaAs layer, we will have an odd number of half periods of the mirror,

$$r_{gt} = \frac{1 - \left(\frac{n_1}{n_2}\right)^{2m} \frac{n_1}{n_3}\frac{1}{n_2}}{1 + \left(\frac{n_1}{n_2}\right)^{2m} \frac{n_1}{n_3}\frac{1}{n_2}} => m = \frac{\ln \frac{1-r_{gt}}{1+r_{gt}} \frac{n_2 \cdot n_3}{n_1}}{2 \ln \frac{n_1}{n_2}}.$$

Plugging in the parameter values, we have

$$m = \frac{\ln \left(\frac{1-\sqrt{0.996}}{1+\sqrt{0.996}} \frac{3.52 \cdot 3.39}{2.95}\right)}{2 \ln \frac{2.95}{3.52}} = 15.6.$$

Because m must be a half-integer, we need to select either 15.5 or 16.5. As we are designing a bottom emitter, we will select $m = 16.5$. With that selection, we need to recalculate the total top mirror reflectance,

$$r_{gt} = \frac{1 - \left(\frac{2.95}{3.52}\right)^{33} \frac{2.95}{3.39 \cdot 3.52}}{1 + \left(\frac{2.95}{3.39 \cdot 3.52}\right)^{33} \frac{2.95}{3.39 \cdot 3.52}} = 0.9985.$$

To compute the differential efficiency through the bottom mirror, given by Eq. (3.70), because we already know the injection efficiency, the internal loss, and the mirror loss, we need to determine the fraction of the output power F_2, and for it, we need to know the effective mirror length L_{eff}. L_{eff} can be calculated using the approximation of Eq. (3.63) for a strongly reflecting grating,

$$L_{eff} = \frac{\lambda_0}{4\Delta\bar{n}} = \frac{980 \text{ nm}}{4 \cdot 0.57} = 429.8 \text{ nm}.$$

To compute the power fraction emitted from the bottom mirror, F_2, we need to express the grating's lossless transmission coefficient $|t'_g|$ from the lossless reflection coefficient $|r'_g|$, using Eqs. (3.67) and (3.68),

$$F_2 = \frac{(1 - |r'_{gb}|^2)e^{-\alpha_i L_{eff}}}{(1 - |r'_{gb}|^2) + \frac{|r'_{gb}|}{|r'_{gt}|}(1 - |r'_{gb}|^2)} = = \frac{(1 - 0.993^2)e^{-0.00086}}{(1 - 0.993^2) + \frac{0.993}{0.9985}(1 - 0.9985^2)}$$

$$= 0.823.$$

Differential efficiency through the back mirror can be calculated using $A_i = \alpha_i \cdot L_{DBR}$, where $L_{DBR} = L_c + 2L_{eff} = 1146$ nm. From Eq. (3.70),

$$\eta_d = F_2 \eta_i \frac{T_m}{T_m + A_i} = 0.349 \cdot 0.9 \frac{\ln \frac{1}{\sqrt{0.9985 \cdot 0.993}}}{1146 \cdot 10^{-8} \cdot 20 + \ln \frac{1}{\sqrt{0.9985 \cdot 0.993}}} = 0.703.$$

Figure 3.21 illustrates the results of grading the material composition (Al-fraction) and modulating the doping ($N_a - N_d$) within a period of a p-doped DBR VCSEL mirror to reduce series resistance. Also indicated is the relative position of the $|\mathscr{E}|^2$-field standing wave and the resulting hole concentration. A bi-parabolic

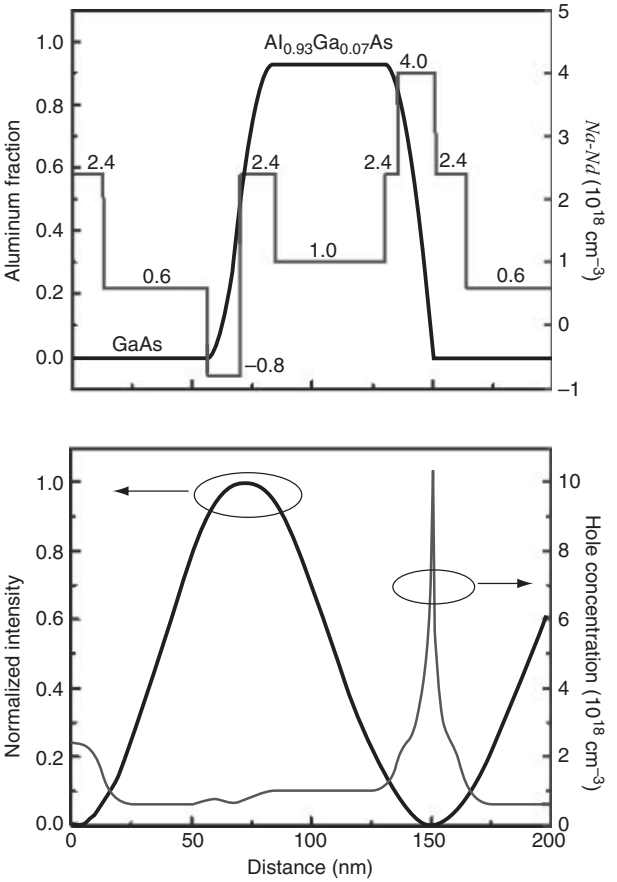


FIGURE 3.21: Illustration of bandgap engineering in a VCSEL DBR mirror—grading of composition and doping to reduce series resistance while avoiding large increases in optical losses or reduction in reflection.

compositional grade is used near the standing-wave peak, and a uni-parabolic grade is used near the null. As the internal loss is greatly increased by a large hole concentration, one aspect of the design is to provide for a relatively flat valence band for good conduction without placing a large free hole concentration near the optical standing wave maximum. Of course, the enhancement factor is also important in this loss analysis, but here we seek to place the lossy regions where the enhancement factor is close to zero. The compositional grading does reduce the reflection per period slightly, but for the compositional, thus index, grading illustrated here, the reduction is only about 6%, and this can be approximately compensated by adding one extra mirror period in a typical design.

3.6.5 In-Plane DBR Lasers and Tunability

The potential tunability of in-plane DBR lasers is one of the main reasons they are of great importance. As indicated in Fig. 3.16, there are usually three sections to a DBR, one active, one passive, and one passive grating. For the in-plane DBR it is convenient to place three separate control electrodes over these regions as shown in Fig. 3.22. One section provides gain, one allows independent mode phase control, and one can shift the mode-selective grating filter, respectively. By applying a control current or voltage to the grating section, its index, $\bar{n}_{DBR}$, changes, causing a change in the optical length of each grating period $\bar{n}_{DBR}\Lambda$ and its center wavelength λ_g. Because the optical length of a single grating period is defined in relationship to the Bragg wavelength as $\bar{n}_{DBR}\Lambda = \lambda_g/2$, the center wavelength of the grating moves according to

$$\frac{\Delta\lambda_g}{\lambda_g} = \frac{\Delta\bar{n}_{DBR}}{\bar{n}_{DBR}}. \tag{3.75}$$

Alternate axial modes can be selected as the mirror loss curve, $\alpha_m(\lambda)$, moves relative to the gain and modes. This is referred to as *mode hop* tuning. Also, the modes will move slightly in wavelength because part of the net cavity length (L_{eff}) is in the grating.

By applying a current or voltage to the phase control electrode, the index of the passive cavity section, $\bar{n}_p$, changes, shifting the axial modes of the cavity. Thus, by applying a combination of control signals to the grating and phase control sections, a broad range of wavelengths are accessible. Because the carrier density is clamped in the active region, changes in current there only have a second-order effect on its index, $\bar{n}_a$, and only small changes in mode wavelength result. We can see more explicitly how the continuous mode shift occurs by solving for the relative shift, $\Delta\lambda_m/\lambda_m$, from Eq. (2.29):

$$\frac{\Delta\lambda_m}{\lambda_m} = \frac{\Delta\bar{n}_a L_a + \Delta\bar{n}_p L_p + \Delta\bar{n}_{DBR} L_{eff}}{\bar{n}_{ga} L_a + \bar{n}_{gp} L_p + \bar{n}_{gDBR} L_{eff}}. \tag{3.76}$$

The "g" subscript in the denominator indicates that the group indices must be used here for the same reasons they appear in prior equations where changes in

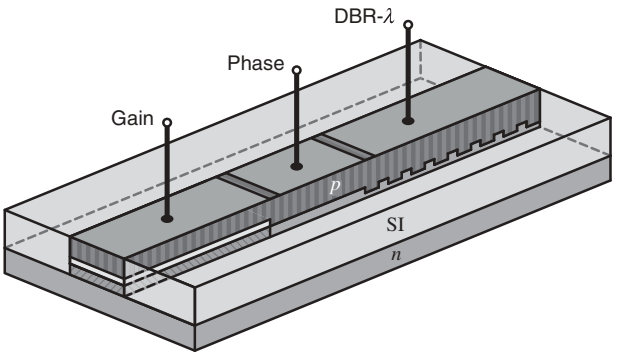


FIGURE 3.22: Illustration of a tunable single-frequency three-section DBR laser. The laser consists of a gain section, phase section, and DBR mirror section.

the wavelength are involved. From Eq. (2.30), we can see how the indices are shifted by carrier injection. For example, with a transverse confinement factor, $\Gamma_{xy} = 10\%$, the effective index shifts by $\Delta\bar{n}/\bar{n} \approx -0.1\%$ for $\Delta N = 10^{18}$ cm^{-3}. This can occur in the phase control and grating regions above threshold. Because these lengths typically account for about half of L_{DBR} in an in-plane laser, the wavelength would be continuously tuned by ~0.05%, or ~8 nm at 1.55 μm. The injected carrier density in the passive regions can be calculated from Eq. (2.16) with $g = 0$ and $dN/dt = 0$. Thus, the effective index change in the jth section can be written as

$$\Delta\bar{n}_j = \frac{\partial\bar{n}}{\partial N}\frac{\eta_i \tau I_j}{qV_j}, \tag{3.77}$$

where $[\partial\bar{n}/\partial N] \approx -\Gamma_{xy} 10^{-20}$ cm^3.

A reverse bias voltage can also change the index of refraction via linear and quadratic electro-optic effects. Effective index shifts of ~0.1% are possible, but this is a little lower than possible with high injection currents. However, this reverse biased effect can have a much faster response time than current injection because the carrier lifetime limits changes in the carrier density to a few hundred megahertz, similar to LEDs. Also, current injection leads to local heating, which can change the index with time constants in the tens of microseconds range. Reverse bias also is more practical in the multisection VCSEL case.

More complex structures than the basic three-section DBR are possible, and some of these will be discussed in Chapter 8 on photonic integrated circuits.

Example 3.5 An InGaAsP/InP tunable 1.55 μm DBR laser, as illustrated in Figure 3.22, is constructed. The length of the gain, phase and DBR sections are 300 μm, 100 μm, and 100 μm, respectively. Average transverse-lateral internal losses in the gain and phase sections are 15 cm^{-1} and 5 cm^{-1}, respectively, and

the injection efficiency is 0.8. The DBR power reflectivity is $|r_g|^2 = 0.2$, and its power transmission is $|t_g|^2 = 0.7$. The other end is a simple cleaved facet. Assume the modal index of $\bar{n} = 3.4$ for the entire structure.

Problem: (1) Determine the differential efficiencies out both ends. (2) Determine the mode spacing, neglecting the change in DBR penetration depth with wavelength. (3) If the modal index of the DBR section is tuned by 0.01, by how much is the new lasing mode shifted from the original lasing wavelength?

Solution: Because the DBR laser in this problem is described using reflection and transmission scattering parameters, we will use Case (a) from Section 3.6.3 to solve it. To determine the differential efficiencies from both ends, we first need to calculate the fractions of the power coming out of both ends, F_1 and F_2. Assuming the cleaved facet introduces no losses, $|t_1|^2 = 1 - |r_1|^2 = 0.68$. From Eq. (3.67),

$$F_1 = \frac{t_1^2}{(1 - r_1^2) + \frac{r_1}{|r_g|}(1 - |r_g|^2)} = \frac{0.68}{0.68 + \sqrt{\frac{0.32}{0.2}}(1 - 0.2)} = 0.4019$$

$$F_2 = \frac{t_g^2}{(1 - r_g^2) + \frac{|r_g|}{r_1}(1 - |r_1|^2)} = \frac{0.7}{(1 - 0.2) + \sqrt{\frac{0.2}{0.32}}(1 - .32)} = 0.5233$$

The mirror loss is given by Eq. (3.66),

$$\alpha_m = \frac{1}{L_a + L_p} \ln\left[\frac{1}{r_1 |r_g|}\right] = \frac{1}{0.04 \text{ cm}} \ln \frac{1}{\sqrt{0.32 \cdot 0.2}} = 34.36 \text{ cm}^{-1}$$

Average internal loss, $\langle \alpha_i \rangle$ is given by

$$\langle \alpha_i \rangle_{a+p} = \frac{\langle \alpha_{ia} \rangle L_a + \langle \alpha_{ip} \rangle L_p}{L_a + L_p} = \frac{300 \cdot 15 \text{ cm}^{-1} + 100 \cdot 5 \text{ cm}^{-1}}{400} = 12.5 \text{ cm}^{-1}$$

Differential efficiencies are now given from Eq. (3.32),

$$\eta_{d1} = F_1 \eta_i \frac{\alpha_m}{\langle \alpha_i \rangle_{a+p} + \alpha_m} = 0.4019 \frac{34.36}{34.36 + 12.5} 0.8 = 0.236$$

$$\eta_{d2} = F_2 \eta_i \frac{\alpha_m}{\langle \alpha_i \rangle_{a+p} + \alpha_m} = 0.5233 \frac{34.36}{34.36 + 12.5} 0.8 = 0.307$$

To calculate the mode spacing, we need to take into account the effective length of the DBR mirror. The mode spacing is given by Eq. (3.40), and for a weakly reflecting grating, we assume $L_{eff} \approx \frac{1}{2} L_{DBR} = 50$ μm.

$$d\lambda = \frac{\lambda^2}{2(\bar{n}_{ga} L_a + \bar{n}_{gp} L_p + \bar{n}_{gp} L_{eff})} = \frac{(1550 \text{ nm})^2}{2(3.8)(300 + 100 + 50 \text{ um})} = 0.702 \text{ nm}.$$

When the DBR mirror section is tuned, two effects come into play, as shown in Fig. 3.23. First, the grating Bragg wavelength will change, due to the changed grating index. Then, the cavity mode comb position will change as well because the effective optical cavity length is now different. For the modal index change of $\Delta\bar{n}_{DBR}$, the center wavelength of the grating moves in direct proportion to the index according to Eq. (3.75),

$$\Delta\lambda_g = \lambda_g \frac{\Delta\bar{n}_{DBR}}{\bar{n}_{DBR}} = 1550\frac{0.01}{3.4}\ \text{nm} = 4.56\ \text{nm}.$$

At the same time, the cavity modes will shift, and we can use Eq. (3.76) to determine the amount of shift,

$$\Delta\lambda_m = \frac{\Delta\bar{n}_{DBR}L_{eff}\lambda_m}{\bar{n}_{ga}L_a + \bar{n}_{gp}L_p + \bar{n}_{gDBR}L_{eff}} = \frac{0.01 \cdot 50\ \mu\text{m}\ \cdot 1550\ \text{nm}}{2(3.8)(300 + 100 + 50\ \mu\text{m})}$$
$$= 0.227\ \text{nm}$$

Finally, the new lasing mode will be the cavity mode that is the closest to the new Bragg wavelength. To compute it, we need to compute by how many cavity modes the Bragg wavelength shifted.

$$\Delta\lambda_{lasing} = \Delta\lambda_m + \left[\frac{\Delta\lambda_g - \Delta\lambda_m}{\delta\lambda}\right]\delta\lambda = 0.227\ \text{nm} + 6 \cdot 0.702\ \text{nm} = 4.439\ \text{nm}.$$

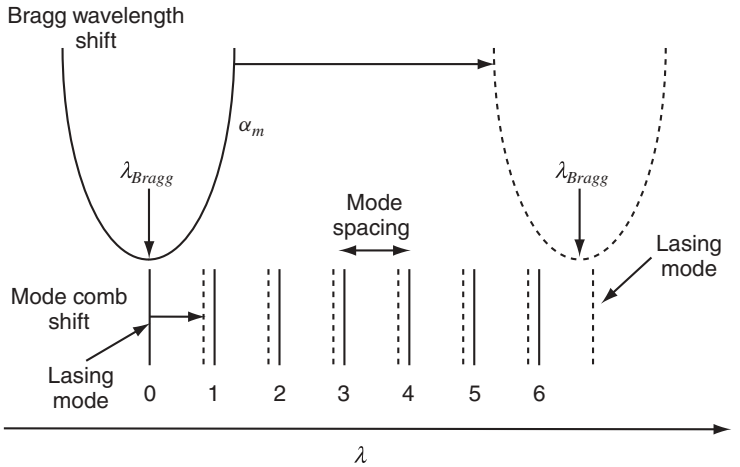


FIGURE 3.23: Illustration of mode tuning in a single-frequency three-section DBR laser. Dashed curves and lines apply after index is increased in the mirror. Two tuning effects can be observed: (1) DBR Bragg wavelength tuning, due to mirror index change and (2) cavity mode comb shift, due to its penetration into the DBR with an increased index.

3.6.6 Mode Suppression Ratio in DBR Laser

A primary reason that people are interested in DBR, coupled-cavity, and DFB lasers (covered later in this chapter) is their potential for single-frequency operation. However, in this context, we must realize that "single-frequency" is a relative term. The measure of single-frequency purity, the mode suppression ratio (MSR), is thus discussed in this section. Mode suppression ratio is simply the ratio of the output power in the primary laser mode to that in the next strongest mode from one end of the laser:

$$\text{MSR} = \frac{P(\lambda_0)}{P(\lambda_1)}, \tag{3.78}$$

where we have dropped the subscripts from the output power as given by Eq. (3.33) and labeled the primary mode as the one at λ_0. More fundamentally, the output power from one end of the laser at the nth mode is given by Eq. (2.38) multiplied by the fraction out one end, $F_1(\lambda_n)$ as given by Eq. (3.31). That is,

$$P(\lambda_n) = F_1(\lambda_n) v_g \alpha_m(\lambda_n) N_p(\lambda_n) h\nu V_p. \tag{3.79}$$

From Eq. (2.17) we can express the steady-state ($dP/dt = 0$) photon density as

$$N_p(\lambda_n) = \frac{\Gamma \beta_{sp} R_{sp}(\lambda_n)}{1/\tau_p(\lambda_n) - \Gamma v_g g(\lambda_n)}. \tag{3.80}$$

Note that in Chapter 2, we used Eq. (2.16) to solve for N_p in terms of the terminal current, but here we are interested in expressing it for the various modes in terms of the net gain margin, the denominator of Eq. (3.80). In this form we can see how the noise injected into a particular mode, $\Gamma \beta_{sp} R_{sp}(\lambda_n)$, is amplified to a large steady-state value as the denominator approaches zero. (But it never actually goes to zero for any finite power out.)

Now, we can plug Eq. (3.80) into Eq. (3.79) and form the ratio given in Eq. (3.78) for the desired pair of modes. That is,

$$\text{MSR} = \frac{F_1(\lambda_0)\alpha_m(\lambda_0)[\alpha_i + \alpha_m(\lambda_1) - \Gamma g(\lambda_1)]}{F_1(\lambda_1)\alpha_m(\lambda_1)[\alpha_i + \alpha_m(\lambda_0) - \Gamma g(\lambda_0)]}, \tag{3.81}$$

where we have assumed that the spontaneous emission is coupled equally into both modes, and that the modes are similar in frequency, volume, and velocity. We have also used Eq. (2.28) for the cavity lifetime (i.e., $1/\tau_p(\lambda_n) = v_g(\alpha_i + \alpha_m)$). Before trying to simplify this any further, we wish to review a generic schematic of loss and gain vs. wavelength shown in Fig. 3.24, similar to Figs. 3.10, 3.17, and indirectly 3.18 for different types of lasers.

By reference to the Figure 3.24, we can simplify Eq. (3.81) by calling the denominator bracket, which is the separation between the mirror loss and the net modal gain for the main mode, $\delta_G = \alpha_m(\lambda_0) - [\Gamma g(\lambda_0) - \alpha_i]$, the loss margin,

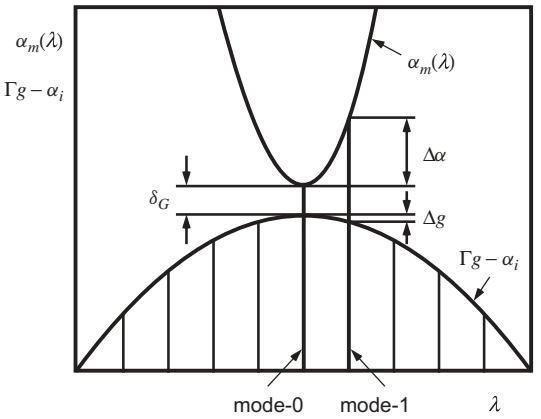


FIGURE 3.24: Definition of gain and loss margins for use in MSR calculations.

$\Delta\alpha = \alpha_m(\lambda_1) - \alpha_m(\lambda_0)$, and the modal gain margin, $\Delta g = \Gamma g(\lambda_0) - \Gamma g(\lambda_1)$. In one special case, if the back mirror provides the frequency-dependent loss, then the fraction of light coupled out of the front mirror will be reduced as the mirror loss is increased. As a result, the coupling fraction ratio out of the front mirror times the mirror loss ratio is $\sim$1. We are then left with the simple result

$$\text{MSR} \approx \frac{\Delta\alpha + \Delta g}{\delta_G} + 1. \tag{3.82}$$

More commonly, the MSR is expressed in terms of decibels (dB) of optical power:

$$\text{MSR(dB)} \approx 10\log_{10}\left[\frac{\Delta\alpha + \Delta g}{\delta_G} + 1\right]. \tag{3.83}$$

However, most generally, one cannot make the last assumption regarding the back mirror, which tends to apply mainly to DBR lasers, and we then must include the prefactors in front of the brackets in Eq. (3.81).

If the spectrum is observed by direct detection, the photodiode current is directly proportional to the optical power, whereas the electrical power is proportional to its square. Thus, if *electrical* power is displayed in decibels on a spectrum analyzer, the observed MSR will appear twice as large as (3.83) would predict.

The value of δ_G can be calculated in terms of cavity parameters and the drive current by combining a number of existing equations. First, we solve Eqs. (3.79) and (3.80) for δ_G at λ_0,

$$\delta_G = \frac{F_1\alpha_m h\nu V_p \Gamma\beta_{sp} R_{sp}}{P_{01}}. \tag{3.84}$$

Then, using Eqs. (2.5) and (2.6) for R_{sp}, and Eqs. (3.31) to (3.33) for P_{01}, we get an expression valid for $I > I_{th}$,

$$\delta_G = (\alpha_i + \alpha_m)\beta_{sp}\eta_r \frac{I_{th}}{(I - I_{th})}, \tag{3.85}$$

and the spontaneous emission factor, β_{sp}, is given by Eq. (A4.10). For typical values of the parameters $\delta_G \sim 10^{-3} I_{th}/(I - I_{th})$ cm^{-1}.

Even though we have developed the mode suppression ratio calculation formalism for DBR lasers, it can be applied to other types of single-mode lasers as well. An example of MSR calculation is thus given at the end of the next section about DFB lasers.

3.7 DFB LASERS

3.7.1 Introduction

A distributed feedback laser (DFB) also uses grating mirrors, but gain is included in the gratings. Thus, it is possible to make a laser from a single grating, although it is usually desirable to have at least a fraction of a wavelength shift near the center to facilitate lasing at the Bragg frequency. Historically, the DFB laser preceded the DBR, primarily because of its simplicity and relative ease of fabrication; that is, no active-passive transitions are necessary. Fig. 3.25 gives a schematic of in-plane versions. Vertical cavity versions are also possible, but there is no advantage over the DBR, and the fabrication is not any easier.

The basic characteristic equation can be cast in the same form as for other lasers (i.e., Eq. (2.21)), but the gain is now in the complex mirror reflectivities, r_1 and r_2. Also, to avoid the troublesome active–passive transitions, there is no passive cavity ($L_p = 0$), and the additional active cavity length (L_a) is typically only a fraction of a wavelength. The complex mirror reflectivities are given by Eq. (3.54) if antireflection (AR) coatings are used at the ends. If no AR coatings are used, one more **T**-matrix must be multiplied by that of Eq. (3.49) before calculating the grating's S_{11}.

For no shift in the gratings, the cavity can be taken to be anywhere within the DFB because all periods look the same. The active length is then a quarter-wavelength long because we have chosen mirror reference planes to fall at a downstep in index looking both to the left and right. (As discussed earlier, this yields a zero grating reflection phase at the Bragg frequency.) Thus, at the Bragg frequency, we can see that this DFB is antiresonant. Because the cavity is so short, we can neglect any phase change in it over the reflection band of the mirrors. Thus, inserting $L_a = \lambda/4\bar{n}$ in Eq. (2.21) and assuming uniform pumping, the threshold condition becomes,

$$r_{g1}(\tilde{\beta}_{th}) r_{g2}(\tilde{\beta}_{th}) = -1. \qquad \text{(unshifted)} \tag{3.86}$$

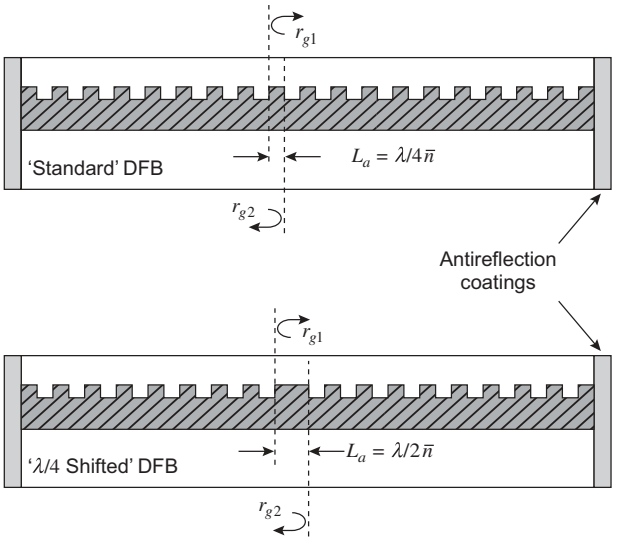


FIGURE 3.25: Standard and quarter-wave shifted DFB lasers. The entire length of a DFB laser is filled with active material embossed with a grating.

Because there is no solution at the Bragg frequency, the wavelength must be scanned for each gain until Eq. (3.86) is satisfied.

The exception to this case is the *gain-coupled* DFB, in which the deviations in refractive index are purely imaginary, as for the case of added gain or loss. For example, the grating could consist of alternate sections of index n_1 and index $n_2 = n_1 + jn_i$; (by definition, $n_i = g\lambda/(4\pi)$ for added gain). Then, the reflection at each discontinuity, $r = jn_i/(2n_1)$, and the net grating reflection, r_g, at the Bragg frequency would be purely imaginary (reflection phase of $\pi/2$) for the selected reference planes indicated in Fig. 3.12. Thus, the fundamental solutions to Eq. (3.86) do occur at the Bragg frequency in this case.

If we again consider only real index perturbations and the cavity is *half* a wavelength long, we can see that the device is resonant at the Bragg frequency where the reflection phase is zero. Actually, as can be seen in Fig. 3.25, a half-wavelength mirror spacing corresponds to a quarter-wave shift between the two gratings. Thus, this configuration is usually referred to as a *quarter-wave-shifted* DFB. In this case we use $L_a = \lambda/2\bar{n}$ in Eq. (2.21), and the threshold condition for uniform pumping becomes

$$r_{g1}(\tilde{\beta}_{th})r_{g2}(\tilde{\beta}_{th}) = 1. \qquad \text{(quarter-wave-shifted)} \tag{3.87}$$

The threshold gain and wavelength solutions to the quarter-wave shifted case are a little easier because we know they occur near the Bragg condition. In both cases, if the pumping is different on the two sides of the cavity, the complex propagation

constants, $\tilde{\beta}_{th}$, will be different in each grating section. In such cases, there will be pairs of threshold gains (one in each section) that satisfy the threshold condition.

For the unshifted DFB, it should be realized that two modes equally spaced on each side of the Bragg wavelength reach threshold simultaneously, if there exist no additional perturbing reflections, such as from uncoated cleaves at the end. This is illustrated in Fig. 3.26a. This simplest of DFBs must rely on such additional reflections to destroy the unwanted degeneracy. In practice, at least one cleaved mirror will do the job if the gratings are not too strong ($\kappa L_g \leq 1$) so the net reflection phase from one end is shifted from that of the grating alone. However, there is still a yield problem because the reflection from the cleave will have a random relative phase, but optimally it should be in quadrature to shift the net phase from that laser end by the maximum amount. One more **T**-matrix must be multiplied by that of Eq. (3.49) to obtain the net grating S_{11} in this case, as mentioned earlier.

3.7.2 Calculation of the Threshold Gains and Wavelengths

The threshold gain and wavelength can also be calculated by observing the net transfer function, $S_{21}(\omega) = 1/T_{11}(\omega)$, through the DFB laser (or any other laser) as the gain is increased, rather than solving Eq. (3.86) or Eq. (3.87). The poles of one of the S-parameters (or zeros of the T_{g11} parameter) *for the entire system* give another form of the characteristic equation. From Eq. (3.50), the characteristic equation of a DFB with AR coated facets can be written as

$$m_{eff}\,\Delta = j, \tag{3.88}$$

where the effective number of mirror periods, m_{eff}, and the detuning parameter, Δ, are defined in Eq. (3.51).

In a numerical calculation, as the gain is increased, the transmission spectrum (S_{21}) of the device under study will develop a strong maximum. The gain required for this maximum to reach some large value and its wavelength are the desired threshold values. This technique is particularly useful to determine the threshold gain margin for spurious modes because even after one mode reaches threshold, the gain can still be increased to look for the next mode to blow up at another wavelength. Figure 3.26 shows example S_{21} spectra together with normalized threshold gains and wavelengths of the first few modes and for several κL_g values for both the standard (a) and quarter-wave-shifted (b) DFB cases. AR coatings are assumed to avoid any additive end reflections.

Figure 3.27 explores the ranges of DFB behavior between the standard and the quarter-wave-shifted designs. Parts (a) and (b) look at the case of gradually introducing the extra length, ΔL_a in the center of the cavity to transition from a standard to a quarter-wave-shifted design as suggested by the inset, while part (c) introduces a fixed quarter-wave-shift from one end and moves it to the center of the DFB grating. Part (a) is really just the locus of points between the end points given by data in Fig. 3.26 for $\kappa L_g = 1$ for the standard and quarter-wave-shifted extremes.

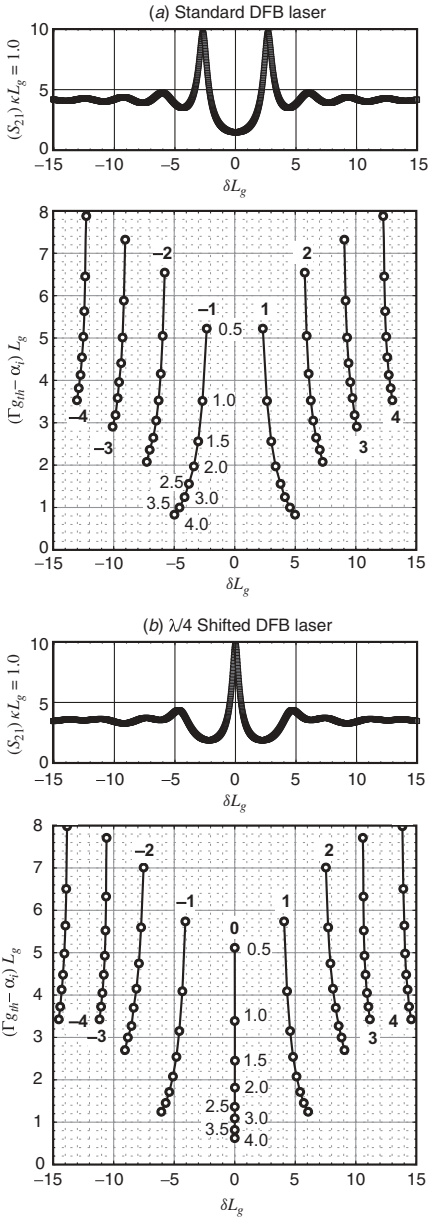


FIGURE 3.26: Example S_{21} spectra and normalized plots of threshold modal gain and threshold wavelength for different modes of (a) standard and (b) quarter-wave shifted DFB lasers. S_{21} spectra shown for $\kappa L_g = 1$ at a gain just below threshold; plots summarize analogous threshold points for this lowest order and some higher order modes with $\kappa L_g\,(\equiv 2mr)$ ranging from 0.5 to 4 in 0.5 increments. Here $\delta = \beta - \beta_0$, where β is the average propagation constant of the grating.

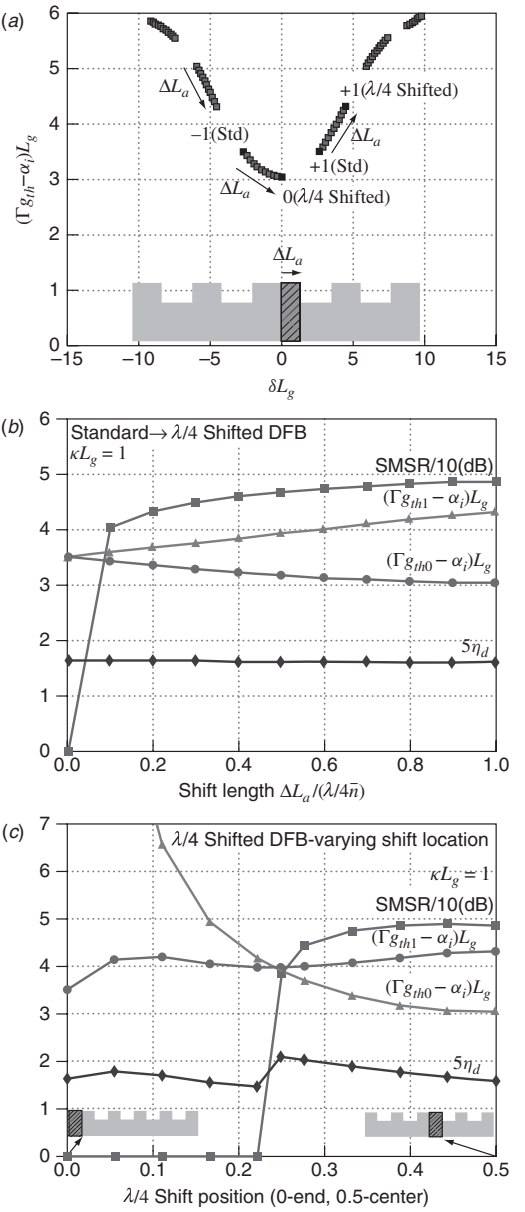


FIGURE 3.27: Characteristics of AR-coated DFBs in transition between the standard and quarter-wave-shifted designs. (a) Normalized threshold modal gain and wavelength as the shift in the center of the cavity, ΔL_a, is increased in increments of 10% of a quarter wavelength of the guided mode from 0 to full quarter-wave shift. (b) & (c) Plots of side-mode-suppression ratio, normalized modal gains for the first two modes, and differential efficiency out of the end nearest the shift; for (b) vary ΔL_g, for (c) vary position of fixed quarter-wave shift from end. Assumed cavity parameters are: $\alpha_i = 10\ \text{cm}^{-1}$, $\eta_i = 0.75$, $\beta_{sp} = 1.25 \cdot 10^{-5}$ and $\eta_r = 0.8$ (per Eq. (3.85)), $I = 4I_{th}$.

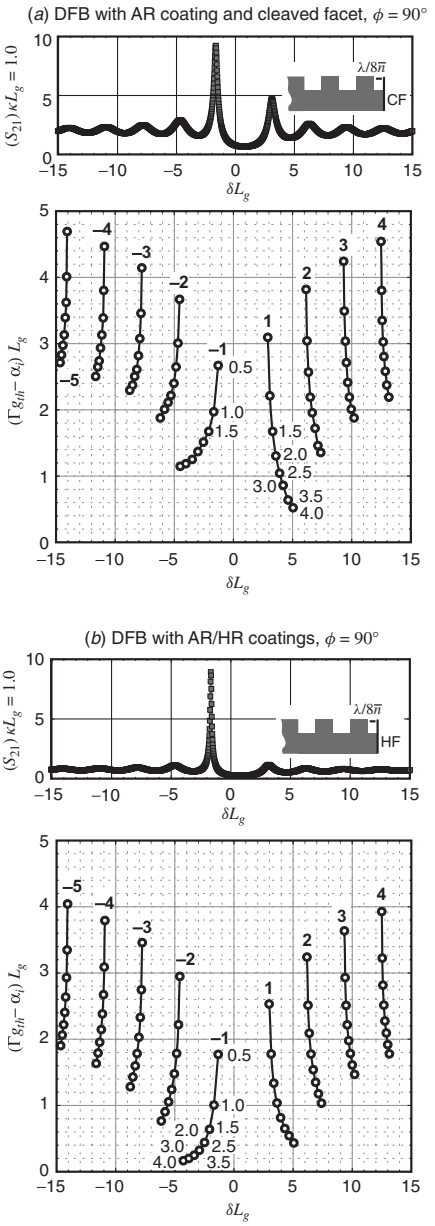


FIGURE 3.28: Example S_{21} spectra and normalized plots of threshold modal gain and threshold wavelength for different modes of a standard DFB laser with one end AR-coated and the other end either (a) a simple cleave, $R = 0.32$, or (b) HR-coated with $R = 0.9$. S_{21} spectra shown for $\kappa L_g = 1$ at a gain just below threshold; plots summarize analogous threshold points for this lowest order and some higher-order modes with $\kappa L_g (\equiv 2mr)$ ranging from 0.5 to 4 in 0.5 increments. The additive end reflections are optimally placed to provide orthogonally phased reflections (as illustrated by the inset).

To calculate example side-mode suppression ratios and differential efficiencies, for parts (b) and (c), it was necessary to assume some cavity parameters as listed in the caption. As can be seen from Fig. 3.27b, a physical shift (or other equivalent phase shift) of somewhat less than 1/10 of a quarter wavelength in the medium is sufficient to break the modal degeneracy and provide good single-mode operation with high side-mode suppression ratio, so the amount of this shift is not so critical. Figure 3.27c deals with the influence of the relative position of the quarter-wave shift inside a laser cavity. It shows that shift needs to be relatively close to the center of the device for good modal purity. In addition, if we are trying to optimize the cavity design to increase the output power from one end of the cavity, Fig. 3.27c shows that only marginal improvement in efficiency can be gained by moving the shift toward the output side.

Figure 3.28 gives S_{21} spectra and normalized threshold modal gains vs. wavelengths for the cases of standard DFB lasers with one end AR-coated, but the other having an additive reflection, either from (a) a simple cleave, $R = 0.32$, or (b) from an HR-coating, $R = 0.9$. In these cases, the reflections are assumed to have phases orthogonal to those of the gratings to provide maximal perturbation of the resultant. Figure 3.29 gives the side-mode-suppression-ratio, normalized modal gains of the first two modes, and the differential efficiency from the AR-coated end for both cases as the location, and therefore phase, as the reflection is shifted relative to the grating. As can be seen, when the additive reflection is either exactly in-phase or out-of-phase with the grating reflection, the laser still behaves as a standard DFB with two degenerate modes. On the other hand, when the end reflection has a somewhat different reflection phase from the grating, it provides a net reflection vector that shifts the mode from the point of this degeneracy, so that reasonable SMSR can be obtained. However, again we note that the SMSR is good over a relatively wide range of cleave phases.

The corrugations of the grating can also cause significant periodic loss and gain variations. That is, r_1 (and κ) may be complex (have a nonzero phase angle) even with the reference planes chosen earlier, which makes r_1 real for real index variations. In the extreme of pure gain modulation and no index variation, Eqs. (3.43) through (3.54) show that r_g would have an angle of $\phi = \pi/4$ at the Bragg wavelength. Thus, as already discussed, the unshifted DFB laser characterized by Eq. (3.86) provides a mode at the Bragg wavelength of the gratings. It performs as the quarter-wave-shifted DFB with real index variations. For these reasons and other potential benefits of this *gain-coupled* DFB, there has been continued active research in this direction.

Example 3.6 An InGaAsP/InP 1550 nm quarter-wave-shifted multiple quantum well buried-heterostructure DFB laser was fabricated from a structure whose confinement factor is $\Gamma = 0.06$. The injection efficiency for this structure was determined to be $\eta_i = 75\%$, and the internal loss $\alpha_{ia} = 10\ \text{cm}^{-1}$. The grating coupling constant is $\kappa = 20\ \text{cm}^{-1}$, the average effective index of the waveguide is 3.21, and the laser length is 500 μm. The laser facets are AR coated.

Problem: (1) Determine the threshold modal gain of this laser. (2) Determine the differential efficiency for this laser. (3) Determine the lasing wavelength of this laser. (4) Determine the threshold modal gain, differential efficiency and the lasing wavelength for an HR-AR standard DFB laser with the same parameters, assuming optimal HR mirror phase.

Solution: To solve this problem, we need to use Figure 3.26b, which provides us with the normalized solutions for the quarter-wave-shifted DFB mirror loss versus wavelength detuning, in function of the grating length and coupling coefficient. From the problem formulation, $\kappa L_g = 500 \cdot 20 \cdot 10^{-4} = 1.0$. Using the chart for $\kappa L_g = 1.0$,

$$A = (\Gamma g_{th} - \langle \alpha_i \rangle) L_g = 3.1 => \Gamma g_{th} = \frac{3.1}{L_g} + \langle \alpha_i \rangle$$

$$= \frac{3.1}{500 \cdot 10^{-4}} \text{ cm}^{-1} + 10 \text{ cm}^{-1} = 72 \text{ cm}^{-1}.$$

For a DFB laser, the differential efficiency is given by

$$\eta_d = \eta_i \frac{A}{\Gamma g_{th} L_g} = 0.75 \frac{3.1}{72 \cdot 0.05} = 0.646,$$

where half of the power would be emitted from each end of the grating, yielding a single-sided differential efficiency of $\eta_d = 0.323$. In a quarter-wave-shifted DFB laser, the detuning of the lasing wavelength from the Bragg wavelength is zero (as seen from the chart in Figure 3.26b), so the lasing wavelength is 1550 nm.

For the HR-AR coated DFB laser, we now need to use the chart from Fig. 3.28b. We follow the same procedure. Using the chart for $\kappa L_g = 1.0$,

$$A = (\Gamma g_{th} - \langle \alpha_i \rangle) L_g = 1.2 => \Gamma g_{th} = \frac{1.2}{L_g} + \langle \alpha_i \rangle$$

$$= \frac{1.2}{500 \cdot 10^{-4}} \text{ cm}^{-1} + 10 \text{ cm}^{-1} = 34 \text{ cm}^{-1}.$$

$$\eta_d = \eta_i \frac{A}{\Gamma g_{th} L_g} = 0.75 \frac{1.2}{34 \cdot 0.05} = 0.529,$$

where in this case, nearly all the power is coming out of the AR-coated facet.

Using the normalized plot of threshold modal gain and wavelength from Fig. 3.28, the normalized detuning for the lasing mode is $\delta L_g = 0.35$, therefore the lasing wavelength is

$$\beta = \frac{2\pi}{\lambda} \bar{n} = \frac{2\pi}{\lambda_0} \bar{n} + \frac{\delta L_g}{L_g} => \lambda = \frac{2\pi \bar{n}}{\frac{2\pi}{\lambda_0} \bar{n} + \frac{\delta L_g}{L_g}}$$

$$= \frac{2\pi}{0.013 + 0.0000007} \text{ nm} = 1551.087 \text{ nm}.$$

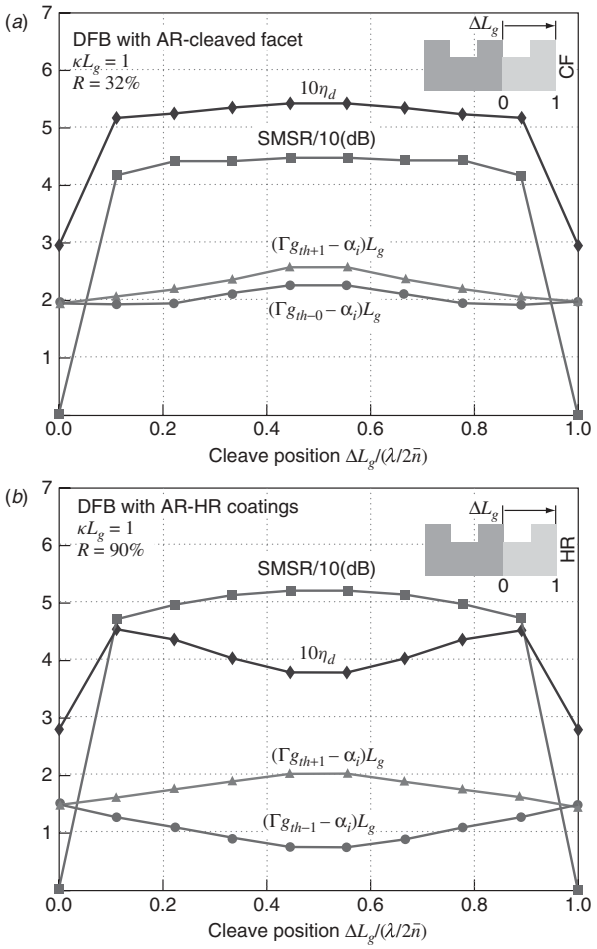


FIGURE 3.29: Characteristics of standard DFB lasers with one end AR-coated and the other end having an additive reflection as the location of this additive reflection is shifted. Case (a) a simple cleave, $R = 0.32$, and (b) HR-coated with $R = 0.9$. Plots of side-mode-suppression ratio (SMSR), normalized modal gains for the first two modes, and differential efficiency out of the AR-coated end. Assumed cavity parameters are: $\alpha_i = 10$ cm^{-1}, $\eta_i = 0.75$, $\beta_{sp} = 1.25 \cdot 10^{-5}$ and $\eta_r = 0.8$ (per Eq. (3.77)), $I = 4I_{th}$.

3.7.3 On Mode Suppression in DFB Lasers

Mode suppression ratio of DFB lasers can be very large, as has been discussed and shown in the previous parts of this section. To calculate it, we utilize the same definition, based on Eq. (3.81),

$$\mathrm{MSR} = \frac{F_1(\lambda_0)\alpha_m(\lambda_0)[\alpha_i + \alpha_m(\lambda_1) - \Gamma g(\lambda_1)]}{F_1(\lambda_1)\alpha_m(\lambda_1)[\alpha_i + \alpha_m(\lambda_0) - \Gamma g(\lambda_0)]}, \tag{3.89}$$

However, for the case of a DFB laser, we again need to recall that $\alpha_m \equiv \Gamma g_{th} - \alpha_i$ and use the values either determined from the normalized modal gain versus detuning charts in the last section or calculated accordingly. Therefore, the following expression is more useful for the case of DFB lasers (expressing the SMSR in dB),

$$\text{MSR} = 10 \log \frac{F_1(\lambda_0)\alpha_m(\lambda_0)}{F_1(\lambda_1)\alpha_m(\lambda_1)} \left[\frac{\Delta\alpha}{\delta_G} + 1 \right].$$

Figures 3.27, 3.28, and 3.29 illustrate SMSR values for various DFB laser types, with certain typical values of cavity parameters.

Example 3.7 Two DFB lasers from Example 3.6, quarter-wave-shifted and standard HR-AR-coated, are operating at four times their respective threshold currents. Assume the values of $\beta_{sp} = 1.25 \cdot 10^{-5}$ and $\eta_r = 0.8$.

Problem: Determine the side mode suppression ratio for both laser types.

Solution: To determine the side mode suppression ratio, we need to calculate the differences between modal gains and mirror losses for the two adjacent DFB modes and use the following equation (which neglects the gain change Δg), between the two modes,

$$\text{MSR} = 10 \log \frac{F_1(\lambda_0)\alpha_m(\lambda_0)}{F_1(\lambda_1)\alpha_m(\lambda_1)} \left[\frac{\Delta\alpha}{\delta_G} + 1 \right]$$

In the case of a quarter-wave-shifted DFB laser, exactly half of the output power will come out of each facet, and this will not be wavelength dependent. Therefore, $F_1(\lambda_0) = F_1(\lambda_1) = 0.5$. From Fig. 3.26b, for the mode $+1$ and $\kappa L_g = 1.0$, the normalized mirror loss is

$$A_{+1} = (\Gamma g_{th} - \langle \alpha_i \rangle) L_g = 4.4.$$

From Example 3.4, we have the value for this same parameter for the fundamental mode, $A_0 = 3.1$.

We compute δ_G for the fundamental mode, using Eq. (3.85),

$$\delta_G = (\alpha_i + \alpha_m)\beta_{sp}\eta_r \frac{I_{th}}{(I - I_{th})} = \left(10 + \frac{3.1}{500 \cdot 10^{-4}} \right) \text{ cm}^{-1}\ 10^{-5} \frac{I_{th}}{4I_{th} - I_{th}}$$

$$= 72 \cdot 10^{-5} \cdot \frac{1}{3} \text{ cm}^{-1} = 24 \cdot 10^{-5} \text{ cm}^{-1}.$$

Finally, the MSR is given by

$$\text{MSR} = 10 \log \frac{1 \cdot A_0/L_g}{1 \cdot A_1/L_g} \left[\frac{\frac{A_1 - A_0}{L_g}}{\delta_G} + 1 \right] = 10 \log \frac{3.1}{4.4} \left[\frac{26}{24 \cdot 10^{-5}} + 1 \right]$$

$$= 48.82 \text{ dB}.$$

We note that this is a very large value. For the HR-AR-coated laser, we can assume that all the output power will be coming out of one facet, and that there is no wavelength dependence in this behavior. Therefore, $F_1(\lambda_0) = F_1(\lambda_1) = 1.0$. From Fig. 3.28b, and Example 3.4, we have $A_{+1} = (\Gamma g_{th} - \langle \alpha_i \rangle)L_g = 1.8$ and $A_0 = 1.2$. This yields a δ_G of

$$\delta_G = (\alpha_i + \alpha_m)\beta_{sp}\eta_r \frac{I_{th}}{(I - I_{th})} = (10 + \frac{1.2}{500 \cdot 10^{-4}})\ \text{cm}^{-1}\ 10^{-5}\frac{I_{th}}{4I_{th} - I_{th}}$$

$$= 34 \cdot 10^{-5} \cdot \frac{1}{3}\ \text{cm}^{-1} = 11.33 \cdot 10^{-5}\ \text{cm}^{-1}.$$

Therefore, the MSR can be calculated as

$$\text{MSR} = 10\log\frac{1 \cdot A_0/L_g}{1 \cdot A_1/L_g}\left[\frac{\frac{A_1 - A_0}{L_g}}{\delta_G} + 1\right] = 10\log\frac{1.2}{1.8}\left[\frac{12}{11.33 \cdot 10^{-5}} + 1\right]$$

$$= 48.49\ \text{dB}.$$

REFERENCES

1. Morgan RA, Hibbs-Brenner MK, Walterson RA, Lehman JA, Marta TM, Bounnak S, Kelweit EL, Akinwade T, Nohava JC. Producible GaAs-based MOVPE-grown vertical-cavity top-surface emitting lasers with record performance. *Electronics Letters* 1995; **31**: 462.
2. Nise N. *Control Systems Engineering*. Hoboken, NJ: John Wiley; 2011.
3. Coldren LA, Corzine SW. *Diode lasers and photonic integrated circuits*. New York: John Wiley; 1995.

READING LIST

Agrawal GP and Dutta NK, 2d ed. *Semiconductor Lasers*, Chs. 7 and 8, Van Nostrand Reinhold, New York (1993).

Haus HA, *Waves and Fields in Optoelectronics*, Ch. 3, Prentice Hall, Englewood Cliffs, NJ (1984).

Ramo S, Whinnery JR, and Vanduzer T, *Fields and Waves in Communication Eectronics*, Ch. 11, Wiley, New York (1984).

Björk G and Nilsson O, *IEEE J. Lightwave Technol.*, **5**, 140 (1987).

PROBLEMS

These problems include material from Appendix 7.

1. (a) Write the **S**- and **T**-matrices between two ports bounding a section of transmission line and a dielectric interface as shown in Fig. 3.30.

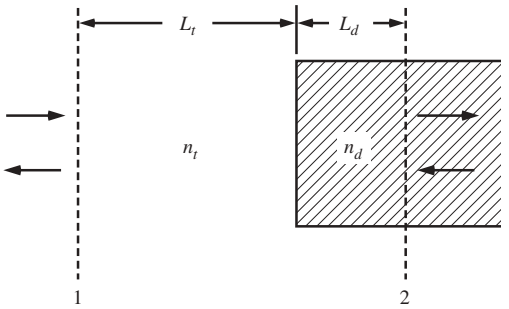


FIGURE 3.30: Two-port scattering junction consisting of a section of transmission line and a section of dielectric.

(b) For $n_t = 1, n_d = 3.5, L_t = 10$ μm, and $L_d = 5$ μm, plot S_{11} and S_{21} vs. wavelength for $900 < \lambda < 1000$ nm.

2. Verify Eqs. (3.24) to (3.27) by solving for the appropriate ratios of scattering amplitudes with the appropriate boundary conditions.

3. Verify Eq. (3.31) by showing that $F_1 + F_2 = 1$ and that $F_1/F_2 = P_{01}/P_{02}$.

4. Write the characteristic equation for a three-mirror laser using only its mirror reflectivities and dimensions (i.e., no r_{eff} in the answer).

5. Plot the effective mirror reflectivity, r_{eff}, vs. wavelength near 1.3 μm over two full cycles of oscillation for an external cavity laser of the form illustrated in Fig. 3.9. Assume the external cavity medium has an index of 1.6, a loss of 5 cm^{-1}, a length $L_p = 200$ μm, and that the mirrors, $r_3 = -0.9, r_2 = 0.5, t_2 = 0.3$ (mode mismatch loss).

6. In Problem 5, if we tune the index of the external cavity by 1%, by how many nanometers do the maxima of r_{eff} shift?

7. In a four-mirror coupled-cavity laser, it is desired to have the maxima of r_{eff} be narrower in wavelength than the minima for best mode selectivity.

A device is fabricated by etching a deep and narrow groove across 1.55 μm InGaAsP/InP DH material to form two active sections. Assume the minimum accurately controlled groove width is 1.2 μm.

(a) Find the minimum groove width to accomplish the desired goal.

(b) Assuming a diffraction power loss of 50% per pass in coupling from one section to the other, what are the values of S_{s11}, R', σ, and r_{eff} in Eq. (3.42) for this case?

(c) Plot r_{eff} vs. wavelength over two periods for this case.

8. Using Mason's rule, confirm the expression for r_{eff} for a complex 3 mirror cavity, as shown in Eq. (3.35).

9. Verify Eq. (3.44).

10. Show that Eq. (3.54) reduces to Eq. (3.58) in the low reflection limit.

11. A VCSEL mirror consists of three grating periods backed by a metallic reflector. The position of the metallic reflector is adjusted so that its reflection adds in-phase with the grating's. Assume the amplitude reflection at each discontinuity of the grating is 0.1, and the metallic layer has an amplitude reflection of 0.95.

(a) What is the net amplitude reflectivity at the Bragg frequency?

(b) What is the effective penetration depth measured in wavelengths? Be sure to define a reference plane.

12. An $In_{0.2}$ $Ga_{0.8}As$/AlGaAs VCSEL as in Fig. 3.16 has the following parameters: $L_a = 0.02$ μm GaAs DH active region placed at standing wave peak. $L_p + L_a = 1$ wavelength in the medium.
Two DBR mirrors: AlAs/GaAs quarter-wave stacks 18 periods each; top (rear) mirror metalized to give a net mirror reflectivity of 99.9%.
Average internal loss, $\langle \alpha_i \rangle = 20$ cm^{-1}. Lasing wavelength 980 nm.

(a) What is the bottom (front) mirror reflectivity?

(b) What is the effective penetration depth into each mirror?

(c) What is the threshold modal gain?

(d) What is the differential efficiency, assuming $\eta_i = 100\%$?

13. Consider the bottom (front) DBR mirror of Problem 12.

(a) Plot its power reflectivity vs. wavelength relative to the Bragg value. Show two minima on each side of the central maximum.

(b) Plot ln(1/R) for this grating vs. wavelength.

14. A tunable three-section DBR as in Fig. 3.22 is constructed to operate near 1.55 μm from InGaAsP/InP materials. Above threshold, the wavelength is tuned by changing the effective indices in the phase and DBR passive sections by injecting current. For no current injection, the operating wavelength is 1.57 μm, the effective index in all sections is 3.4, $\partial \bar{n}/\partial N = 10^{-21}$ cm^3, $\eta_i = 70\%$, and the carrier lifetime is independent of carrier density and equals 3 ns in all sections. The waveguide cross section in all regions is 0.2×3 μm; the gain, phase shift, and grating regions are each 200 μm long; and the grating has a reflectivity per unit length of 100 cm^{-1}. The other mirror is a cleaved facet. Plot the wavelength vs. current to the grating:

(a) Assume no current is applied to the phase shift region and show at least three axial mode jumps.

(b) Repeat for a phase shift current sufficient to maintain operation at the grating's Bragg wavelength.

(c) In (b) also plot the required phase shift current on the opposite axis. Stop plots when any current reaches 50 mA.

15. A 1.55 μm InGaAsP DBR laser consists of an active section 500 μm long butted to a passive grating section 500 μm long. The coupling constant–length product, $\kappa L_g = 1$ for the grating. The active section is terminated with a cleaved facet on the opposite end. The active section injection efficiency is 70%, and the average internal modal losses are 20 cm^{-1} throughout both sections.

(a) What is the effective DBR mirror length?

(b) What are η_{d1} and η_{d2}?

16. A quarter-wave shifted DFB laser has an internal quantum efficiency of 60%, a modal loss of 10 cm^{-1}, AR coated facets, $L_g = 500$ μm, $\kappa L_g = 1, \eta_r = 1$.

(a) What is the threshold modal gain?

(b) For operation at twice threshold with $\beta_{sp} = 10^{-4}$, what is the MSR?

(c) What is the differential efficiency from each end?

(d) What κL_g gives the best MSR, and what is it?

17. The same material and structure from Example 3.5 has been used to produce 500 μm long standard DFB lasers. Calculate the threshold modal gain for:

(a) The standard laser whose one facet has been AR coated, and the other facet is a cleaved facet with 90° out-of-phase additive reflection.

(b) The standard laser whose one facet has been AR coated, and the other facet has been HR coated, with 90° out of phase additive reflection.

18. Standard DFBs with internal modal losses of 10 cm^{-1} and various κL_g's are fabricated with one end AR coated and one end cleaved. Assume that the cleave provides a reflection in quadrature with that of the grating and that the output is from the AR coated end. Use $L_g = 500$ μm and other parameters from Problem 16.

(a) What κL_g gives the best MSR, and what is η_{d1} in this case?

(b) What κL_g gives the worst MSR, and what is η_{d1} in this case?

19. Two different types of DFB lasers are fabricated from the same 1.55-μm wavelength wafer that contains uniform grating and 2-μm-wide buried heterostructure lateral waveguides. The first are standard AR-coated DFB lasers, with 500-μm cavity length. The threshold current on these is measured to be 30 mA, and the differential efficiency, considering both ends, is found to be 0.5. As expected, two degenerate axial modes are observed.

The second type of laser is formed by cleaving out 250-μm-long devices from the same wafer, and then applying AR coatings to only one end. In this case. the spectra of some devices show marginal MSRs, while others show relatively good MSRs. The devices with the best MSRs are measured, and these provide threshold currents of 30 mA and differential efficiencies, considering both ends, of 0.6. From other measurements, we have determined that the threshold modal gain of these shorter lasers is twice that of the longer lasers.

(a) What is the injection efficiency for these lasers?

(b) What is the average internal loss?

20. For DFB lasers from the previous example,

(a) What is the grating coupling coefficient κ?

(b) Assuming a gain characteristic of the form $\Gamma g = \Gamma g_0 \ln(J/J_{tr})$, what are the values of Γg_0 and J_{tr}?

(c) What MSR would be expected in the better 250-μm devices, if operated at twice the threhold, and assuming $\beta_{sp}\eta_r = 10^{-5}$?

21. For a 1/8 wavelength shifted DFB laser, with parameters listed in the caption of Fig. 3.27, what is the threshold modal gain and the differential efficiency? How do these compare to the values obtained for a quarter-wave-shifted version of this laser?

CHAPTER FOUR

Gain and Current Relations

4.1 INTRODUCTION

In Chapter 1, various transitions responsible for the generation and recombination of carriers within the semiconductor were introduced. In Chapter 2, the rates at which these transitions occur were shown to provide the fundamental description of LED and laser operation through the development of the rate equations. The optical gain, for example, was defined in terms of the difference between the stimulated emission and absorption rates. Radiative efficiency was defined in terms of the spontaneous and nonradiative recombination rates. Simple relationships between these rates and the carrier density were assumed in Chapter 2 to provide a feel for how semiconductor lasers generally behave. In the present chapter, we would like to delve a little deeper into the fundamentals of these transitions.

We will first develop a quantitative description of radiative transitions, from which we will be able to determine both the optical gain and the corresponding radiative current density as a function of injection level. Then, we will consider nonradiative transitions and see how they compare to the radiative transition rates in different material systems. Finally we will provide a set of example gain calculations for common materials to quantify the various relationships between the gain, carrier density, and current density.

In Appendix 6, an alternative description of radiative transitions traditionally applied to discrete energy level lasers is adapted for use in semiconductors. The reader is encouraged to examine this appendix, for not only does the analysis bridge the gap between Einstein's approach and the treatment provided here, but it is also

Diode Lasers and Photonic Integrated Circuits, Second Edition.
Larry A. Coldren, Scott W. Corzine, and Milan L. Mašanović.
© 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

hoped that by covering the same material from a different perspective, the reader will gain a deeper understanding of radiative processes in semiconductors.

More in-depth discussions of many of the relations used in this chapter can be found in Appendices 8 through 11. These discussions are presented at a higher level and are not required for the basic understanding of material presented in this chapter. In brief, the envelope function approximation and the calculation of the valence subband structure in quantum wells with and without strain can be found in Appendix 8. Fermi's Golden Rule, a key relation for estimating gain in semiconductors, is derived from first principles in Appendix 9. The resulting transition matrix element and polarization-dependent effects related to it are considered in Appendix 10. Finally, the effects of strain on the bandgap of semiconductors are discussed in Appendix 11. The latter part of Appendix 8 and Appendix 11 are recommended reading for anyone particularly interested in strained materials.

4.2 RADIATIVE TRANSITIONS

4.2.1 Basic Definitions and Fundamental Relationships

In Chapter 1, different types of transitions were discussed in reference to Fig. 1.6. Here we would like to concentrate on the radiative transitions. Specifically, there are three types of radiative transitions between the conduction and valence bands that are important in modern semiconductor lasers. These *band-to-band* radiative transitions are sketched in Fig. 4.1.

In the first diagram, the energy of the photon is transferred to an electron, elevating it from some state 1 in the valence band to some state 2 in the conduction band. Such *stimulated absorption* events generate new carriers and are also responsible for the disappearance of photons. In the second diagram, the incoming photon stimulates the electron to liberate energy in the form of a new photon, lowering it from state 2 in the conduction band to state 1 in the valence band. Such *stimulated emission* events provide a recombination path for carriers and are more importantly the source of new photons. The third diagram is really no different from the second diagram except that the field that stimulates the electron to emit a photon and make a downward transition is not a real field, but a *vacuum-field* (as it is commonly referred to in the quantum world). Because vacuum-field-induced transitions can occur with no classical field stimulation, we refer to them as *spontaneous emission*

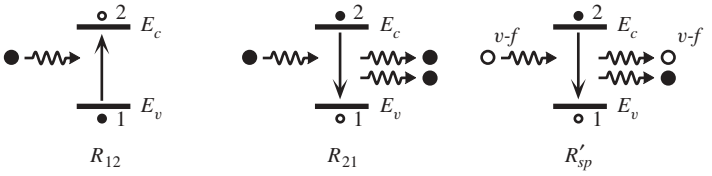


FIGURE 4.1: Band-to-band radiative transitions: stimulated absorption, stimulated emission, and spontaneous emission. (All rates are defined per unit volume.)

events. In the absence of classical fields, spontaneous emission serves as one of the dominant recombination paths of carriers in direct bandgap semiconductors, and is by far the most common source of photons provided by nature.

The rates at which the three radiative processes in Fig. 4.1 occur depend on a number of factors. Two primary factors are the density of photons and the density of *available* state pairs. As we will find later, the dependence on the photon density enters through the local electric field strength, $|\mathscr{E}|^2$. Thus, R_{12}, $R_{21} \propto |\mathscr{E}|^2$ and $R'_{sp} \propto |\mathscr{E}^{v-f}|^2$, where $|\mathscr{E}^{v-f}|^2$ is the vacuum-field strength. We will have more to say about vacuum-fields and how to evaluate $|\mathscr{E}^{v-f}|^2$ in Section 4.4.

The dependence of the transition rates on the density of available state pairs can be broken down into two components: one that is strictly material dependent, and the other that depends on the injection levels. The first component is the density of *total* state pairs, which is found by taking the appropriate average between the density of states in the conduction and valence bands. This *reduced* density of states function, ρ_r, is introduced in Appendix 6 for the case of parabolic bands, and it will be derived in general a little later in this chapter. The second component is the *fraction* of state pairs available to participate in the transition, as also introduced in Appendix 6. For upward transitions, this fraction is maximized when all carriers are placed in the valence band. For downward transitions, this fraction is maximized when all carriers are placed in the conduction band. The former population of carriers occurs naturally, whereas the latter *inverted* population can only be achieved by providing energy, which pumps the carriers into the conduction band (for example, by current injection into the center of a *pn* junction). The fraction of available state pairs will be quantified later.

Now let's consider the electromagnetic field a little more carefully. First of all, it is important to appreciate that downward transitions not only create a new photon, but they also a new photon *into the same optical mode as the stimulating photon* (whether it is a real or vacuum-field photon). For this reason, it is important to distinguish photons in one optical mode from photons in another (see Appendices 3 and 4 for a discussion of optical modes). By associating $|\mathscr{E}|^2$ with the field strength of one optical mode, we can interpret the transition rates in Fig. 4.1 as *single-mode* transition rates (in fact, the prime on R'_{sp} is used to distinguish this single-mode spontaneous emission rate from the total band-to-band spontaneous emission rate, R_{sp}). The total transition rates are then found by summing over all optical modes.

Another interesting feature of downward transitions is that in addition to appearing in the same optical mode, the newly created photon also contributes to the existing field constructively due to the harmonic nature of the perturbation. This feature allows the optical mode to build up a very coherent field. Unfortunately, the vacuum-field phase is not correlated with the phase of the real fields in the optical mode. As a result, new photons introduced into the mode through spontaneous emission have random phases relative to the coherent fields created through stimulated emission. And although the number of photons introduced into the mode through spontaneous emission can be made small relative to the photons introduced through

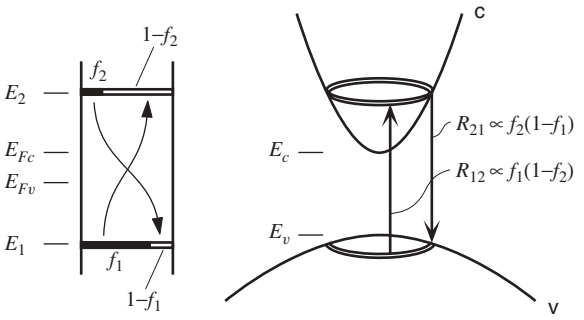


FIGURE 4.2: State pairs which interact with photons at E_{21}. Energy and momentum conservation reduce the set of state pairs to the annulus shown in the plot of energy versus momentum in two dimensions. The occupation probabilities, f_1 and f_2, reduce this set even further.

stimulated emission, they can never be removed completely, implying that perfect coherence in a laser can never be achieved. Chapter 5 considers the implications of spontaneous emission as a phase noise source in more detail.

So far we have not specified the electron states 1 and 2 in any detail. As will be shown next, photons with energy $h\nu$ induce upward and downward transitions only between those electron state pairs that conserve both energy and momentum in the course of the transition. In other words, we must have $E_2 - E_1 = E_{21} = h\nu$ and $\mathbf{k}_2 = \mathbf{k}_1$. These conservation laws reduce the interaction to a very particular region of the $E - k$ diagram of the semiconductor, as illustrated in Fig. 4.2 for two dimensions of k-space. Furthermore, within this region only vertical transitions are allowed.

Now electrons typically only spend about 0.1 ps in any given state due to collisions with phonons and other electrons. As a result, their energy is uncertain, making the annulus shown in Fig. 4.2 appear fuzzy. To properly account for this, the total transition rates should include an integration over the energy uncertainty. In Appendix 6, this integration is included right from the start. However, because the integral tends to clutter the math, we will defer this procedure to the very end, where a more thorough discussion will be included.

Another restriction we must consider is that transitions occur only between filled initial states and empty final states, as outlined in Appendix 6. Figure 4.2 illustrates the fraction of state pairs that satisfy this criterion for both upward and downward transitions. Writing out the Fermi factors explicitly, the three radiative transition rates become

$$\begin{aligned} R_{12} &= R_r \cdot f_1(1 - f_2), \\ R_{21} &= R_r \cdot f_2(1 - f_1), \\ R'_{sp} &= R_r^{v-f} \cdot f_2(1 - f_1). \end{aligned} \tag{4.1}$$

In these equations, R_r represents the radiative transition rate that would exist if all state pairs were available to participate in the transition. For the spontaneous emission rate, we must use $R_r^{v-f} = R_r$ with $|\mathscr{E}|^2 \rightarrow |\mathscr{E}^{v-f}|^2$. We will derive an explicit expression for R_r later, but from the preceding discussions we already know that R_r is proportional to the field strength and the reduced density of states function.

Because R_{21} and R_{12} are competing effects in that one generates new photons and the other takes them away, we would also like to know the *net* generation rate of photons in the semiconductor, or

$$R_{st} \equiv R_{21} - R_{12} = R_r \cdot (f_2 - f_1). \tag{4.2}$$

We will show in Section 4.3 as we have by a more phenomenological route in Chapter 2, that the *net stimulated emission rate*, R_{st}, is directly proportional to the optical gain in the material.

The occupation probabilities in Eqs. (4.1) and (4.2) can usually be described using Fermi statistics even under nonequilibrium conditions by using a separate Fermi level for the conduction and valence bands:

$$f_1 = \frac{1}{e^{(E_1 - E_{Fv})/kT} + 1} \quad \text{and} \quad f_2 = \frac{1}{e^{(E_2 - E_{Fc})/kT} + 1}, \tag{4.3}$$

where E_{Fc} and E_{Fv} are the conduction and valence band *quasi-Fermi levels*. Under nonequilibrium forward bias conditions, E_{Fc} and E_{Fv} are separated by slightly less than the applied voltage to the junction.

Simple relations between the transition rates are easily derived by substituting Eq. (4.3) for the occupation probabilities:

$$\frac{R_{21}}{R_{12}} = \frac{f_2(1-f_1)}{f_1(1-f_2)} = e^{(\Delta E_F - E_{21})/kT}, \tag{4.4}$$

$$\frac{R'_{sp}}{R_{st}} = \frac{|\mathscr{E}^{v-f}|^2}{|\mathscr{E}|^2}\frac{f_2(1-f_1)}{f_2 - f_1} = \frac{|\mathscr{E}^{v-f}|^2}{|\mathscr{E}|^2}\frac{1}{1 - e^{(E_{21} - \Delta E_F)/kT}}. \tag{4.5}$$

The first ratio Eq. (4.4) reveals that the stimulated emission rate will be larger than the absorption rate only when

$$E_{Fc} - E_{Fv} \equiv \Delta E_F > E_{21}. \tag{4.6}$$

Stated another way, the *net* stimulated emission rate (and hence the optical gain) will become positive only when the quasi-Fermi level separation is greater than the photon energy of interest. And because the photon energy must at the very least be equal to the bandgap energy, we conclude that to achieve gain in the semiconductor, we must have

$$\Delta E_F > E_g. \tag{4.7}$$

This condition demands that the voltage across a *pn* junction must be greater than the bandgap to achieve gain in the active region.

The second ratio Eq. (4.5) reveals a fundamental relationship between the single-mode spontaneous emission rate and the net stimulated emission rate. This relation will be developed further in Section 4.4. We will find there that the ratio of field strengths is just equal to the reciprocal of the number of photons in the optical mode.

4.2.2 Fundamental Description of the Radiative Transition Rate

To fully quantify all three radiative transition rates, we need only evaluate the one transition rate, R_r, appearing in Eq. (4.1). The treatment in Appendix 6 also concludes that Einstein's stimulated rate constant, B_{21}, is all that is necessary to determine the three radiative transition rates. However, Einstein's approach does not provide a means of determining B_{21} in semiconductors. Fortunately, the transition rate, R_r can be estimated using a relation known as Fermi's Golden Rule, derived in Appendix 9. To evaluate Fermi's Golden Rule, we need to provide an accurate description of the interaction that occurs between the electron in the crystal and the electromagnetic field.

To describe the electron fully we must provide a model for the electron's wavefunction in both states 1 and 2. To be rigorous, ψ_1 and ψ_2 must be found by solving the Schrödinger equation with the appropriate crystal potential. Such an exact solution, however, would be difficult to find and inconvenient to work with. Fortunately, a useful approximation can be made that decomposes the crystal potential into (1) a complex atomic-scale potential that is periodic with the crystal lattice, and (2) a macroscopic potential that follows the spatial dependence of the conduction or valence band edge (created by either doping or material composition variations) as illustrated in Fig. 4.3. Appendix 8 shows that the corresponding electron wavefunctions can then be written as the product of two functions:

$$\psi_1 = F_1(\mathbf{r}) \cdot u_v(\mathbf{r}) \qquad \text{and} \qquad \psi_2 = F_2(\mathbf{r}) \cdot u_c(\mathbf{r}). \tag{4.8}$$

The *envelope* function, $F(\mathbf{r})$, is a slowly varying function satisfying Schrödinger's equation using the macroscopic potential and an appropriate *effective* mass. The *Bloch* function, $u(\mathbf{r})$, is a complex periodic function that satisfies Schrödinger's equation using the atomic-scale potential. Each energy band in the crystal has its own Bloch function. Fortunately, one never really needs to determine $u(\mathbf{r})$ precisely. Only the symmetry properties of these functions are necessary for most calculations, as discussed in Appendix 8. Thus we can concentrate our attention on the simpler envelope function.

In uniformly doped bulk material, the flat energy bands imply a constant macroscopic potential that leads to simple plane wave solutions for the envelope functions. In a quantum-well potential, plane wave solutions exist along directions within the plane of the well. However, along the confinement direction, $F(\mathbf{r})$ takes on either a cosine or sine wave distribution inside the well and decays exponentially outside

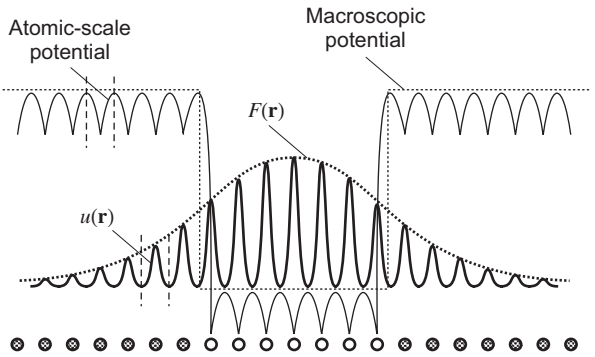


FIGURE 4.3: Illustration of a quantum-well potential and the corresponding lowest energy electron wavefunction.

the well. The bulk, quantum-well, and (by extension) quantum-wire and quantum dot envelope functions therefore take the following form:

$$F(\mathbf{r}) = e^{-j\mathbf{k}\cdot\mathbf{r}}/\sqrt{V}, \qquad \text{(bulk)} \tag{4.9}$$

$$F(\mathbf{r}) = F(z)\cdot e^{-j\mathbf{k}\cdot\mathbf{r}_{||}}/\sqrt{A}, \quad \text{(quantum well)} \tag{4.10}$$

$$F(\mathbf{r}) = F(x,y)\cdot e^{-jk_z z}/\sqrt{L}, \quad \text{(quantum wire)} \tag{4.11}$$

$$F(\mathbf{r}) = F(x,y,z), \qquad \text{(quantum dot)} \tag{4.12}$$

where V (A, L) is the volume (area, length) of the crystal, and appears for normalization purposes. For the quantum well, the position vector $\mathbf{r}_{||}$ is parallel to the quantum-well plane, and $F(z)$ is the simple one-dimensional solution to the quantum-well potential considered in Appendix 1. For the quantum wire, the length of the wire runs along z, and $F(x,y)$ is the two-dimensional solution within the quantization plane. For the quantum dot, $F(x,y,z)$ is the 3-D wavefunction of the box, and we assume the states of individual boxes do not interact significantly. The quantum-well electron wavefunction illustrating both envelope and Bloch function components is superimposed over the crystal potential in Fig. 4.3. Because we often need only the envelope function of the electron for many calculations, it is common to associate $F(r)$ with the complete wavefunction of the electron, ψ. This association is oftentimes harmless; however, in the present case it is necessary to emphasize that it represents only the slowly varying envelope of the complete electron wavefunction.

With the electron wavefunctions defined, we can move on to describing the interaction between the electron and the electromagnetic wave. The wave's interaction with the electron enters into Schrödinger's equation through the vector potential:

$$\mathbf{A}(\mathbf{r},t) = \hat{\mathbf{e}}\mathrm{Re}\{\mathscr{A}(\mathbf{r})e^{j\omega t}\} = \hat{\mathbf{e}}\tfrac{1}{2}[\mathscr{A}(\mathbf{r})e^{j\omega t} + \mathscr{A}^*(\mathbf{r})e^{-j\omega t}], \tag{4.13}$$

where $\hat{\mathbf{e}}$ is the unit polarization vector in the direction of $\mathbf{A}$, and $\omega(\hbar\omega)$ is the angular frequency (energy) of the photon. The vector potential is related to the electric field via $\mathscr{E} = -\partial\mathbf{A}/\partial t$. The kinetic energy term of Schrödinger's equation describing the electron in the crystal (given by Eq. (A8.1)) is now modified by the substitution

$$\mathbf{p}^2 \rightarrow (\mathbf{p} + q\mathbf{A})^2 \approx \mathbf{p}^2 + 2q\mathbf{A}\cdot\mathbf{p}, \tag{4.14}$$

where q is the *magnitude* of the electron charge. This modification accounts for the electromagnetic field's ability to accelerate and/or decelerate charged particles (and hence modify the electron's kinetic energy). In expanding the square, we can neglect the squared vector potential term because it does not affect our final results (orthogonality of the wavefunctions ensures us that the operator $\mathbf{A}^2$ does not perturb the system, assuming we can neglect the spatial variation of $\mathbf{A}$ within one unit cell of the crystal). Substituting Eq. (4.13) into (A8.1), we can write the new Hamiltonian as

$$H = H_0 + [H'(\mathbf{r})e^{j\omega t} + h.c.], \quad H'(\mathbf{r}) = \frac{q}{2m_0}\mathscr{A}(\mathbf{r})\hat{\mathbf{e}}\cdot\mathbf{p}. \tag{4.15}$$

The *h.c.* stands for Hermitian conjugate and simply means that we take the complex conjugate of all terms except the Hermitian momentum operator $\mathbf{p}$. The term in brackets can be viewed as a time-dependent perturbation to the original Hamiltonian, H_0. This perturbation term is the driving force for transitions between the conduction and valence bands.

By studying the time evolution of some electron wavefunction initially in a valence band state, for example, as it makes an upward transition to the conduction band in the presence of the time-harmonic perturbation, it is possible to determine the rate at which such transitions will occur. This procedure is carried out in Appendix 9. The resulting transition rate per unit volume of active material is given (in units of s^{-1} cm^{-3}) by

$$R_r = \frac{2\pi}{\hbar}|H'_{21}|^2\rho_f(E_{21})|_{E_{21}=\hbar\omega}, \tag{4.16}$$

$$H'_{21} \equiv \langle\psi_2|H'(\mathbf{r})|\psi_1\rangle = \int_V \psi_2^* H'(\mathbf{r})\psi_1 d^3\mathbf{r}. \tag{4.17}$$

Equation (4.16) is known as Fermi's Golden Rule. It reveals that the number of transitions per unit active volume; V, occurring per second, is dependent on (1) the density of final states, $\rho_f(E_{21})$, (in units of energy^{-1} cm^{-3}) available to the electron, and (2) the spatial overlap of the initial and final electron wavefunctions with the harmonic perturbation defined in Eq. (4.14) – the integral being defined as H'_{21} (in units of energy). It is important to appreciate that the field will only invoke a response from electrons that exist in states that have possible final states separated by $E_{21} \approx \hbar\omega$. That is, the two electron states must be in resonance with the oscillating field. As a result, $\rho_f(E_{21})$ and H'_{21} in Eq. (4.15) must be evaluated at

$E_{21} = \hbar\omega$. This resonance condition derived in Appendix 9 is a statement of energy conservation.

Using Eq. (4.15), the job of determining R_r is reduced to providing explicit relations for both the density of final states and the overlap integral (or *matrix element* as it is commonly called). The next two subsections tackle this chore.

4.2.3 Transition Matrix Element

The matrix element $|H'_{21}|^2$ determines the *strength of interaction* between two states. This interaction can be strong, negligible, or identically zero, all depending on the wavefunctions describing the two electron states. For example, in a quantum well only transitions between subbands with the same quantum number are *allowed*; all others are *forbidden*. The wavefunction overlap also leads to the **k**-selection rule, which dictates that transitions between plane wave states are forbidden unless the k-vectors of the two states are equal (the two electron states must propagate along the same direction). In addition to these considerations, the interaction strength can also depend on the polarization of the incident light, if the material has some preferential axis of symmetry. For example, the interaction strength between conduction and heavy-hole states in a quantum well is much stronger for electric fields in the plane of the well than perpendicular to the well.

To derive an expression for $|H'_{21}|^2$, we insert the definition of $H'(\mathbf{r})$ in Eq. (4.15) into the definition of $|H'_{21}|^2$ given by Eq. (4.17). We can reduce Eq. (4.17) by expressing the electron and hole wavefunctions in terms of the envelope/Bloch function formalism using Eq. (4.8). Because the momentum operator when operating on a product can be written as $\mathbf{p}AB = B\mathbf{p}A + A\mathbf{p}B$, the overlap integral can be expressed as the sum of two terms

$$\begin{aligned} H'_{21} &= \frac{q}{2m_0}\int_v F_2^* u_c^* (\mathscr{A}(\mathbf{r})\hat{\mathbf{e}}\cdot\mathbf{p}) F_1 u_v d^3\mathbf{r} \\ &= \frac{q}{2m_0}\left[\int_v u_c^* u_v F_2^* (\mathscr{A}(\mathbf{r})\hat{\mathbf{e}}\cdot\mathbf{p}) F_1 d^3\mathbf{r} + \int_v [F_2^* \mathscr{A}(\mathbf{r}) F_1] u_c^* \hat{\mathbf{e}}\cdot\mathbf{p} u_v d^3\mathbf{r}\right]. \end{aligned} \quad (4.18)$$

In transitions from the conduction band to the valence band, the first integral within the brackets vanishes[1] due to the orthogonality condition expressed in Eq. (A16.5) and due to the fact that the other terms in the integrand are, to a good approximation, constant in any one unit cell. To evaluate the second integral, we break up the integration over the crystal volume into a sum of integrations over each unit cell. The terms collected in brackets in the second integral can again be taken as constant over the dimensions of a unit cell, and we can write

$$H'_{21} = \frac{q}{2m_0}\sum_j [F_2^* \mathscr{A}(\mathbf{r}) F_1]_{\mathbf{r}=\mathbf{r}_j} \int_{\text{unit cell}} u_c^* \hat{\mathbf{e}}\cdot\mathrm{p} u_v d^3\mathbf{r}, \quad (4.19)$$

[1]For transitions within the *same* energy band, the Bloch function overlap is equal to unity and the first integral may or may not be zero, depending on the envelope function overlap.

where j sums over all units cells in the crystal, and $\mathbf{r}_j$ is a position vector to the jth cell. Because the Bloch functions, u, repeat themselves in each unit cell, the integral can be pulled out of the summation to obtain

$$H'_{21} = \frac{q}{2m_0}\left[\frac{1}{V_{uc}}\int_{\text{unit cell}} u_c^*\hat{\mathbf{e}}\cdot\mathbf{p}u_v d^3\mathbf{r}\right]\sum_j [F_2^*\mathscr{A}(\mathbf{r})F_1]_{\mathbf{r}=\mathbf{r}_j}V_{uc}$$

$$= \frac{q}{2m_0}\langle u_c|\hat{\mathbf{e}}\cdot\mathbf{p}|u_v\rangle\int_V F_2^*\mathscr{A}(\mathbf{r})F_1 d^3\mathbf{r}, \tag{4.20}$$

where, by assuming the volume of a unit cell to be very small, we have converted the summation back into an integral. We have also used Dirac notation to express the Bloch function overlap integral.

The envelope function overlap integral in Eq. (4.20) can be further simplified by recognizing that the spatial variation of $\mathscr{A}(\mathbf{r})$ is typically much slower than that of the envelope functions allowing us to pull it out of the integration. Assuming $\mathscr{A}(\mathbf{r})$ to be a plane wave of the form $\mathscr{A}_0 e^{-j\mathbf{k}\cdot\mathbf{r}}$, and ignoring the spatial dependence (i.e., the exponential term), we obtain

$$\int_V F_2^*\mathscr{A}(\mathbf{r})F_1 d^3\mathbf{r} \approx \mathscr{A}_0\int_V F_2^*F_1 d^3\mathbf{r} \equiv \mathscr{A}_0\langle F_2|F_1\rangle. \tag{4.21}$$

We will consider this integral in more detail a little later in Eqs. (4.23) through (4.25). Substituting Eq. (4.21) into Eq. (4.20), we finally obtain

$$|H'_{21}|^2 = \left(\frac{q\mathscr{A}_0}{2m_0}\right)^2|M_T|^2, \quad \text{where} \quad |M_T|^2 \equiv |\langle u_c|\hat{\mathbf{e}}\cdot\mathbf{p}|u_v\rangle|^2|\langle F_2|F_1\rangle|^2. \tag{4.22}$$

The prefactor in the first equality comes directly from the perturbation Hamiltonian Eq. (4.15). The second term, $|M_T|^2$, is referred to as the *transition matrix element* and is given special attention in Appendix 10. The first component, $|\langle u_c|\hat{\mathbf{e}}\cdot\mathbf{p}|u_v\rangle|^2$, contains the polarization dependence of the interaction, which will depend on the particular symmetries of the conduction and valence band Bloch functions. Aside from the polarization dependence (which can be a function of photon energy), we can consider this *momentum matrix element* to be a constant, $|M|^2$, for a given material.

As shown in Appendix 8, the constant $|M|^2$ can be determined experimentally. Table 4.1 tabulates the most accurately reported values for several important materials systems. (Note that $2|M|^2/m_0$ has units of energy.) Appendix 10 shows how $|M_T|^2$ can be expressed in terms of $|M|^2$. This involves expanding the dot product as well as considering the overlap of the envelope functions in $|M_T|^2$ given by Eq. (4.22). Table 4.2 summarizes the results for bulk and quantum-well materials for either transverse electric (TE: electric field in the quantum-well plane) or transverse magnetic (TM: electric field perpendicular to quantum-well plane) polarizations. Due to band mixing effects, the values for quantum wells shown are only valid for small transverse k-vectors, k_t. A full dependence of the $|M_T|^2/|M|^2$ as k_t

TABLE 4.1: Magnitude of $|M|^2$ for Various Material Systems

Material system	$\frac{2\|M\|^2}{m_0}$ (in eV)	Reference
GaAs	28.8 ± 0.15	[1,2]
$Al_xGa_{1-x}As(x < 0.3)$	$29.83 \pm 2.85x$	[3]
$In_xGa_{1-x}As$	$28.8 - 6.6x$	[1,2]
InP	19.7 ± 0.6	[1,2]
$In_{1-x}Ga_xAs_yP_{1-y}(x = 0.47y)$	$19.7 + 5.6y$	[2,4]
GaN	14.0	[13]
InN	14.6	[13]
AlN	14.5	[13]
$In_{0.24}Ga_{0.76}N$	0.274	[14]
$In_{0.15}Ga_{0.85}N$	0.823	[14]

TABLE 4.2: Magnitude of $|M_T|^2/|M|^2$ for Different Transitions and Polarizations

	Bulk		Quantum-well ($k_t \sim 0$)	
Polarization	C–HH	C–LH	C–HH	C–LH
TE	1/3	1/3	1/2	1/6
TM	1/3	1/3	0	2/3

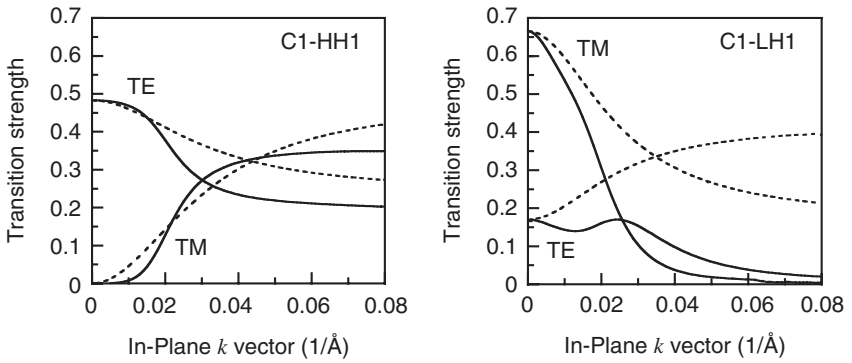


FIGURE 4.4: Relative transition strengths for both TE and TM light polarization for the two lowest subband transitions in an unstrained $GaAs/Al_{0.2}Ga_{0.8}As$ 80 Å QW. The dashed curves represent what one would calculate assuming parabolic subbands. The transition strength as plotted here is defined as $|M_T|^2/|M|^2$ (bulk value is 1/3).

is varied is given in Figs. 4.4 and 4.5, for a GaAs/AlGaAs and InGaAs/InGaAsP quantum well, respectively. Figure 4.6 shows the dependence of $|M_T|^2|$ as k_t is varied for InGaN and AlGaN quantum wells.

In addition to creating a polarization sensitivity, the transition matrix element also restricts the types of states that can interact. For transitions *between two*

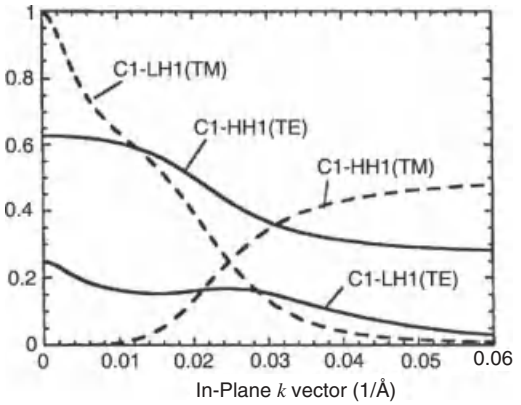


FIGURE 4.5: Relative transition strengths for both TE and TM light polarization for the transitions in an unstrained $In_{0.53}Ga_{0.47}As/InGaAsP$ 60 Å QW. The dashed curves represent the TM transitions and the solid curves TE. The transition strength as plotted here is defined as $|M_T|^2/|M|^2$ (bulk value is 1/3). ((© John Wiley and Sons, [15].)

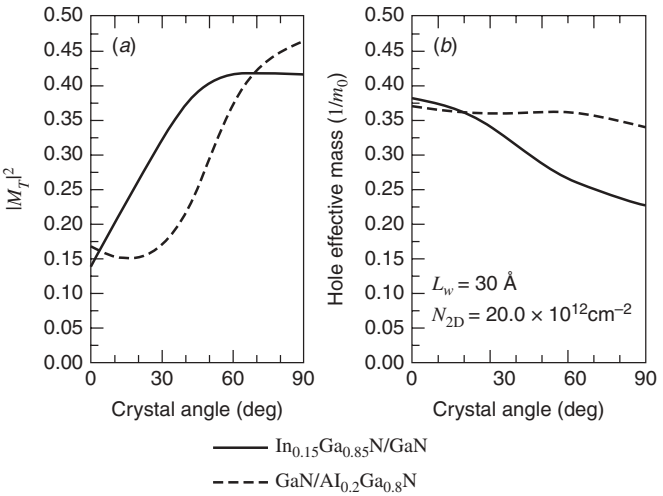


FIGURE 4.6: (a) Normalized y-polarized optical matrix elements and (b) average hole effective masses as functions of crystal orientation of WZ $In_{0.15}Ga_{0.85}N/GaN$ QWs (30 Å). For comparison, results (dashed line) of $GaN/Al_{0.2}Ga_{0.8}N$ QW structure are also plotted. *Reprinted by permission from Jpn. J. Appl. Phys.* [16].

plane wave states in a "bulk" active medium (i.e., $V \to \infty$), we can use Eq. (4.9) to set

$$\langle F_2|F_1\rangle = \frac{1}{V}\int_V e^{j\mathbf{k}_2\cdot\mathbf{r}}e^{-j\mathbf{k}_1\cdot\mathbf{r}}d^3\mathbf{r} = \delta_{\mathbf{k}_1,\mathbf{k}_2}, \tag{4.23}$$

where the Kronecker delta, $\delta_{\mathbf{k}_1,\mathbf{k}_2}$, is zero unless $\mathbf{k}_2 = \mathbf{k}_1$, in which case it equals one. This spatial *phase-matching* condition, known as the **k**-selection rule, is a statement of momentum conservation. Thus, only states propagating along the same direction in the crystal can interact. If the spatial dependence of the field were not ignored in deriving Eq. (4.22), then an additional plane wave term, $e^{-j\mathbf{k}\cdot\mathbf{r}}$, would appear in Eq. (4.23), and the **k**-selection rule for upward transitions would become: $\mathbf{k}_2 = \mathbf{k}_1 + \boldsymbol{\kappa}$. However, the wavevector of the field, $\boldsymbol{\kappa}$, is typically orders of magnitude smaller than the electron wavevector and can usually be ignored (justifying our earlier simplifying step in Eq. (4.21)).

In a quantum well, the envelope functions are given by Eq. (4.10). Transitions between two such quantum-well states are governed by the following overlap:

$$\begin{aligned}\langle F_2|F_1\rangle &= \frac{1}{A}\int_V F_2^*(z)e^{j\mathbf{k}_2\cdot\mathbf{r}_\parallel}F_1(z)e^{-j\mathbf{k}_1\cdot\mathbf{r}_\parallel}d^3\mathbf{r} \\ &= \int_z F_2^*(z)F_1(z)\,dz. \quad (\text{with } \mathbf{k}_2 = \mathbf{k}_1)\end{aligned} \tag{4.24}$$

Thus, we can assume **k**-selection in the plane of the quantum well, but we still need to evaluate $|\langle F_2|F_1\rangle|^2$ perpendicular to the plane, where again F_2 and F_1 are simply the particle-in-a-box envelope functions found for the quantized energy levels of the quantum well in both conduction and valence bands.

Due to orthogonality between the quantum-well wavefunction solutions, the overlap integral in Eq. (4.24) reduces to the following rule for subband transitions:

$$|\langle F_2|F_1\rangle|^2 \approx \delta_{n_c,n_v}. \tag{4.25}$$

This means that transitions can only occur between quantum-well subbands that have the same quantum number, $n_c = n_v$. These are referred to as *allowed* transitions. Transitions between subbands with dissimilar quantum numbers are *forbidden* transitions. Both are illustrated in Fig. 4.7. The allowed transitions are usually referred to as the $n = 1$ transition, the $n = 2$ transition, etc. The "nearly equal to" sign is used in Eq. (4.25) because the different effective mass and barrier height in the conduction and valence bands means that the wavefunctions of the two bands are not completely orthogonal to each other. Nevertheless, allowed transition overlaps are usually (but not always) close to unity (0.9–1), and forbidden transition overlaps are usually very small (0 – 0.1).

These considerations can be extended to quantum wires, which have potential barriers in two dimensions. For example, in a quantum wire, F_2 and F_1 are functions of both quantized directions, such that $|\langle F_2|F_1\rangle|^2$ represents an integration over the "quantization plane" (see Eq. (4.12)). In this case, **k**-selection is obeyed only along the length of the wire.

In heavily doped materials, electron states that are bound to charged donors or acceptors can exist. If the concentrations are high enough, it is conceivable that many transitions near the band edge will be either *band-to-bound* or *bound-to-bound* transitions. In these cases, the **k**-selection rule cannot be assumed

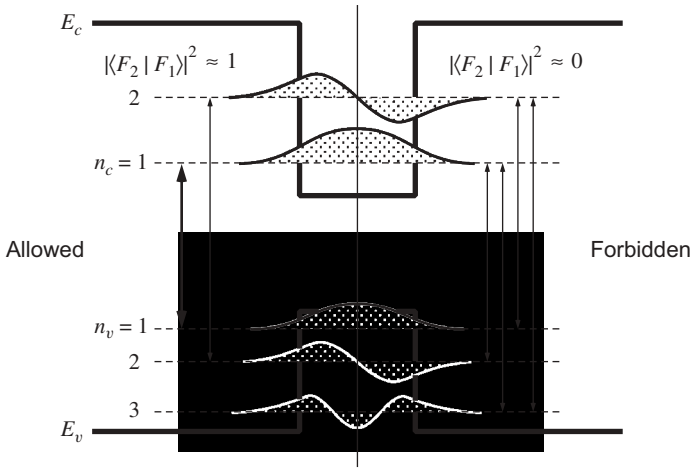


FIGURE 4.7: Allowed and forbidden transitions in a quantum well. The most important "$n = 1$" transition is highlighted in bold.

in any direction and $|\langle F_2|F_1\rangle|^2$ must be evaluated explicitly for the envelope functions that correspond to these bound states. In early treatments [5], such band-to-bound transitions were given considerable attention. However, these analyses were provided when most active regions were heavily doped. More recently, heavily doped active regions have faded in popularity in favor of undoped quantum-well active regions (or sometimes *modulation doped* quantum wells, where the doping ions are physically separated from the quantum wells). As a result, most current semiconductor lasers operate on the physics of band-to-band transitions. The rest of this chapter will concentrate on transitions occurring between plane wave states such that the **k**-selection rule can be assumed.

With $|H'_{21}|^2$ given by Eq. (4.22) and the **k**-selection rule established, we now need to define the final density of states more carefully.

4.2.4 Reduced Density of States

The derivation of Fermi's Golden Rule in Appendix 9 assumes the electron initially occupies a single state, which makes a transition to one of a large number of final states. In a semiconductor, both final *and* initial states of the electron are immersed within a large number of nearby states, as illustrated in Fig. 4.8. For this case, the final density of states appearing in Eq. (4.16) should actually be interpreted as the *density of transition pairs per unit transition energy*, δE_{21}. This density of transition pairs is referred to as the *reduced* density of states function, $\rho_r(E_{21})$. One specific form for $\rho_r(E_{21})$ is derived in Appendix 6; however, here we would like to determine the general form.

If we assume the **k**-selection rule applies, then only states with identical k-vectors can form a transition pair, and only vertical transitions in k-space can

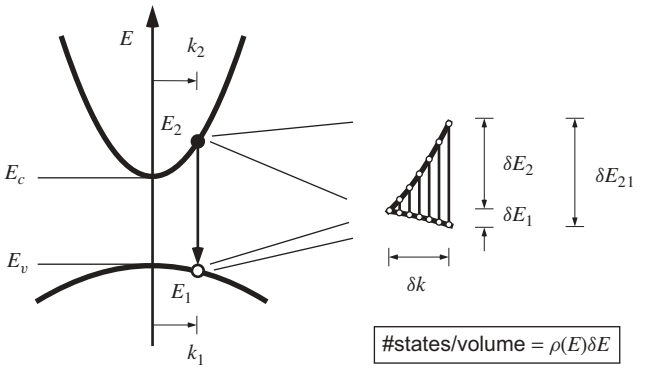


FIGURE 4.8: Relationship between the energy ranges in the conduction and valence bands for a given dk in k-space, assuming **k**-selection applies.

occur. Because of this restriction, the number of transition pairs within δk is equal to the number of states in either the conduction or valence band, and $\rho_r \delta E_{21} = \rho_c \delta E_2 = \rho_v \delta E_1$. This allows us to set $\delta E_2 = (\rho_r/\rho_c)\delta E_{21}$ and $\delta E_1 = (\rho_r/\rho_v)\delta E_{21}$. Summing these relations and setting $\delta E_{21} = \delta E_2 + \delta E_1$, we immediately obtain

$$\frac{1}{\rho_r} = \frac{1}{\rho_c} + \frac{1}{\rho_v}. \tag{4.26}$$

This is the more general form for the final density of states to be used in Fermi's Golden Rule. Note that as $\rho_v \to \infty$, we have $\rho_r \to \rho_c$. This is the case we solved in Appendix 9, with ρ_c interpreted as the final density of states. For a finite density of states in the valence band, such as $\rho_v = \rho_c$, ρ_r is reduced from ρ_c to $\rho_c/2$ (hence the name *reduced* density of states). In typical semiconductors, $\rho_v \geq 5\rho_c$, or $(5/6)\rho_c \leq \rho_r \leq \rho_c$. Thus, ρ_r is generally very close to ρ_c. However, in strained materials, ρ_v can be reduced significantly, as shown in Appendix 8, bringing ρ_r closer to $\rho_c/2$.

For general use, it turns out that Eq. (4.26) is not that practical. An alternate definition of $\rho_r(E_{21})$ can be found by relating it to the density of states in k-space. From Fig. 4.8, we can set $\rho_r(E_{21})\delta E_{21} = \rho(k)\delta k$. Rearranging, we find

$$\frac{1}{\rho_r(E_{21})} = \frac{1}{\rho(k)}\frac{dE_{21}(k)}{dk} = \frac{1}{\rho(k)}\left[\frac{dE_2(k)}{dk} - \frac{dE_1(k)}{dk}\right]. \tag{4.27}$$

This definition allows $\rho_r(E_{21})$ to be evaluated at any given point in k-space once the derivatives of the electron and hole energies with respect to k are known at that point. This definition is especially useful when $E_2(k)$ and $E_1(k)$ are not parabolic functions (see Figs. A8.4 and A8.6 for examples of nonparabolic subbands in QWs). The density of states in k-space, $\rho(k)$, for various dimensional structures is

TABLE 4.3: Density of States for Bulk (3D), Quantum Well (2D), Quantum Wire (1D) and Quantum Dot (0D) Structures (Including Spin)

Dimension	$\rho(k)$	$\rho(E)$
3	$\dfrac{k^2}{\pi^2}$	$\dfrac{\sqrt{E}}{2\pi^2}\left[\dfrac{2m}{\hbar^2}\right]^{3/2}$
2	$\dfrac{k}{\pi d_z}$	$\dfrac{m}{\pi\hbar^2 d_z}$
1	$\dfrac{2}{\pi d_x d_y}$	$\dfrac{\rho(k)}{\sqrt{E}}\left[\dfrac{2m}{\hbar^2}\right]^{1/2}$
0	–	$2\delta(E-E_0)$

summarized in Table 4.3. (Note that the z-direction is taken as the narrow dimension of the quantum well, and the axis of the quantum wire to be consistent with Appendices 8 and 10.)

If both bands involved in the transition are parabolic, an even more straightforward definition of $\rho_r(E_{21})$ can be used. We can generally state that the transition energy is equal to the bandgap energy $E_g = E_c - E_v$ plus the kinetic energies of the electron and hole. If the electrons and holes follow parabolic dispersion curves, we have

$$E_{21} = E_g + \frac{\hbar^2 k^2}{2m_c} + \frac{\hbar^2 k^2}{2m_v} = E_g + \frac{\hbar^2 k^2}{2m_r}, \quad \text{where} \quad \frac{1}{m_r} \equiv \frac{1}{m_c} + \frac{1}{m_v}. \tag{4.28}$$

In other words, the dispersion of E_{21} with k also follows a parabolic curve with a curvature characterized by a *reduced* mass, m_r. As a result, the density of transition states along E_{21} is entirely analogous to the density of states function in either the conduction or valence band, with the following associations:

$$\begin{aligned} \rho_r(E_{21}) &\leftrightarrow \rho_c(E_2), \rho_v(E_1) \\ E_{21} - E_g &\leftrightarrow E_2 - E_c, E_v - E_1 \quad \text{(parabolic bands)} \\ m_r &\leftrightarrow m_c, m_v \end{aligned} \tag{4.29}$$

The derivation of $\rho_r(E_{21})$ provided in Appendix 6 for bulk material confirms these associations. More generally, the density of states per unit energy given in Appendix 1 and summarized in Table 4.3 for various dimensional structures can be applied directly to $\rho_r(E_{21})$ using Eqs. (4.29) and (4.28) as long as the energy bands are parabolic.

Finally, to evaluate the Fermi occupation probabilities in Eq. (4.3), we need the individual electron and hole energies. With parabolic bands and assuming **k**-selection, the individual electron and hole energies in terms of the transition

energy can be found using Eq. (4.28):

$$E_2 = E_c + (E_{21} - E_g)\frac{m_r}{m_c}, \quad E_1 = E_v - (E_{21} - E_g)\frac{m_r}{m_v}. \tag{4.30}$$

When the bands are not parabolic, we must in general use $E_2(k)$ and $E_1(k)$ evaluated at the k-vector, which yields the desired transition energy, E_{21}.

Example 4.1 In a semiconductor optical amplifier, the TE and the TM mode gains are found to be the same for a particular wavelength of incident light with sufficient current applied. Consider only the C-HH1 transition in an 8-nm-wide unstrained GaAs quantum well, surrounded by the 20% Al-AlGaAs barriers, having the lowest transition wavelength of 840 nm.

Problem: What in-plane k-vector and wavelength would satisfy the equal TE and TM gain condition?

Solution: To solve this problem, we need to utilize the material from Appendix 10, and we need to recall that the polarization dependence in the gain of the quantum wells comes from the matrix element M_T. From Fig. 4.4, which applies for this material system, at the bottom of the first subband, for $k_t = 0$, all of the gain is going to the TE polarized photons. If we can tailor the value of the in-plane k-vector k_t, we can achieve the conditions under which the values of the matrix elements for TE and TM polarization will be the same. This condition is fulfilled for $k_t = 0.03$ Å^{-1}, in which case

$$\left|\frac{M_T}{M}\right|^2_{TE} = \left|\frac{M_T}{M}\right|^2_{TM} = 0.27.$$

The transition energy is equal to the bandgap energy, plus kinetic energies of electrons and holes, as expressed by Eq. (4.28). Therefore,

$$\Delta E_t = \frac{\hbar^2 k_t^2}{2m_c} + \frac{\hbar^2 k_t^2}{2m_v} = \frac{\hbar^2 k_t^2}{2m_r}.$$

Using the values from Table 1.1, we can compute the reduced mass as $m_r = \frac{m_c m_v}{m_c + m_v} = \frac{0.067 \cdot 0.38}{0.067 + 0.38} m_0 = 0.057 \cdot m_0$. From there, $\Delta E_t = 5.98 \cdot 10^{-2}$ eV. Finally, the wavelength can be obtained from the relationship

$$\frac{\Delta\lambda}{\lambda} = -\frac{\Delta E}{E_g},$$

where E_g corresponds to the transition wavelength of $\lambda = 840$ nm. From here, we have

$$\Delta\lambda = -\frac{\lambda}{E}\Delta E = -\frac{840 \text{ nm}}{1.24/0.840 \text{ eV}} \cdot 5.98 \cdot 10^{-2} \text{ eV} = 34 \text{ nm},$$

and $\lambda = 840 - 34$ nm $= 806$ nm.

4.2.5 Correspondence with Einstein's Stimulated Rate Constant

Einstein's original description of radiative transitions outlined in Appendix 6 defines the downward transition rate in terms of a stimulated rate constant, B_{21}, weighted by the radiation spectral density, $W(\nu)$, and the differential number of state pairs available for downward transitions, dN_2. This rate constant can be related to R_r, allowing us to quantify B_{21} using Fermi's Golden Rule.

Assuming the lineshape broadening function to be a delta function, the stimulated emission rate in Eq. (A14.11) becomes $dR_{21} = B_{21}\ W(\nu)dN_2$. This differential transition rate must still be integrated over all transition pairs affected by $W(\nu)$. Using Eq. (A14.8) for dN_2, setting $W(\nu) \rightarrow h\nu N_p \delta(\nu - \nu_0)$, and integrating over the transition energy, the stimulated emission rate becomes

$$R_{21} = B_{21} \cdot h\nu_0 N_p h\rho_r f_2(1 - f_1). \tag{4.31}$$

Comparing this to Eq. (4.1), we conclude that

$$B_{21} = \frac{R_r}{h\nu_0 N_p h\rho_r} = \frac{1}{h\nu_0 N_p} \frac{|H'_{21}|^2}{\hbar^2}. \tag{4.32}$$

The second equality uses Fermi's Golden Rule (4.16) to expand R_r. With this and Eq. (4.37) defined later, B_{21} and all relations dependent on B_{21} in Appendix 6 can be quantified.

4.3 OPTICAL GAIN

4.3.1 General Expression for Gain

The explicit relation between the net stimulated emission rate and the optical gain was derived in Chapter 2. For reference, we repeat that derivation here in a slightly different way. As discussed in Chapter 2, we can define the *material* gain per unit length as the proportional growth of the photon density as it propagates along some direction in the crystal. This definition can be related to the transition rates as follows:

$$g = \frac{1}{N_p}\frac{dN_p}{dz} = \frac{1}{v_g N_p}\frac{dN_p}{dt} = \frac{1}{v_g N_p}(R_{21} - R_{12}). \tag{4.33}$$

The second equality uses the group velocity, v_g, to transform the spatial growth rate to the growth rate in time. The growth rate in time is then linked to the net generation rate of photons per unit volume. Finally, using Eq. (4.2) we obtain

$$g = \frac{R_{st}}{v_g N_p} = \frac{R_r}{v_g N_p}(f_2 - f_1). \tag{4.34}$$

Using Fermi's Golden Rule (4.16) for R_r, we have

$$g_{21} = \frac{2\pi}{\hbar}\frac{|H'_{21}|^2}{v_g N_p}\rho_r(E_{21})\cdot(f_2 - f_1). \tag{4.35}$$

The electromagnetic perturbation is proportional to the field strength. Thus, to evaluate the ratio, $|H'_{21}|^2/N_p$, we need to relate the field strength to the photon density. The energy density in terms of the photon density is $\hbar\omega N_p$. The energy density in terms of the electric field strength is $\frac{1}{2}n^2\varepsilon_0|\mathscr{E}|^2$. If the material is dispersive, this becomes $\frac{1}{2}nn_g\varepsilon_0|\mathscr{E}|^2$. The electric field is related to the vector potential through a time derivative. For time-harmonic fields, we can set $|\mathscr{E}|^2 = \omega^2|\mathscr{A}|^2$. By equating the two versions of the energy density, we obtain the desired relation:

$$\tfrac{1}{2}nn_g\varepsilon_0\ \omega^2|\mathscr{A}_0|^2 = \hbar\omega N_p \rightarrow |\mathscr{A}_0|^2 = \frac{2\hbar}{nn_g\varepsilon_0\,\omega}N_p. \tag{4.36}$$

Using this relation and the definition of the matrix element Eq. (4.22), we can set[2]

$$\frac{|H'_{21}|^2}{N_p} = \frac{1}{N_p}\left(\frac{q\mathscr{A}_0}{2m_0}\right)^2|M_T|^2 = \frac{q^2\hbar}{2nn_g\varepsilon_0 m_0^2\,\omega}|M_T|^2. \tag{4.37}$$

The material gain per unit length Eq. (4.35) then becomes

$$g_{21} = g_{max}(E_{21})\cdot(f_2 - f_1)$$

where

$$g_{max}(E_{21}) = \frac{\pi q^2\hbar}{n\varepsilon_0\ cm_0^2}\frac{1}{h\nu_{21}}|M_T(E_{21})|^2\rho_r(E_{21}). \tag{4.38}$$

The maximum gain, g_{max}, is a property of the material, whereas the Fermi factor, $f_2 - f_1$, is dependent on the injection level. Per example, in GaAs 80–100 Å quantum wells, the maximum gain of each subband transition is $g_{max} \sim 10^4$ cm^{-1} (or 1 μm^{-1}).

[2]The transition matrix element in the expression for gain is occasionally written instead as the *dipole moment* matrix element, $q^2|x|^2$, where $|x|^2 = |\langle u_c|\hat{\mathbf{e}}\cdot\mathbf{x}|u_v\rangle|^2|\langle F_2|F_1\rangle|^2$, and $\mathbf{x}$ is the position operator. The relationship between the two is given by $q^2|M_T|^2 = m_0^2\ \omega^2 q^2|x|^2$. Thus, Eq. (4.37) can alternatively be written as

$$\frac{|H'_{21}|^2}{N_p} = \frac{\hbar\omega}{2nn_g\varepsilon_0}q^2|x|^2.$$

This ratio can then be used to define the gain using Eq. (4.35); however, we will not make use of this alternate expression for gain in this chapter.

Gain in Reduced Dimensionality Structures In reduced-dimensional structures such as a quantum well, ρ_r corresponds to the reduced density of states between two quantized subbands. The total gain at E_{21} is found by summing over all possible subband pairs:

$$g_{21} = \sum_{n_c} \sum_{n_v} g_{21}^{sub}(n_c, n_v). \quad \text{(quantum well)} \tag{4.39}$$

The double sum indicates that all subband combinations should be considered. In practice however, the selection rules arising from the envelope function overlap expressed in Eq. (4.25) and illustrated in Fig. 4.7 suggest that the gain from $n_c = n_v$ subband pairs will dominate the gain spectrum. In particular, the $n = 1$ gain is usually the largest, and hence most important, transition in quantum-well lasers. We will use the following example to demonstrate this point as well as to highlight many other basic considerations involved with determining the gain for a given injection level in a quantum well.

Example 4.2 An optical probe beam is transmitted through an AR-coated GaAs epi-wafer normal to the surface to measure the absorption properties of an active region that lies in the plane of the wafer. The active region contains a single strained InGaAs quantum well, 8 nm in thickness within a GaAs/AlGaAs waveguide, with the confinement factor of 0.015. The probe beam has a wavelength of 970 nm, and the lowest energy level in the quantum well provides an absorption edge at 980 nm. It is determined that the one-pass absorption of the probe beam through the unpumped quantum well is 1.5%. Also, the threshold modal gain of a 3-μm-wide, 500-μm-long ridge laser made in this material is 29.8 cm^{-1}.

Problem: (1) What is the maximum material gain at 970 nm for a very strongly pumped active region? (2) What is the Fermi function difference $(f_2 - f_1)$ at the lasing threshold?

Solution: From the transmission measurements, we have that the single pass absorption is 1.5%. That is,

$$\frac{P_{out}}{P_{in}} = e^{-g_{max}d},$$

where d is the thickness of the well. From here, the maximum gain is

$$g_{max} = \frac{0.015}{8 \cdot 10^{-7}} \text{ cm}^{-1} = 1.875 \cdot 10^4 \text{ cm}^{-1}.$$

To calculate the Fermi function difference $(f_2 - f_1)$, we utilize the relationship for gain, Eq. (4.38),

$$g = g_{max}(f_2 - f_1).$$

Therefore, we need to calculate the threshold gain, which can be obtained from the parameters given in the problem,

$$g_{th} = \frac{\Gamma g_{th}}{\Gamma} = \frac{29.8}{0.015} = 1986.67 \text{ cm}^{-1}.$$

Finally,

$$(f_2 - f_1) = \frac{g_{th}}{g_{max}} = \frac{1986.67}{18750} = 0.106.$$

The left side of Fig. 4.9 illustrates the lowest two energy subbands of a quantum well in both conduction and valence bands (neglecting the light-hole subbands and assuming parabolic subbands for simplicity). Under strong forward-bias conditions, the equilibrium Fermi level is separated into two quasi-Fermi levels, one for all conduction subbands and one for all valence subbands (the quasi-Fermi functions are indicated by the dashed curves). The separation of the quasi-Fermi levels is constrained by the requirement that *charge neutrality* be maintained within the quantum well (if there were a charge imbalance, band-bending would occur in the diode junction in such a way as to neutralize the imbalance). Thus, we must have $N(E_{Fc}) = P(E_{Fv})$ in the quantum well (see Eq. (A14.20) for a more explicit version of this relation). Because the valence band typically has many more states per unit energy, the valence quasi-Fermi function does not have to penetrate nearly as deeply as

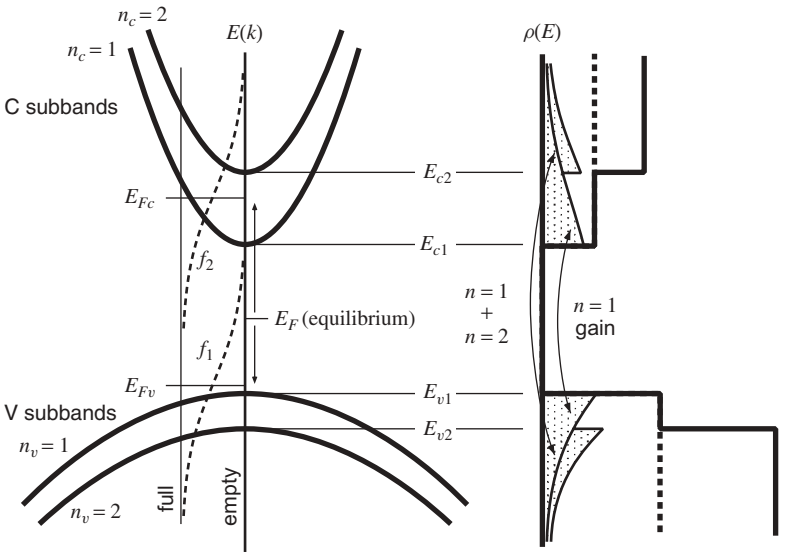


FIGURE 4.9: QW subbands and corresponding density of states illustrating the relationships between the carrier populations, the quasi-Fermi levels, and the gain at the subband edges.

the conduction quasi-Fermi function to obtain the same overall carrier density. As a result, the quasi-Fermi levels separate asymmetrically as indicated in the figure.

The right side of Fig. 4.9 shows the constant density of states functions of each subband, which when added together produce a staircase density of states in each band. When $\rho(E)$ is multiplied by the fraction of filled (empty) states in the band, we obtain the electron (hole) distribution as a function of energy (the contributions from the $n = 1$ and $n = 2$ subbands are indicated separately in the figure). The shaded area under the carrier distribution curves then yields the total carrier density. Because charge neutrality requires that $N = P$, we conclude that the total shaded area must be the same in the conduction and valence bands for any injection level. Thus, the quasi-Fermi levels must always adjust themselves to ensure that this requirement is met (with the larger step heights in the valence band, this again explains why the valence quasi-Fermi function does not have to penetrate as deeply).

With the relationship between the quasi-Fermi levels and the carrier density qualitatively defined, we can now proceed to analyze the gain of the quantum well. Many optical properties of the quantum well depicted in Fig. 4.9 can be qualitatively determined by simple inspection. For example, from Eq. (4.6) we immediately know that at the injection level indicated, the quantum well provides gain at the lowest subband edge, simply because $E_{Fc} - E_{Fv} > E_{c1} - E_{v1}$. Thus, despite the fact that the electron density at E_{v1} is slightly larger than the electron density at E_{c1}, a population inversion at the band edge has in fact been achieved. This is because a "population inversion" requires $f_2 > f_1$, *not* $\rho_c f_2 > \rho_v f_1$ (as the name might lead one to conclude). Furthermore, the fact that we have gain at the band edge even though $E_{Fv} > E_{v1}$ reinforces the concept that it is the *relative difference* and not the absolute positions of E_{Fc} and E_{Fv} that determines the gain (i.e., it is not a requirement to have *both* $E_{Fc} > E_{c1}$ and $E_{Fv} < E_{v1}$ to achieve gain).

We can take the analysis a step further by using Eq. (4.38) to estimate the gain at the band edge. Evaluating terms at the appropriate energies, we find the band edge gain to be $g_{max1}(E_{g1}) \cdot (f_2(E_{c1}) - f_1(E_{v1}))$, where g_{max1} uses the envelope function overlap and reduced density of states between the two $n = 1$ subbands. Estimating from the figure that $f_2(E_{c1}) - f_1(E_{v1}) \approx 0.8 - 0.6 = 0.2$, we conclude that the band edge gain is roughly 20% of its maximum possible value. If we assume $g_{max1}(E_{g1}) \sim 10^4$ cm^{-1} (a typical number), then the band edge gain in the figure is roughly 2000 cm^{-1}.

As we move to higher photon energies, the population inversion clearly declines, implying that the gain from $n = 1$ transitions is largest at the band edge. However, at high enough photon energies a second population of carriers starts contributing to the transition process. In principle, this added supply of carriers starts contributing to the transition process. In principle, this added supply of carriers could increase the gain significantly. However in this case, at photon energies close to the second subband edge, we know the quantum well is absorbing because $E_{Fc} - E_{Fv} < E_{c2} - E_{v2}$. The absorption comes from two contributions: the carrier populations in the $n = 1$ subbands and the carrier populations in the $n = 2$ subbands, as the summation in Eq. (4.39) implies. Estimating from the figure

that $f_2(E_{c2}) - f_1(E_{v2}) \approx 0.3 - 0.8 = -0.5$, we find that the absorption from the $n = 2$ subband transitions is $0.5g_{max2}(E_{g2})$. For the $n = 1$ subband transitions, it's a little trickier because the Fermi energies must be estimated for electrons and holes separated by E_{g2} in the $n = 1$ subbands. Using Eq. (4.30), we must evaluate $f_2[E_{c1} + (E_{g2} - E_{g1})(m_r/m_c)] - f_1[E_{v1} - (E_{g2} - E_{g1})(m_r/m_v)]$, which from the figure is approximately $\approx 0.2 - 0.7 = -0.5$. The overall absorption at the second subband edge is therefore $0.5(g_{max1}(E_{g2}) + g_{max2}(E_{g2}))$ (cross-population transitions such as g_{max12} and g_{max21} also exist but as mentioned earlier, their contributions are typically small). So we find that the absorption at E_{g2} has not yet been converted into gain but has at least been reduced to half its maximum absorption value with the application of a forward bias. To achieve gain at the $n = 2$ subband edge, the forward bias must be increased to the point where $E_{Fc} - E_{Fv} > E_{g2}$. To surpass the $n = 1$ subband edge gain, the forward bias must be even stronger such that

$$(g_{max1}(E_{g2}) + g_{max2}(E_{g2}))(f_2(E_{c2}) - f_1(E_{v2})) > (g_{max1}(E_{g1})f_2(E_{c1}) - f_1(E_{v1})).$$

So for all but very strong forward-bias conditions, the $n = 1$ gain dominates.

Quantum Dots To apply the gain theory to quantum dots, we need to take into account their special physical and quantum properties. If we assume that we have a uniform distribution of quantum dots, as shown in Fig. 4.10, 1 dot will occupy a unit volume equal to $V = L_x L_y L_z$. Each dot occupies a volume equal to $V_{dot} = d_x d_y d_z$ inside this unit volume. The three-dimensional dot density is then given by

$$\rho_{dot}^{3D} = \frac{1}{V} = \frac{1}{L_x L_y L_z}$$

The conduction band energy is given by

$$E = E_{c0} + \frac{\hbar^2 \pi^2}{2m^*}\left[\left(\frac{n_x}{d_x}\right)^2 + \left(\frac{n_y}{d_y}\right)^2 + \left(\frac{n_z}{d_z}\right)^2\right]. \tag{4.40}$$

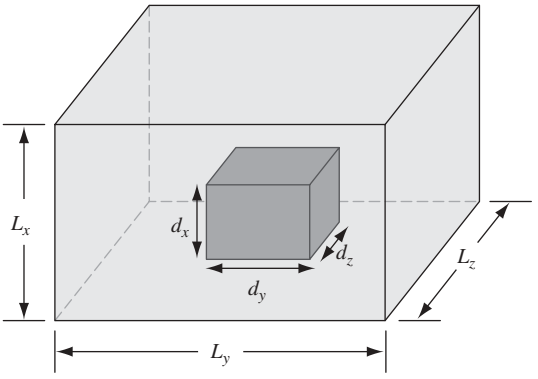


FIGURE 4.10: Quantum dots and their geometric parameters.

where E_{c0} is the energy of the conduction band edge. The electron density can be determined from the quantum dot density, and the probability of occupation for each of the dot's states, defined by Fermi's distribution,

$$N = \frac{2}{V} \sum_{n_x, n_y, n_z} f_c(E). \tag{4.41}$$

Exactly analoguous equation and relationships hold for the holes in the valence band. Using Eq. (4.35), and accounting for the fact that only discrete transitions between states with identical quantum numbers can occur, similarly to the quantum-well case discussed, the expression for the gain spectrum of the quantum dots is given by

$$g_{21} = \sum_c \sum_v \frac{2\pi}{\hbar} \frac{|H'_{21}|^2}{v_g N_p} 2 \cdot \rho_{dot}^{3D} \cdot (f_2 - f_1). \tag{4.42}$$

Using the substitution from Eq. (4.37), we can write the gain in the form

$$g_{21} = g_{max}(E_{21}) \cdot (f_2 - f_1), \tag{4.43}$$

where g_{max} is given by

$$g_{max} = \sum_{n_{cx}, n_{cy}, n_{cz}} \sum_{n_{vx}, n_{vy}, n_{vz}} \frac{\pi q^2 \hbar}{n \varepsilon_0 c m_0^2} \frac{1}{h\nu_{21}} \cdot \frac{2}{V} |\langle u_c | \hat{\mathbf{e}} \cdot \mathbf{p} | u_v \rangle|^2 \times \delta(E_2(n_{cx}, n_{cy}, n_{cz}) - E_1(n_{vx}, n_{vy}, n_{vz}) - E_{21}). \tag{4.44}$$

Here, we have written the value of the transition matrix element $|M_T|$ explicitly. If we define $E_{21}(n_x, n_y, n_z) = E_2(n_{cx}, n_{cy}, n_{cz}) - E_1(n_{vx}, n_{vy}, n_{vz})$, we can further simplify Eq. (4.43), to give

$$g_{max} = \sum_{n_x, n_y, n_z} \frac{\pi q^2 \hbar}{n \varepsilon_0 c m_0^2} \frac{1}{h\nu_{21}} \cdot \frac{2}{V} |\langle u_c | \hat{\mathbf{e}} \cdot \mathbf{p} | u_v \rangle|^2 \delta(E_{21}(n_x, n_y, n_z) - E_{21})). \tag{4.45}$$

As further discussed in the next section, lineshape broadening has significant influence on the final gain spectrum, which is particularly true for quantum dots, where the delta function in the gain expression will be replaced by a lineshape function. In particular, inhomogenous broadening, to be discussed in the next section, tends to be a particularly large factor in quantum dots formed by most experimental growth technologies.

To summarize, by invoking charge neutrality to link both quasi-Fermi levels to a given carrier density, evaluating the quasi-Fermi functions at the appropriate

energies, and summing over all subband transition pairs, we can determine the gain at any given photon energy relative to the maximum value. In Section 4.3.3, we will return to a more general discussion of the gain spectrum and its properties in bulk and quantum-well materials.

4.3.2 Lineshape Broadening

Equation A6.24 in Appendix 6 expresses the gain in terms of B_{21}. Replacing B_{21} with Eq. (4.32) and using Eq. (4.37), we find that Eq. (A6.24) is indeed equivalent to Eq. (4.33) derived here. However, the gain more generally defined in Eq. (A6.23) has an additional integration which takes into account the energy uncertainty of the electron states. This *energy broadening* of electron states ultimately limits the resolution of features we can observe in the gain spectrum and is therefore particularly important to consider in reduced dimensional structures where the reduced density of states function contains very sharp features. To recover Eq. (A6.23) from Eq. (4.38), we need to consider how the broadening of electron states affects the gain.

Figure 4.11 reveals that when the energy states are broadened, many different transition pairs contribute to gain at a particular photon energy. These transition pairs are primarily clustered within the energy uncertainty width of the *lineshape function* describing the probable energy distribution of each transition pair. To determine the *total* gain at $h\nu_0$, we must integrate g_{21} over all transition energies

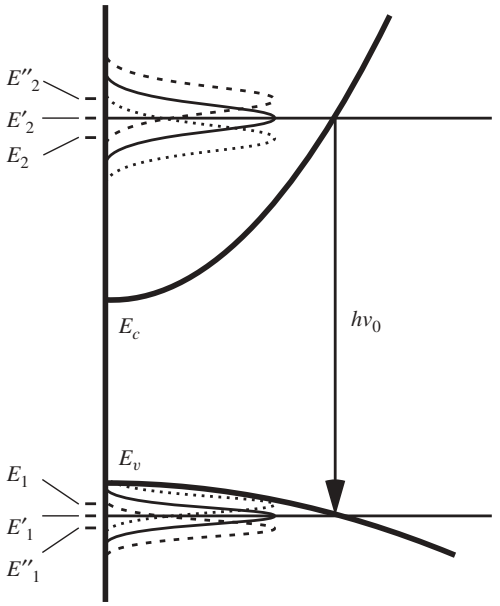


FIGURE 4.11: Three (of many) transition pairs that contribute to gain at $h\nu_0$.

weighted by the appropriate lineshape function, $\mathscr{L}(h\nu_0 - E_{21})$.[3] The gain including lineshape broadening therefore takes the form:

$$g(h\nu_0) = \int g_{21}\mathscr{L}(h\nu_0 - E_{21})dE_{21}. \tag{4.46}$$

In Eq. (4.38), $h\nu_{21}$ should be set equal to $h\nu_0$, whereas all other terms dependent on E_{21} should be considered variables of the integration. This expression for the gain now agrees with Eq. (A6.23) derived in Appendix 6.

The specific form for the lineshape function to be used in Eq. (4.46) can be determined by attempting to study the time evolution of an electron state, taking into account its interaction with phonons and other electrons. In a first-order approximation, we might assume that the probability of finding the electron in a given state decays exponentially, as we found for the electron's interaction with photons in Appendix 9. This simplistic time dependence when Fourier transformed to the energy domain immediately leads to a Lorentzian lineshape function:

$$\mathscr{L}(E - E_{21}) = \frac{1}{\pi}\frac{\hbar/\tau_{in}}{(\hbar/\tau_{in})^2 + (E - E_{21})^2}. \tag{4.47}$$

The *intraband relaxation* time, τ_{in}, is the time constant associated with the exponential decay of the electron. The energy full-width is related to τ_{in} via $\Delta E_{21} = 2\hbar/\tau_{in}$ (compare Eq. (4.36) to Eq. (A6.27)).[4] Early investigations, which attempted curve fits of gain and spontaneous emission spectra to measurements in bulk material, lead to values of $\tau_{in} \approx 0.1$ ps [7]. However, the gain and emission spectra did not match very well, particularly on the low-energy side of the spectrum near the band edge, where the details of the lineshape function are most apparent. Thus, other more sophisticated theoretical methods of determining the lineshape function have since been employed.

Using a quantum mechanical density matrix approach, Yamanishi and Lee [8] have suggested that the electron state decays initially as a Gaussian but then takes on exponential behavior for larger times. This leads to less energy in the tails of the lineshape function than the Lorentzian function, which is more in line with experimental observations. Asada [9] has also performed a detailed analysis of intraband

[3]This lineshape function is actually a combination of the individual electron and hole lineshapes comprising the transition pair:

$$\mathscr{L}(h\nu_0 - E_{21}) = \int \mathscr{L}_2(E - E_2)\mathscr{L}_1((E - h\nu_0) - E_1)\,dE.$$

[4]In Appendix 9, it is shown that if a state decays exponentially with a time constant, $\tau(= 1/W)$, then the energy uncertainty of the state is $\Delta E_{FWHM} = \hbar/\tau$. The combined energy uncertainty of states 1 and 2 is $\Delta E_{21} = \hbar(1/\tau_2 + 1/\tau_1)$. With τ_{in} defined as the *average* time constant: $1/\tau_{in} = \frac{1}{2}(1/\tau_2 + 1/\tau_1)$, we end up with $\Delta E_{21} = 2\hbar/\tau_{in}$.

scattering in quantum wells, arriving at an asymmetrical lineshape function, which falls off much faster than a Lorentzian on the low-energy side of the transition, similar to the findings of Yamanishi and Lee. Kucharska and Robbins [10] have kept with a Lorentzian lineshape, but have theoretically derived an energy-dependent lifetime, arguing that the scattering rate out of a state is dependent on where the state is in the band, and on how full the band is.

In an attempt to keep the lineshape function as simple as possible while maintaining some degree of accuracy, Chinn et al. [11] approximated the numerical lineshape function derived by Yamanishi and Lee with a simple curve fit that describes the time dependence of the electron state. This time dependence is given by

$$e^{-l(t)},$$

where

$$\log_{10} l(t[ps]) = 2 + 1.5 \log_{10} t - 0.5\sqrt{(2 + \log_{10} t)^2 + 0.36.} \tag{4.48}$$

For long times, $l(t) \rightarrow t$, reproducing the exponential decay. For short times, $l(t) \rightarrow t^2$ implying that the state initially decays as a Gaussian. The most efficient way to make use of the Chinn lineshape function is to inverse fast Fourier transform (inverse FFT) the gain spectrum, multiply it by Eq. (4.48), and FFT back to the energy domain. This proves to be the fastest method of evaluating the convolution contained in Eq. (4.46). We will make use of this simplified Chinn lineshape function in later calculations.

Some example lineshape functions are shown in Fig. 4.12. The Chinn lineshape has less energy in the tails than the Lorentzian lineshape. However, the Gaussian lineshape, which has been included for comparison, has significantly less energy in the tails than either of the other two. The effects of convolving these lineshapes with a typical quantum-well gain spectrum are shown in the lower part of Fig. 4.12. The dramatic smoothing of the sharp features of the gain spectrum can reduce the peak gain substantially. However at higher gains, the reduction is not as significant. Note that due to the energy in the tails, the Lorentzian lineshape function creates absorption of almost 100 cm^{-1} below the band edge. The other two lineshapes do not suffer this problem. Of the three, the Chinn lineshape is perhaps closest to representing the actual complex lineshape function of the semiconductor. However, in reality the lineshape function is a complex function of both transition energy and injection level.

For quantum dots, lineshape broadening will change the delta function into the appropriate lineshape. To account for this, the new expression for maximum gain g_{max} of quantum dots becomes

$$g_{max} = \sum_{n_x, n_y, n_z} \frac{\pi q^2 \hbar}{n \varepsilon_0 c m_0^2} \frac{1}{h\nu_{21}} \cdot \frac{2}{V} |\langle u_c | \hat{\mathbf{e}} \cdot \mathbf{p} | u_v \rangle|^2 \mathscr{L}(E_{21}(n_x, n_y, n_z) - E_{21})). \tag{4.49}$$

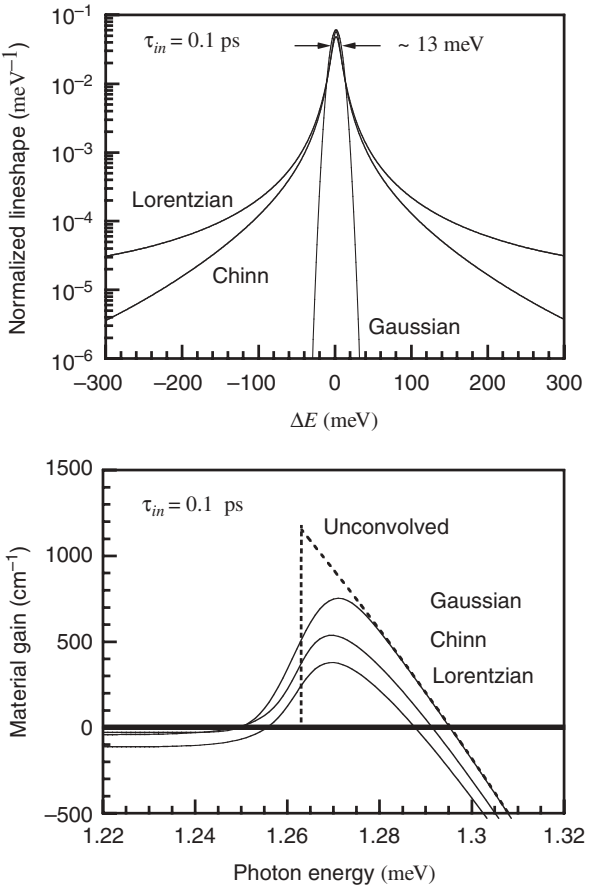


FIGURE 4.12: Comparison of various lineshape functions and the resulting convolved gain spectra.

Another contribution to the lineshape broadening comes from the effect of physical dimension variation in the epitaxial structure, leading to the effect of inhomogenous broadening. Thickness variations in quantum wells will always exist but can generally be controlled and neglected. However, in self-formed quantum dots, the variations in physical size of the quantum dots are inevitable, and they must be accounted for in the final expression for gain. This variation in size leads to the variation in energies of the dot states, as can be seen from Eq. (4.40). This type of broadening can also be modeled by introducing another broadening factor into the gain expression, this time with a Gaussian distribution,

$$B_I(E) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(E - E_{21}(n_x, n_y, n_z))}. \tag{4.50}$$

To add this additional broadening parameter into the expression for quantum dot gain, we need to integrate over energy,

$$g_{21} = \frac{\pi q^2 \hbar}{n\varepsilon_0 c m_0^2} \frac{1}{h\nu_{21}} \sum_{n_x, n_y, n_z} \int_0^\infty \frac{2}{V} |\langle u_c | \hat{\mathbf{e}} \cdot \mathbf{p} | u_v \rangle|^2 \mathscr{L}(E - E_{21}) B_I(E)(f_2 - f_1) dE. \tag{4.51}$$

4.3.3 General Features of the Gain Spectrum

From the general gain Eq. (4.38), it is clear that the gain/absorption spectrum is bounded to a maximum value of $|g_{max}|$. With no carrier injection, the material is strongly absorbing with an absorption spectrum equal to $-g_{max}(E_{21})$. With carrier injection, we can invert the carrier population near the band edge and change the Fermi factor, $f_2 - f_1$, from -1 toward $+1$, converting the absorption into gain. As considered earlier in relation to Fig. 4.9, the carrier inversion is highly concentrated near the band edge. Therefore, as the photon energy increases away from the band edge, $f_2 - f_1$ must steadily reduce back to -1, and the gain spectrum must reduce back to the unpumped absorption spectrum. This conversion of absorption into gain is depicted in Fig. 4.13 for both bulk and quantum-well material.

In both bulk and quantum-well cases, the material is transparent below the bandgap. Just above the bandgap, a region of positive gain exists. Beyond this, the material becomes strongly absorbing. The crossing point from positive gain to absorption occurs when $f_2 = f_1$ and was considered earlier in discussions related to Eqs. (4.4) through (4.7). There it was found that the stimulated emission and absorption rates exactly cancel (making the material transparent) when $h\nu = \Delta E_F$. Thus the region of positive gain extends, between the bandgap and the quasi-Fermi level separation:

$$E_g < h\nu < \Delta E_F. \qquad \text{(positive gain)} \tag{4.52}$$

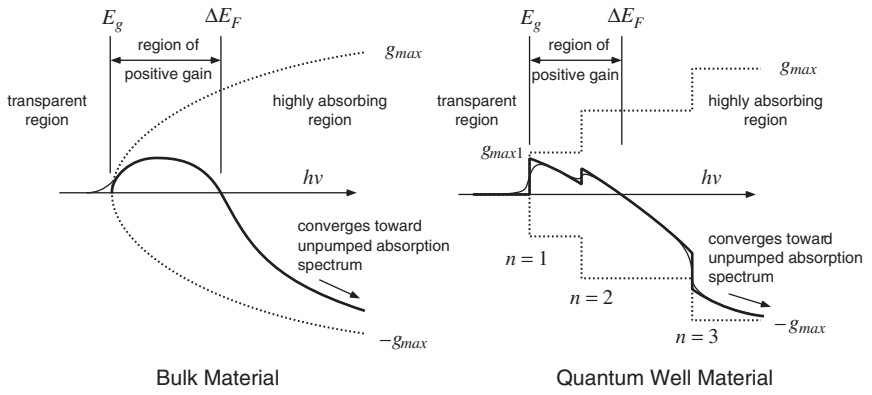


FIGURE 4.13: Gain spectra in bulk and quantum-well materials. The thin smoothed curves indicate the effect of lineshape broadening.

In other words, gain is achieved in the material only when the carrier injection is high enough to create a quasi-Fermi-level separation exceeding the bandgap—or the separation between the lowest allowed quantum states in the case of quantum wells. Also, the larger the quasi-Fermi-level separation we can create, the wider the gain bandwidth we can achieve in the material.

The shape of the bounding limits in Fig. 4.13 representing g_{max} and $-g_{max}$ is primarily governed by the reduced density of states function[5] from Eq. (4.38). Hence in bulk material, g_{max} follows a square root dependence, while in quantum-well material, g_{max} follows a steplike dependence where each step corresponds to the addition of a new subband transition pair. As a result, the bulk gain spectrum is quite smooth, whereas the quantum-well gain spectrum is rather jagged. However in the latter case, lineshape broadening tends to smooth out the discontinuous features into rounded "bumps," one for each subband transition pair.

In bulk material, the peak of the gain spectrum increases and shifts to higher photon energies with increasing ΔE_F. In contrast, the peak of the quantum-well gain spectrum remains fixed at the $n = 1$ gain peak near the band edge under most conditions. With increased carrier injection, the $n = 1$ gain saturates at g_{max1} as the states in the first subband reach complete inversion, while the $n = 2$ gain continues to increase to a value twice as high as the $n = 1$ gain (twice as high because it contains contributions from both $n = 1$ and $n = 2$ subband transition pairs). Thus, under very high carrier injection, the overall peak gain can jump to the $n = 2$ gain peak as discussed in relation to Fig. 4.9. Experiments on quantum-well lasers do in fact show a discrete jump in the lasing wavelength from the $n = 1$ to the $n = 2$ peak as the cavity length is reduced (i.e., as the threshold gain is increased).

In drawing the quantum-well gain spectrum in Fig. 4.13, a few features have been idealized. First of all, an additional set of steps spaced differently from the ones shown should be included to account for both conduction-to-heavy hole (C – HH) *and* conduction-to-light hole (C – LH) subband transitions. In other words, there should be an $n = 1$ step for both types of transitions, etc. In practice, for the more common situation where the electric field lies in the plane of the quantum well, the matrix element for C – HH transitions is three times larger than the matrix element for C – LH transitions, implying that the steps related to C – LH transitions are much smaller and not as important. We have also neglected steps due to forbidden transitions. Inclusion of these transitions introduces small peaks in between the major peaks due to the small but finite overlap of forbidden transitions. However, in practice such small peaks are not usually observed.

Another simplification we have made in Fig. 4.13 is that all steps comprising g_{max} have been drawn with equal height. In actuality, the overlap integral defining the matrix element is typically smaller for higher subband transitions, implying that the $n = 1$ step height is usually the largest (sometimes by as much as a factor

[5]There are other energy-dependent terms comprising $g_{max}(E)$. For example, the $1/\hbar\omega$ dependence modifies the shape of $g_{max}(E)$ slightly, but not noticeably. In reduced-dimensional structures, the matrix element also has an energy dependence; however, we will ignore this dependence for the present discussion.

of 2, depending on the barrier height). Finally, we have completely neglected the energy dependence of the matrix element. This energy dependence is illustrated in Appendix 10. Generally for polarizations of interest, the matrix element is a maximum at the band edge and decays to less than half of its peak value at higher energies. Thus, the flat plateaus in between the steps of g_{max} should actually slope toward zero, modifying the gain spectrum accordingly (the $1/\hbar\omega$ dependence will also contribute to this sloping toward zero). Gain calculations using practical material systems presented at the end of this chapter will reveal these more subtle features in the spectrum.

4.3.4 Many-Body Effects

The preceding theory of gain involving Fermi's Golden Rule considers each electron in isolation as it interacts with the electromagnetic field. In other words, we have used a single-particle theory to obtain the gain spectrum. In reality, there is a large density of both electrons and holes present in our system. The mutual interactions between these particles are generally referred to as *many-body* effects. We have already considered one consequence of such many-body effects in our discussion of lineshape broadening, which is related to collisions between particles and/or phonons in the crystal. In addition to this important effect, there are two other significant consequences of many-body effects: exciton states and bandgap shrinkage. Exciton states exist primarily at low carrier densities and low temperatures, whereas bandgap shrinkage becomes noticeable at high carrier densities.

Under conditions of low carrier density and low temperature, it is possible for an electron and hole to orbit each other for an extended period of time (in analogy with a hydrogen atom), forming what is referred to as an *exciton* pair. Such exciton pairs have a *binding energy* associated with them that is equal to the energy required to separate the electron and hole. As a result, electrons that are elevated from the valence band to one of these exciton states will absorb radiation at energies equal to the bandgap *less the binding energy* (the bandgap will appear to be red-shifted). More significantly however, the overlap integral (and hence the matrix element) of these two-particle states can be quite large. As a result, band-to-exciton transitions tend to dominate the absorption spectrum. However, exciton states are limited to states near $k = 0$, and hence band-to-exciton transitions are clustered at the band edge (or subband edge). The overall effect is the appearance of very strong absorption peaks near the subband edges in quantum-well material, and near the band edge in bulk material.

Exciton absorption peaks are clearly visible in quantum wells at room temperature, as seen in Fig. 4.14 for a typical GaAs QW. The first two steps in the "staircase" absorption spectrum predicted from the density of states (see Fig. 4.13) can be seen, with each step contributing about 10^4 cm^{-1} absorption. However, the exciton peaks riding on top of the steps, particularly the $n = 1$ peaks, dominate the absorption spectrum. Each observed exciton peak corresponds to one of the subband transitions illustrated earlier in Fig. 4.7. The allowed transitions ($n_c = n_v$) are

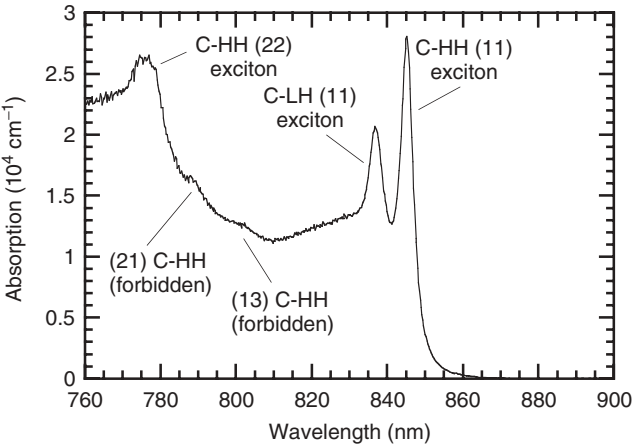


FIGURE 4.14: Absorption spectrum of a GaAs/$Al_{0.3}Ga_{0.7}As$ 100 Å QW at room temperature.

clearly dominant; however, traces of absorption can also be seen from forbidden transitions ($n_c \neq n_v$).

In quantum dot structures, excitonic recombination processes are exclusively observed because electron and hole wavefunctions are tightly confined within the same small volume. Besides simple two particle excitons, higher complexes such as trions, bi-excitons and multiexitons are found and can be observed in cathodoluminescence studies. Different exciton peaks from a single InAs quantum dot in the GaAs matrix are shown in Fig. 4.15.

At room temperature, exciton absorption peaks are not nearly as dramatic in bulk material. The reason is that reduced-dimensional structures confine and hold the electron and hole more closely together, producing a higher binding energy. However, the larger QW binding energy is still only on the order of a few meV. Hence exciton states even in QW material are somewhat "fragile," and collisions with phonons and carriers can easily break the exciton apart. For this reason, exciton absorption peaks are strongest and sharpest at low temperatures and low carrier densities. As the temperature and/or carrier density increases, the exciton lifetime diminishes and the exciton absorption peak broadens. Eventually at high enough temperatures and/or carrier densities, the exciton peak disappears altogether. At carrier densities required to achieve gain in the material, exciton states completely vanish, and the absorption/emission spectrum becomes dominated by band-to-band transitions. Thus, excitons have little effect on the gain spectrum of the material.

The second many-body effect occurs at high carrier densities, where the charges actually screen out the atomic attractive forces. With a weaker effective atomic potential, the single-atom electron wavefunctions of interest become less localized and the nearest-neighbor electron overlap becomes higher. From discussions in Appendix 1, the larger overlap increases the width of the energy bands (ΔE is larger in Fig. A1.7), reducing the gap between bands. Although this description

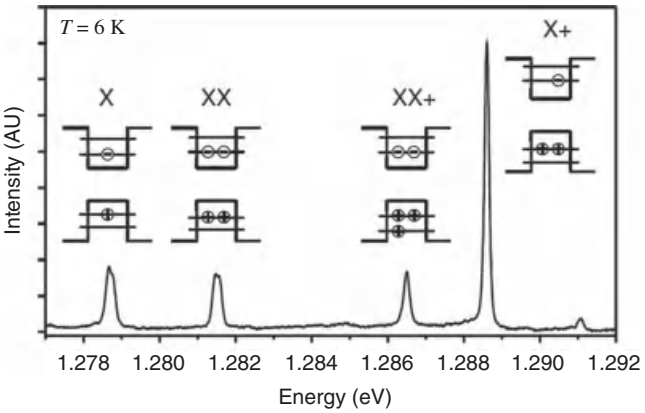


FIGURE 4.15: Cathodoluminescence spectra of a single-InAs quantum dot in a GaAs matrix. The emission lines are from different excitonic complexes: X, exciton; XX, biexciton; XX+, positively charged biexciton; X+, positively charged exciton. In all cases, an electron recombines with a hole, and the Coulomb interaction with additional charge carriers leads to a shift of the emission relative to the neutral exciton. *Reprinted with permission from Elsevier* [6].

is only qualitative, it does reveal that the bandgap should shrink with increasing carrier density.

It can also be argued theoretically that the *bandgap shrinkage* is inversely related to the average spacing between carriers, or $\Delta E_G \propto -1/r_s$ (the closer the carriers are, the more their own Coulomb potentials screen out the atomic potential). In bulk material, the average volume occupied by one carrier is inversely related to the carrier density, and hence $V \propto 1/N \propto r_s^3$. As a result, we conclude that $\Delta E_G \propto -N^{1/3}$ [5]. In a quantum well, the average area occupied by one carrier in the quantum-well plane is inversely related to the sheet density, or $A \propto 1/N \propto r_s^2$. So in a quantum well, we might expect that $\Delta E_G \propto -N^{1/2}$ [12]. However, taking into account the finite thickness of the quantum well in determining the electron Coulomb potential, the theoretical power dependence on carrier density has been estimated at closer to the one-third power law [17].

Experimental measurements of the power dependence of the bandgap shrinkage on carrier density yield numbers between 0.32 and 0.38 [5, 12, 17] for *both* bulk and quantum-well material. As for the absolute shift, at a density of 10^{18} cm^{-3} in bulk or 10^{12} cm^{-2} in a quantum well, the bandgap is reduced by anywhere from 22 to 32 meV, assuming $N = P$ (for measurements on p-doped material, the shift has been doubled assuming electrons would contribute equally to the shift in a laser where $N = P$). In light of the spread in measured data, it is common practice to simply assume the bandgap shrinks with the one-third power of carrier density in both bulk and quantum-well material, or

$$\Delta E_g = -cN^{1/3}, \tag{4.53}$$

where N can be either the two- or three-dimensional carrier density. A common value used for the bandgap shrinkage constant in bulk material (assuming $N = P$), which also falls within the measured range for quantum well material, is

$$c \approx 32 \text{ meV}/(10^{18} \text{ cm}^{-3})^{1/3}, \quad \text{(bulk GaAs)} \tag{4.54}$$

$$c \approx 32 \text{ meV}/(10^{12} \text{ cm}^{-2})^{1/3}. \quad \text{(GaAs/AlGaAs QW)} \tag{4.55}$$

$$c \approx 20 - 25 \text{ meV}/(10^{12} \text{ cm}^{-2})^{1/3}. \quad \text{(InGaAs/InP QW)} \tag{4.56}$$

$$c \approx 10 - 20 \text{ meV}/(10^{12} \text{ cm}^{-2})^{1/3}. \quad \text{(AlGaAs/GaAs QW)} \tag{4.57}$$

$$c \approx 2.1 \text{ meV}/(10^{8} \text{ cm}^{-2})^{1/3}. \quad \text{(GaN QW)} \tag{4.58}$$

Equations (4.54) to (4.58) are not entirely accurate because there is some experimental uncertainty in both the one-third power law and the value of c. However, they do provide a simple and reasonable estimate of the extent of bandgap shrinkage. Less data exist for other material systems, so it is common to assume the same values in the InGaAsP system as for the GaAs/AlGaAs system.

The net effect of bandgap shrinkage is that as carrier density increases, the entire gain spectrum redshifts by a noticeable amount. In principle, the shift is accompanied by a slight distortion (i.e., reshaping and enhancement) of the spectrum. However, to first order we can neglect the distortion and simply assume that high carrier densities produce a *rigid* shift of the entire gain spectrum to longer wavelengths. This phenomenon is observable in quantum-well lasers where the high threshold carrier density shifts the lasing wavelength beyond the known band edge wavelength of the quantum well. Bandgap shrinkage is a particularly important factor in situations where there is some critical alignment between a desired cavity mode of the laser and the gain spectrum (as in a short-cavity VCSEL, for example).

4.3.5 Polarization and Piezoelectricity

For certain crystal lattice types, with particular symmetry properties and the type of bonds between atoms, large built-in electric fields are possible along certain crystal directions. This phenomenon, called piezoelectricity, occurs in the Group III nitrides arranged in the wurzite type crystal structures. III nitride wurzite crystals have a unique structure where different atom species alternate between different atomic planes, thus forming fixed charge densities at the interfaces. Spontaneous polarization, caused by atomic bond assymetry between the nearest neighbors, occurs along the growth direction of a III nitride wurtzite crystal, dictated by the substrate (this growth direction is called c axis). In addition, strain induced in these types of crystals creates a separate piezoelectric polarization field along the same axis.

For electronic devices, this polarization is largely seen as beneficial, as it gives rise to 2-D sheet concentrations at interfaces an order of magnitude higher than in other III-V systems. For LEDs and lasers (an example of a GaN barrier and InGaN quantum-well structure was shown in Fig. 1.19), the fixed sheet charges at the

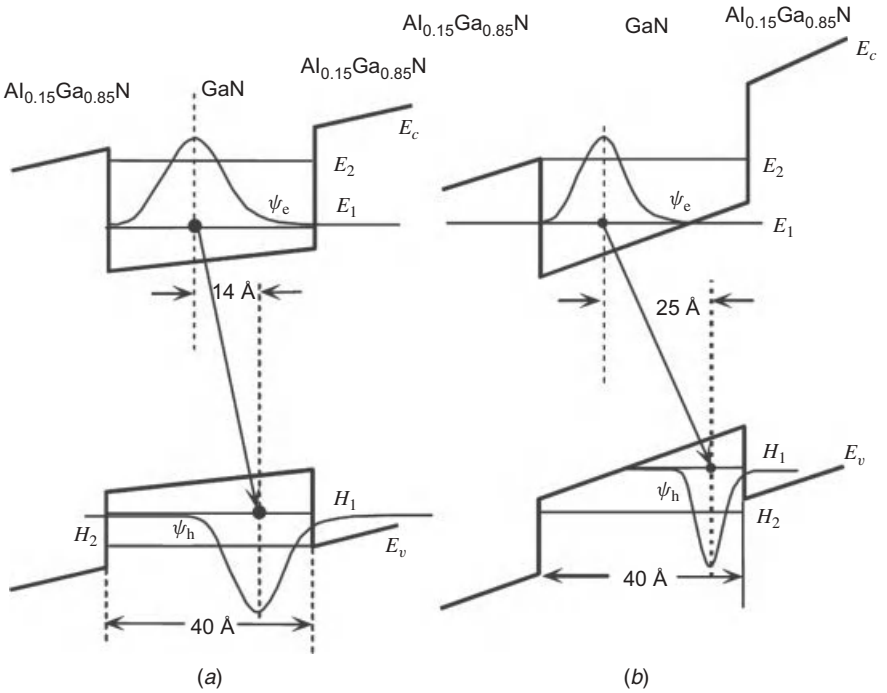


FIGURE 4.16: Schematic energy band diagrams of the GaN/$Al_xGa_{1-x}N$ MQW sample (a) with the piezoelectric field under the influence of screening by either photoexcited carriers or other free carriers in the system, and (b) the original piezoelectric field in the absence of carrier screening (or after long delay times in the case of time resolved luminescence experiments). *Reprinted by permission from Applied Physics Letters* [35].

heterointerfaces cause large electric fields normal to the QW plane (1–2 MV/cm). These internal polarization-related electric fields cause the bands to tilt and the bandgap to redshift through the quantum confined Stark effect (discussed in Chapter 8), causing spatial separation of electron and hole wave functions, as illustrated in Fig. 4.16. This spatial separation, in turn, reduces the overlap between the electron and hole wavefunctions, and thus probability of radiative recombination, reducing the efficiency in nitride LEDs and laser diodes. The energy band diagram of a GaN/AlGaN quantum-well sample, shown in Fig. 4.16 shows (a) the band diagram and the piezoelectric field under the influence of screening of other free carriers in the system (as in forward bias), and (b) the band diagram under the original piezoelectric field. The challenges associated with polar heterostructures in the nitrides can be significantly reduced by growth in semipolar orientations and avoided entirely through material growth in nonpolar orientations (m-plane). Both polar and nonpolar GaN heterostructures are an area of active research, in particular with regard to obtaining large area, high quality growth substrates that would yield economic LED and LD devices with higher performance.

4.4 SPONTANEOUS EMISSION

4.4.1 Single-Mode Spontaneous Emission Rate

As we inject a high enough carrier density to achieve gain in the material, we also inevitably increase the spontaneous emission rate. The spontaneous emission rate is important to consider because for every spontaneous photon emitted, a new carrier must be injected into the active region, as discussed in Chapter 2. In short-wavelength materials, this carrier recombination mechanism represents the largest component of the current we must inject into the active region.

To determine the spontaneous emission rate per unit active volume *into one optical mode*, R'_{sp}, we return to Eq. (4.5), which relates the single-mode spontaneous and stimulated emission rates:

$$R'_{sp} = \frac{|\mathscr{E}^{v-f}|^2}{|\mathscr{E}|^2} n_{sp} R_{st}. \tag{4.59}$$

The first term replaces the field strength with the vacuum-field strength of the mode, $|\mathscr{E}^{v-f}|^2$. The second term adjusts the Fermi factor. The last term, R_{st}, can be related to the gain through Eq. (4.33). The *population inversion factor*, n_{sp}, discussed in Appendix 6 is defined as

$$n_{sp} \equiv \frac{f_2(1-f_1)}{f_2 - f_1} = \frac{1}{1 - e^{(E_{21} - \Delta E_F)/kT}}. \tag{4.60}$$

Its value is typically between 1 and 2 at gain thresholds commonly encountered in lasers. Figure 4.17 displays the dependence of n_{sp} on gain for various active materials.

To evaluate $|\mathscr{E}^{v-f}|^2$, we turn to a quantum mechanical description of the optical mode. Without going into the details, it can be shown that an optical mode can be described quantum mechanically using the mathematical formalism developed for harmonic oscillators. One of the basic properties of harmonic oscillators is that the probability of elevating the state n to $n+1$ (via the creation operator) is proportional to $n+1$. In describing the optical mode, n refers to the number of photons in the mode. Thus, the probability of adding a new photon to the mode is proportional to the number of photons in the mode *plus one*—as if an imaginary photon were present in the mode. In terms of transition rates, it is the field strength of this imaginary photon that induces "spontaneous" downward transition events. Thus, the vacuum-field strength of the mode is equivalent to the field strength generated by one photon in the mode: $|\mathscr{E}^{v-f}|^2 = |\mathscr{E}_1|^2$. This conclusion is consistent with Einstein's approach, which establishes that the equivalent spectral density inducing spontaneous emission is equal to one photon per optical mode (see Appendix 6).

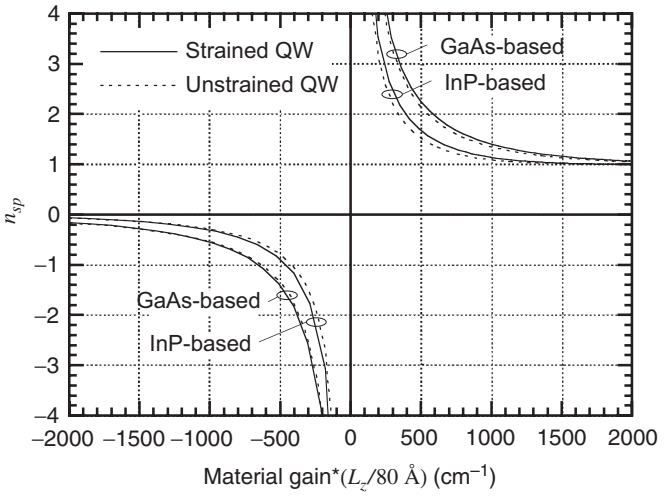


FIGURE 4.17: The theoretical population inversion factor versus gain for several relevant materials. GaAs based: strained InGaAs/GaAs 80 Å QW and unstrained GaAs/AlGaAs 80 Å QW. InP based: compressively strained InGaAs/InP 30 Å QW and unstrained InGaAs/InP 60 Å QW. See Section 4.6 for more details on these structures. In each case, n_{sp} and g are evaluated at a fixed energy slightly larger (11 meV) than the band gap.

We can express the classical field strength as: $|\mathscr{E}|^2 = N_p V_p |\mathscr{E}_1|^2$, where V_p is the mode volume (i.e., $N_p V_p$ is the number of photons in the mode). Thus, in Eq. (4.59) we can set

$$|\mathscr{E}^{v-f}|^2 = |\mathscr{E}_1|^2, \qquad (4.61)$$
$$|\mathscr{E}|^2 = N_p V_p |\mathscr{E}_1|^2.$$

Then using Eq. (4.34) to set $R_{st} = v_g g N_p$, Eq. (4.59) becomes

$$R'_{sp} = \frac{v_g g n_{sp}}{V_p} = \frac{\Gamma v_g g n_{sp}}{V}. \qquad (4.62)$$

Setting $1/V_p = \Gamma/V$ in the second equality, we conclude that the spontaneous emission rate into the mode is fundamentally related to the *modal* gain.

4.4.2 Total Spontaneous Emission Rate

To find the total amount of spontaneous emission occurring in the active region, we must sum the single-mode rate Eq. (4.44) over all optical modes. Let's denote the *total* spontaneous emission occurring within the energy range $dh\nu$, as $R_{sp}^{21}, dh\nu$, where R_{sp}^{21} represents the total spontaneous emission rate *per unit energy* per unit

active volume. Equating this with a sum over modes near $h\nu_{21}$ within $dh\nu$, we have

$$R_{sp}^{21}\, dh\nu = \sum_{\substack{\text{modes} \\ \text{in} dh\nu}} R'_{sp} = n_{sp} \sum_{\substack{\text{modes} \\ \text{in} dh\nu}} \frac{v_g g_{21}}{V_p}. \tag{4.63}$$

Because n_{sp} only depends on $h\nu$, it has been pulled out of the sum and evaluated at $h\nu_{21}$. However, the remaining three terms depend on the specifics of each mode (either through modal dispersion, polarization dependence, or mode volume). If we define average values for the three terms, the sum reduces to

$$\sum_{\text{modes}} \frac{v_g g_{21}}{V_p} = \frac{\bar{v}_g \bar{g}_{21}}{\bar{V}_p} N_{modes}, \tag{4.64}$$

where N_{modes} is the number of modes within $dh\nu$. Defining *individual* average values for each term is justified as long as the variations of each term in the sum are not correlated, and there is no reason to believe otherwise.

Considering that the average is over modes going in all directions, it is probably safe to say that the average group velocity, $\bar{v}_g$, is somewhere close to the material group velocity of the active region, v_g. The dependence of the quantum-well gain is through the polarization state of the mode (in bulk material there is no dependence and $\bar{g}_{21} = g_{21}$). Assuming the polarization states are isotropically distributed over the two in-plane TE and one perpendicular TM polarizations, the average material gain in a quantum well becomes

$$\bar{g}_{21} = \tfrac{1}{3}[2g_{21}^{TE} + g_{21}^{TM}]. \tag{4.65}$$

To handle $1/\bar{V}_p$ (or equivalently, the average confinement factor, $\bar{\Gamma}$), we need to evaluate N_{modes}. In large cavities, N_{modes} is most easily found using the mode density concept. If we assume the cavity is a large rectangular metal box of volume, V_{box}, the density of optical modes per unit frequency per unit volume can be derived. The procedure is outlined in Appendix 4, and the result is given by Eq. (A4.5). We repeat it here for reference:

$$\rho_0(\nu)\, d\nu = \frac{8\pi}{c^3} n^2 n_g \nu^2\, d\nu. \tag{4.66}$$

The density of modes per unit *energy* is equal to $\rho_0(\nu) d\nu / dE = \rho_0(\nu)/h$. With this definition, the total number of modes within $dh\nu$ becomes

$$N_{modes} = \rho_0(\nu)/h \cdot V_{box}\, dh\nu. \tag{4.67}$$

Inserting this into Eq. (4.64), we are left with the volume ratio $V_{box}/\bar{V}_p$. If the laser cavity were a large metal box, then the mode volume would be equivalent to the

volume of the box (aside from standing wave effects, which should average out over the spontaneous emission bandwidth), and we could simply set

$$V_{box}/\bar{V}_p = 1. \tag{4.68}$$

Of course, laser cavities are typically much more complex than a simple metal box, and V_p is potentially different for every optical mode. Thus, there is no guarantee that Eq. (4.50) holds in real laser cavities, or that we can even define a mode density as given in Eq. (4.66).

Fortunately, if the cavity is much larger than the wavelength of bandgap radiation in the active region, or more specifically if $V_{cav} \gg \lambda^3$, it turns out that a more rigorous treatment usually averages out to the simple metal box treatment (especially when the emission bandwidth is large relative to resonances in the cavity). To get a feel for the numbers, a typical $0.2 \times 4 \times 200$ μm^3 GaAs in-plane laser has $V_{cav} \sim 10,000\lambda^3$. Thus, the simple metal box assumption contained in Eqs. (4.66) through (4.68) is expected to work well in this case.

In a VCSEL, the simple box assumption is more questionable. However, the volume of a typical GaAs VCSEL with dimensions of $1 \times 10 \times 10$ μm^3, is smaller than a typical in-plane laser by only a factor of two, and $V_{cav} \sim 5000\lambda^3$. Thus, for VCSELs of this size, the simple box assumption should still hold. However, when the lateral dimensions of the VCSEL are reduced below 1 μm, significant deviations from the simple box assumption are expected to occur. Under these circumstances, the mode density concept must be abandoned and more sophisticated mode-counting techniques must be used to estimate the total spontaneous emission rate. In addition to mode counting, the average value for $1/\bar{V}_p$ (or equivalently $\bar{\Gamma}$) is also required. Numerous researchers interested in microcavity lasers have investigated such numerical exercises for a number of cavity geometries [18, 19].

Combining Eqs. (4.63) through (4.68) assuming the simple box assumption holds, the total spontaneous emission rate per unit energy per unit active volume (in units of s^{-1} cm^{-3} eV^{-1}) becomes

$$R_{sp}^{21} = \frac{1}{h}\rho_0(\nu_{21}) \cdot v_g n_{sp} \bar{g}_{21}. \tag{4.69}$$

This result is identical to Eq. (A6.32) derived in Appendix 6, with the exception that the gain here is more correctly defined as an average over all polarizations. Using the explicit expression for gain Eq. (4.33) and mode density Eq. (4.48), the general expression for spontaneous emission becomes

$$R_{sp}^{21} = \frac{4n\pi q^2}{\varepsilon_0 h^2 c^3 m_0^2} h\nu_{21} |\bar{M}_T(E_{21})|^2 \rho_r(E_{21}) \cdot f_2(1 - f_1), \tag{4.70}$$

where

$$|\bar{M}_T(E_{21})|^2 = \frac{1}{3} \sum_{\substack{\text{all three} \\ \text{polarizations}}} |M_T(E_{21})|^2.$$

The transition matrix element is the only factor dependent on polarization in the expression for gain. This is why the averaging over polarizations only needs to include $|M_T|^2$.

The actual spontaneous emission spectrum will be affected by lineshape broadening in the same way the gain is affected. In analogy with Eq. (4.46), the spontaneous emission spectrum taking lineshape broadening into account is related to R_{sp}^{21} through the following:

$$R_{sp}^{h\nu}(h\nu) = \int R_{sp}^{21} \mathscr{L}(h\nu - E_{21}) dE_{21}. \tag{4.71}$$

The spectrum is generally peaked just above the bandgap energy because the electrons and holes are concentrated at the band edges. The spectrum gradually decays to zero at higher energies. To determine the total emission rate, we must integrate over all photon energies. However, in practice integrating over a limited range near the bandgap energy is usually sufficient to account for all of the spontaneous emission. It is shown in Appendix 6 that the integration over $R_{sp}^{h\nu}(h\nu)$ is essentially the same as integrating over the simpler R_{sp}^{21}. Thus, the total band-to-band spontaneous emission rate is given by

$$R_{sp} = \int R_{sp}^{21}\, dE_{21} = \eta_i \eta_r \frac{I}{qV}. \tag{4.72}$$

The second equality allows us to determine the radiative component of the current (excluding stimulated emission) required by the active region to obtain a given gain. Example calculations in different material systems will be given later in this chapter.

Example 4.3 For the ridge lasers from Example 4.2, we would like to know the ratio of the spontaneous emission power to the output power at 40 mA bias. Assume that the value of the population inversion factor $n_{sp} = 1.5$.

Problem: (1) Calculate the spontaneous emission power into the mode above threshold. (2) Calculate the optical mode density for this laser. (3) What is the total spontaneous emission power within a 1 nm bandwidth near the lasing wavelength?

Solution: The spontaneous emission power into a mode can be calculated from the known spontaneous emission rate, active region confinement factor Γ, and photon energy. The spontaneous emission rate into a mode is given by Eq. (4.62). From here, we have

$$\begin{aligned} P_{sp} &= R'_{sp} \cdot (h\nu) \cdot V = \frac{\Gamma g_{th} v_g n_{sp}}{V} \cdot (h\nu) \cdot V \\ &= (29.8\ \text{cm}^{-1}) \left(\frac{3}{4.5} \cdot 10^{10} \frac{\text{cm}}{\text{s}} \right) (1.5) \left(\frac{1.24\ \text{eV}}{0.97} \right) (1.6 \cdot 10^{-19}) \frac{\text{J}}{\text{eV}} = 61.0\ \text{nW}. \end{aligned}$$

To determine the total spontaneous emission power in a given wavelength range, we need know the total spontaneous emission rate over all modes that exist in that wavelength range, as well as the photon energy and the active region volume.

The total spontaneous emission per unit energy and per unit volume is given by Eq. (4.69) as $R_{sp}^{21} = \frac{1}{h}\rho_0(\nu_{21}) \cdot v_g n_{sp} \bar{g}_{21}$. To compute it, we need to compute the mode density ρ_0 and the average gain $\bar{g}_{21}$. From Appendix 4, the expression for the density of optical modes is given by Eq. (A4.5),

$$\rho_0(\nu) = \frac{8\pi}{c^3} n^2 n_g \nu^2 \Rightarrow \frac{1}{h}\rho_0(\nu) = \frac{8\pi}{(3 \cdot 10^{10}\ \text{cm/s})^3} \frac{(3.6)^2 (4.5)\left(\frac{1.24\ \text{eV}\mu\text{m}}{0.97\ \mu\text{m}}\right)^2}{\left(\frac{6.626 \cdot 10^{-34}\text{Js}}{1.6 \cdot 10^{-19}\text{J/eV}}\right)^3}$$

$$= 1.25 \cdot 10^{15}\ \text{eV}^{-1}\ \text{cm}^{-3}.$$

The average gain is given by

$$\bar{g}_{21} = \frac{1}{3}(2g_{21}^{TE} + g_{21}^{TM}) = \frac{2}{3}g_{21}^{TE} = \frac{2}{3}1986.67\ \text{cm}^{-1} = 1324\ \text{cm}^{-1},$$

because the material has been optimized for TE polarization gain only, as shown in Figure 4.4. Now, we can calculate the total spontaneous emission rate per unit volume as

$$R_{sp}^{21} = \left(\frac{1}{h}\rho_0(\nu_{21})\right) \cdot (v_g)(n_{sp})(\bar{g}_{21})$$

$$= (1.25 \cdot 10^{15}\ \text{eV}^{-1}\ \text{cm}^{-3})\left(\frac{3}{4.5} \cdot 10^{10}\frac{\text{cm}}{\text{s}}\right)(1.5)(1324\ \text{cm}^{-1})$$

$$= 1.65 \cdot 10^{28}\ \text{eV}^{-1}\ \text{cm}^{-3}\text{s}^{-1}.$$

To calculate the total power in the 1-nm bandwidth, we need to convert the bandwidth into energy and calculate the active region volume,

$$\Delta E = \Delta\lambda \frac{dE}{d\lambda} = \Delta\lambda\left(-\frac{E_p}{\lambda_p}\right) = 1\ \text{nm}\left(-\frac{1.24\ \text{eV}\mu\text{m}}{(0.97\ \mu\text{m})^2}\right) = 1.32 \cdot 10^{-3}\ \text{eV}$$

$$V_a = w \cdot d \cdot L = (3\ \mu\text{m})(8\ \text{nm})(500\ \mu\text{m}) = 12000 \cdot 10^{-15}\ \text{cm}^3.$$

Finally, the total spontaneous power emitted in the 1-nm wavelength range around the lasing wavelength is

$$P_{sptotal} = R_{sp}^{21} \cdot \Delta E \cdot h\nu \cdot V_a$$

$$= (1.65 \cdot 10^{28}\ \text{eV}^{-1}\ \text{cm}^{-3}\text{s}^{-1})(1.32 \cdot 10^{-3}\ \text{eV})\left(\frac{1.24\ \text{eV}\mu\text{m}}{0.97\ \mu\text{m}}\right)$$

$$\times (12000 \cdot 10^{-15}\ \text{cm}^3) = 41.87\ \mu\text{W}.$$

4.4.3 Spontaneous Emission Factor

With the preceding description of spontaneous emission, we can derive a simple expression for the spontaneous emission factor used in the rate equations to express the fraction of total spontaneous emission that enters the mode of interest. Using Eq. (4.62) for the emission rate into one mode and Eq. (4.72) for the total emission rate, we have

$$\beta_{sp} \equiv \frac{R'_{sp}}{R_{sp}} = \frac{\Gamma v_g g n_{sp}}{\eta_i \eta_r I/q}, \tag{4.73}$$

where I as used here does not include stimulated emission current.

With Eq. (4.73), parameters readily accessible experimentally can be used to estimate β_{sp}. For example, assume we have a laser with a threshold current of 10 mA, a modal threshold gain of 50 cm^{-1}, and an internal efficiency of 75%. If nonradiative recombination is minimal, then $\eta_r \approx 1$. If we assume the group index is ~4 and the population inversion factor is ~1.5, we quickly find $\beta_{sp} \approx 1.2 \times 10^{-5}$, typical of experimental values measured with in-plane lasers. To get a better estimate of β_{sp}, we need to determine $c/v_g, n_{sp}$, and η_r more accurately. However, c/v_g is almost always in the range of 4–5, n_{sp} is usually between 1.25 and 1.75 for gains commonly required in lasers, and η_r is typically between 50% and 80%, implying that simple estimates of $c/v_g, \eta_{sp}$, and η_r will get us within a factor of 2 of the correct value of β_{sp}.

In a typical VCSEL, the percent gain per pass is in the range of 0.5%–1% and the cavity length is ~1 μm. This gives a modal gain of 50–100 cm^{-1}. *Radiative* threshold currents are in the range of 0.5–2.0 mA for 10 μm diameter devices. If we assume a 1 mA radiative threshold current and 100 cm^{-1} threshold modal gain, we find $\beta_{sp} \approx 1.8 \times 10^{-4}$ (assuming again that c/v_g and n_{sp} are equal to 4 and 1.5). This is about an order of magnitude higher than β_{sp} observed in typical in-plane lasers. The main difference lies in the reduction of the threshold current, which is in turn predominantly due to the reduction in active-region volume possible with VCSEL structures.

Equation (4.73) also reveals that β_{sp} is not a constant, but is dependent on the injection level. As the injection level increases, the spontaneous emission into one mode saturates at a maximum value just as the gain saturates at g_{max}. Meanwhile the current and total spontaneous emission rate continue to increase as the spectrum broadens to include more modes in the emission process. The net effect is that β_{sp} decreases with increasing injection level. In rate equation analyses, β_{sp} is often approximated as a constant (see Chapters 2 and 5). This is justified for near- and above-threshold analyses, as long as the value assumed for β_{sp} is the actual value that would exist near the threshold injection level.

4.4.4 Purcell Effect

As the cavity dimensions become small, the optical mode density departs from its simple bulk form, just as in the case of quantum wells and dots for the electronic

density of states. As spontaneous emission can only be made into allowed modes, which are now limited for a very small cavity, the spontaneous emission factor into each of these increases accordingly. Thus, the rate of spontaneous recombination can be modified by placing the active material inside a resonant cavity. This effect was first described by Purcell in 1946. For this effect to be noticeable, the linewidth of a spontaneous transition between two states has to be equal or smaller in value to the linewidth of a resonant cavity mode. In this case, the spontaneous emission derived from a bulk photon mode density will be enhanced by the Purcell enhancement factor,

$$F_p = \frac{3}{4\pi^2}\left(\frac{\lambda_c}{n}\right)^3 \frac{Q}{V},$$

where Q is the cavity quality factor, defined as the ratio of the stored and dissipated energy in an optical cavity, V is the volume of the cavity, and λ_c/n is the wavelength in the medium. The magnitude of this effect will be dominated by the quality factor Q and volume V—the smaller the volume, and the stronger the cavity, the larger the effect. The factor can be both less and higher than unity. The Purcell effect has been experimentally studied using single quantum dot type structures. It is very important in the work with microcavities, which can be used for quantum encryption and optical logic elements, to name a couple applications.

4.5 NONRADIATIVE TRANSITIONS

With radiative processes defined, we now need to consider nonradiative transitions to determine their relative importance in the overall carrier recombination process. Three common types of nonradiative transitions are depicted in Fig. 4.18. These processes were briefly discussed in Chapter 1 in reference to Fig. 1.6. In the following sections we will provide a more detailed description of each process.

4.5.1 Defect and Impurity Recombination

The first type of nonradiative transition appearing in Fig. 4.18 depicts an energy level in the middle of the gap, which serves to trap an electron from the conduction

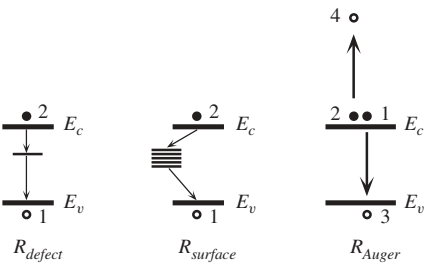


FIGURE 4.18: Various types of nonradiative recombination paths.

band temporarily before releasing it to the valence band. Energy levels of this sort can arise from a variety of causes. Defects in the lattice structure are one source. For example, a void at an atomic site or an extra atom lodged in between the lattice structure can produce insufficient or extra orbitals, which leads to a mismatch in the covalent bonding pattern. Such localized *dangling bonds* give rise to discrete energy levels, which can possibly appear in the middle of the bandgap. Another common source of midgap energy levels are impurities. Just as dopants create impurity levels, near the band edge, other types of atoms can create impurity levels, which are closer to the middle of the gap. Oxygen, for example, is a particularly insidious impurity because it is an abundant element and is known to create large recombination rates in aluminum-containing compounds.

The defect or impurity recombination rate, R_{defect}, or R_d for short, was first analyzed four decades ago in a classic paper by Shockley and Read [20]. Hall simultaneously arrived at a similar result [21]. Thus it is known as the Shockley–Read–Hall recombination theory. The theory begins by writing down the four possible transition rates into and out of the trap (up and down from the conduction band, and up and down to the valence band). Setting the rates equal in thermal equilibrium to determine relationships between the up and down rates (similar to the derivation of radiative transition rates in Appendix 6), an expression for the nonequilibrium recombination rate can be derived. Within the Boltzmann nondegenerate carrier density regime discussed in Appendix 2, the defect recombination rate takes the form

$$R_d = \frac{NP - N_i^2}{(N^* + N)\tau_h + (P^* + P)\tau_e}, \tag{4.74}$$

where N_i is the intrinsic carrier concentration, τ_e is the time required to capture an electron from the conduction band assuming all traps are empty, and τ_h is the time required to capture a hole from the valence band assuming all traps are full. As might be expected, the capture rate is proportional to the density of traps, $1/\tau_{e,h} \propto N_t$, such that the higher the trap density, the shorter the capture times. N^* and P^* appearing in the denominator are the electron and hole densities that would exist if the Fermi level were aligned with the energy level of the trap. The important point here is that if the trap level is close to either band edge, then either N^* or P^* will become large, substantially reducing the recombination rate. Thus, the most effective recombination centers are those with energy levels close to the middle of the gap, so-called deep-level traps. Great care must be taken to avoid introducing such deep-level impurities (such as oxygen) into the crystal lattice.

It is apparent that Eq. (4.74) has a nontrivial dependence on both electron and hole densities. However, for laser applications we are primarily interested in the high-level injection regime where $P = N \gg N_i, N^*, P^*$. Under these conditions, Eq. (4.74) simplifies to

$$R_d = \frac{N}{\tau_h + \tau_e}. \quad \text{(high-level injection)} \tag{4.75}$$

Thus, defect or impurity recombination follows a linear dependence on carrier density in the active region of lasers, and we can define the defect component of the linear recombination constant as: $A_d = 1/(\tau_h + \tau_e)$, using the notation of Chapter 2.

In the low-level injection limit, the form of the recombination rate changes. If we have deep-level traps, then N^* and P^* are negligible ($N^* \approx P^* \approx N_i$), and they can be removed from Eq. (4.74). With $N = N_0 + \delta N$ and $P = P_0 + \delta N$ (assuming equal numbers of excess electrons and holes) and assuming δN is small, the recombination rate becomes

$$R_d = \delta N \frac{N_0 + P_0}{N_0 \tau_h + P_0 \tau_e}. \quad \text{(low-level injection)} \tag{4.76}$$

The recombination rate is again linear in the excess carrier density; however, the rate depends on whether the material is doped n-type or p-type. For n-type material, $A_d = 1/\tau_h$ (it is limited by the capture of holes), and for p-type material $A_d = 1/\tau_e$ (it is limited by the capture of electrons). Thus, while the defect or impurity recombination rate is linear with the excess carrier density at either low or high injection levels, the lifetime does increase from either τ_h or τ_e to $\tau_h + \tau_e$ as the carrier density is increased.

Recombination via defects and impurities is primarily a problem for excess minority carriers or carriers injected into a region under nonequilibrium conditions. Majority carrier current flow in heavily doped materials is relatively unaffected by defects and impurities (aside from the possibility that defects and impurities can affect the ionization of doping species, which can affect the doping efficiency). In other words, material that contains a very high density of defects or impurities may not be a good choice for a laser's active region, but it still may carry majority carrier current just fine if doped heavily enough.

With modern MBE and MOCVD growth technologies, the crystal quality of semiconductor devices is such that the defect density and the density of impurity atoms are at most 10^{16} cm^{-3}, and are more often below 10^{15} cm^{-3}. With such low trap densities, the defect and impurity recombination rates are negligible in typical laser applications. However, there are instances where such recombination can become large and problematic.

In the early stages of MBE growth technology (in the late 1970s), it was a common belief that growth temperatures should be kept below 580°C to retain good surface morphology. However, a puzzling concern among researchers was that GaAs/AlGaAs lasers grown by MBE invariably had much higher threshold current densities than equivalent structures grown by LPE. Finally, Tsang et al. [22] found that increasing the growth temperature to 650°C led to dramatic reductions in the threshold current density, lower than the best LPE material. Their explanation was that impurities (perhaps oxygen) were being incorporated into the AlGaAs cladding layers at densities high enough to significantly increase the threshold current density. At the higher growth temperatures, the incorporation of impurities was minimized, and dramatic improvements in the threshold current were observed. More recently, the purity of aluminum sources used in MBE systems has significantly improved, allowing the growth temperature to be reduced back

down to 600–620°C. However, unusually high threshold current densities in just-grown lasers, which normally would yield low thresholds, can often indicate that the MBE or MOCVD machine is introducing unwanted impurities, either through a leak in the system or corrupted material sources.

An example of where defect recombination is important is in the aging of lasers. As lasers are used, they inevitably go through dramatic thermal variations. Such repeated thermal cycles can cause stress to the crystal. This stress can cause small defects within the crystal lattice to spread and grow large, just as a small crack in the windshield of your car can propagate across your entire field of view. A particularly interesting example of this is the emergence of dark-line defects, which tend to appear after lasers have been in use for thousands of hours. Such defects show up as dark lines when viewed from the surface due to the absence of carriers and hence spontaneous emission in regions where the "cracks" propagate. If the original quality of the crystal is high, such dark-line defects can be minimized. However, if they do become large with age, the threshold current of the laser will suffer over time, and eventually the laser will die.

Another common example of where defect recombination can become important is in the area of strained-layer research. This research is interested in growing epitaxial layers of materials that have a different "native" lattice constant than the substrate material. In a lattice-mismatched growth, the epitaxial layer will attempt to deform to the substrate lattice structure. However, as the layer becomes thicker, atomic forces building up within the epitaxial lattice structure will at some critical thickness break discontinuously with the deformed lattice and begin to force the lattice back to its native form. Strained layers with thickness below the critical thickness can be of very high quality (and are indeed superior in many respects to their unstrainted counterparts). However, strained layers with thicknesses much greater than the critical thickness will inevitably be laden with severe lattice defects.

As the strained layer is grown thicker still, the lattice defects unfortunately tend to propagate along the growth direction. As a result, thick strained layers have very high defect recombination rates and are unusable for any type of carrier injection applications such as lasers. However, techniques do exist that can block defect propagation, allowing the upper portion of a thick epitaxial layer to provide a high-quality lattice structure with low defect densities and hence low defect recombination rates. For example, good-quality epitaxially grown GaAs/AlGaAs lasers have been successfully grown on silicon substrates using such techniques, even though the lattice constants of the two semiconductors are very different. Furthermore, high doping of the defect-laden transition region can allow high currents to pass across the interface, allowing for the integration and interconnection of silicon circuits with GaAs lasers.

4.5.2 Surface and Interface Recombination

The second type of nonradiative transition in Fig. 4.18 depicts electrons recombining via *surface states* of the crystal. The two-step recombination mechanism is analogous to the defect and impurity recombination mechanism. However, in this

case the number of traps is characterized by a *two-dimensional* sheet density at an exposed surface of the crystal or at an interface between two materials. These surface states primarily arise from the termination of the lattice, which inevitably leaves a few unmatched bonds on one side of every exposed unit cell. Such *dangling bonds* occur in very high densities, forming a miniband as opposed to individual energy levels, as depicted in the figure. Surface recombination is most damaging when the exposed surface-to-volume ratio is large—in other words, when the device size is reduced. Interface recombination is damaging when the interface quality is poor. With modern growth technologies, interface recombination is minimal in common materials, but can be very detrimental in more experimental materials research efforts. In addition, devices that make use of regrowth technology (i.e., devices that are put back into the growth chamber after etching or some other processing step is performed) can suffer from poor interfaces and hence high interface recombination.

The description of surface recombination is accomplished via the Shockley–Read–Hall theory in analogous manner to defect and impurity recombination theory. However, instead of defining a capture *time*, τ, of carriers distributed throughout a volume of material, we define a capture *rate* of carriers located within some capture length of the surface: L_c/τ. A larger capture length allows surface states to capture more carriers per unit time and hence leads to a higher capture rate. Because of the units, this capture rate is referred to as a velocity. The association with velocity does have a physical significance. For example, if the capture velocity of the trap is larger than the average thermal velocity of the carriers, then the capture rate will be limited by the thermal velocity. This maximum arrival rate of carriers at the surface sets an upper limit on the capture velocity of somewhere near 10^7 cm/s. In most semiconductors, the capture velocity is at least an order of magnitude smaller than this.

Redefining the electron and hole capture times as capture *velocities* (i.e., setting $\tau_e \to 1/v_e$ and $\tau_h \to 1/v_h$), the analogous version of Eq. (4.74), neglecting N^* and P^*, for surface recombination becomes[6]

$$R_{sr} = \frac{a_s}{V} \cdot \frac{NP - N_i^2}{N/v_h + P/v_e}. \tag{4.77}$$

The first term effectively distributes the exposed surface area, a_s, over the volume of the active region, V, because we have defined R_{sr} as the rate per unit active volume, not rate per unit surface area. This geometrical factor is shown in Fig. 4.19 for two common laser geometries. As can be seen, a_s/V makes surface recombination important when the pillar diameter or stripe width is small.

Under high-level injection in the active region, $P = N \gg N_i$, and Eq. (4.77) reduces to

$$R_{sr} = \frac{a_s}{V} v_s N \quad \text{(high-level injection)} \tag{4.78}$$

[6]A more rigorous version of Eq. (4.77) would integrate over all surface states within the miniband, with possibly energy-dependent capture velocities. However, for our purposes, we will assume a discrete energy level for the surface states.

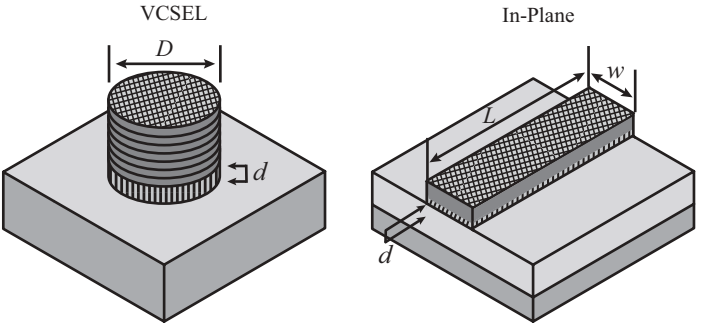


FIGURE 4.19: Exposed surface area and surface-to-volume ratio in both pillar-type *VCSEL* and in-plane ridge laser geometries.

with

$$\frac{1}{v_s} = \frac{1}{v_h} + \frac{1}{v_e}. \tag{4.79}$$

The linear relationship to the carrier density allows us to define the surface recombination component of the linear recombination constant as $A_{sr} = (a_s/V)v_s$. The *surface recombination velocity*, v_s, which controls the surface recombination rate, is seen to be an average of the individual electron and hole capture velocities. The use of inverse velocities in Eq. (4.61) reflects the fact that if, for example, the electron capture velocity is very high, then surface recombination will be limited by the capture rate of holes and $v_s \approx v_h$.

Using Eq. (4.60), the surface recombination current and current density in the laser can be written as

$$I_{sr} = qa_s v_s N \quad \text{and} \quad J_{sr} = q v_s N \left(\frac{a_s}{V} d\right). \tag{4.80}$$

To estimate I_{sr}, it is important to realize that we need to know the carrier density *which exists at the surface*. As a first-order, upper-bound estimate of I_{sr} or J_{sr}, we can naively use the carrier density in the center of the active region (which can be roughly estimated from the threshold gain, if we know the gain as a function carrier density). However, in reality the heavy recombination at the surface will deplete the surface carrier density to some level that balances the recombination rate with the gradient-driven lateral diffusion current directed from the center to the surface of the active region (see Problem 6 for more details). To determine this carrier density requires solving the carrier diffusion equation.[7]

[7]Such a procedure can take on various degrees of complexity. The simplest approach is to assume an ambipolar diffusion coefficient representing the effective diffusion of both electrons and holes (which

Under low-level injection conditions, the surface recombination rate changes just as Eqs. (4.57) and (4.58) are different. For laser applications we are not particularly interested in this case. However, experimental measurements of the surface recombination velocity are often made under such conditions. Using Eq. (4.78) to define v_s, we find that Equation 4.59 evaluated under small carrier density perturbations away from the equilibrium values, N_0 and P_0, gives

$$v_s \equiv \frac{V}{a_s} \cdot \frac{R_{sr}}{\delta N} = \frac{N_0 + P_0}{N_0/v_h + P_0/v_e}. \quad \text{(low-level injection)} \tag{4.81}$$

If the surface recombination velocity is measured under low-level injection conditions, then the measured value will be dependent on the doping present. For n-type material, $v_s = v_h$ (it is limited by the capture rate of holes), and for p-type material, $v_s = v_e$ (it is limited by the capture rate of electrons). Eq. (4.81) also assumes that $\delta N = \delta P$, which may not necessarily be true, yielding still different values for v_s. The most reliable values for v_s relevant for laser applications must be made under strong high-level injection conditions. For example, measuring the threshold current dependence on the laser geometry is the most direct method. However, geometry-dependent optical losses, which change the threshold gain, can skew these results as well.

Surface recombination tends to be much more of a problem in the short-wavelength GaAs system than in the long-wavelength InGaAsP system. This is because v_s is two orders of magnitude larger in GaAs. Values measured under low-level injection conditions on GaAs and GaAs/AlGaAs quantum wells give [24]

$$v_s \approx 4 - 6 \times 10^5 \text{ cm/s}. \quad \text{(GaAs, bulk and QW)} \tag{4.82}$$

In n-type GaAs, values closer to $2{-}3 \times 10^6$ cm/s have been reported [25], suggesting that v_h is greater than v_e by close to an order of magnitude. Measurements on strained InGaAs/GaAs quantum-well lasers show a reasonable improvement over GaAs [26]:

$$v_s \approx 1 - 2 \times 10^5 \text{ cm/s}. \quad \text{(InGaAs/GaAs, QW)} \tag{4.83}$$

In InGaAsP materials, surface recombination tends to be lower. Measured values on InP give [27]

$$v_s \leq 10^4 \text{ cm/s}. \quad \text{(InP, bulk)} \tag{4.84}$$

within the model sets the carrier densities equal). The recombination then follows Eq. (4.60) using the surface recombination velocity given in Eq. (4.61), and a single carrier diffusion equation can be used. A more extensive model would allow the electron and hole densities to be different, each with their own diffusion constant. The recombination rate at the surface for both electrons and holes would then take on the form Eq. (4.59), where the individual electron and hole capture velocities must be known or fitted. The solution would then have to satisfy the drift-diffusion equations for electrons and holes. To be complete, the drift-diffusion equations must then be coupled with Poisson's equation taking into account the charge distribution of the surface states [23].

For quaternary InGaAsP, such low values are also expected. For GaN, a measured value is

$$v_s = 5 \times 10^4 \text{ cm/s.} \quad \text{(GaN, bulk)} \tag{4.85}$$

Example 4.4 Etched air post VCSELs have been fabricated in two different materials: one on a GaAs substrate lasing at 850 nm, the other on InP lasing at 1550 nm. Both have diameters of 3 μm. The GaAs device's active region consists of 3–8 nm wide quantum wells; the InP device has an active region consisting of 5–6 nm unstrained quantum wells—both characterized by the parameters in the Tables 4.4 and 4.5. The etch has proceeded through the active regions of both devices, resulting in surface recombination velocities of $5 \cdot 10^5$ and $5 \cdot 10^3$cm/s, respectively, for GaAs and InP devices. VCSEL mirrors are constructed in such a way that the one-pass threshold gain of the active regions in both cases is 1%, including the enhancement factor of 1.8.

Problem: What are the surface recombination currents for the two cases assuming uniform carrier distributions?

Solution: From the definition of the one pass threshold gain, we can calculate the material threshold gain,

$$G_{th} = 0.01 = \Gamma \cdot g_{th} \cdot L = \Gamma_{xy} \cdot \Gamma_{enh} \cdot g_{th} \cdot L_a.$$

For the GaAs VCSEL, we have

$$g_{th} = \frac{G_{th}}{\Gamma_{xy}\Gamma_{enh} \cdot L_a} = \frac{0.01}{(1)(1.8)(24 \cdot 10^{-7} \text{ cm})} = 2315 \text{ cm}^{-1}.$$

Now, using the relationship between material gain and carrier density, from Table 4.4, we have that

$$N_{\text{GaAs}} = N_{tr} \cdot e^{\frac{g_{th}}{g_0}} = 2.6 \cdot 10^{-18} \text{ cm}^{-3} \cdot e^{\frac{2315}{2400}} = 6.82 \cdot 10^{18} \text{ cm}^{-3}.$$

Finally, from the definition of the surface recombination current, we obtain

$$\begin{aligned} I_{SR} &= (1.6{\cdot}10^{-19}\text{C})(24{\cdot}\pi{\cdot}3{\cdot}10^{-7}{\cdot}10^{-4} \text{ cm}^2) \\ &\quad \times (5{\cdot}10^5 \text{ cm/s})(6.82{\cdot}10^{18} \text{ cm}^{-3}) = 1.23 \text{ mA}. \end{aligned}$$

Similarly, for InP, we get

$$g_{th} = 1852 \text{ cm}^{-1}; \quad N_{InP} = 6.15 \cdot 10^{18} \text{ cm}^{-3};$$

and

$$\begin{aligned} I_{SR} &= (1.6{\cdot}10^{-19} \text{ C})(30{\cdot}\pi{\cdot}3{\cdot}10^{-7}{\cdot}10^{-4} \text{ cm}^2) \\ &\quad \times (5{\cdot}10^3 \text{ cm/s})(6.15{\cdot}10^{18} \text{ cm}^{-3}) = 13.9 \text{ μA}. \end{aligned}$$

In GaAs/AlGaAs and InGaAs/GaAs lasers, surface recombination can be quite severe for small devices. To get a feeling for how small, Fig. 4.20 plots the threshold current density of a typical InGaAs QW VCSEL and an in-plane laser for different values of v_s (using the gain as a function of carrier density and current density given later in the chapter). At $v_s = 1 \times 10^5$ cm/s, the VCSEL threshold current density doubles from ~500 A/cm^2 to ~1000 A/cm^2 when the pillar diameter is reduced to 10 μm. This characteristic is common for InGaAs QW VCSELs. If we were to use GaAs QWs with v_s perhaps equal to 5×10^5 cm/s, the threshold current density would double at 50 μm! In long wavelength materials, $v_s \leq 10^4$ cm/s and pillar diameters could be reduced to 1 μm before the threshold current density would double. Similarly for the in-plane laser, problems become severe for stripe widths less than 10 μm using InGaAs QWs (with $v_s = 1 \times 10^5$ cm/s). However, it is important to realize that these estimates assume the carrier profile is flat across the active region. A more realistic carrier profile will produce threshold current

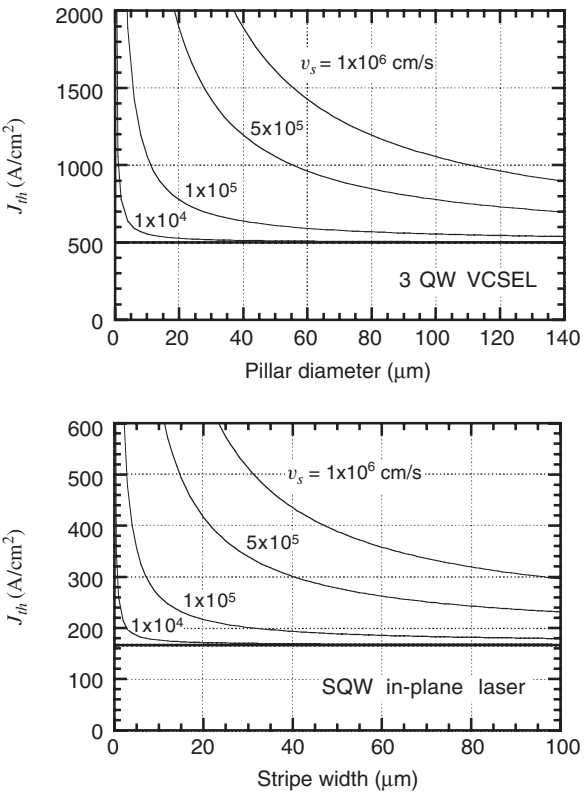


FIGURE 4.20: Threshold current density versus lateral device dimensions for different surface recombination velocities (assuming $D_{np} = \infty$). Both lasers use $In_{0.2}Ga_{0.8}As$/GaAs 80 Å QWs and have a 1500 cm^{-1} threshold *material* gain (Tables 4.4 and 4.5 are used to obtain $N_{th} = 3.62 \times 10^{18}$ cm^{-3} and $J_{th} = 166.4$A/cm^2 per QW). The length of the in-plane laser is assumed to be 250 μm.

densities somewhat lower than indicated, particularly at the smaller dimensions (see Problem 7 and Example 4.5).

Because there is much interest in reducing the active volume well below the limits imposed in Fig. 4.20, much attention has focused on ways to reduce or eliminate surface recombination. One obvious method would be to define the stripe or pillar *without etching through* the active region. However, one must now contend with lateral outdiffusion of carriers at the stripe or pillar edges. In Problem 7, it is shown that the ambipolar diffusion of carriers out of the active region is equivalent to a surface recombination velocity of magnitude

$$v_{sD} \equiv \frac{D_{np}}{L_{np}} = \sqrt{\frac{D_{np}}{\tau_{np}}} = 1 \times 10^5 \text{ cm/s} \cdot \left[\frac{D_{np}}{20 \text{ cm}^2/\text{s}} \frac{2 \text{ μm}}{L_{np}} \right] \tag{4.86}$$

D_{np} and L_{np} are the ambipolar diffusion constant and related diffusion length, and τ_{np} is the average carrier lifetime. The last equality evaluates v_{sD} for common values, revealing that carrier outdiffusion "velocities" and surface recombination velocities are unfortunately of similar magnitude. To minimize carrier outdiffusion, we must bury the active region in a higher-bandgap material that provides a lateral potential barrier to carriers. However, such regrowth technologies, although popular and effective in long-wavelength systems, have remained problematic in AlGaAs-containing compounds.

Example 4.5 The surface recombination velocity can be estimated using the simple "broad-area" (i.e., infinite stripe width) threshold carrier density; however, in reality, the carrier density profile will vary over the cross section of the active region, particularly when the active width is narrow. In this problem, the effects of a finite diffusion constant for carriers in the active region will be examined. Assume that the carrier densities in the active region are high enough that any differences in the diffusion profiles of electrons and holes will set up an electric field that will pull the two densities to nearly the same profile. In this *ambipolar* diffusion limit, the hole diffusion rate is enhanced by a factor of ~2 by the forward pull of the electrons, and the electron diffusion rate is limited to approximately twice the normal hole diffusion rate by the backward pull of the holes. The overall effect is that we can assume the electron and hole densities are equal everywhere in the active region and are characterized by a single ambipolar diffusion constant, D_{np}. The lateral profile of carriers is then governed by the simple diffusion equation:

$$D_{np} \frac{d^2 N(x)}{dx^2} = -\frac{I(x)}{qV} + \frac{N(x)}{\tau_{np}}. \tag{4.87}$$

The carrier lifetime is in general a function of N; however, to obtain analytic solutions, we can evaluate the lifetime at the broad-area threshold value, $\tau_{np}|_{th} = qL_z N_{th}/J_{th}$. The problem we wish to solve is the carrier density profile across the

width of the active region in the in-plane laser depicted in Fig. 4.19. For this case, we can define two distance regions: one beneath the contact within w where we assume a uniform current injection profile, and the region outside w where there is no current injection. Mathematically, with $x = 0$ defined as the center of the stripe, we have $I(x) = I_0$ for $x < w/2$, and $I(x) = 0$ for $x > w/2$. In fabricating the laser we can either leave the active region in place outside the stripe, or we can remove it by etching through the active region outside the contact area. The first case leads to carrier outdiffusion, whereas the second case leads to surface recombination. We would like to compare these two cases.

Problem: (1) With the active region in place away from the contact, carriers are free to diffuse outside the stripe width. Solve for the concentration of the carriers $N(x)$ in and out of the stripe assuming the carrier density and its derivative (i.e., the diffusion current) are constant across the $x = w/2$ boundary. Solve for the carrier profile in this case. (2) With the active region etched away, the carriers recombine at the surface, solve for the concentration of the carriers $N(x)$ under the stripe assuming the diffusion current (defined by the slope of the carrier density) is equal to the surface recombination current, $D_{np}dN/dx = -v_sN$, at the $x = w/2$ boundary. Place your result in terms of the diffusion equivalent surface recombination velocity, $v_{sD} = \sqrt{D_{np}/\tau_{np}}$. Solve for the carrier profile in this case.

Solution: From the carrier diffusion equation, (4.111), we can express the carrier concentration as

$$\frac{d^2N(x)}{dx^2} - \frac{N(x)}{D_{np}\tau_{np}} = -\frac{I(x)}{qVD_{np}}. \tag{4.88}$$

If we define $x = 0$ as the lateral center of the laser stripe, we then have that

$$I(x) = \begin{cases} I_0 & |x| \leq \left|\frac{W}{2}\right| \\ 0 & |x| > \left|\frac{W}{2}\right|. \end{cases} \tag{4.89}$$

If we define $L = \sqrt{D_{np}\tau_{np}}$ and $G = \frac{I(x)}{qVD_{np}}$, then the carrier diffusion equation can be rewritten as

$$\frac{d^2N(x)}{dx^2} - \frac{N(x)}{L^2} = -G.$$

This is a special simplified form of the second order linear partial differential equation, whose solutions can be expressed analytically.

For $|x| < \left|\frac{W}{2}\right|$, within the stripe width, the carrier diffusion equation has a solution of

$$N(x) = A\cosh\left(\frac{x}{L}\right) + GL^2$$

because the solution needs to be symmetric about $x = 0$, due to the statement of the problem. A will be determined by matching the boundary conditions.

For $|x| > \left|\frac{W}{2}\right|$, withing the stripe width, the carrier diffusion equation has a solution of

$$N(x) = Be^{\frac{-|x|}{L}},$$

where B will be determined by matching the boundary conditions. The other mathematically possible solution, $e^{\frac{|x|}{L}}$, is not physical because the carrier density would go to infinity with $x- > \infty$.

Case 1—active region left outside the stripe

In this case, the boundary conditions at $x = \frac{w}{a}$ are that $N(x)$ is continuous, and $\frac{dN(x)}{dx}$ is continuous. Applying these boundary conditions, we have

$$\begin{aligned} A &= \frac{-1}{2} GL^2 e^{-w/2L} \\ B &= -A\left(e^{w/L} - 1\right) \\ &= \frac{GL^2}{2}\left(e^{w/2L} - e^{-w/2L}.\right) \end{aligned}$$

Therefore, the carrier density profile is given by

$$N(x) = \begin{cases} GL^2\left[1 + \frac{-1}{2}\left(e^{(x-w/2)/L} + \left(e^{(-x-w/2)/L}\right)\right)\right] & |x| \le \left|\frac{W}{2}\right| \\ \frac{GL^2}{2}\left(e^{(w/2-|x|)/L} - \left(e^{(-|x|-w/2)/L}\right)\right) & |x| > \left|\frac{W}{2}\right| \end{cases}$$

Case 2—active region etched outside the stripe

In this case, there will be no carriers outside the active strip. The solution for the carrier density equation inside the strip is the same as previously derived; however, there will be a different boundary condition to satisfy at the surface. Basically, the diffusion current at the boundary needs to be equal to the surface recombination current, $-q v_s N = qD_{np}\frac{dN}{dx}$. If we define surface recombination velocity as $v_{sD} = \sqrt{\frac{D_{np}}{\tau_{np}}}$, we have

$$N(x) = GL^2\left[1 - \frac{e^{x/L} + e^{-x/L}}{\left(1 + \frac{v_{sD}}{v_s}e^{\frac{w}{2L}}\right) + \left(1 - \frac{v_{sD}}{v_s}e^{\frac{-w}{2L}}\right)}\right].$$

Other techniques for reducing the effects of surface recombination are also being considered. One interesting method involves surface passivation. As might be expected, the density and character of surface states is very dependent on the surface chemistry. For example, an oxidized surface provides a different recombination velocity than a freshly etched surface. If a GaAs surface is soaked for a while in a water solution containing Na_2S, $(NH_4)_2S$, or any one of various salts that yield aqueous sulfur species, it has been demonstrated that the surface recombination

velocity can be reduced dramatically, yielding $v_s \approx 10^3$ cm/s under high-level injection conditions [28]. Unfortunately, the sulfur-containing compounds that tie up and passivate the dangling bonds are somewhat volatile, and the surface passivation effect can fade within an hour or two of exposure to air. Thus, efforts are under way to make the passivation more permanent by placing an oxide or nitride cap over the surface immediately after it has been passivated to lockin and retain the passivating layer.

4.5.3 Auger Recombination

The last type of transition in Fig. 4.18 depicts what is essentially a collision between two electrons that knocks one electron down to the valence band and the other to a higher energy state in the conduction band. The high-energy electron eventually thermalizes back down to the bottom of the conduction band, releasing the excess energy as heat to the crystal lattice. An analogous collision can occur between two holes in the heavy hole (HH) band; in this case, the hole that is knocked deeper into the valence band is transferred to either the split-off (SO) or light hole (LH) band. In all, there are three types of transitions, collectively referred to as *Auger* processes, which are relevant in III-V semiconductors. These are shown in Fig. 4.21. For quantum-well material, additional types of subband-to-subband and bound-to-unbound transitions can be defined, but they still fall into these three general categories.

Because Auger processes depend on carriers colliding with one another, the Auger recombination rate, R_A, should increase rapidly as carrier density increases. The CCCH process involves three electron states and one heavy-hole state, and hence is expected to become important when the electron density is high. The CHHS and CHHL processes involve one electron state, two heavy-hole states, and one split-off or light-hole state. Thus, they are expected to become important when the hole density is high. In lasers, the electron and hole densities in the active region are equal (if the active region is not heavily doped) implying that all three processes are potentially important.

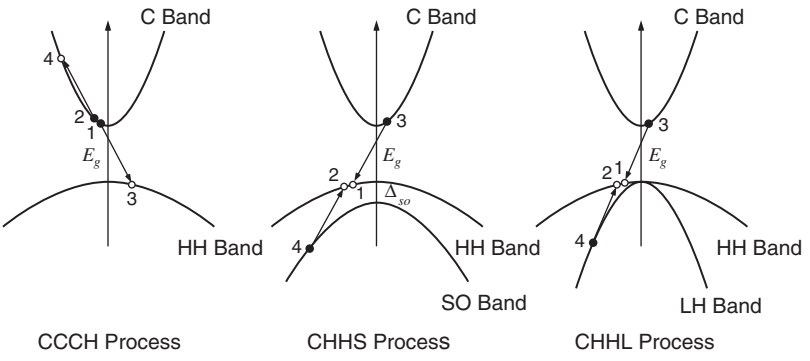


FIGURE 4.21: Auger processes in III–V semiconductors.

The Auger recombination rate, like the radiative transition rates, is dependent on the probability of finding the various states occupied or empty, as the case may be. For the Auger processes in Fig. 4.21, the relevant Fermi factors are

$$P_{1\to 3} = f_{c1} f_{c2} (1 - f_{v3})(1 - f_{c4}), \quad \text{(CCCH)} \tag{4.90a}$$

$$P_{3\to 1} = (1 - f_{v1})(1 - f_{v2}) f_{c3} f_{v4}. \quad \text{(CHHS and CHHL)} \tag{4.90b}$$

The subscript on the probability, P, indicates the significant electron recombination path. The additional c and v subscripts on the Fermi functions identify the quasi-Fermi level to be used in each case: E_{Fc} or E_{Fv}.

To obtain simple indicators of how these Fermi factors vary with carrier density, we use the Boltzmann approximation (which is strictly valid only for low carrier densities). From Table 10.1 we have

$$\frac{N}{N_c} \approx e^{-(E_c - E_{Fc})/kT} \qquad \text{and} \qquad \frac{P}{N_v} \approx e^{-(E_{Fv} - E_v)/kT}, \tag{4.91}$$

which hold as long as $E_{Fc} \ll E_c$ and $E_{Fv} \gg E_v$. Under these conditions, the 1 in the denominator of the Fermi functions defined in Eq. (4.3) can be neglected, and

$$\begin{aligned} f_c &\approx e^{-(E - E_{Fc})/kT} = \frac{N}{N_c} e^{-(E - E_c)/kT}, \\ 1 - f_v &\approx e^{-(F_{Fv} - E)/kT} = \frac{P}{N_v} e^{-(E_v - E)/kT}. \end{aligned} \tag{4.92}$$

With these relations, the transition probabilities can be approximated by

$$P_{1\to 3} \approx \frac{N^2 P}{N_c^2 N_v} e^{-(\Delta E_1 + \Delta E_2 + \Delta E_3)/kT}, \quad \text{(CCCH)} \tag{4.93a}$$

$$P_{3\to 1} \approx \frac{N P^2}{N_c N_v^2} e^{-(\Delta E_1 + \Delta E_2 + \Delta E_3)/kT}, \quad \text{(CHHS and CHHL)} \tag{4.93b}$$

where $\Delta E_i = E_i - E_c$ for conduction band states, and $\Delta E_i = E_v - E_i$ for valence band states. The fourth, high-energy state is assumed to be completely empty (CCCH) or completely full (CHHS and CHHL). These probability factors suggest that CCCH processes are $\propto N^2 P$ whereas, CHHS and CHHL processes are $\propto N P^2$. Furthermore, for a given electron and hole density, the probability of Auger recombination increases exponentially with temperature. The strong dependence on both carrier density and temperature makes Auger recombination a potentially devastating recombination path for carriers in laser applications.

The Auger recombination rate is also strongly dependent on the bandgap of the material. To understand why this is so, we need to examine the consequences of energy and momentum conservation. With Auger transitions, as with radiative transitions, the initial energy and momentum of the system must be conserved.

This constrains Auger transitions to specific regions of the bands. In particular, the lowest possible energy configuration does not occur right at the band edge, as it does with radiative transitions, but at a slightly higher energy. This "lowest energy" configuration for the CCCH process is depicted in Fig. 4.22 for two different bandgaps. The shading illustrates the location of electrons assuming a zero temperature distribution. The presence of holes at state 3 is greatly enhanced for the smaller bandgap material; this allows for a much higher Auger transition rate *for the same overall carrier density*. In fact at zero temperature, such lowest-energy Auger transitions would not occur in the higher-bandgap material because state 3 would be fully occupied. The two initial electron states, 1 and 2, are also closer to the band edge in the smaller-bandgap material, further enhancing the Auger transition rate, because at finite temperatures more electrons can be found there.

To gain a more quantitative feel for how the bandgap affects the Auger transition rate, we need to estimate how the transition probabilities in Eq. (4.93b) are affected by the bandgap. Figure 4.22 demonstrates how states 1, 2, and 3 move closer to the band edge as the bandgap is reduced. In mathematical terms, the sum of energy offsets from the band edge, $\Delta E_1 + \Delta E_2 + \Delta E_3$, is reduced as the bandgap is reduced. This sum can be described more conveniently in terms of ΔE_4 using energy conservation. Setting the initial energy equal to the final energy for the CCCH process in Fig. 4.21 or 4.16, we must have

$$\Delta E_1 + \Delta E_2 = -(E_g + \Delta E_3) + \Delta E_4, \tag{4.94}$$

where E_c is used as the energy reference. Rearranging, we obtain

$$\Delta E_1 + \Delta E_2 + \Delta E_3 = \Delta E_4 - E_g. \tag{4.95}$$

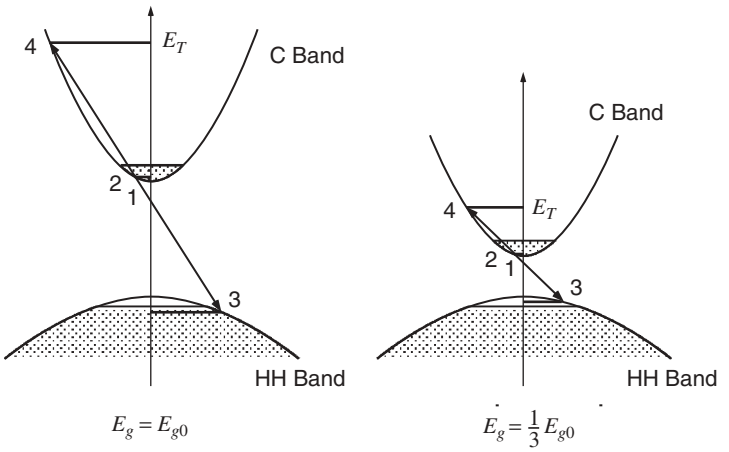


FIGURE 4.22: Lowest-energy CCCH Auger transition for two different bandgaps. The four states are drawn roughly to scale, with $\mu = m_C/m_H \approx 5$.

Thus, the probabilities in Eq. (4.93b) are maximized when ΔE_4 is minimized. This minimum value for ΔE_4 is referred to as the *threshold* energy, E_T, of the Auger process. The lowest-energy CCCH configuration in Fig. 4.22 corresponds to this threshold energy process and is therefore the most probable configuration for CCCH Auger transitions.

Appendix 12 details the math involved with minimizing Eq. (4.78) subject to momentum conservation. This exercise provides expressions for all energies and k-vectors associated with the threshold Auger process. In particular, the threshold energy for all three Auger processes in Fig. 4.21 are found to be

$$E_T = \frac{2m_C + m_H}{m_C + m_H} E_g, \quad \text{(CCCH)} \tag{4.96}$$

$$E_T = \frac{2m_H + m_C}{2m_H + m_C - m_S} (E_g - \Delta_{so}), \quad \text{(CHHS)} \tag{4.97}$$

$$E_T = \frac{2m_H + m_C}{2m_H + m_C - m_L} E_g. \quad \text{(CHHL)} \tag{4.98}$$

For the CHHS process, E_T is measured from the SO band edge. The effective mass prefactors typically fall in the range of 1.1–1.2.[8] The important point to notice is that the threshold energy is proportional to the bandgap. Using Eq. (4.95) in Eq. (4.93b), the maximum probability for CCCH Auger transitions becomes

$$P_{1\to 3} \approx \frac{N^2 P}{N_c^2 N_v} e^{-(E_T - E_g)/kT}, \quad \text{(CCCH)} \tag{4.99}$$

and similarly for the other two processes. With $E_T \propto E_g$, it is clear that the maximum probability for Auger transitions increases exponentially as the bandgap is decreased. For example, if we were to reduce the bandgap of a given material by a factor of two, the probability for CCCH Auger recombination would increase by

$$\frac{P(E_g/2)}{P(E_g)} \approx \frac{e^{-(1/2)(E_T - E_g)/kT}}{e^{-(E_T - E_g)/kT}} = e^{(1/2)aE_g/kT}, \tag{4.100}$$

where

$$a = E_T/E_g - 1 \quad (\approx 0.1 - 0.2). \tag{4.101}$$

InGaAs, $a = 0.15$ for the CCCH process, giving an increase of $e^{4.14} \approx 63$ at room temperature, if the bandgap were reduced by a factor of 2. With $a = 0.2$, this factor would be $e^{5.52} \approx 250$, indicating a strong dependence on $E_T/E_g - 1$. This example, although not quantitatively accurate (in that it only considers the transition probability of one set of states, and assumes the Boltzmann

[8]For GaAs (0.87 μm) : (CCCH, CHHS, CHHL) = (1.15, 1.22, 1.12).
For InGaAsP (1.3 μm) : (CCCH, CHHS, CHHL) = (1.12, 1.13, 1.07).
For InGaAsP (1.55 μm) : (CCCH, CHHS, CHHL) = (1.11, 1.11, 1.06).

approximation), nevertheless serves to illustrate the severity of the problem for long-wavelength lasers, which must deal with Auger recombination rates much larger than in short-wavelength lasers.

A type of Auger transition we have not yet considered involves an additional particle, a phonon of the crystal lattice, which can absorb much of the momentum in the transition. Such *phonon-assisted* transitions are familiar from indirect gap materials such as silicon, where strong phonon-assisted absorption across the indirect gap can occur, even though the electron's momentum changes dramatically (it is transferred from the phonon). With phonon-assisted *Auger* processes, the final momentum of the two electrons does not have to equal the initial momentum (but the *overall* momentum including the phonon is still conserved). In reference to Fig. 4.22, this implies that states 1, 2, and 3 are free to move closer to the band edge than allowed with momentum conservation. Thus, they become more probable and less sensitive to the bandgap and temperature because a threshold energy (which depends on bandgap) no longer exists. However, phonon-assisted Auger processes do involve an additional particle and hence are less likely to occur overall.

In general, phonon-assisted Auger transitions become important in situations where normal Auger recombination is minimal: that is, in large-bandgap materials and/or at low temperatures. For example, Auger recombination in GaAs is dominated by phonon-assisted processes, whereas InGaAsP long-wavelength materials are dominated by normal Auger processes at room temperature. However, at low temperatures below 100–150°C, normal Auger processes are suppressed, and the less temperature sensitive, phonon-assisted processes begin to dominate even in long-wavelength materials.

To quantify the *total* Auger recombination rate, one would in principle use Fermi's Golden Rule to estimate the transition rate for a given set of states. This would involve evaluating the overlap integral of the four states with a coulombic potential perturbation, in addition to evaluating the Fermi factors defined in Eq. (4.90b). With Fermi's Golden Rule known for each set of four states, one would then have to sum over all possible sets of states that obey energy and momentum conservation. With four states involved in the transition, this would involve summing independently over four k-vectors. However, energy and momentum conservation constrain the sums to two independent k-vectors.

Example 4.6 In a compressively strained, 1550-nm quantum-well material, the Auger coefficient was measured to be $C = 6 \cdot 10^{-29}$ cm^6/s at 300 K. The Auger threshold energy can be assumed to be 10% higher than the bandgap energy corresponding to the 1550-nm lasing wavelength.

Problem: If the temperature were increased to 340 K, what would the value of the new Auger coefficient C be?

Solution: From the problem statement, we can establish the relationship between the Auger coefficients at different temperatures,

$$C(340\text{ K}) = C_0 \cdot e^{-\frac{0.1 \cdot E_g}{kT}}.$$

For this active material, the bandgap energy is $E_g = 800$ meV. At 300 K,

$$C(300\text{ K}) = 6 \cdot 10^{-29}\text{ cm}^6/\text{s} = C_0 \cdot e^{-\frac{80}{26}},$$

where $kT = 26$ meV at room temperature. Finally,

$$C(340\text{ K}) = C(300\text{ K}) \cdot e^{\frac{300}{340}} = 1.436 \cdot C(300\text{ K}) = 8.617 \cdot 10^{-29}\text{ cm}^6/\text{s}.$$

Theoretical models attempting the preceding procedure were first considered by Beattie and Landsberg in their pioneering 1959 paper on Auger recombination [29]. Since that time researchers have applied more refined versions of the theory to various material systems. Dutta and Nelson [30] analyzed the InGasP system, whereas Takeshima [32] also analyzed AlGaAs and other systems. Van de Walle analyzed InGaN-based systems using density-functional theory with many body perturbation theory [31]. Taylor et al. [39] as well as others have attempted to extend the theory to quantum wells and quantum-wire material. Unfortunately, the difficulty with theories of Auger recombination is that information of the band structure at more than a bandgap away from the band edge must be known accurately. Overlap intergrals of "k-space distant" Bloch functions must also be known. Such experimental information is sparse, and theories are inevitably led to making very simplifying assumptions. In contrast, the spontaneous emission rate considered in Section 4.4 can be obtained from the band edge Bloch function overlap and band edge curvatures, data that is experimentally abundant. Hence, spontaneous emission rate calculations can be quite accurate.

In general, theories can predict the Auger rate to within an order of magnitude. That is not to say they are not important. On the contrary, they remain valuable for predicting trends in the Auger recombination rate, such as the temperature and bandgap dependence. Also, the *relative* effects of material composition variations and of reduced dimensionality can be estimated theoretically. In addition, the relative importance of the three Auger processes can be determined. Most Auger theories predict that the CHHS process dominates in common III–V semiconductors, with the CCCH process almost an order of magnitude smaller (however, some theories estimate comparable magnitudes for these two processes). The CHHL process is orders of magnitude smaller than either the CHHS or CCCH process and its contributions are negligible.

Continued refinements in the theory can hopefully produce more accurate predictions. For example, Takeshima [32] has enhanced the accuracy of his Auger theory by using realistic band structures. His model extends deep into both conduction and valence bands, accounting for the nonparabolicity in all bands. In addition the band model includes the anisotropy of the band structure (the change in the band curvature along different directions of the crystal), which he considers a very important factor. His predictions include phonon-assisted Auger processes implicitly, which he finds also to be important in InGaAsP material, contrary to the results of other

researchers. His predictions, in particular, appear to match well with experimentally obtained results on InGaAsP material.

In light of the above theoretical challenges, the most common method of estimating Auger recombination is to use experimentally obtained Auger coefficients in combination with the calculated or experimentally measured carrier density, assuming a recombination rate per unit volume of the form:

$$R_A = C_n N^2 P + C_p N P^2, \tag{4.102}$$

where the first term is due to the CCCH Auger process and the second term is due to the CHHS process. These dependencies are strictly valid only under nondegenerate conditions where Eq. (4.93b) can be used. For degenerate carrier densities ($\geq 10^{18}$ cm^{-3}), this functionality overestimates the Auger recombination rate somewhat. However, use of Eq. (4.102) is widespread and convenient.

In laser applications with lightly doped active regions, $N = P$ at high injection levels and the Auger recombination rate simplifies to

$$R_A = CN^3 = \frac{I_{Auger}}{qV}, \tag{4.103}$$

where C is a generic experimentally determined Auger coefficient that lumps together CCCH, CHHS, and phonon-assisted Auger processes. In long-wavelength InGaAsP materials, various carrier lifetime measurements that can extract the cubic dependence on carrier density place C anywhere from 10^{-29} cm^6/s to 10^{-28} cm^6/s at room temperature, depending on the particular method of measurement and the material used. Representative values at room temperature are:

$$C \approx 2 - 3 \times 10^{-29} \text{ cm}^6/\text{s}, \quad \text{(bulk1.3 } \mu\text{m InGaAsP)} \tag{4.104}$$

$$C \approx 7 - 9 \times 10^{-29} \text{ cm}^6/\text{s}. \quad \text{(bulk1.55 } \mu\text{m InGaAsP)} \tag{4.105}$$

Equation (4.104) agrees well with the predictions of Takeshima [32]. For GaAs, less data exist. Experimental data from Takeshima gives an Auger coefficient that is about an order of magnitude less than in long-wavelength systems at room temperature:

$$C \approx 4 - 5 \times 10^{-30} \text{ cm}^6/\text{s}. \quad \text{(bulk GaAs)} \tag{4.106}$$

For InGaN, based on the work by Van de Walle [31], the Auger coefficient is given by

$$C \approx 2 \times 10^{-30} \text{ cm}^6/\text{s}. \quad \text{(InGaN quantum wells, } \lambda = 0.496 \ \mu\text{m)} \tag{4.107}$$

For other material systems, the Auger recombination rate is not well characterized, but is expected to be similar to the values given in Eqs. (4.104) through (4.107), depending on the bandgap of the material.

In quantum wells, the band structure is converted into subbands and it is expected that the Auger rate is modified. Discussions by Smith et al. [33]

and Taylor et al. [34] predict that the Auger coefficient should be reduced by $\sim\sqrt{kT/aE_g}$ in quantum-well material, where a is defined in Eq. (4.101), for $aE_g > kT$. For long-wavelength materials, this factor is on the order of 1.5–2.0. Experiments by Hausser et al. [39] indicate a reduction of about 3 in the Auger coefficient when comparing measurements of quantum-well and bulk material. Thus, reduced dimensionality appears to help with Auger recombination, but does not completely solve the problem.

Another possible method of minimizing Auger recombination involves using strain, which it has been suggested should reduce the Auger coefficient (see Problem 7). Strained quantum-well lasers in long-wavelength material systems in fact have approached threshold current densities comparable to larger-bandgap quantum-well lasers, suggesting that the Auger recombination is much lower in these devices.

4.6 ACTIVE MATERIALS AND THEIR CHARACTERISTICS

In this last section, we are going to run through a number of example calculations in bulk and quantum-well material in the two most popular material systems, GaAs/AlGaAs and InGaAs/InP. The examples will serve to quantitatively illustrate the various dependencies between the gain, current, and carrier density in different materials. The order of presentation will proceed as follows. We will first look at various gain spectra for different carrier densities. The peak gain and differential gain will then be determined as a function of carrier density. The spontaneous emission spectrum and its relation to carrier density will then be explored. From this, the current as a function of carrier density and ultimately the gain as a function of current density will be obtained. Experimental comparisons will then be made to estimate the accuracy of the model. Finally, we will explore the parameter space by varying the width of the quantum well, the doping, and the temperature.

4.6.1 Strained Materials and Doped Materials

Before getting started with the examples, we need to briefly consider the concept of strained QWs, because a number of the examples considered include them. Strained QWs use a material that has a different native latice constant than the surrounding barrier material. If the QWs native lattice constant is larger than the surrounding lattice constant, the QW lattice will compress in the plane, and the lattice is said to be under *compressive* strain. If the opposite is true, the QW is under *tensile* strain. In the GaAs system, adding a little indium to GaAs to form InGaAs can increase the native lattice constant, allowing for the construction of InGaAs/GaAs *compressively* strained QWs. In the InP system, the InGaAs ternary (or InGaAsP quaternary) can be either smaller or larger than the lattice constant of InP. Thus, both tensile and compressive strain can be achieved with InGaAs/InP QWs. However, in any lattice-mismatched system, it is important to realize that there is a critical thickness beyond which the strained lattice will begin to revert back to its native state, causing high densities of lattice defects. For typical applications, this critical thickness is on the

order of a few hundred angstroms, thus limiting the thickness of strained active layers to a few QWs.

The effects of strain on the bandgaps are considered in detail in Appendix 11. The corresponding effects on the band curvature are discussed in the latter part of Appendix 8. Essentially strain of either type increases the curvature of the valence band structure, greatly reducing the effective mass. However, compressive strain is better at doing this. In relation to the gain, the implications are many. For example, Fig. 4.23 shows a typical band structure in the center. Due to charge neutrality under high-injection conditions, and the asymmetry in the effective masses, the quasi-Fermi levels separate more toward the conduction band, as discussed in relation to Fig. 4.9. Strain can reduce the valence band effective mass allowing the quasi-Fermi levels to separate more symmetrically, as shown to the left.

The advantage of the left plot can be seen in the following. Concentrating on the conduction band, it is clear that *for a given quasi-Fermi level separation*, the electron density is always lower in the left plot than in the center plot. Thus, at transparency when the quasi-Fermi levels are separated by the bandgap in both cases, the left plot inevitably has a lower carrier density. In other words, the transparency carrier density, N_{tr}, can be reduced substantially in strained materials. In addition, because the current goes roughly as N^2, the transparency current, J_{tr}, can also be reduced. In practical structures, the reduction can be as much as a factor of 2.

An additional advantage of the arrangement on the left in Fig. 4.23 is that the differential gain is also enhanced. So not only does the material reach transparency faster, but the gain also increases faster with carrier density. To understand this, it is important to realize that the differential gain, *dg/dN*, depends on how quickly the *band edge carrier density* changes in response to movements in the quasi-Fermi levels. Because the slope of the Fermi occupation probability function with energy is maximized at the location of the quasi-Fermi level, the band edge carrier density will be affected most when the quasi-Fermi level is aligned with the band edge. Thus, for increased differential gain, it is critical to bring *both* quasi-Fermi levels as close to the band edges as possible. The left plot clearly accomplishes this task

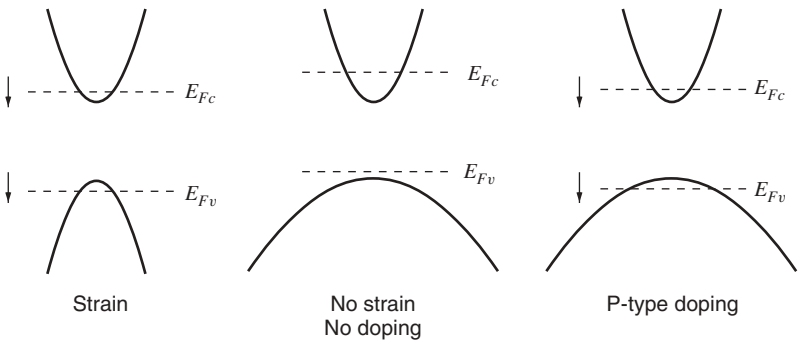


FIGURE 4.23: Illustration of how the quasi-Fermi levels are affected by strain and *p*-type modulation doping.

(particularly near transparency). The improvement in differential gain in strained materials can be as much as a factor of 2.

Another way to improve the differential gain is by p-type doping of the active region. As shown in the plot on the right in Fig. 4.23, the addition of doping can pull the quasi-Fermi levels down to a more symmetrical position just as strain does. Similarly, because of the alignment of the quasi-Fermi levels with the band edges, the differential gain can again be improved. However, in this case, while the *electron* density can be reduced at transparency for similar reasons as with the strained example, the *hole* density is increased. Now because the electron density is degenerate while the hole density is nondegenerate, the downward shift in quasi-Fermi levels increases the hole density faster than it decreases the electron density (P increases approximately exponentially while N decreases approximately linearly). Thus, the NP product actually increases in this case, resulting in a higher transparency current density. So unlike strained materials, where both transparency and differential gain are improved, p-type doping increases the differential gain at the expense of increased transparency current densities. Later we will examine this relationship more quantitatively.

4.6.2 Gain Spectra of Common Active Materials

The following gain spectra were calculated using Eqs. (4.38) and (4.39) with the reduced density of states functions derived in Appendix 8. Lineshape broadening which smooths the spectral features has been included using Eq. (4.46) with Eq. (4.48) as the lineshape function. The two quasi-Fermi levels defining the gain have been connected by assuming charge neutrality in the QW region, including filling of barrier states directly above the QW states. The carrier density therefore refers to all carriers within the QW region. Also for the QWs, the polarization is assumed to be in the plane of the well (TE polarization) because the matrix element is much larger for this polarization. And unless otherwise specified, all active materials are undoped.

Figure 4.24 shows gain spectra for three common active materials in the GaAs system: (1) bulk GaAs, (2) an unstrained GaAs/$Al_{0.2}Ga_{0.8}As$ 80 Å QW, and (3) a compressively strained $In_{0.2}Ga_{0.8}As$/GaAs 80 Å QW. Figure 4.25 shows gain spectra for three common active materials in the InP system: (1) lattice-matched bulk $In_{0.53}Ga_{0.47}As$, (2) an unstrained $In_{0.53}Ga_{0.47}As$/(Q1.08) 60 Å QW, and (3) a compressively strained $In_{0.68}Ga_{0.32}As$/(Q1.08) 30 Å QW. The strained QW has a +1% lattice mismatch to InP. Both QWs have bandgap wavelengths of 1.5 μm. The (Q1.08) notation implies quaternary InGaAsP with 1.08 μm bandgap wavelength.

The bulk gain spectra in both Figs. 4.24 and 4.25 are much smoother than the staircase QW gain spectra. However, the QW gain spectra tend to provide higher gain near the band edge. In comparing the QW gain spectra, the unstrained QWs have a higher maximum gain, g_{max}, due to the larger density of states in the valence band available to participate in transitions. However, it takes more carriers to reach near complete inversion in the unstrained QWs. Other features in the QW

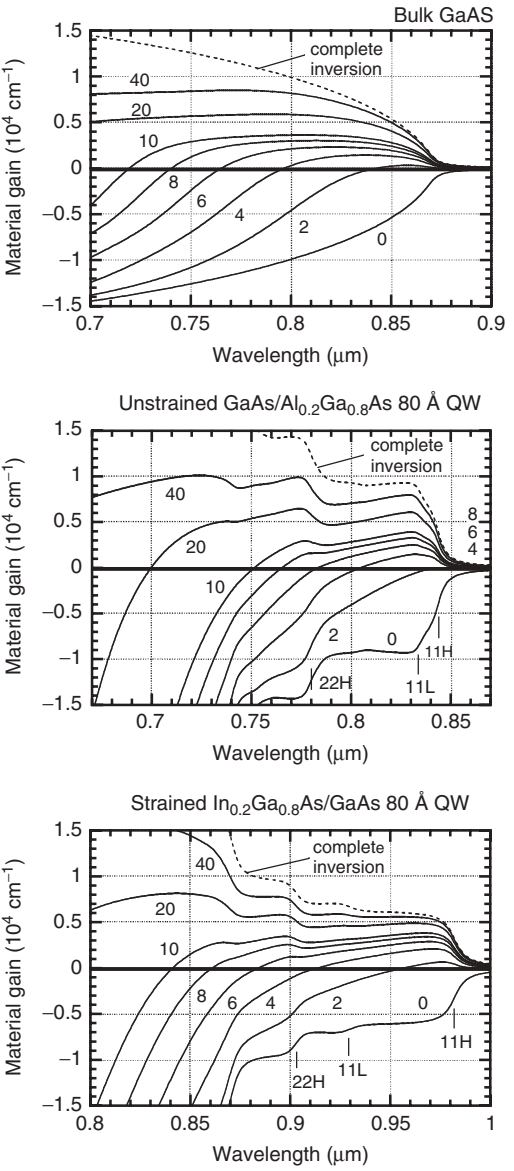


FIGURE 4.24: TE gain spectrum versus carrier density in GaAs based materials. Indicated values are the sheet carrier densities: $\times 10^{12}$ cm^{-2} (the bulk "sheet" density assumes an 80 Å width).

gain spectra include various subband transitions. These features are to be compared with the idealized version in Fig. 4.13. At shorter wavelengths in the GaAs based QWs, a bulk contribution from the 100 Å AlGaAs or GaAs barrier layers on either side of the QW appears. In fact, the barriers also become inverted for very high

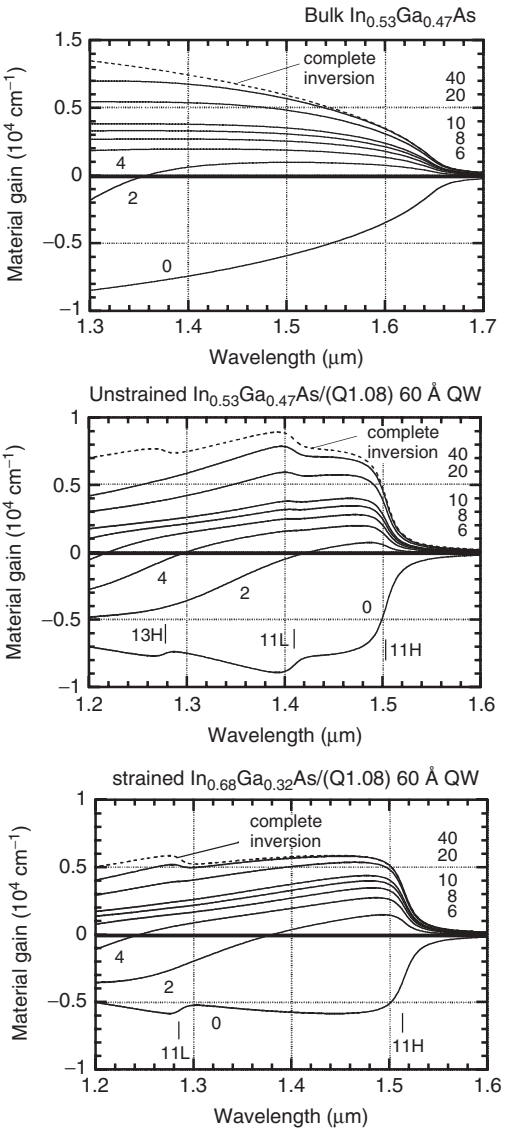


FIGURE 4.25: TE gain spectrum versus carrier density in InP-based materials. Indicated values are the sheet carrier densities: $\times 10^{12}$ cm^{-2} (the bulk "sheet" density assumes a 60 Å width). The material gain in the 30 Å QW is divided by 2 to account for the smaller optical confinement in comparison to the 60 Å QW.

carrier densities, particularly in the InGaAs/GaAs QW. In the InP-based QWs, the bulk contributions are not within the wavelength range shown. The sheet carrier density (NL_z) required to achieve gain in all cases appears to be slightly less than 2×10^{12} cm^{-2}.

4.6.3 Gain versus Carrier Density

To get a better feel for how the gain varies with carrier density, the peak of the gain spectrum is plotted as a function of carrier density in Fig. 4.26. The upper plot includes the GaAs-based active materials. Generally, N_{tr} is clustered near 2×10^{12} cm^{-2}. However, the strained QW has the lowest N_{tr} and increases the fastest with N, as expected from earlier discussions in Section 4.6.1. Curiously, the bulk GaAs and the GaAs QW have similar characteristics. In other words, it would appear that an 80 Å bulk GaAs layer would behave similarly to an 80 Å

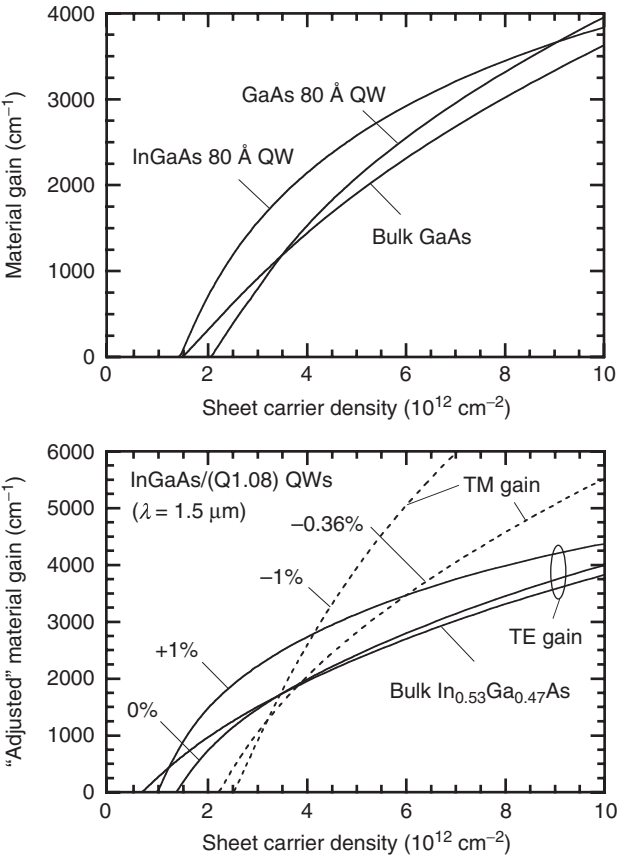


FIGURE 4.26: Peak TE gain versus sheet carrier density in GaAs (upper) and InP (lower) based materials (bulk "sheet" densities assume an 80 Å width for GaAs and a 60 Å width for $In_{0.53}Ga_{0.47}As$). In the lower plot, the strained QWs have 1.5 μm bandgaps where: $1\% = (30\ \text{Å}, x_{ind} = 0.68), 0\% = (60\ \text{Å}, x_{ind} = 0.53), -0.37\% = (120\ \text{Å}, x_{ind} = 0.48)$, and $-1\% = (150\ \text{Å}, x_{ind} = 0.38)$. Also in the lower plot, the dashed curves are TM gain, and the "adjusted" material gain is the material gain multiplied by $L_z/60$ Å to account for the difference in confinement for QWs of different width.

GaAs QW, and that the major improvement to be gained from QWs is simply a volume effect rather than a "quantum size" effect.

The lower plot in Fig. 4.26 includes the InP-based active materials with the addition of two strained QWs: (1) a −0.37% tensile strained $In_{0.48}Ga_{0.52}As$/(Q1.08) 120 Å QW, and (2) a −1% tensile strained $In_{0.38}Ga_{0.62}As$/(Q1.08) 150 Å QW, both with bandgap wavelengths close to 1.5 μm. For these QWs, the C–LH transition (which provides higher gain for polarizations perpendicular to the well plane–TM polarization) is dominant, and hence for these two cases, the peak TM gain is plotted. In this system, N_{tr} for the unstrained and compressive materials is clustered closer to 1×10^{12} cm^{-2}. Again, the +1% strained QW increases the quickest with carrier density. However, it saturates faster than the tensile strained QWs. Bulk $In_{0.53}Ga_{0.47}As$ has the lowest N_{tr} and remains comparable to the unstrained QW performance. The two tensile strained QWs have a higher N_{tr}, but do not saturate nearly as quickly, allowing for lower threshold carrier densities at high threshold gains.

Curve fits have been applied to all gain curves in Fig. 4.26. The results appear in Table 4.4. The two-parameter (N_{tr}, g_0) logarithmic functionality works reasonably well. However, for more linear gain curves (such as the bulk gain curves), the fit is not so good. To correct for this, a third linearity parameter, N_s, has been added. For this three-parameter model, the logarithmic functionality converts to linear functionality as $N_s \to \infty$ and $g_0/N_s \to$ constant. Thus, the larger N_s is in the table, the more linear the curve is, and the worse the two-parameter fit. In general, the strained materials follow much closer to a pure logarithmic functionality. Both two- and the more accurate three-parameter fits appear in the table. All carrier densities in the table have been converted to volume carrier densities, and all material gains

TABLE 4.4: Three- and Two-Parameter Gain versus Carrier Density Curve Fits

	$g = g_0 \ln\left[\frac{N+N_s}{N_{tr}+N_s}\right]$			$g = g_0 \ln[N/N_{tr}]$	
Active Material	N_{tr}	N_s	g_0	N_{tr}	g_0
Bulk GaAs	1.85	6	4200	1.85	1500
$GaAs/Al_{0.2}Ga_{0.8}As$ 80 Å QW	2.6	1.1	3000	2.6	2400
$In_{0.2}Ga_{0.8}As$/GaAs 80 Å QW	1.8	−0.4	1800	1.8	2100
Bulk $In_{0.53}Ga_{0.47}As$	1.1	5	3000	1.1	1000
InGaAs 30 Å QW (+1%)	3.3	−0.8	3400	3.3	4000
InGaAs 60 Å QW (0%)	2.2	1.3	2400	2.2	1800
InGaAs 120 Å QW (−0.37%)	1.85	0.6	2100	1.85	1800
InGaAs 150 Å QW (−1%)	1.7	0.6	2900	1.7	2300
$In_{0.15}Ga_{0.85}N$ 25 Å QW [42]	—	—	—	8	2400
Inverse Relation			*Differential Gain*		
$N = (N_{tr}+N_s)e^{g/g_0} - N_s$			$\frac{dg}{dN} = \frac{g_0}{N+N_s}$		

$[N] = 10^{18}$ cm^{-3}, [g] = cm^{-1}.

are for each particular well width ("adjusted" material gains as used in Fig. 4.26 are not used here).

The differential gain, *dg/dN*, is an important parameter in high-speed laser applications. This is primarily because the relaxation resonance frequency of the laser depends on the square root of the differential gain as we learned in Chapter 2 and will discuss more fully in Chapter 5. Thus, higher differential gains can ideally improve the modulation response of the laser (in practice, however, other damping factors are just as important in determining the overall modulation bandwidth). Figure 4.27 plots *dg/dN* for the gain curves in Fig. 4.26, dramatically revealing the importance of working close to transparency when designing high-speed lasers. In general, numbers in the mid- 10^{-16} cm^2 are expected, and this is what is generally observed in QW lasers (however, there can be a large variation in

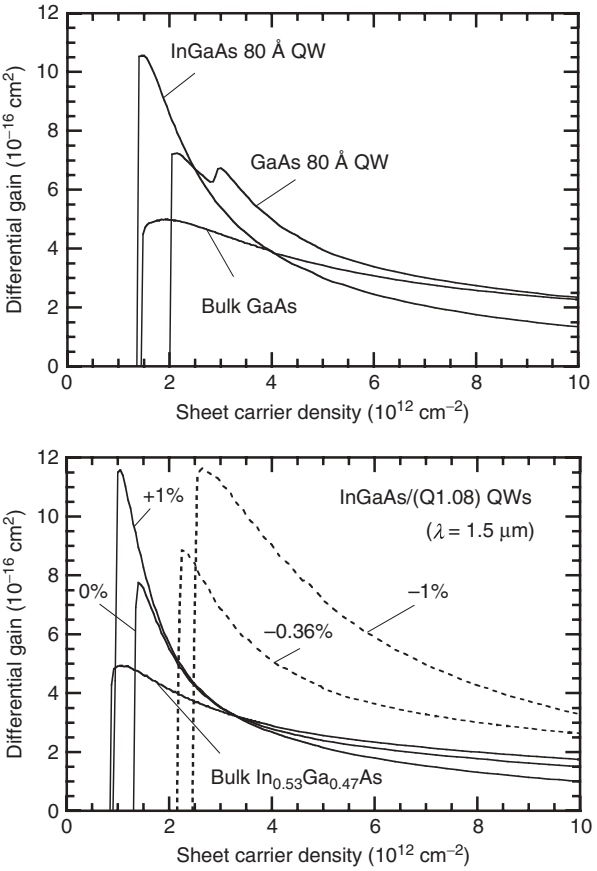


FIGURE 4.27: Peak differential gain, *dg/dN*, versus sheet carrier density in GaAs (upper) and InP-(lower) based materials (bulk "sheet" densities assume an 80 Å width for GaAs and a 60 Å width for $In_{0.53}Ga_{0.47}As$). The compositions of the strained QWs in the lower plot can be found in the caption of Fig. 4.26.

the measured differential gain depending on the threshold carrier density of the laser). In both material systems, the strained QWs yield a ∼50% improvement over unstrained QWs. The bulk values in both cases are the lowest, coming in at below 5×10^{-16} cm^2. In the InP system, *dg/dN* is highest at both extremes of strain and dips by about one-third in between these extremes. The secondary peak on the GaAs QW *dg/dN* curve corresponds to the peak gain switching to the C–LH (11) transition, which rides on top of the C–HH (11) transition at slightly shorter wavelengths (see Fig. 4.24).

Example 4.7 The GaAs/AlGaAs quantum-well active material from Table 4.4 is part of a 5 quantum-well-stack-based material used to fabricate an 800-μm-long and 3-μm-wide ridge laser. The transverse confinement factor Γ_{xy} is 5.5%. From the parameter extraction, the average internal losses are determined to be $\alpha_i = 14$ cm^{-1}, and the injection efficiency is $\eta_i = 0.75$.

Problem: Determine the threshold current for this laser and compare the contributions from spontaneous and nonradiative threshold current components.

Solution: The threshold modal gain of this all-active laser is given by

$$\Gamma g_{th} = \langle \alpha_i \rangle + \frac{1}{L} \ln\left(\frac{1}{R}\right) = 14 \text{ cm}^{-1} + \frac{10^4}{800} \ln\left(\frac{1}{0.32}\right) \text{cm}^{-1} = 28.24 \text{ cm}^{-1}.$$

To calculate the threshold current, we first need to calculate the threshold carrier density using the Eq. (2.48). Because we are dealing with a GaAs-based material system, we only need to take into account the radiative threshold current component, given by Eq. (2.51). However, we need to compare the two current components and thus need to calculate the nonradiative threshold current as well, given by Eq. (2.52). Using the two-parameter fit from Table 4.4 and Eq. (2.48), the threshold carrier density is

$$N_{th} = N_{tr} e^{\frac{g_{th}}{g_{0N}}} = 2.6\, \mathrm{e}^{\frac{28.24}{0.055 \cdot 2400}} \cdot 10^{18} \text{ cm}^{-3} = 3.22 \cdot 10^{18} \text{ cm}^{-3}.$$

The volume of the active region consists of five 8-nm-wide quantum wells, bound by the laser facets and the ridge width,

$$V = L \cdot W \cdot N_w \cdot d = 800 \cdot 3 \cdot 5 \cdot 0.008 \cdot 10^{-12} \text{ cm}^{-3} = 96 \cdot 10^{-12} \text{ cm}^{-3}.$$

Finally, the threshold current is given by

$$I_{th} = \frac{qV}{\eta_i}(BN_{th}^2 + CN_{th}^3) = \frac{1.602 \cdot 10^{-19} \cdot 96 \cdot 10^{-12} \text{ cm}^{-3}}{0.75} \times (10.37 \cdot 10^{26} + 1.334 \cdot 10^{26})\text{A} = 24 \text{ mA},$$

where $B = 1 \cdot 10^{-10}$ cm^3/s and $C = 4 \cdot 10^{-30}$ cm^6/s.

The contribution from the Auger nonradiative current is $\frac{1.334}{11.704} = 11.4\%$.

4.6.4 Spontaneous Emission Spectra and Current versus Carrier Density

Having determined the gain as a function of carrier density, we now need to understand how the spontaneous emission spectrum that generates radiative recombination varies with carrier density. The spontaneous emission spectrum is found using Eqs. (4.70) and (4.71). An example spectrum is shown in Fig. 4.28 for a strained InGaAs/GaAs QW for different sheet carrier densities. The staircase spectrum can be observed and the saturation of the $n = 1$ transition at very high carrier densities is noticeable. However, the second step is much smaller than the first step. The reason is that the second electron state in the QW is barely confined, leading to a poor overlap between the C2 state and the strongly confined HH2 state (this is also visible in the gain spectrum in Fig. 4.24). The result is a smaller matrix element and hence smaller step height.

The third step indicated by the dashed curves in the spectrum of Fig. 4.28 is actually spontaneous emission occurring in the GaAs barrier layers. Such barrier contributions to the radiative current density are particularly large in this active material. In all other QWs considered, barrier recombination is minimal. The reason is that in this material, the QW valence band states allow E_{Fv} to penetrate deeper than normal due to the strained light band structure. As a result, both N and P in the barrier regions can become large, as opposed to other QWs where P in the barriers usually remains very small. The NP product in the barriers can therefore become large leading to a high recombination there. The barrier heights are also relatively small, allowing the barrier regions to populate quickly as carrier density increases. The barrier recombination could be reduced by increasing the GaAs

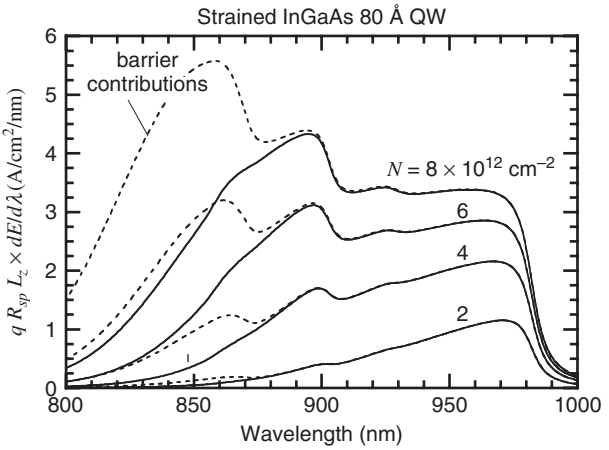


FIGURE 4.28: Spontaneous emission spectrum (per unit wavelength) versus sheet carrier density in a strained $In_{0.2}Ga_{0.8}As$/GaAs 80 Å QW. The dashed curves include spontaneous emission from the 100 Å thick GaAs barrier layers on either side of the well (total of 200 Å). The emission is normalized such that peak × bandwidth (in nm) yields the spontaneous current density in A/cm^2.

barriers to AlGaAs, however, materials growth issues of InGaAs grown directly on AlGaAs make InGaAs/AlGaAs QWs difficult to grow.

The area under the spontaneous emission spectrum in Fig. 4.28 represents the *total* spontaneous emission rate. For example at 4×10^{12} cm^{-2}, the peak is $\sim$2 and the bandwidth is $\sim$100 nm. The (peak) $\times$ (bandwidth) therefore yields a spontaneous current density of $\sim$200 A/cm^2 based on the units of the vertical axis. Performing a more exact integration of the area under the curve via Eq. (4.72), we can determine the spontaneous current density as a function of carrier density. This relationship is plotted in Fig. 4.29 for the 80 Å GaAs-based materials and the 60 Å InP-based materials.

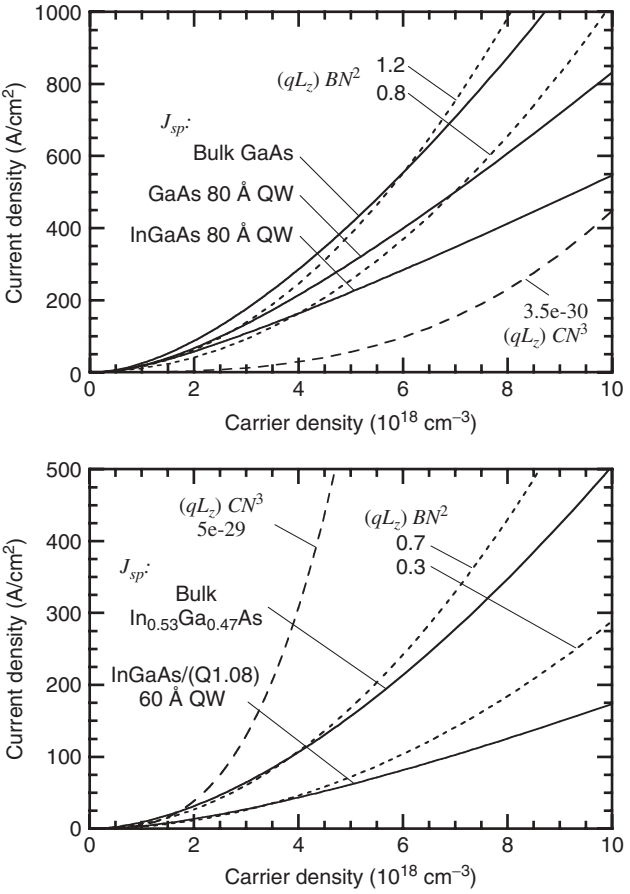


FIGURE 4.29: Spontaneous recombination current density versus carrier density in GaAs (upper) and InP (lower) based materials (bulk currents per unit area assume an 80 Å width for GaAs and a 60 Å width for $In_{0.53}Ga_{0.47}As$). Indicated values for B have units of 10^{-10} cm^3/s, and values for C have units of cm^6/s.

For low carrier densities in Fig. 4.29, bulk GaAs roughly follows a BN^2 law with $B = 1.2 \times 10^{-10}$ cm^3/s. For bulk $In_{0.53}Ga_{0.47}As$, B is closer to 0.7×10^{-10} cm^3/s. These values are in good agreement with measured values for these two materials. The smaller coefficient in the latter case is the result of a lower optical mode density at longer wavelengths (less modes, less spontaneous emission). For higher carrier densities, the BN^2 law overestimates the recombination rate. This is because the square law is strictly valid only within the Boltzmann approximation. As the injection level becomes highly degenerate, the recombination rate does not increase as rapidly as N^2. To account for this, the bimolecular recombination rate is occasionally written as $B_0N^2 - B_1N$. This trend is observed in all cases plotted in Fig. 4.29. Interestingly, the QWs have lower B coefficients than their bulk counterparts.

The Auger recombination current is also plotted in Fig. 4.29 assuming typical values for the two material systems. In GaAs, Auger recombination compared to radiative recombination is small for normal carrier densities ($2 - 4 \times 10^{18}$ cm^{-3}). Thus, the radiative efficiency in GaAs based materials is quite high and can approach unity at low carrier densities. In the InP system, the Auger current dominates for $N > 2 \times 10^{18}$ cm^{-3}. Thus, in most InP-based lasers, the radiative efficiency is below 50% and can often be much lower than this. In this system, where the optical mode density is low enough to yield very low radiative currents, it is unfortunate that another mechanism such as Auger recombination has to turn on and ruin everything.

4.6.5 Gain versus Current Density

With the current known, we can now proceed to determine the peak gain as a function of current density as shown in Fig. 4.30. The upper plot shows the GaAs-based materials with and without the current contributions from Auger recombination and barrier recombination (assuming $C = 3.5 \times 10^{-30}$ cm^6/s). These latter currents can degrade the active material performance for material gains larger than 2000 cm^{-1}. Plotted versus current density, it is clear that the InGaAs/GaAs QW provides much better gain performance, particularly at currents below $300 - 400$ A/cm^2. The transparency current density for the strained QW is near 50 A/cm^2, in good agreement with experimentally measured values. The unstrained GaAs QW has a transparency current density slightly higher than 100 A/cm^2. And in general for a given gain, the strained QW requires roughly half the current required by the unstrained QW: Thus, strained InGaAs QWs are the active region of choice in GaAs-based materials (as long as the application can tolerate the 0.98 μm wavelength). In comparing bulk GaAs to the GaAs QW, we see somewhat better performance out of the QW; however, the "quantum size" effects are not that significant. In fact, the bulk GaAs 80 Å well has a lower transparency current density.

In the InP system, only the radiative currents are included in Fig. 4.30. The trends are generally similar to those seen with the gain versus carrier density. However in this case, the unstrained QW clearly outperforms the bulk $In_{0.53}Ga_{0.47}As$. For the compressively strained and unstrained materials, transparency radiative current densities are near 10–15 A/cm^2. For the tensile strained QWs, these values

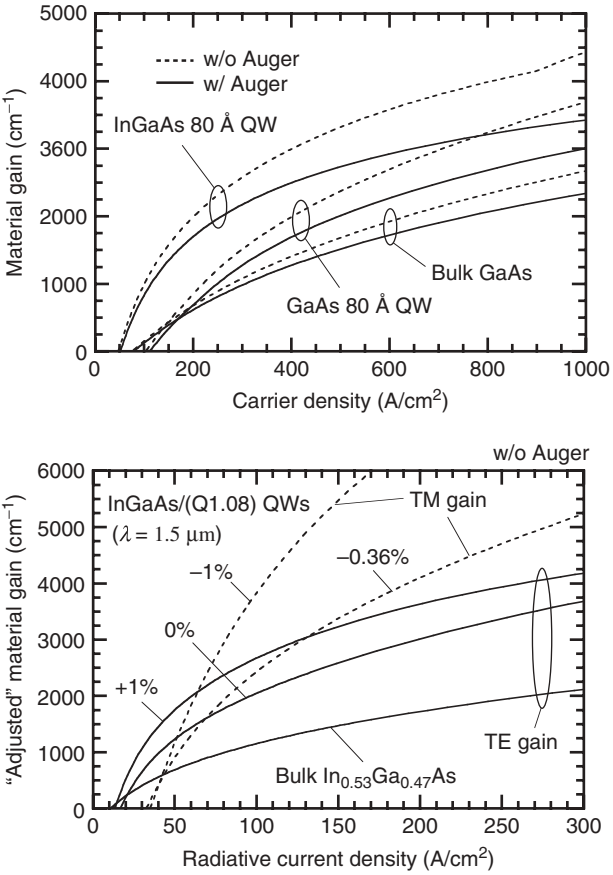


FIGURE 4.30: Peak TE gain versus current density in GaAs (upper) and InP (lower) based materials (bulk currents per unit area assume an 80 Å width for GaAs and a 60 Å width for $In_{0.53}Ga_{0.47}As$). The effects of Auger and barrier recombination are included in the upper plot only. Other comments regarding the lower plot can be found in the caption of Fig. 4.26.

are closer to 30–35 A/cm^2. Unfortunately, Auger recombination prevents such low threshold current densities from being realized in lasers.

Curve fits analogous to the gain versus carrier density fits are listed in Table 4.5. For GaAs-based active materials, curve fits are provided for both radiative (spontaneous) current as well as the current including barrier and Auger recombination (assuming 100 Å barrier widths on either side of the QW and $C = 3.5 \times 10^{-30}$ cm^6/s). For InP-based active materials, only radiative current is fitted (the Auger current can be included by using the carrier density curve fits along with an appropriate value for C). Again, all gains are not "adjusted" material gains, but actual material gains that can be used directly in optical mode overlap integrals to determine the modal gain.

In Chapter 2, the current was empirically related to the carrier density through a polynomial fit (i.e., $J \propto AN + BN^2 + CN^3$). Combining Tables 4.4 and 4.5, we have an alternative description of this relationship:

$$J = (J_{tr} + J_s)\left[\frac{N + N_s}{N_{tr} + N_s}\right]^{g_{0N}/g_{0J}} - J_s. \tag{4.108}$$

Here g_{0N} and g_{0J} refer to the fitting parameter g_0 used with N and J, respectively. Althoug this is a more complex six-parameter fit, it does also provide the gain versus carrier and current density relations. The fit is most useful in converting theoretical gain calculations into practical working models. We will make use of it in the next chapter. It should however be noted that in using Eq. (4.108) below transparency, it is better to use the two-parameter curve fits (i.e., $N_s = J_s = 0$) in Tables 4.4 and 4.5 to ensure that J goes to zero when N does. Note also that in the two-parameter limit, $J = J_{tr}[N/N_{tr}]^{g_{0N}/g_{0J}}$, and hence the power relation is governed by the ratio g_{0N}/g_{0J}. Thus for a simple BN^2 relation we would expect

TABLE 4.5: Three- and Two-Parameter Gain versus Current Density Curve Fits

	$g = g_0 \ln\left[\frac{J + J_s}{J_{tr} + J_s}\right]$			$g = g_0 \ln[J/J_{tr}]$	
Active Material	J_{tr}	J_s	g_0	J_{tr}	g_0
$J_{sp} + J_{bar} + J_{Aug}$					
Bulk GaAs	80	140	1400	80	700
GaAs/$Al_{0.2}Ga_{0.8}$As 80 Å QW	110	50	1600	110	1300
$In_{0.2}Ga_{0.8}$As/GaAs 80 Å QW	50	−10	1100	50	1200
InGaAsP/InGaAsP 65 Å QW [36]	—	—	—	44	1040
$In_{0.15}Ga_{0.85}$N 25 Å QW [37]*	—	—	—	300	1200
$In_{0.15}Ga_{0.85}$N 25 Å QW [38]*	900	31 K	5403	—	—
J_{sp}					
Bulk GaAs	75	200	1800	75	800
GaAs/$Al_{0.2}Ga_{0.8}$As 80 Å QW	105	70	2000	105	1500
$In_{0.2}Ga_{0.8}$As/GaAs 80 Å QW	50	0	1440	50	1440
J_{sp}					
Bulk $In_{0.53}Ga_{0.47}$As	11	30	1000	11	500
InGaAs 30 Å QW (+1%)	13	2	2800	13	2600
InGaAs 60 Å QW (0%)	17	11	1500	17	1200
InGaAs 120 Å QW (−0.37%)	32	18	1400	32	1100
InGaAs 150 Å QW (−1%)	35	10	1700	35	1500

Inverse Relation

$J = (J_{tr} + J_s)e^{g/g_0} - J_s$

$[J] = \text{A/cm}^2, [g] = \text{cm}^{-1}$

Differential Gain

$$\frac{dg}{dJ}L_z = \frac{g_0}{J + J_s}L_z$$

*Values obtained for a ridge structure.

g_{0N} to be twice as large as g_{0J}. In comparing the two-parameter fits in Tables 4.4 and 4.5, we see that this is true to an extent.

Example 4.8 Table 4.5 gives two-parameter fit data for the material gain of 1% compressively strained 3-nm-thick 1.55-μm quantum-well on Indium Phosphide that neglect Auger recombination. Now, we would like to include an Auger coefficient, determined in Example 4.5.

Problem: Determine the new J_{tr} and g_0 that would best model the gain curve for the gains between 500 and 2500 cm^{-1}.

Solution: To solve this problem, we need to compute g_{02} and J_{tr2} of the new logarithmic relationship between the gain and current density. To accomplish this, we need to calculate the current densities for two different values of gain. The new current density is given by $J_2 = J_1 + J_A$, where $J_A = \frac{q \cdot C \cdot N^3 \cdot V}{area} = q \cdot d \cdot C \cdot N^3$. From Table 4.5, we have that $g_{01} = 2600$ cm^{-1} and $J_{tr1} = 13$ A/cm^2, and the quantum well width $d = 3$ nm. Because we also know the carrier density versus gain dependence for this material from Table 4.4, $g = g_{0N} \ln \frac{N}{N_{tr}}$, and $g_{0N} = 4000$ cm^{-1} and $N_{tr} = 3.3 \cdot 10^{18}$ cm^{-3}, we are in a position to calculate the total current density. At transparency, $g = 0$, we have

$$J_{tr2} = J_{tr1} + J_{Atr} = 13 \text{ A/cm}^2 + q \cdot d \cdot CN_{tr}^3 = (13 + 103) \text{ A/cm}^2.$$

At $J_1 = e \cdot J_{tr1}$, we have

$$g = g_{01} \ln \frac{e \cdot J_{tr}}{J_{tr}} = g_{01} = 2600 \text{ cm}^{-1}, N = N_{tr} e^{\frac{g}{g_{0N}}}$$
$$= 3.3 \cdot 10^{18} e^{\frac{2600}{4000}} \text{ cm}^{-3} = 6.32 \cdot 10^{18} \text{ cm}^{-3}.$$

From here,

$$J_A = q \cdot d \cdot C \cdot N^3$$
$$= (1.6 \cdot 10^{-19}\text{C})(3 \cdot 10^{-7} \text{ cm})(6 \cdot 10^{-29} \text{ cm}^6/s)(6.32 \cdot 10^{18} \text{ cm}^{-3}) = 727 \text{ A/cm}^2$$

and finally

$$J_2 = e \cdot J_{tr1} + J_A = 762 \text{ A/cm}^2.$$

Now, g_{02} can be calculated from the transparency current,

$$g_{02} = \frac{g_{01}}{\ln \frac{J_2}{J_{tr2}}} = \frac{2600 \text{ cm}^{-1}}{\ln \frac{762}{116}} = 1381 \text{ cm}^{-1}.$$

4.6.6 Experimental Gain Curves

With the gain versus current density known, we can examine how well the theory works by comparing it to measurements made on in-plane lasers. By measuring the threshold and differential efficiency variations with cavity length, it is possible to extract the internal loss, the internal injection efficiency, and the gain as a function of injected current density. Results of this type of measurement are shown in Figs. 4.31 and 4.32.

The theoretical curves in Figs. 4.31 and 4.32 are found to match quite well in all three material systems, at 980 nm, 1550 nm and 445 nm. The value for C, which best matches the InP data, also agrees well with other measured values

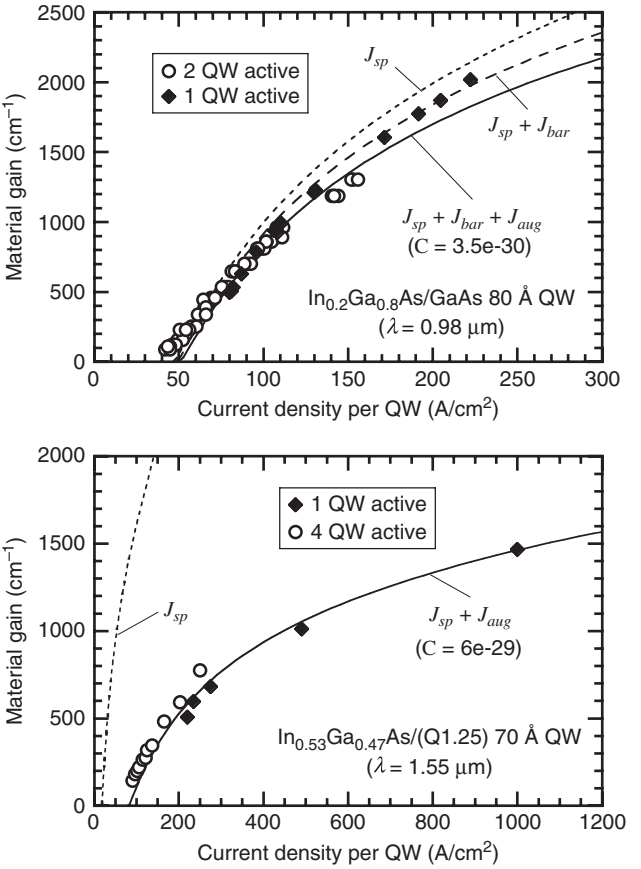


FIGURE 4.31: Experimental material gain versus injected current curves for $In_{0.2}Ga_{0.8}As/GaAs$ 1 and 2 80 Å QW active region 0.98 μm lasers [26] (upper), and $In_{0.53}Ga_{0.47}As/(Q1.25)$ 1 and 4 70 Å QW active region 1.55 μm lasers [45] (lower). Theoretical curves are superimposed on the plots. The gain in the lower plot is well represented by the two-parameter expression, $g = 583 \ln(J/81)$ cm^{-1}, where J is given in A/cm^2.

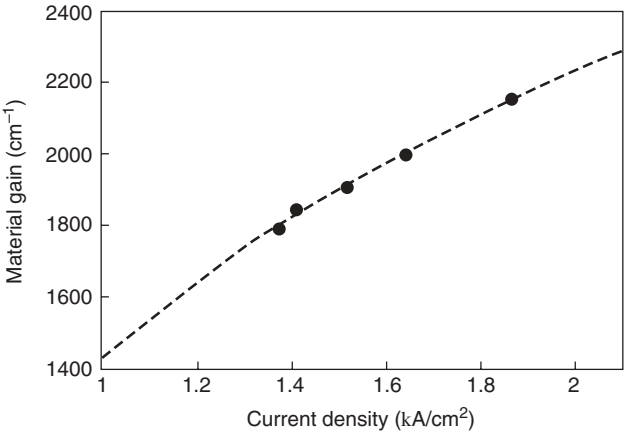


FIGURE 4.32: Experimental material gain versus injected current curves for $In_{0.15}Ga_{0.85}N$ single quantum well blue laser diode at 445 nm. Theoretical curve is superimposed on the plot. The gain is well represented by the two-parameter expression, $g = 1200 \ln(J/300)$ cm^{-1}, where J is given in A/cm^2. *Reprinted with permission from Applied Physics Letters*, [42].

Eqs. (4.104) and (4.105). Although it is true that in the InP system, the choice of C makes the theory more of a curve-fitting procedure, the theory for the InGaAs/GaAs QW involves very few fitting parameters and hence represents very good agreement with experiment within the uncertainties involved with the Auger and barrier recombination rates. Figure 2.15 also reveals excellent agreement between the InGaAs/GaAs QW theory and experiment. Finally, excellent match is observed for an InGaN-based quantum-well active region in Fig. 4.32, where for high-quality material under sufficiently high current injection near threshold, radiative recombination dominates [42].

4.6.7 Dependence on Well Width, Doping, and Temperature

In addition to picking the best active material, there are other choices to make in designing a laser. For a QW active region, an obvious decision is the well width. Aside from having a preference for a particular lasing wavelength, it is useful to have some guidelines in deciding on the optimum well width. Because each subband of a QW is inherently two-dimensional in nature, we should expect that the performance of the $n = 1$ transition has little dependence on the width of the well. However, in practice it is the spacing between subbands that interferes and causes problems.

For example, the current required to reach transparency and 30 cm^{-1} modal gain in an $In_{0.2}Ga_{0.8}As$/GaAs QW is plotted as a function of well width on the left in Fig. 4.33. For midrange values between 50 and 100 Å, there is indeed little change in the QW performance. However, at wider well widths the subband spacing becomes smaller (see inset). As a result, other subbands become populated

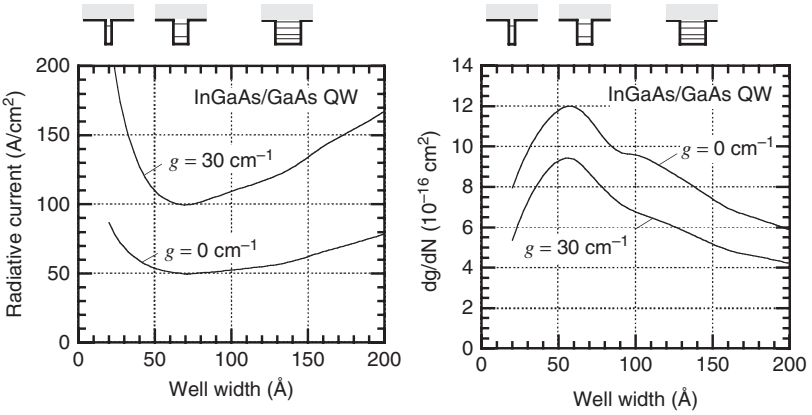


FIGURE 4.33: Radiative current density and differential gain versus well width in an $In_{0.2}Ga_{0.8}As/GaAs$ QW (the GaAs barrier on either side of the QW is 100 Å thick). The curves are for the values at transparency and 30 cm^{-1} *modal* gain (assuming an optical mode width of 2500 Å as given by some SCH structure).

at typical injection levels, degrading the QW performance. In fact, the current starts to become linear with well width as you would expect in bulk material. At narrower well widths, the quantized state is squeezed to the top of the QW, very close to the barrier states (see inset). As a result, the barriers begin to populate significantly at typical injection levels, and again the QW performance is compromised. Thus, there is a window of optimum performance; however, for the current density, the window is fairly broad. For the differential gain, on the other hand, the functionality is more sharply peaked near 60 Å. Away from this peak, the wider well widths reduce the subband spacing, whereas the narrower widths bring the QW state too close to the barrier states.

Another design consideration involves the doping of the active region. Now doping within the active material itself can alter the band structure near the band edges, degrading the performance. However, by placing the doping in the barrier regions (modulation doping), the charge neutrality condition can be altered. In this case, the quasi-Fermi levels are connected through the following equation:

$$N + N_A^- = P + N_D^+, \tag{4.109}$$

where in quantum wells, the acceptor and donor densities are most easily defined as sheet densities.

The basic effects of p-type doping were discussed in Section 4.6.1. Figure 4.34 provides a more quantitative description by plotting the gain and differential gain versus current density for a GaAs/AlGaAs 80 Å QW for n-type, p-type, and undoped material. The downward shift of the quasi-Fermi levels for p-type doping increases the NP product (because N is degenerate and P is nondegenerate at transparency), whereas the upward shift of the quasi-Fermi levels for n-type doping reduces the NP product. Thus, J_{tr} is increased for p-type doping and is

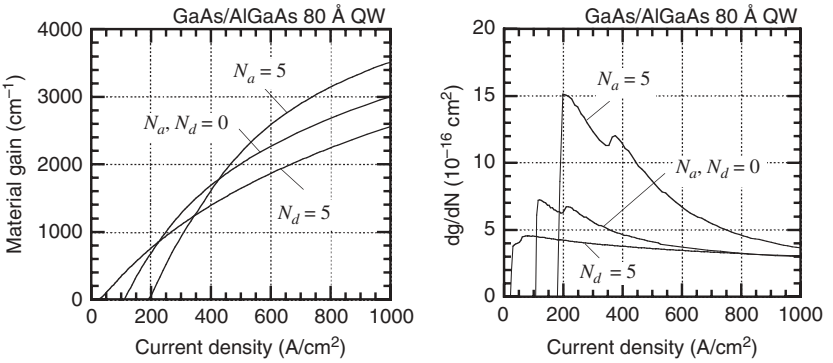


FIGURE 4.34: Peak TE gain and differential gain versus current density for three different doping levels in a GaAs/$Al_{0.2}Ga_{0.8}As$ 80 Å QW. The donor and acceptor sheet densities are in units of 10^{12} cm^{-2}.

reduced for *n*-type doping. However, due to the alignment of the quasi-Fermi levels with respect to the band edges, *p*-type doping increases the differential gain dramatically, whereas *n*-type doping reduces the differential gain. Thus, depending on the application (ultra-high-speed laser, or ultra-low-threshold current), *p*-type and *n*-type doping can add interesting variations to the gain curve, although neither is as interesting as strain, which can simultaneously reduce J_{tr} and increase *dg/dN*.

Another parameter that affects the active material performance is an external parameter, the operating temperature. Figure 4.35 illustrates the effects temperature has on the gain versus current density in both unstrained GaAs QWs and unstrained InGaAs/InP QWs including Auger recombination. In general, the current required to reach a given gain increases with increasing temperature. The dominant cause of this is the broadening of the Fermi occupation probability function, which spreads

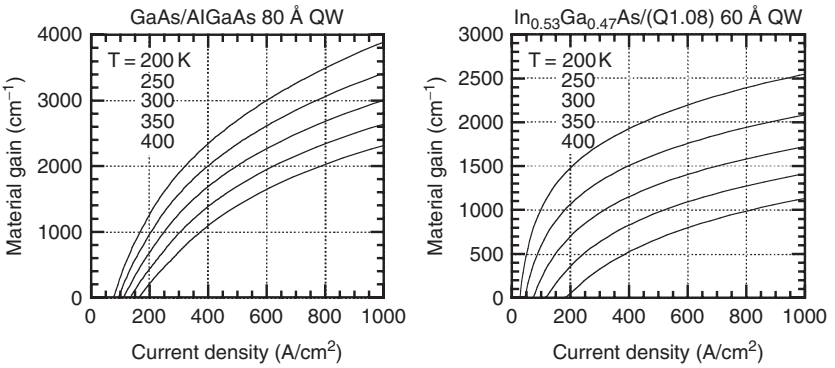


FIGURE 4.35: Peak TE gain versus current density for different temperatures in a GaAs/$Al_{0.2}Ga_{0.8}As$ 80 Å QW (left) and an $In_{0.53}Ga_{0.47}As$/(Q1.08) 60 Å QW (right, with $C = 5 \times 10^{-29}$ cm$^6/s$).

the carriers over a larger energy range for a given overall carrier density. The result is a lower spectral concentration of inverted carriers, which leads to a broadening and flattening of the gain spectrum.

In in-plane semiconductor lasers, the threshold current generally increases exponentially with temperature. As discussed in Chapter 2, this dependence is commonly characterized by

$$I_{th} \sim e^{T/T_0} \rightarrow T_0 = \frac{T_2 - T_1}{\ln(I_2/I_1)}, \tag{4.110}$$

where T_0 is a parameter that characterizes the thermal behavior of the threshold current of the laser. The second equation allows one to estimate T_0 from two threshold measurements at two different temperatures. By comparing the two sets of gain curves in Fig. 4.35, it is evident that InP-based materials are affected by temperature more than GaAs-based materials. Measurements yield a value of $T_0 \sim$50–100 K for InP-based in-plane lasers and a value of $T_0 \sim$ 100–150 K for GaAs-based in-plane lasers. These values are smaller than the theoretical curves in Fig. 4.35 would suggest. However, there are other temperature-dependent effects. For example, the Auger coefficient itself is temperature-dependent, and as the latter part of Appendix 2 reveals, carrier leakage out of the active region is also very sensitive to temperature.

Example 4.9 For an InP the laser based on the material from Example 4.6, we would like to estimate the performance at 340 K, knowing that the radiative efficiency, given as the ratio between the radiative and total threshold current, is $\eta_r = \frac{I_{RAD}}{I_A + I_{RAD}} = 0.134$ at 300 K.

Problem: (1) Determine the ratio of the threshold current at 340 K/300 K, assuming only Auger nonradiative recombination, and that the required threshold carrier density increases by 10% as the temperature is raised from 300 to 340 K. (2) What is the characteristic temperature T_0 for this laser's threshold current?

Solution: From the problem statement, we have that the threshold current is given by the sum of the radiative and Auger components, $I_{th} = qV(BN^2 + CN^3) = I_{RAD} + I_A$. At 340 K, $N(340\text{ K}) = 1.1N(300\text{ K}), B(340\text{ K}) = B(300\text{ K})$, and $C(340\text{ K}) = 1.436 \cdot C(300\text{ K})$. At 340 K,

$$\frac{\Delta I_{th}}{I_{th}} = \eta_r \left(\frac{\Delta I_{RAD}}{I_{RAD}} \right) + (1 - \eta_r) \left(\frac{\Delta I_A}{I_A} \right).$$

The change in the radiative current is caused by the change in the carrier concentration, $\Delta I_{RAD} = ((1.1)^2 - 1) \cdot I_{RAD}$, whereas the change in the Auger current is caused by both the carrier concentration, and the Auger coefficient change, $\Delta I_A = ((1.1)^3 \cdot 1.436) \cdot I_A$. Finally,

$$\frac{\Delta I_{th}}{I_{th}} = 0.134 \cdot (0.21) + 0.866 \cdot (1.911) = 1.683.$$

From here, we have that

$$\frac{I_{th}(340\text{ K})}{I_{th}(300\text{ K})} = \frac{\Delta I_{th} + I_{th}}{I_{th}} = 2.683 = e^{\frac{40\text{ K}}{T_0}}$$

and finally $T_0 = \frac{40\text{ K}}{\ln 2.683} = 40.5\text{ K}$.

REFERENCES

1. Hermann C, Weisbuch C. $\vec{k}\cdot\vec{p}$ perturbation theory in III-V compounds and alloys: a reexamination. *Phys Rev B* 1977;**15**: 823.
2. Hermann C, Weisbuch C. *Modern problems in condensed matter sciences*. Volume 8, Optical orentation, ed. Agranovich VM, Maradudin AA. Amsterdam: North-Holland; 1984. p 463–508.
3. Jani B, Gibart P, Portal JC, Aulombard RL. Effective masses in Sn-doped $Ga_{1-x}Al_xAs$ ($x < 0.33$) determined by the Shubnikov–de Haas effect. *J Appl Phys* 1985;**58**: 3481.
4. Nicholas RJ, Portal JC, Houlbert C, Perrier P, Pearsall TP. An experimental determination of the effective masses for $Ga_xIn_{1-x}As_yP_{1-y}$ alloys grown on InP. *Appl Phys Lett* 1979;**34**: 492.
5. Casey HC Jr, Panish MB. *Heterostructure lasers, part A: fundamental principles*. Orlando, FL: Academic Press; 1978. For bandgap shift, see also Camassel J, Auvergne D, Mathieu H. Temperature dependence of the band gap and comparison with the threshold frequency of pure GaAs lasers. *J Appl Phys* 1975;**46**: 2683.
6. Kaminow IP, Li T, Willner AE. *Optical fiber telecommunications V A—components and subsystems*. Amsterdam: Elsevier; 2008.
7. Yamada M, Ishiguro H, Nagato H. Estimation of the intra-band relaxation time in undoped AlGaAs injection laser. *Jpn J Appl Phys* 1980;**19**: 135.
8. Yamanishi M, Lee Y. Phase dampings of optical dipole moments and gain spectra in semiconductor lasers. *IEEE Journal of Quantum Electronics* 1987;**QE-23**: 367.
9. Asada M. Intraband relaxation effect on optical spectra. In: Zory PS Jr, editor. *Quantum well lasers*. San Diego: Academic Press; 1993. Chap. 2.
10. Kucharska AI, Robbins DJ. Lifetime broadening in GaAs-AlGaAs quantum well lasers. *IEEE Journal of Quantum Electronics* 1990;**QE-26**: 443.
11. Chinn SR, Zory P, Reisinger AR. A model for GRIN-SCH-SQW diode lasers. *IEEE Journal of Quantum Electronics* 1988;**QE-24**: 2191.
12. Tarucha S, Kobayashi H, Horikoshi Y, Okamoto H. Carrier-induced energy-gap shrinkage in current-injection GaAs/AlGaAs MQW heterostructures. *Jap J Appl Phys* 1984;**23**: 874. Also see Tomita A, Suzuki A. Carrier-induced lasing wavelength shift for quantum well laser diodes. *IEEE Journal of Quantum Electronics* 1987;**QE-23**: 1155.
13. Piprek J. *Semiconductor optoelectronid devices: introduction to physics and simulation*. San Diego, CA: Academic Press; 2003.
14. Zhao H, Arif RA, Tansu N. Self-consistent gain analysis of type-II "W" InGaN–GaNAs quantum well lasers. *J Appl Phys* 2008;**104**: 043 104–.
15. Chuang SL. *Physics of photonic devices*. Hoboken, NJ: Wiley; 2009.

16. Seoung-Hwan Park J. Crystal orientation effects on many-body optical gain of wurtzite InGaN/GaN quantum well lasers. *Jpn J Appl Phys* 2003;**42**: 170.
17. Kleinman DA, Miller RC. Band-gap renormalization in semiconductor quantum wells containing carriers. *Phys Rev B* 1985;**32**: 2266.
18. Baba T, Hamano T, Koyama F, Iga K. Carrier-induced lasing wavelength shift for quantum well laser diodes. *IEEE Journal of Quantum Electronics* 1990;**QE-27**: 1347; Spontaneous emission factor of a microcavity DBR surface emitting laser. II. Effects of electron quantum confinements. *IEEE Journal of Quantum Electronics* 1992;**QE-28**: 1310.
19. Brorson SD, Yokoyama H, Ippen EP. Spontaneous emission rate alteration in optical waveguide structures. *IEEE Journal of Quantum Electronics* 1990;**QE-26**: 1492.
20. Shockley W, Read WT Jr. Statistics of the recombinations of holes and electrons. *Phys Rev* 1952;**87**: 835
21. Hall RN. Electron-hole recombination in germanium. *Phys Rev* 1952;**87**: 387.
22. Tsang WT, Reinhart FK, Ditzenberger JA. Very low current threshold GaAs-$Al_xGa_{1-x}As$ double-heterostructure lasers grown by molecular beam epitaxy. *Appl Phys Lett* 1980;**36**: 11.
23. Otaredian T. Temperature dependence of the band gap and comparison with the threshold frequency of pure GaAs lasers. *Solid-State Electron* 1993;**36**: 905.
24. Swaminathan V, Freund JM, Chirovsky LMF, Harris TD, Kuebler NA, D'Asaro LA. Evidence for surface recombination at mesa sidewalls of self-electro-optic effect devices. *J Appl Phys* 1990;**68**: 4116.
25. Jastrzebski L, Lagowski J, Gatos HC. Application of scanning electron microscopy to determination of surface recombination velocity: GaAs. *Appl Phys Lett* 1975;**27**: 537.
26. Hy SY, Corzine SW, Law KK, Young DB, Gossard AC, Coldren LA, Merz JL. Lateral carrier diffusion and surface recombination in InGaAs/AlGaAs quantum-well ridge-waveguide lasers. *J Appl Phys* 1994;**76**: 4479.
27. Nolte DD. Lateral carrier diffusion and surface recombination in InGaAs/AlGaAs quantum-well ridge-waveguide lasers. *Solid-State Electron* 1990;**33**: 295.
28. Yablonovitch E, Sandroff CJ, Bhat R, Gmitter T. Nearly ideal electronic properties of sulfide coated GaAs surfaces. *Appl Phys Lett* 1987;**51**: 439.
29. Beattie AR, Landsberg PT. Auger effect in semiconductors. *Proc R Soc London* 1959;**249**: 16.
30. Dutta NK, Nelson RJ. The case for Auger recombination in $In_{1-x}Ga_xAs_yP_{1-y}$. *J Appl Phys* 1982;**53**: 74.
31. Delaney KT, Rinke P, Van de Walle CG. Auger recombination rates in nitrides from first principles. *Appl Phys Lett* 2009;**94**: 191109
32. Takeshima M. Auger recombination in InAs, GaSb, InP, and GaAs. *J Appl Phys* 1972;**43**: 4114.
33. Smith C, Abram RA, Burt MG. Auger recombination in long-wavelength quantum-well lasers. *Electron Lett* 1984;**20**: 893.
34. Taylor RI, Kelsall RW, Abram RA. Theory of Auger recombination in a quantum well wire. *Surface Sci* 1986;**174**: 169.
35. Kim HS, Lin JY, Jiang HX, Chow WW, Botchkarev A, Morkoç H. Piezoelectric effects on the optical properties of GaN/$Al_xGa_{1-x}N$ multiple quantum wells. *Appl Phys Lett* 1998;**73**: 3426.

36. Raring JW. *Advanced indium phosphide based monolithis integration using quantum well intermixing and MOCVD regrowth*. [dissertation]. Santa Barbara (CA): University of California, Santa Barbara; 2006.

37. Ryu HY, Ha KH, Son JK, Lee SN, Paek HS, Jang T, Sung YJ, Kim KS, Kim HK, Park Y, Nam OH. Determination of internal parameters in blue InGaN laser diodes by the measurement of cavity-length dependent characteristics. *Appl Phys Lett* 2008;**93**: 1110.

38. Farrell RM, Haeger DA, Hsu PS, Fugito K, Feezell DF, Denbaars SP, Speck JS, Nakamura S. Determination of internal parameters for AlGaN-cladding-free m-plane InGaN/GaN laser diodes. Forthcoming.

39. Hausser S, Fuchs G, Hangleiter A, Streubel K, Tsang WT. Auger recombination in bulk and quantum well InGaAs. *Appl Phys Lett* 1990;**56**: 913.

40. Heimbuch ME, Holmes AL Jr, Reaves CM, Mack MP, Denbaars SP, Coldren LA. Tertiarybutylarsine and tertiarybutylphosphine for the MOCVD growth of low threshold 1.55 μm $In_xGa_{1-x}As$/InP quantum-well lasers. *J Electron Mat* 1994;**23**: 87.

READING LIST

Agrawal GP, Dutta NK. *Semiconductor lasers*. 2nd ed. New York: Van Nostrand Reinhold; 1993. Chapter 3.

Corzine SW, Yan, R-H, Coldren LA. Optical gain in III-V bulk and quantum well semiconductors. In: Zory PS Jr, editor. *Quantum well lasers*. San Deigo, CA: Academic Press; 1993. Chapter 1.

Ebeling KJ. *Integrated opto-electronics*. Berlin: Sprinter-Verlag; 1993. Chapter 8.

PROBLEMS

1. An edge-emitting laser uses a single InGaAs quantum well. The lasing wavelength is at 950 nm, the threshold current is 1 mA, the active and modal volumes are 10 and 300 μm^3, respectively, the modal index and group index are 3.5 and 4.0, respectively, the injection efficiency is measured to be 0.8, and the threshold inversion factor, $n_{sp} = 1.6$. By making measurements from the back surface of the chip and factoring in appropriate geometrical corrections, the total spontaneous emission power spectrum at and above threshold is found to be as in Fig. 4.36. (Do not use any material data from elsewhere)

(a) What is the radiative efficiency, η_r?

(b) What is the spontaneous emission rate per unit energy per unit active volume at 950 nm?

(c) What is the optical mode density?

(d) What is the threshold modal and material gains for the TE mode?

(e) What is the spontaneous emission factor, β_{sp}?

2. The reduced density of states function determines the total density of state pairs available as a function of photon energy and is therefore critical in determining

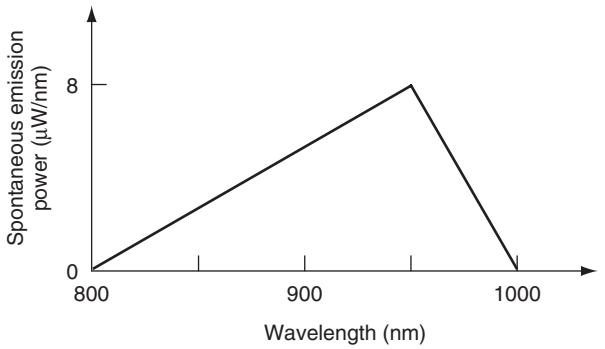


FIGURE 4.36: Total spontaneous emission power spectral density versus wavelength.

the shape of the gain spectrum. Thus, it is worthwhile getting a feel for this function:

(a) Plot $\rho_r(E)$ for bulk GaAs from the GaAs bandgap energy to the $Al_{0.2}Ga_{0.8}As$ bandgap energy. Be sure to include both C–HH and C–LH transitions.

(b) Plot $\rho_r(E)$ for a fictitious 100 Å GaAs QW with infinitely high barriers, assuming parabolic subbands and including all *allowed* C–HH and C–LH subband pairs between the GaAs and $Al_{0.2}Ga_{0.8}As$ bulk bandgaps. Overlay the bulk GaAs $\rho_r(E)$ on top of this curve. Discuss any correspondence you observe between these curves.

(c) Now plot the QW $\rho_r(E)$ in (b) for C–HH transitions *only* and overlay the C–HH component of the bulk GaAs $\rho_r(E)$ on top of this curve. Make a similar comparison for C–LH transitions. Discuss any correspondence you observe between these curves.

(d) Mathematically derive the reasons why the comparisons in (c) work the way they do. *Hint*: Begin by expressing the step edges of the QW $\rho_r(E)$ in terms of the quantum numbers, and then (after some substitutions) compare with the bulk $\rho_r(E)$).

(e) Now plot a realistic QW $\rho_r(E)$ assuming a 100 Å GaAs QW with $Al_{0.2}Ga_{0.8}As$ barriers (assume that 60% of the band discontinuity occurs in the conduction band). In your plot, assume parabolic subbands and include all allowed C–HH and C–LH transitions between the GaAs and $Al_{0.2}Ga_{0.8}As$ bulk bandgaps. Also, add to this plot the contributions to $\rho_r(E)$ from the $Al_{0.2}Ga_{0.8}As$ bulk barriers by extending the energy range another 50%. Overlay the bulk GaAs $\rho_r(E)$ on top of this curve. In comparing the bulk GaAs to the GaAs QW, do you qualitatively find much difference in the *total* density of state pairs existing within the energy range plotted?

3. The measured absorption curve of a GaAs/$Al_{0.2}Ga_{0.8}As$ 100 Å QW (with excitonic effects removed) reveals that the C1–HH1 absorption step has a

magnitude of 10^4 cm^{-1} for light with polarization such that the electric field lies in the plane of the QW (TE polarization).

(a) Determine the differential lifetime, τ_{sp}^{21}, for this transition.

(b) With the TE absorption known, we can estimate the TE gain spectrum for different pumping levels. Plot the expected gain spectrum from the QW band edge to the bulk $Al_{0.2}Ga_{0.8}As$ bandgap energy, for two different injection levels: (1) $E_{Fc} \gg E_{c1}$ (i.e., $E_{Fc} \to \infty$) and $E_{Fv} = E_{hh1}$, and (2) $E_{Fc} \gg E_{c1}$ and $E_{Fv} = E_{hh2}$. Neglect forbidden transitions, C–LH transitions, and bulk barrier transitions in your plot. Also assume that the density of state pairs and the overlap integral is the same for all subband transition pairs. Further assume that the transition matrix element is not a function of energy. Neglect lineshape broadening in your calculation.

(c) What is the *maximum* possible spontaneous emission rate per unit energy at the C1–HH1 subband edge of this QW, assuming the rate is the same for light polarized in any direction?

(d) Considering only C–HH subband transition pairs, what are the radiative currents required to support the two gain spectra plotted in (b), assuming a single QW active region with lateral dimensions of 2 μm × 200 μm? In your estimation, assume that the optical mode density and g_{max} are constants evaluated at the C1–HH1 subband edge. Also assume the emission rate is the same for light polarized in any direction. Be sure to include spontaneous transitions over *all* energies by letting $E_{Fc} \to \infty$.

(e) Can you explain why the quasi-Fermi level positions given in (b) are unrealistic?

4. The previous problem was simplified by allowing $E_{Fc} \to \infty$, while E_{Fv} remained finite. For this situation, the electron density would be much higher than the hole density. In reality, charge neutrality in the active region under high injection conditions requires that the electron and hole densities be equal, making the previous problem unrealistic (but useful for academic purposes). A different, more realistic simplification, which includes charge neutrality, can be made by neglecting *all* but the lowest C1–HH1 subband transition pair in our analysis.

In this problem, we will make use of this approximation to explore the effect of the effective mass asymmetry between the conduction and heavy-hole band. Defining $D = m_{hh}/m_c$, show that charge neutrality leads to the following relationship between the Fermi functions at the C1 and HH1 subband edges (neglecting all other subband and bulk transitions pairs):

$$f_1(E_{hh1}) = (1 - f_2(E_{c1}))^{1/D}. \tag{4.111}$$

Using Eq. (4.111), plot the following quantities as a function of the fraction of filled states at the C1 subband edge (from 0 to 1), for symmetric ($D = 1$) and asymmetric ($D = 5$) bands (assume m_c is the same in both cases):

(a) The fraction of *empty* states at the HH1 subband edge.

(b) The stimulated absorption and stimulated emission Fermi factors.

(c) The subband edge gain normalized to the maximum possible gain in each case.

(d) The carrier density normalized to the density at $f_2(E_{c1}) = 0.5$.
These plots reveal fundamental differences between the symmetric and asymmetric band structures in the way they behave under nonequilibrium conditions.

(e) Which of the two band structures reaches transparency first as a function of the fraction of filled C1 subband edge states? What is the main cause of this?

(f) Does the symmetric or the asymmetric band structure have the lowest transparency carrier density (the carrier density at which $g = 0$)? What are the values for the normalized carrier density in each case? Assuming the radiative current density roughly follows a BN^2 dependence, how different do you expect the transparency current densities to be?

(g) Derive an expression for the differential gain, dg/dN, explicitly including the dependence of g_{max} on D and m_c (neglect lineshape broadening; it may be helpful to begin by setting $dg/dN = (dg/df_2(E_{c1}))/(dN/df_2(E_{c1}))$). Does the symmetric or asymmetric band structure provide a higher differential gain at transparency? By how much? Can you explain this behavior qualitatively?

(h) The application of strain in the plane of the QW is known to reduce the in-plane heavy-hole effective mass dramatically. From the preceding considerations, and what you know about the effective mass asymmetry of unstrained GaAs QWs, can you predict qualitatively what changes in performance in strained QWs can be expected? Should we be using strained QWs in the active regions of lasers? What benefits do you foresee?

5. Assume we have a strained QW that has a completely symmetric subband structure in the conduction and valence bands including subband curvatures and subband spacings (also assume that only heavy-hole subbands exist). Now, the modal absorption loss (w/o excitonic effects) of the lowest subband transition is measured to be 400 cm^{-1} in an unpumped laser. If we want to ensure that the $n = 1$ gain is always larger than the $n = 2$ gain for modal gain 200 cm^{-1}, what is the minimum subband spacing between the two lowest conduction band states that we can tolerate? Assume g_{max} is constant and of equal magnitude for each subband transition, and also neglect lineshape broadening effects.

6. Comparing Appendix 6 to Chapter 4, what is τ_{sp}^{21} in terms of $|M_T|^2$?

7. For the sake of completeness, the setup of this problem repeats the setup of Example 4.5. where we've seen than the surface recombination velocity can be estimated using the simple "broad-area" (i.e., infinite stripe width) threshold carrier density. In reality the carrier density profile will vary over the cross

section of the active region, particularly when the active width is narrow. In this problem, the effects of a finite diffusion constant for carriers in the active region will be examined.

Assume that the carrier densities in the active region are high enough that any differences in the diffusion profiles of electrons and holes will set up an electric field that will pull the two densities to nearly the same profile. In this *ambipolar* diffusion limit, the hole diffusion rate is enhanced by a factor of ~2 by the forward pull of the electrons, and the electron diffusion rate is limited to approximately twice the normal hole diffusion rate by the backward pull of the holes. The overall effect is that we can assume the electron and hole densities are equal everywhere in the active region and are characterized by a single ambipolar diffusion constant, D_{np}. The lateral profile of carriers is then governed by the simple diffusion equation:

$$D_{np}\frac{d^2N(x)}{dx^2} = -\frac{I(x)}{qV} + \frac{N(x)}{\tau_{np}}. \tag{4.112}$$

The carrier lifetime is in general a function of N; however, to obtain analytic solutions, we can evaluate the lifetime at the broad-area threshold value, $\tau_{np}|_{th} = qL_zN_{th}/J_{th}$.

The problem we wish to solve is the carrier density profile across the width of the active region in the in-plane laser depicted in Fig. 4.19. For this case, we can define two distance regions: one beneath the contact within w, where we assume a uniform current injection profile, and the region outside w where there is no current injection. Mathematically, with $x = 0$ defined as the center of the stripe, we have $I(x) = I_0$ for $x < w/2$, and $I(x) = 0$ for $x > w/2$. In fabricating the laser we can either leave the active region in place outside the stripe, or we can remove it by etching through the active region outside the contact area. The first case leads to carrier outdiffusion, whereas the second case leads to surface recombination. We would like to compare these two cases.

(a) With the active region in place away from the contact, carriers are free to diffuse outside the stripe width. Draw a sketch of this configuration and solve Eq. (4.112) for $N(x)$ in and out of the stripe assuming the carrier density and its derivative (i.e., the diffusion current) are constant across the $x = w/2$ boundary. Qualitatively sketch $N(x)$.

(b) With the active region etched away, the carriers recombine at the surface. Draw a sketch of this configuration and solve Eq. (4.112) for $N(x)$ under the stripe, assuming the diffusion current (defined by the slope of the carrier density) is equal to the surface recombination current, $D_{np}dN/dx = -v_sN$, at the $x = w/2$ boundary. Place your result in terms of the diffusion equivalent surface recombination velocity, $v_{sD} = \sqrt{D_{np}/\tau_{np}}$. Qualitatively sketch $N(x)$. In comparing $N(x)$ in (ii) to $N(x)$ in (i), what is the significance of v_{sD}? Show that Eq. (4.78) is recovered in the limit of $D_{np} \to \infty$.

Assuming an injection current of 166.4 A/cm^2 (corresponding to the broad-area threshold in Fig. 4.18) and a stripe width of 2 μm, plot $N(x)$ on

the same graph for four different cases: (1) $N(x)$ found in (i), (2) $N(x)$ found in (ii) with $v_s = 1 \times 10^4$ cm/s, (3) $v_s = 1 \times 10^5$ cm/s, and (4) $v_s = 5 \times 10^5$ cm/s. Assume $D_{np} = 20$ cm^2/s for all cases. Produce similar plots for stripe widths of 10 μm and 20 μm.

The single quantum-well in-plane laser defined in Fig. 4.20 is expected to lase with the above injected current density. However, the finite stripe width reduces N, which reduces the gain, preventing the onset of lasing at 166.4 A/cm^2. Assuming that the injection current must establish a carrier density *in the center of the stripe*, which is equal to the broad-area threshold carrier density in order to lase, calculate and tabulate the new threshold current density for all 12 cases considered above. Compare the threshold current densities for (2) through (4) to the values found using the simple techniques discussed in relation to Eq. (4.80). Answer the following questions:

(a) How does the inclusion of carrier diffusion affect the estimated magnitude of the surface recombination current?

(b) When is carrier diffusion important to consider?

(c) How does the threshold current density with carrier outdiffusion (profile (1)) compare to the three different recombination velocity profiles? Can you use v_{sD} to understand this comparison?

(d) For InGaAs QW active regions, is it better to etch through the active region?

(e) For GaAs QW active regions, is it better to etch through the active region? In reality, the threshold lasing condition will be more complex than simply setting $N(0) = N_{th}$ because the material gain will be relaxed near the stripe edges, reducing the overall *modal* gain.

(f) Qualitatively discuss how the above-threshold current densities would be modified by the inclusion of this effect.

8. The 300 K Auger recombination coefficient, C, is measured to be 10^{-29} cm^6/s for some material with a bandgap of 1 eV and an Auger threshold energy that is 10% higher.

(a) If the temperature were increased to 400 K, what is the new C?

(b) If the material were changed to have a bandgap of 2 eV with everything else remaining the same, and the temperature then increased to 400 K, what is the new C?

9. Discuss the effects of strain on the threshold energy for the CCCH Auger process by constructing a diagram analogous to Fig. 4.22. Assume $D = 5$ in one case and $D = 1$ in the other (see Problem 4 for the definition of D). You may find Appendix 12 helpful in quantifying the exact positions of the four states in each case (use bulk GaAs values for m_c and E_g). How would you expect this difference in the threshold energy to affect the Auger transition rate for a given electron and hole density?

10. The one pass absorption for probe light passing perpendicularly across a single unpumped strained InGaAs 80 Å quantum well is determined to be 1% for a wavelength 20 nm shorter than the absorption edge, which is at 1 μm. A 1000-μm-long broad-area, cleaved-facet edge-emitting laser with a transverse confinement factor of 1% made from this material is found to have threshold and transparency current densities of 100 and 70 A/cm^2, respectively, at about the same wavelength. Additional measurements indicate an internal loss of 5 cm^{-1}.

(a) What is the threshold material gain?

(b) What is the Fermi function difference $(f_2 - f_1)$ at threshold?

(c) What are g_0 and J_{tr} in a two-parameter logarithimic fit to this data?

(d) What is the Fermi function difference $(f_2 - f_1)$ at a current of twice threshold in the laser?

(e) What is the approximate β_{sp}?

(f) If the laser were cut to half its length (i.e., 500 μm) what would the new threshold current density and Fermi function difference at threshold be?

CHAPTER FIVE

Dynamic Effects

5.1 INTRODUCTION

In Chapter 2, the rate equations for both the carrier and the photon density of a given optical mode, in the active region were developed from simple intuitive arguments. The below-threshold and above-threshold limits to the steady-state solutions of the rate equations were considered to give a feel for the operating characteristics of the laser. We also took a first look at the small-signal intensity modulation response of the laser. In this chapter we wish to expand on these simplified discussions considerably, drawing on the results of Chapter 4 for the evaluation of the various generation and recombination rates.

We will first summarize many of the results obtained in Chapter 2. This review will serve as a convenient reference to the relevant formulas. We will then move on to dynamic effects, starting with a differential analysis of the rate equations. We will derive the small-signal intensity and frequency modulation response of the laser as well as the small-signal transient response. We then consider large-signal solutions to the rate equations. The turn-on delay of the laser and the general relationship between frequency chirping and the modulated output power will also be treated here. Next we consider laser noise. Using the Langevin method, we will determine the laser's relative intensity noise (RIN) and the frequency fluctuations of the laser's output from which we can estimate the spectral linewidth of the laser. We will also discuss the role of the injection current noise in determining the RIN at high power and examine the conditions necessary for noise-free laser operation.

Diode Lasers and Photonic Integrated Circuits, Second Edition.
Larry A. Coldren, Scott W. Corzine, and Milan L. Mašanović.
© 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

In the final two sections we consider dynamic effects in specific types of lasers. The first section considers carrier transport limitations in separate-confinement heterostructure (SCH) quantum-well lasers. The second deals with the effects of weak external feedback in extended cavity lasers.

5.2 REVIEW OF CHAPTER 2

Figure 5.1 summarizes the reservoir model used to develop the rate equations in Chapter 2. To ensure particle conservation, each arrow in the flowchart represents the *number of particles* flowing per unit time. This is why the rates per unit volume and the densities are multiplied by the active-region volume V of the carrier reservoir or the mode volume V_p of the photon reservoir. Before deriving the rate equations, we will briefly follow the flowchart from input to output.

Starting with the carrier reservoir, we have the rate of carrier injection into the laser I/q. Of these carriers only $\eta_i I/q$ reach the active region, where η_i is the *injection* or *internal* efficiency of the laser introduced in Chapter 2. The rest recombine elsewhere in the device (this includes all carriers recombining at rates not directly tied to the carrier density in the active region). Once in the carrier reservoir, the carriers have a number of options. Some recombine via nonradiative recombination at the rate $R_{nr}V$. Others recombine spontaneously at the rate $R_{sp}V$, of which a certain fraction emit photons into the mode of interest at the rate $R'_{sp}V$ (or $\beta_{sp}R_{sp}V$ where β_{sp} is the spontaneous emission factor). Other carriers recombine via stimulated emission into the mode of interest at the rate $R_{21}V$. Although photons in other modes can also induce stimulated recombination of carriers, we limit ourselves initially to a single mode. Finally, photons in the photon reservoir

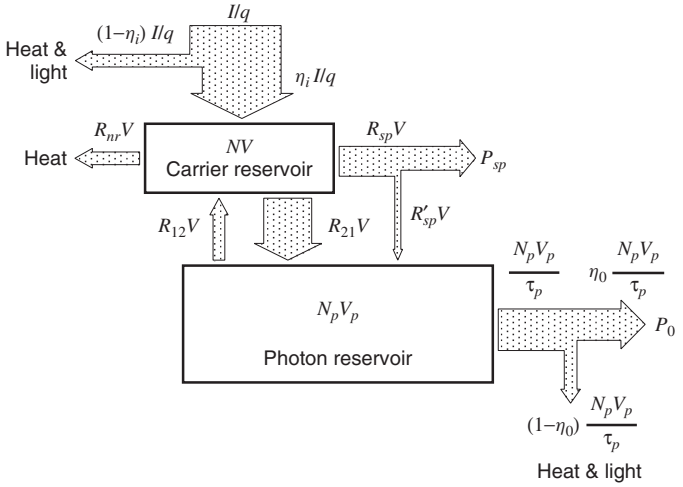


FIGURE 5.1: Model used in the rate equation analysis of semiconductor lasers.

can be absorbed, serving as an additional source of carriers, which are generated at the rate $R_{12}V$. In Chapter 2, we also considered carriers that might leak out of the active region via lateral diffusion and/or thermionic emission at the rate R_lV. However, because R_lV is often negligible and furthermore plays a role identical to $R_{nr}V$ in the rate equations, we will not include it explicitly in this chapter (if we wish to include carrier leakage at any time, we can set $R_{nr} \rightarrow R_{nr} + R_l$).

In the photon reservoir, the stimulated and spontaneous emission rates into the mode provide the necessary generation of photons at the rate $R_{21}V + R'_{sp}V$. Stimulated absorption in the active region depletes photons at the rate $R_{12}V$. All other photons leave the cavity at the rate N_pV_p/τ_p. Of those leaving the cavity, only $\eta_0 N_p V_p/\tau_p$ leave through the desired mirror to be collected as useful output power, P_0, where η_0 is the *optical* efficiency of the laser to be defined below. The rest of the photons exit the cavity through a different mirror or disappear through: (1) free carrier absorption in the active region (which does *not* increase the carrier density), (2) absorption in materials outside the active region, and/or (3) scattering at rough surfaces.

5.2.1 The Rate Equations

By setting the time rate of change of the carriers and photons equal to the sum of rates into, minus the sum of rates out of the respective reservoirs, we immediately arrive at the carrier and photon *number* rate equations:

$$V\frac{dN}{dt} = \frac{\eta_i I}{q} - (R_{sp} + R_{nr})V - (R_{21} - R_{12})V, \tag{5.1}$$

$$V_p\frac{dN_p}{dt} = (R_{21} - R_{12})V - \frac{N_pV_p}{\tau_p} + R'_{sp}V. \tag{5.2}$$

Setting $R_{21} - R_{12} = v_g g N_p$ using Eq. (4.37), dividing out the volumes, and using $\Gamma = V/V_p$ (by definition of V_p), we obtain the *density* rate equations derived in Chapter 2 in slightly more general form:

$$\frac{dN}{dt} = \frac{\eta_i I}{qV} - (R_{sp} + R_{nr}) - v_g g N_p, \tag{5.3}$$

$$\frac{dN_p}{dt} = \left[\Gamma v_g g - \frac{1}{\tau_p}\right] N_p + \Gamma R'_{sp}. \tag{5.4}$$

It is a matter of preference to use either the density or the number rate equations in the analysis of lasers.[1] Throughout this book we choose to use the density versions. This choice forces us to do a little more book-keeping by explicitly including V and V_p in various places (as we will see in Section 5.5) but in the end places results

[1]It is important to note that the gain term in the density versions has a Γ in the photon density rate equation, but not in the carrier density rate equation. While in the number rate equations, the gain term is symmetric. This asymmetry in the density rate equations is often overlooked in the literature.

in terms of the more familiar carrier *density* and related terms (e.g., the differential gain, the *A, B*, and *C* coefficients, etc.).

To complete the description, the output power of the mode and total spontaneous power from all modes are given by

$$P_0 = \eta_0 h\nu \frac{N_p V_p}{\tau_p} \quad \text{and} \quad P_{sp} = h\nu R_{sp} V, \tag{5.5}$$

where

$$\eta_0 = F \frac{\alpha_m}{\alpha_m + \langle \alpha_i \rangle}. \tag{5.6}$$

The optical efficiency defined here times the injection efficiency yields the differential quantum efficiency defined in Chapters 2 and 3, $\eta_d = \eta_i \eta_0$. The prefactor F in Eq. (5.6) is the fraction of power not reflected back into the cavity that escapes as useful output from the output coupling mirror as derived in Chapter 3. The mirror loss can usually be defined as $\alpha_m = (1/L)\ln(1/r_1 r_2)$, and $\langle \alpha_i \rangle$ is the spatial average over any internal losses present in the cavity. The photon lifetime is given by

$$\frac{1}{\tau_p} = v_g(\alpha_m + \langle \alpha_i \rangle) = \frac{\omega}{Q}, \tag{5.7}$$

where v_g is the group velocity of the mode of interest including both material and waveguide dispersion. The more general definition of τ_p in terms of the cavity Q is useful in complex multisection lasers where α_m can prove difficult to define (see the latter part of Appendix 5). In such cases, we can use the definition: $Q \equiv \omega$ (energy stored in cavity)/(total power lost).

We can use Eq. (4.67) to set $R'_{sp} = \Gamma v_g g n_{sp}/V$ in Eq. (5.4). However, a useful approximation to R'_{sp} is found by setting $R'_{sp} \rightarrow \beta_{sp} R_{sp}$ and evaluating β_{sp} at its threshold value:

$$\beta_{sp} \approx \beta_{sp}\Big|_{th} = \frac{\Gamma v_g g n_{sp}}{\eta_i \eta_r I/q}\bigg|_{th} = \frac{q}{I_{th}\tau_p}\left[\frac{n_{sp}}{\eta_i \eta_r}\right]_{th}, \tag{5.8}$$

where Eq. (4.82) was used to define β_{sp} and $\eta_r = R_{sp}/(R_{sp} + R_{nr})$. The latter equality makes use of Eq. (5.11). The term in brackets at threshold is typically between 1.5 and 2 in short-wavelength materials, but it can be as large as 10 in long-wavelength materials due to low radiative efficiencies (high Auger recombination).

5.2.2 Steady-State Solutions

The steady-state solutions of Eqs. (5.3) and (5.4) are found by setting the time derivatives to zero. Solving Eq. (5.4) for the steady-state photon density and

Eq. (5.3) for the DC current, we have

$$N_p(N) = \frac{\Gamma R'_{sp}(N)}{1/\tau_p - \Gamma v_g g(N)}, \tag{5.9}$$

$$I(N) = \frac{qV}{\eta_i}(R_{sp}(N) + R_{nr}(N) + v_g g(N) N_p(N)). \tag{5.10}$$

As the functionality implies, it is useful to think of N as the independent parameter of the system, which we can adjust to determine different values of both current and photon density. Solving for $v_g g N_p$ in Eq. (5.9), and using Eq. (5.11) below to define g_{th}, Eq. (5.10) can alternatively be written as

$$I(N) = \frac{qV}{\eta_i}(R_{sp}(N) - R'_{sp}(N) + R_{nr}(N) + v_g g_{th} N_p(N)). \tag{5.10$'$}$$

In this form, the right-hand side more clearly identifies where the current eventually goes. In fact, Eq. (5.10′) can be derived by simply drawing a box around the *entire* carrier–photon system in Fig. 5.1 and equating the inward flow with the sum of all outward flows (as opposed to Eq. (5.10), which is derived by drawing a box only around the carrier reservoir).

To gain more understanding of how Eqs. (5.9) and (5.10) behave, we will first consider the limiting forms (below and above threshold) and then we will plot the equations over the entire range. The threshold gain and carrier density as defined in Chapter 2 are

$$\Gamma v_g g_{th} \equiv \frac{1}{\tau_p} \quad \text{and} \quad g(N_{th}) = g_{th}. \tag{5.11}$$

The first two cases below summarize (1) what happens when the gain and carrier density are much smaller than Eq. (5.11) and (2) what happens as they approach their threshold values.

Case (i): Well Below Threshold

For $N \ll N_{th}$ and $\Gamma v_g g \ll 1/\tau_p$, we can neglect $\Gamma v_g g$ in the denominator of Eq. (5.9), and the steady-state solutions become

$$N_p(N) = \Gamma R'_{sp}(N)\tau_p \approx 0, \tag{5.12}$$

$$I(N) = \frac{qV}{\eta_i}(R_{sp}(N) + R_{nr}(N)), \tag{5.13}$$

and

$$P_0 \approx 0 \quad \text{and} \quad P_{sp} = \eta_i \eta_r \frac{h\nu}{q} I. \tag{5.14}$$

With negligible power in the mode, the injection current only needs to resupply carriers lost to spontaneous and nonradiative recombination. The output power is close to zero, and the spontaneous power increases approximately linearly with injected current (if Auger recombination is significant, the increase will be sublinear because η_r decreases). These solutions were also derived in Eqs. (2.5) and (2.7). As we approach threshold, the denominator of Eq. (5.9) becomes small, and the power in the mode starts to build up.

Case (ii): Above Threshold

As $N \to N_{th}$ and $\Gamma v_g g \to 1/\tau_p$, we can evaluate all terms in Eqs. (5.9) and (5.10) at N_{th} *except* for the denominator of Eq. (5.9), which contains the difference between the threshold gain and the actual gain:

$$N_p(N) = \frac{R'_{sp}(N_{th})/v_g}{g_{th} - g(N)}, \tag{5.15}$$

$$I(N) = \frac{qV}{\eta_i}(R_{sp}(N_{th}) + R_{nr}(N_{th})) + \frac{qV}{\eta_i} v_g g_{th} N_p(N), \tag{5.16}$$

and

$$P_0 = \eta_i \eta_0 \frac{h\nu}{q}(I - I_{th}) \quad \text{and} \quad P_{sp} = \eta_i \eta_r \frac{h\nu}{q} I_{th}. \tag{5.17}$$

The output power in (5.17) was found by rearranging (5.16) and recognizing that $V v_g g_{th} N_p = P_0/(\eta_0 h\nu)$. The *threshold current* of the laser is defined as

$$I_{th} = \frac{qV}{\eta_i}(R_{sp}(N_{th}) + R_{nr}(N_{th})). \tag{5.18}$$

In this above-threshold limit, the output power increases linearly with current, and the spontaneous power saturates at the level found at threshold. From (5.15), it is clear that N and g never actually reach N_{th} and g_{th} for finite output powers (and finite current). They remain ever so slightly below their "threshold" values. These solutions were also derived in Eqs. (2.35) and (2.41).

In lasers where β_{sp} is relatively large, we need to revise our derivation slightly. For example, comparing Eq. (5.16) to Eq. (5.10) reveals that we have inadvertently dropped a factor, $-R'_{sp}(N_{th})$, in arriving at Eq. (5.16) (it happened when we set $g \to g_{th}$). The more exact threshold current is found by setting $N_p \to 0$ in Eq. (5.10′). Equation (5.18) is then revised to

$$I_{th} = \frac{qV}{\eta_i}((1 - \beta_{sp}(N_{th}))R_{sp}(N_{th}) + R_{nr}(N_{th})). \tag{5.18′}$$

The threshold current as defined here represents the offset in the *LI* curve at output power levels generated by carrier densities close to N_{th}. If nonradiative recombination is negligible, then as $\beta_{sp}(N_{th}) \to 1$, the threshold current reduces to zero. In such a "thresholdless" laser, *all* injected current is funneled into the lasing mode. For typical lasers, $\beta_{sp}(N_{th}) \ll 1$ and (5.18′) reduces to Eq. (5.18).

Using Eq. (5.17), the current above threshold can be written as

$$I = I_{th} + \frac{q}{h\nu}\frac{P_0}{\eta_i \eta_0}. \tag{5.18''}$$

Case (iii): Below and Above Threshold

To observe the transition between below threshold and above, it is useful to plot (5.9) as a function of (5.10). Parameterized light–current ($L-I$) curves found by varying N from 0 to N_{th} are shown in Fig. 5.2 for a typical in-plane laser and three VCSELs. The details of the VCSEL and in-plane structures are given later in Table 5.1. The plots use Table 4.4 for $g(N)$ and the combination of Tables 4.4 and 4.5 for $R_{sp}(N) + R_{nr}(N)$ (see Eq. (4.123)). We can find $R'_{sp}(N)$ directly from gain calculations[2] (solid curves in upper plot) or we can approximate $R'_{sp}(N)$ using Eq. (5.8) to define the threshold value of β_{sp} (dashed curves in upper plot). For these examples, β_{sp} actually decreases by about 30% as the current increases to threshold. As a result, the approximate dashed curves underestimate the power level by about 30% at very low currents. Above threshold the output power becomes independent of the value of β_{sp}, as Eq. (5.17) verifies.

Each of the "idealized" $L-I$ curves[3] in the upper plot of Fig. 5.2 displays a sudden ~30 dB increase in output power (or ~20 dB for the smaller VCSELs). This dramatic change of course corresponds to the lasing threshold of the laser and reveals that the transition to lasing is quite sharp. Below the lasing threshold, the power in the mode comes primarily from the spontaneous emission rate, $\Gamma R'_{sp}$. This power is typically well under 1 μW and is larger for smaller devices because the fraction of spontaneous emission into the lasing mode increases with decreasing mode volume (see Eq. (4.67)). As the current increases, the gain approaches the loss, and the stimulated emission rate increases dramatically. For the largest VCSEL and the in-plane laser this occurs at ~0.7 mA and ~1.1 mA, whereas for the

[2]Setting $R'_{sp}(N) = \Gamma v_g g n_{sp}/V$, the gain calculations can be used to find both g and n_{sp} as a function of N. For the InGaAs/GaAs 80 Å QW considered in Section 4.6, the calculated gain–population inversion factor product can be modeled very well (for $g < 2500\ \text{cm}^{-1}$) by $g n_{sp} = (850\ \text{cm}^{-1}) \times \ln[1 + N^2/2]$ where N is in units of $10^{18}\ \text{cm}^{-3}$ (if desired we can combine this with the curve fit for g to estimate n_{sp} for $g > 0$).

[3]The $L-I$ curves are ideal in the sense that surface recombination and heating at high injection levels have been neglected. In realistic devices, surface recombination will require additional current to reach threshold (particularly in the smaller devices), and heating will limit the maximum output power of the devices.

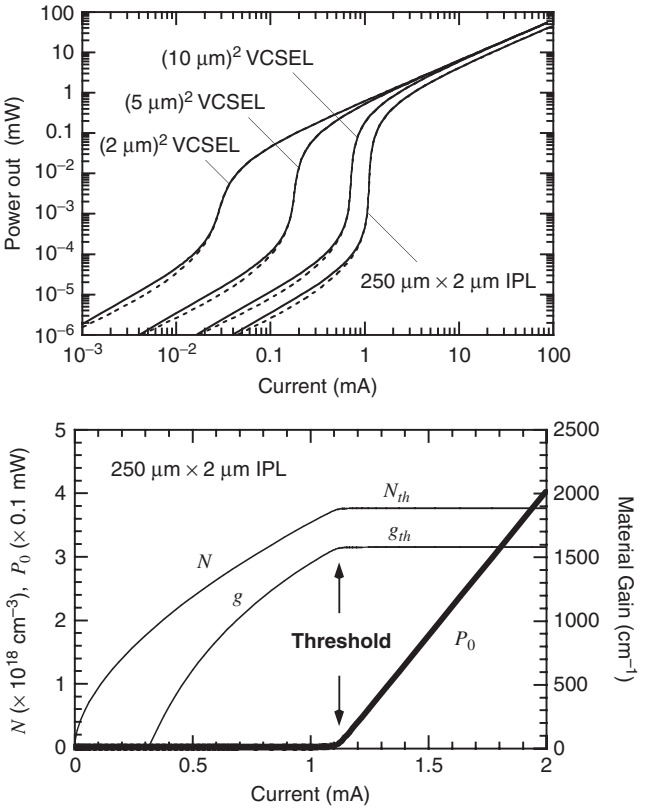


FIGURE 5.2: Upper plot: light vs. current in two different lateral-size 3-QW VCSELs and in a single QW in-plane laser (IPL) (all lasers use InGaAs/GaAs 80 Å QWs). Lower plot: Light vs. current on a linear scale for the same in-plane laser. Plot also shows carrier density and material gain vs. current.

smaller VCSELs only ∼0.2 mA and ∼30 μA are required due to the smaller active volumes. Above the lasing threshold all curves approach a unity slope, which on a log-log plot implies a linear relationship between power and current.

The linear $L - I$ relationship above threshold for the in-plane laser is more clearly illustrated in the lower plot in Fig. 5.2. The $L - I$ curve near threshold, which was approximated in Chapter 2 as a discontinuous change in the $L - I$ slope, is more accurately a *knee* in the curve. This knee becomes softer for smaller devices due to the higher spontaneous emission rate into the mode. However, when measured experimentally, the light collected into a detector often captures spontaneous emission from more than just the lasing mode, and this can also give the appearance of a softer transition to lasing.

The lower plot in Fig. 5.2 also displays the carrier density and material gain as a function of current. Below threshold, the approximately quadratic carrier density–current relationship and the logarithmic gain–current relationship can be observed.

For this example, transparency in the active material occurs at about one-third of the way toward threshold. At threshold, both the carrier density and gain become clamped near their threshold values. Beyond threshold, the carrier density and gain continue to increase; however, the increase is in the fractions of a percent range. Thus for all practical purposes, above threshold we can set $N = N_{th}$ and $g = g_{th}$ as we did in Chapter 2. The only time we cannot use this approximation is when we need the difference $N_{th} - N$ or $g_{th} - g$.

The carrier clamping mechanism illustrated so dramatically in Fig. 5.2 is perhaps best understood by defining a *stimulated* carrier lifetime τ_{st} and writing the total carrier recombination rate as

$$R_{tot}(N) = \frac{N}{\tau_{sp}} + \frac{N}{\tau_{nr}} + \frac{N}{\tau_{st}}, \tag{5.19}$$

where

$$\frac{1}{\tau_{sp}} + \frac{1}{\tau_{nr}} = \frac{R_{sp} + R_{nr}}{N}, \tag{5.20}$$

$$\tau_{st} = \frac{N}{v_g g N_p} \approx \frac{N_{th}}{v_g g_{th}} \cdot \frac{1}{N_p}. \tag{5.21}$$

The inverse dependence of the stimulated carrier lifetime on photon density sets up a negative feedback loop, which prevents N from increasing beyond its threshold value. For example, as the current increases and the clamped carrier density increases closer to N_{th}, the gain moves incrementally closer to g_{th}, which increases the output power through Eq. (5.15). This increase in output power reduces τ_{st} just enough to prevent any further increase in N. Carrier clamping is therefore maintained via the photon density's control over the stimulated carrier lifetime. As a result of this interplay, all injected current above threshold $\eta_i(I - I_{th})$ is simply eaten up by the self-adjusting N/τ_{st}. For this reason, $\eta_i(I - I_{th})$ is referred to as the *stimulated emission current*. Note that when $1/\tau_{st} = 1/\tau_{sp} + 1/\tau_{nr}$ (i.e., when $\eta_i(I - I_{th}) = \eta_i I_{th}$), we are at twice threshold.

5.2.3 Steady-State Multimode Solutions

As discussed in Chapter 3, if the laser cavity is not carefully designed, it is inevitable that a number of different resonant modes of the cavity (either axial or lateral) will require similar amounts of gain to reach threshold. To make matters worse, the gain spectrum in semiconductor materials is very broad and can typically provide almost uniform gain across at least a few mode spacings. In such cases, the gain difference $g_{th} - g$ in Eq. (5.15) can be similar for different modes of the cavity at a given injection level. As a result, the photon density can build up in more than one mode.

In such multimode situations, the steady-state solution for the photon density in each mode remains the same as before; however, the steady-state current must

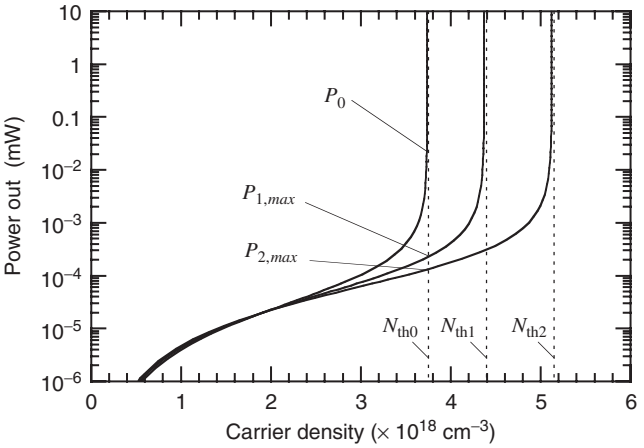


FIGURE 5.3: Light vs. carrier density for three different modes in an SQW in-plane laser. The threshold modal gains for modes 0, 1, and 2 are 50, 60, and 70 cm^{-1}, respectively.

include contributions from all m modes:

$$N_{pm} = \frac{\Gamma_m R'_{spm}}{1/\tau_{pm} - \Gamma_m v_{gm} g_m}, \tag{5.22}$$

$$I = \frac{qV}{\eta_i}\left(R_{sp} + R_{nr} + \sum_m v_{gm} g_m N_{pm}\right). \tag{5.23}$$

Figure 5.3 plots the steady-state photon density Eq. (5.23) for three modes of the in-plane laser considered earlier, assuming the threshold modal gains for mode 0, 1, and 2 are 50, 60, and 70 cm^{-1}, respectively. From earlier discussions we know the carrier density cannot increase beyond N_{th0} for finite current levels. As a result, the power in modes 1 and 2 cannot increase beyond the levels indicated in the figure. Thus the power in the *side modes* saturates just as the spontaneous power saturates at threshold.[4] However, the closer N_{th1} and N_{th2} are to N_{th0}, the higher this maximum saturated power level will be. The mode suppression ratio (defined as the ratio of power in the main mode to the power in either of the side modes) can be determined from Eq. (5.22) as worked out in Chapter 3. Note that due to the clamping of power in the side modes, the MSR increases linearly with the power in the main mode.

[4]Actually, gain compression created by high photon densities can force N to increase in order to maintain $g_0 \sim g_{th0}$. As a result, with increasing power in the main mode, N can increase slightly beyond N_{th0}, and the side mode power can to some degree increase beyond the levels indicated in Fig. 5.3.

5.3 DIFFERENTIAL ANALYSIS OF THE RATE EQUATIONS

When we want to observe how lasers behave dynamically in response to some perturbation to the system such as a modulation of the current, we must analyze Eqs. (5.3) and (5.4) with the time derivatives included. Unfortunately, exact analytical solutions to the full rate equations cannot be obtained. Therefore if analytical solutions are desired we must make some approximations. In this section we analyze the rate equations by assuming that dynamic changes in the carrier and photon densities away from their steady-state values are small. Such *small-signal* responses of one variable in terms of a perturbation to another can be accommodated by taking the *differential* of both rate equations. Considering I, N, N_p, and g as dynamic variables, the differentials of Eqs. (5.3) and (5.4) become

$$d\left[\frac{dN}{dt}\right] = \frac{\eta_i}{qV} dI - \frac{1}{\tau_{\Delta N}} dN - v_g g dN_p - N_p v_g dg, \tag{5.24}$$

$$d\left[\frac{dN_p}{dt}\right] = \left[\Gamma v_g g - \frac{1}{\tau_p}\right] dN_p + N_p \Gamma v_g \; dg + \frac{\Gamma}{\tau'_{\Delta N}} dN, \tag{5.25}$$

where

$$\frac{1}{\tau_{\Delta N}} = \frac{dR_{sp}}{dN} + \frac{dR_{nr}}{dN} \approx A + 2BN + 3CN^2, \tag{5.26}$$

$$\frac{1}{\tau'_{\Delta N}} = \frac{dR'_{sp}}{dN} \approx 2\beta_{sp} BN + \frac{d\beta_{sp}}{dN} BN^2. \tag{5.27}$$

The *differential* carrier lifetime $\tau_{\Delta N}$ depends on the *local* slope dR/dN, whereas the total carrier lifetime τ depends on the overall slope R/N. Due to the mixed quadratic and cubic dependence of R on N, $\tau_{\Delta N}$ is typically a factor of two to three smaller than τ. The differential lifetime of carriers that radiate photons into the lasing mode $\tau'_{\Delta N}$ is typically in the tens of microseconds range and is negligible in most cases.[5]

Equations (5.24) and (5.25) can also be found by performing an expansion about the steady-state of the form $x(t) = x_0 + dx(t)$ for all dynamic variables. Neglecting product terms involving two or more small-signal terms and canceling out the steady-state solutions, we would obtain a set of approximate rate equations for $dN(t)$ and $dN_p(t)$ identical to Eqs. (5.24) and (5.25). This procedure was in fact carried out in Chapter 2. The differential approach simply provides a more direct path to the desired equations.

[5]A maximum limit can be placed on $1/\tau'_{\Delta N}$ by defining it as follows:

$$\frac{1}{\tau'_{\Delta N}} = \frac{v_g n_{sp}}{V_p} \frac{dg}{dN}\left[1 + \frac{g}{n_{sp}} \frac{dn_{sp}}{dg}\right] \leq \frac{v_g n_{sp}}{V_p} \frac{dg}{dN}.$$

The second term within brackets goes from -1 at transparency to 0 at infinite pumping levels. Thus, the latter inequality defines the largest possible value for $1/\tau'_{\Delta N}$ at positive gains.

The gain variation dg can be further expanded by assuming it is affected by both carrier and photon density variations:

$$dg = adN - a_p dN_p. \tag{5.28}$$

The sign convention reflects the fact that gain increases with increasing carrier density while it decreases or is *compressed* with increasing photon density. In Chapter 4 it was shown that the gain versus carrier density could be well approximated by a logarithmic function. The gain is also known to be inversely proportional to $(1 + \varepsilon N_p)$, where ε is a constant known as the *gain compression factor*. Thus, we can approximate the gain by

$$g(N, N_p) = \frac{g_0}{1 + \varepsilon N_p} \ln\left(\frac{N + N_s}{N_{tr} + N_s}\right). \tag{5.29}$$

With this expression the gain derivatives become

$$a = \frac{\partial g}{\partial N} = \frac{g_0}{(N + N_s)(1 + \varepsilon N_p)} \equiv \frac{a_0}{(1 + \varepsilon N_p)}, \tag{5.30}$$

$$a_p = -\frac{\partial g}{\partial N_p} = \frac{\varepsilon g}{(1 + \varepsilon N_p)}. \tag{5.31}$$

In Eq. (5.30), a_0 is defined as the *nominal* differential gain—the value of a with zero photon density (with no gain compression). Also note that neither a nor a_p are constants—both tend to get smaller at higher densities.

Replacing dg with Eq. (5.28), collecting like terms and defining a set of *rate* coefficients, the differential rate equations become

$$\frac{d}{dt}(dN) = \frac{\eta_i}{qV}\, dI - \gamma_{NN}\, dN - \gamma_{NP} dN_p, \tag{5.32}$$

$$\frac{d}{dt}(dN_p) = \gamma_{PN}\, dN - \gamma_{PP} dN_p, \tag{5.33}$$

where

$$\gamma_{NN} = \frac{1}{\tau_{\Delta N}} + v_g a N_p, \quad \gamma_{NP} = \frac{1}{\Gamma \tau_p} - \frac{R'_{sp}}{N_p} - v_g a_p N_p, \tag{5.34}$$

$$\gamma_{PN} = \frac{\Gamma}{\tau'_{\Delta N}} + \Gamma v_g a N_p, \quad \gamma_{PP} = \frac{\Gamma R'_{sp}}{N_p} + \Gamma v_g a_p N_p.$$

We have replaced g in these definitions using the steady-state relation: $1/\tau_p - \Gamma v_g g = \Gamma R'_{sp}/N_p$ (see Eq. (5.9)). To interpret the subscripts of the rate coefficients, it is helpful to remember "effect precedes cause." For example, γ_{NP} defines

the effect on N caused by changes in P (i.e., N_p). This mnemonic aid allows us to quickly associate γ_{NN} with the differential carrier lifetime, γ_{PP} with the effective photon lifetime, γ_{NP} with the gain, and γ_{PN} with the differential lifetime of carriers, which radiate into the mode. The change in gain Eq. (5.29) also adds a differential gain term $\propto a$ to the rates caused by N, γ_{NN} and γ_{PN}, and a gain compression term $\propto a_p$ to the rates caused by N_p, γ_{PP} and γ_{NP}. Aside from the small-signal assumptions, these rate coefficients contain no approximations. Well above threshold, N_p is large enough that a few terms can be dropped and the rate coefficients simplify to

$$\gamma_{NN} = 1/\tau_{\Delta N} + v_g a N_p, \quad \gamma_{NP} = 1/\Gamma\tau_p - v_g a_p N_p, \text{(above threshold)} \qquad (5.35)$$
$$\gamma_{PN} = \Gamma v_g a N_p, \qquad \gamma_{PP} = \Gamma v_g a_p N_p.$$

Table 5.1 gives values of these and related parameters for two specific laser structures (the values assume $N = N_{th}|_{\varepsilon=0}$ at and above threshold).

The rate coefficients have been introduced to allow us to conveniently describe the differential rate equations in a compact matrix form:

$$\frac{d}{dt}\begin{bmatrix} dN \\ dN_p \end{bmatrix} = \begin{bmatrix} -\gamma_{NN} & -\gamma_{NP} \\ \gamma_{PN} & -\gamma_{PP} \end{bmatrix}\begin{bmatrix} dN \\ dN_p \end{bmatrix} + \frac{\eta_i}{qV}\begin{bmatrix} dI \\ 0 \end{bmatrix}. \qquad (5.36)$$

In this form, the current is seen as the driving term or forcing function. Later when we treat noise in semiconductor lasers, the current forcing function will be replaced by noise sources. If necessary, we could also choose any other parameter to be the forcing function. For example, if we could somehow modulate the mirror loss, then the analysis would proceed by replacing the last term in (5.38) with $v_g N_p \begin{bmatrix} 0 \\ -d\alpha_m \end{bmatrix}$. The rate coefficients would remain unchanged.

For multimode small-signal analysis, we can extend Eq. (5.36) in a natural way:

$$\frac{d}{dt}\begin{bmatrix} dN \\ dN_{p1} \\ \vdots \\ dN_{pm} \end{bmatrix} = \begin{bmatrix} -\gamma_{NN} & -\gamma_{NP1} & \cdots & -\gamma_{NPm} \\ \gamma_{PN1} & -\gamma_{PP1} & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ \gamma_{PNm} & 0 & 0 & -\gamma_{PPm} \end{bmatrix}\begin{bmatrix} dN \\ dN_{p1} \\ \vdots \\ dN_{pm} \end{bmatrix} + \frac{\eta_i}{qV}\begin{bmatrix} dI \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \qquad (5.37)$$

The three rate coefficients γ_{PP}, γ_{PN}, and γ_{NP} for each mode m are still given by Eq. (5.36) with N_p replaced by N_{pm}. However, N_p in the rate coefficient γ_{NN} is replaced by a sum over all photon densities. The solution of Eq. (5.37) for two modes is considered in one of the problems at the end of the chapter. The zeros in the interaction matrix of Eq. (5.37) imply zero interaction or zero coupling between modes. In reality the gain compression of one mode might affect the gain experienced by another mode. Such coupling could be included by replacing the zeros with some *intermodal* gain compression term.

TABLE 5.1: List of Common Parameters for Two Laser Structures

Parameter	In-Plane Laser	VCSEL
d	80 Å	10 μm
w	2 μm	10 μm
L_a	250 μm	3×80 Å
L	250 μm	1.15 μm
Γ_{xy}	0.032	1
Γ_z	1	$1.83 \times 0.0209 = 0.0382$
V	4×10^{-12} cm^3	2.4×10^{-12} cm^3
V_p	1.25×10^{-10} cm^3	0.628×10^{-10} cm^3
$\sqrt{R_1 R_2}$	0.32	0.995
α_m	45.6 cm^{-1}	43.6 cm^{-1}
α_i	5 cm^{-1}	20 cm^{-1}
F_1	0.5	0.9
$\eta_{d1} = \eta_i \eta_0$	$0.8 \times 0.45 = 0.36$	$0.8 \times 0.617 = 0.494$
g_{th}	1580 cm^{-1}	1665 cm^{-1}
τ_p	2.77 ps	2.20 ps
At Threshold:		
N_{th}	3.77×10^{18} cm^{-3}	3.93×10^{18} cm^{-3}
J_{th}	$178.3/0.8 = 223$ A/cm^2	$575.2/0.8 = 719$ A/cm^2
I_{th}	1.11 mA	0.719 mA
η_r	0.840	0.829
a	5.34×10^{-16} cm^2	5.10×10^{-16} cm^2
a_p	2.37×10^{-14} cm^2	2.50×10^{-14} cm^2
τ	2.71 ns	2.63 ns
$\tau_{\Delta N}$	1.57 ns	1.52 ns
$\tau'_{\Delta N}$	44.3 μs	23.0 μs
n_{sp}	1.13	1.11
R'_{sp}	1.02×10^{23} cm^{-3}/s	2.09×10^{23} cm^{-3}/s
$\beta_{sp}\|_{th}$	0.869×10^{-4}	1.69×10^{-4}
At $P_{01} = 1$ mW:		
I	3.31 mA	2.32 mA
N_p	2.43×10^{14} cm^{-3}	2.80×10^{14} cm^{-3}
γ_{NP}	1.12×10^{13} s^{-1}	1.18×10^{13} s^{-1}
γ_{PN}	2.95×10^{7} s^{-1}	3.88×10^{7} s^{-1}
γ_{NN}	1.56×10^{9} s^{-1}	1.67×10^{9} s^{-1}
γ_{PP}	1.32×10^{9} s^{-1}	1.93×10^{9} s^{-1}
f_R Eq. (5.49)	2.907 GHz	3.423 GHz
$\sim f_R$ Eq. (5.51)	2.904 GHz	3.418 GHz
γ	2.88×10^{9} s^{-1}	3.60×10^{9} s^{-1}
γ_0	0.651×10^{9} s^{-1}	0.686×10^{9} s^{-1}
K	0.265 ns	0.250 ns
$(\Delta\nu)_{ST}$	1.07 MHz	2.27 MHz
RIN peak	−112 dB/Hz	−111 dB/Hz

TABLE 5.1: (*Continued*)

Active Material	Material Parameters For Both Laers $\ln_{0.2}Ga_{0.8}As/GaAs$ 80 Å QWs @ 980 nm)
Curve fits (from Chapter 4):	
N_{tr}, N_s, g_{0N}	1.8×10^{18} cm^{-3}, -0.4×10^{18} cm^{-3}, 1800 cm^{-1}
J_{tr}, J_s, g_{0J}	50 A/cm^2, -10 A/cm^2, 1100 cm^{-1}
J_{tr}, J_s, g_{0J} (J_{sp} only)	50 A/cm^2, 0, 1440 cm^{-1}
gn_{sp}	850 cm^{-1} $\times$ ln[1 + $(N/10^{18}$ cm$^{-3})^2/2$]
v_g	$3/4.2 \times 10^{10}$ cm/s
η_i	0.8
ε	1.5×10^{-17} cm^3
A	assumed negligible
B	$\sim 0.8 \times 10^{-10}$ cm^3/s (see Fig. 4.29)
C	3.5×10^{-30} cm^6/s

5.3.1 Small-Signal Frequency Response

To obtain the small-signal responses $dN(t)$ and $dN_p(t)$ to a sinusoidal current modulation $dI(t)$, we assume solutions of the form

$$\begin{aligned} dI(t) &= I_1 e^{j\omega t}, \\ dN(t) &= N_1 e^{j\omega t}, \\ dN_p(t) &= N_{p1} e^{j\omega t}. \end{aligned} \tag{5.39}$$

Setting $d/dt \rightarrow j\omega$ and rearranging Eq. (5.36), we obtain

$$\begin{bmatrix} \gamma_{NN} + j\omega & \gamma_{NP} \\ -\gamma_{PN} & \gamma_{PP} + j\omega \end{bmatrix} \begin{bmatrix} N_1 \\ N_{p1} \end{bmatrix} = \frac{\eta_i I_1}{qV} \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{5.40}$$

The determinant of the matrix is given by

$$\begin{aligned} \Delta &\equiv \begin{vmatrix} \gamma_{NN} + j\omega & \gamma_{NP} \\ -\gamma_{PN} & \gamma_{PP} + j\omega \end{vmatrix} \\ &= (\gamma_{NN} + j\omega)(\gamma_{PP} + j\omega) + \gamma_{NP}\gamma_{PN} \\ &= \gamma_{NP}\gamma_{PN} + \gamma_{NN}\gamma_{PP} - \omega^2 + j\omega(\gamma_{NN} + \gamma_{PP}). \end{aligned} \tag{5.41}$$

With this, we can apply Cramer's rule to obtain the small-signal carrier and photon densities in terms of the modulation current:

$$N_1 = \frac{\eta_i I_1}{qV} \cdot \frac{1}{\Delta} \begin{vmatrix} 1 & \gamma_{NP} \\ 0 & \gamma_{PP} + j\omega \end{vmatrix}, \tag{5.42}$$

$$N_{p1} = \frac{\eta_i I_1}{qV} \cdot \frac{1}{\Delta} \begin{vmatrix} \gamma_{NN} + j\omega & 1 \\ -\gamma_{PN} & 0 \end{vmatrix}. \tag{5.43}$$

Expanding the determinants, the small-signal solutions can be written as

$$N_1 = \frac{\eta_i I_1}{qV} \cdot \frac{\gamma_{PP} + j\omega}{\omega_R^2} H(\omega), \tag{5.44}$$

$$N_{p1} = \frac{\eta_i I_1}{qV} \cdot \frac{\gamma_{PN}}{\omega_R^2} H(\omega), \tag{5.45}$$

where the modulation response is conveniently described in terms of the following two-parameter *modulation transfer function*:

$$H(\omega) = \frac{\omega_R^2}{\Delta} \equiv \frac{\omega_R^2}{\omega_R^2 - \omega^2 + j\omega\gamma}. \tag{5.46}$$

We define ω_R as the *relaxation resonance frequency* and γ as the *damping factor*. The physical significance of these parameters was discussed in Chapter 2. Comparing (5.46) to (5.41), the following associations can be made:

$$\omega_R^2 \equiv \gamma_{NP}\gamma_{PN} + \gamma_{NN}\gamma_{PP}, \tag{5.47}$$

$$\gamma \equiv \gamma_{NN} + \gamma_{PP} \tag{5.48}$$

The densities N_1 and N_{p1} both follow the frequency response of the modulation transfer function; however, the carrier density has an additional zero in the complex frequency plane at $\omega = j\gamma_{PP}$. The general behavior of $H(\omega)$ is shown in Fig. 5.4. It is essentially a second-order low-pass filter with a damped resonance appearing near the cutoff frequency. The intensity modulation can follow the current modulation up to frequencies near ω_R, with an enhancement in the response existing at the relaxation resonance. Beyond the resonance, the response drops off dramatically. The actual peak frequency of the resonance, ω_P, is slightly less than ω_R depending on the damping. The frequency at which the electrical power response drops to half its DC value, $\omega_{3\text{ dB}}$, is somewhat higher than ω_R for small damping. The relations for both ω_P and $\omega_{3\text{ dB}}$ are included in the figure. To understand how we can maximize the modulation bandwidth, or $\omega_{3\text{ dB}}$, we need to evaluate Eqs. (5.47) and (5.48).

The relaxation resonance frequency and the damping factor that characterize $H(\omega)$ can be expanded using Eq. (5.34):

$$\omega_R^2 = \frac{v_g a N_p}{\tau_p} + \left[\frac{\Gamma v_g a_p N_p}{\tau_{\Delta N}} + \frac{\Gamma R'_{sp}}{N_p \tau_{\Delta N}}\right]\left(1 - \frac{\tau_{\Delta N}}{\tau'_{\Delta N}}\right) + \frac{1}{\tau'_{\Delta N}\tau_p}, \tag{5.49}$$

$$\gamma = v_g a N_p \left[1 + \frac{\Gamma a_p}{a}\right] + \frac{1}{\tau_{\Delta N}} + \frac{\Gamma R'_{sp}}{N_p}. \tag{5.50}$$

In practice, the expression for ω_R^2 can be simplified dramatically. For example, the last term is small compared to the first for $N_p V_p > n_{sp}$ (see footnote 5), and

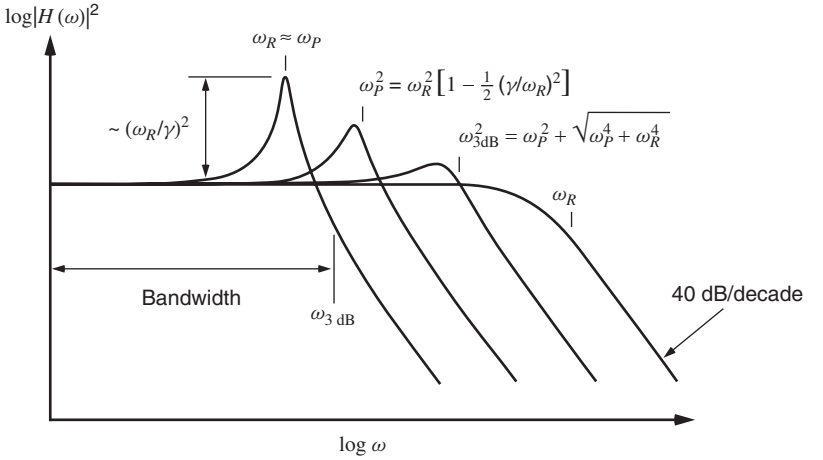


FIGURE 5.4: Sketch of the modulation transfer function for increasing values of relaxation resonance frequency and damping factor, including relationships between the peak frequency, ω_P, the resonance frequency, ω_R, and the 3 dB down cutoff frequency, $\omega_{3\text{dB}}$.

hence can be ignored above threshold where $n_{sp} \sim 1 - 2$ and $N_p V_p \gg 1$. Of the remaining terms, two are $\propto N_p$, and one is $\propto 1/N_p$. The former will dominate at some point above threshold as the photon density increases. For typical numbers (see Table 5.1), this crossover occurs very close to threshold. Hence, we can also neglect the $1/N_p$ term. Comparing the coefficients of the two terms $\propto N_p$, with $\tau_{\Delta N} \gg \tau_p$ and $a \sim \Gamma a_p$, we conclude that $a/\tau_p \gg \Gamma a_p/\tau_{\Delta N}$. Thus, the first term dominates over all other terms and ω_R^2 reduces to

$$\omega_R^2 \approx \frac{v_g a N_p}{\tau_p}. \qquad \text{(above threshold)} \tag{5.51}$$

This is the same result we found in Chapter 2 with the exception that in this case $a = a_0/(1 + \varepsilon N_p)$. As mentioned in Section 2.7, ω_R can be enhanced by increasing the photon density or output power. This increase continues until the photon density approaches $1/\varepsilon$, at which point the differential gain falls off appreciably due to gain compression. For photon densities well beyond $1/\varepsilon$, ω_R^2 becomes independent of output power, saturating at $v_g a_0/\tau_p \varepsilon$ (however, in practice, such high photon densities are not typically encountered, and furthermore, it is not clear that the form assumed here for gain compression is even valid at such high photon densities).

Because we will be concentrating on the laser performance above threshold, we will specifically define ω_R using Eq. (5.51) unless otherwise stated. This simplified definition of ω_R can be used to rewrite the damping factor as

$$\gamma = K f_R^2 + \gamma_0, \tag{5.52}$$

where

$$K = 4\pi^2 \tau_p \left[1 + \frac{\Gamma a_p}{a}\right] \quad \text{and} \quad \gamma_0 = \frac{1}{\tau_{\Delta N}} + \frac{\Gamma R'_{sp}}{N_p}. \tag{5.53}$$

For large resonance frequencies, the K-factor describes the damping of the response, and as such is an important parameter in the characterization of high-speed lasers. The damping factor offset γ_0 is important at low powers where the relaxation resonance frequency is small. In practice, K and γ_0 are used as fitting parameters to be extracted from the laser modulation response.

Example 5.1 A small signal modulation is applied to a VCSEL whose diameter is 4 μm. Resonant frequency, f_R, damping, γ, and ratio of FM/IM, M/m, measured for three different bias values above threshold are shown in Table 5.2.

TABLE 5.2:

$I - I_{th}$(mA)	f_R(GHz)	γ(GHz)	$\frac{M}{m}$(@ 50 MHz)
0.5	6	5.5	30
1.0	8.5	10.5	60
2.0	12	20.5	120

Problem: What are the K factor and the damping factor offset, γ_0?

Solution: The relationship between the damping factor and the resonance frequency is given by Eq. (5.52),

$$\gamma = K f_R^2 + \gamma_0.$$

Taking two points from the experimental data, we can calculate K and γ_0,

$$\gamma_1 - \gamma_2 = K(f_{R1}^2 - f_{R2}^2), \quad \rightarrow \quad K = \frac{\gamma_1 - \gamma_2}{f_{R1}^2 - f_{R2}^2} = \frac{5}{36} = 0.1389 \text{ ns},$$

and

$$\gamma_0 = \gamma_1 - K f_{R1}^2 = 0.5 \text{ GHz}.$$

The solution can be verified by making sure that the third data point satisfies the equation.

Because the damping increases in proportion to ω_R^2, as we attempt to drive the laser harder to increase ω_R, the response flattens out as illustrated in Fig. 5.4. At

some point, the damping becomes large enough that the response drops below the 3 dB cutoff at frequencies less than ω_R. As a result, there is a maximum bandwidth that can be achieved. Using the formulas in Fig. 5.4 combined with the definition of γ (neglecting γ_0), we can determine the modulation bandwidth for low damping as well as the maximum possible bandwidth:

$$f_{3\text{ dB}} \approx f_R\sqrt{1+\sqrt{2}} \approx 1.55 f_R, \quad (\gamma/\omega_R \ll 1) \tag{5.54}$$

$$f_{3\text{ dB}}|_{\max} = \sqrt{2}\frac{2\pi}{K}. \qquad (\gamma/\omega_R = \sqrt{2}) \tag{5.55}$$

The modulation bandwidth increases linearly with the relaxation resonance frequency and remains about 50% larger than f_R until damping becomes strong. With strong damping the bandwidth is compromised and eventually decreases with further increases in ω_R and γ. The optimum damping and maximum bandwidth occur when $\omega_P = 0$ and $\omega_R = \omega_{3\text{ dB}}$ as defined in Fig. 5.4. This point is determined by the K-factor, which in turn is determined by the photon lifetime of the cavity. The K-factor therefore defines the intrinsic modulation bandwidth capabilities of the laser.

Returning to the AC photon density modulation response Eq. (5.45), we can use the explicit definition of ω_R to simplify the relation. With the above-threshold version of γ_{PN} in Eq. (5.36), we can set $\gamma_{PN} \approx \Gamma v_g a N_p = \Gamma\omega_R^2\tau_p$. We then obtain

$$\frac{N_{p1}}{I_1} = \frac{\eta_i}{qV}\Gamma\tau_p H(\omega). \tag{5.56}$$

Using Eq. (5.5) to set $\tau_p = \eta_0 h\nu N_{p1} V_p/P_1$, the AC output power modulation response finally becomes

$$\frac{P_1}{I_1} = \eta_i\eta_0\frac{h\nu}{q}H(\omega). \tag{5.57}$$

This result was also found in Eq. (2.54). The electrical power received, which is of more fundamental interest is given by the absolute square of Eq. (5.57). In decibels, the frequency response is therefore $\propto 10\log_{10}|H(\omega)|^2$ or $20\log_{10}|H(\omega)|$. The phase shift of the modulation is given by $\angle H(\omega)$.

Figure 5.5 gives an example of experimental modulation responses of an InGaAs/GaAs VCSEL at different biases. The resonance peak, which occurs at $\sim\omega_R$ is clearly visible in each curve. Note that the resonance shifts to higher frequencies with increasing bias current. This is because the output power and photon density increase with increasing current above threshold which increases ω_R via Eq. (5.51). We can write this relationship more directly by plugging Eqs. (5.5) and (5.17) into Eq. (5.51) to obtain

$$\omega_R^2 = \frac{v_g a}{qV_p}\eta_i(I - I_{th}). \tag{5.58}$$

For high photon densities, it is important to appreciate that $a = a_0/(1+\varepsilon N_p)$, and hence gain compression can affect the resonance frequency. However, for practical

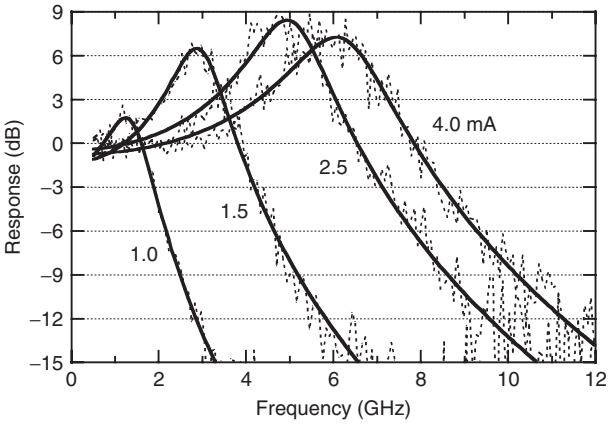


FIGURE 5.5: Small-signal intensity modulation of a mesa-etched intracavity-contacted 7 μm diameter 3-quantum-well InGaAs/GaAs VCSEL at the DC biases indicated. Its threshold current is 0.75 mA, the powers out are 0.2, 0.4, 0.9 and 1.4 mW, respectively, and the MSR is greater than 30 dB over the entire range [1]. Experimental curves are dashed and curve fits are solid.

situations its effect is minimal. The curves in Fig. 5.5 also reveal that the resonance peak flattens and broadens out with increasing bias. This is due to the increase in damping present through γ which goes as ω_R^2 from Eq. (5.53). And unlike the resonance frequency, the K-factor is enhanced by $\sim$2 due to gain compression, *independent of the magnitude of the photon density*.

5.3.2 Small-Signal Transient Response

To find the transient response to the *linearized* system characterized by Eq. (5.36), it is useful to rewrite the double-pole modulation transfer function as the product of two single-pole transfer functions:

$$H(\omega) = \frac{\omega_R^2}{\omega_R^2 - \omega^2 + j\omega\gamma} = \frac{\omega_R^2}{(j\omega + s_1)(j\omega + s_2)}. \tag{5.59}$$

The roots $s_{1,2}$ in the complex frequency plane are given by

$$s_{1,2} = \tfrac{1}{2}\gamma \pm j\omega_{osc}, \qquad \omega_{osc} = \omega_R\sqrt{1 - (\gamma/2\,\omega_R)^2}. \tag{5.60}$$

Note from Fig. 5.4 that $\omega_{osc}^2 = \frac{1}{2}(\omega_P^2 + \omega_R^2)$. From linear systems theory, these complex roots suggest solutions of the following form:

$$e^{-\gamma t/2}(e^{j\omega_{osc}t} \pm e^{-j\omega_{osc}t}), \tag{5.61}$$

plus any additional constants required to satisfy initial and final conditions.

In general we can say that in response to an abrupt change in the system (e.g., a step increase in current, a reduction in the photon lifetime, etc.), the transient carrier and photon densities will oscillate sinusoidally or "ring" at the rate ω_{osc} before eventually decaying to new steady-state values, the decay being characterized by the damping factor (hence the name). When the system is underdamped (i.e., when $\gamma/2\ \omega_R \ll 1$), the oscillations are characterized by the relaxation resonance frequency ω_R from Eq. (5.60).[6] However, $\gamma \propto \omega_R^2$ so that as ω_R increases with increasing output power, the system approaches critical damping, or $\gamma/2\ \omega_R \to 1$ and $\omega_{osc} \to 0$. Under these conditions, the densities will simply rise/fall exponentially to their new steady-state values. At even larger output powers the system becomes overdamped, $\gamma/2\omega_R > 1$, and the oscillation frequency becomes imaginary, $\omega_{osc} \to j\gamma_{osc}$, which slows the system down, increasing the rise and fall times of the transient solutions. The maximum modulation bandwidth of the laser is obtained when $\gamma/2\omega_R = 1/\sqrt{2}$ (which is slightly underdamped), yielding $\omega_{osc} = \omega_R/\sqrt{2}$ (and $\omega_P = 0$). Thus, for practical applications the laser is always underdamped, and oscillations toward steady-state are expected.

Now let us consider a specific example. We want to know the effect of suddenly increasing the current by dI. If we define the transient responses as $dN(t)$ and $dN_p(t)$, then Eq. (5.61) can be used to obtain their general forms (i.e., damped sines or damped cosines), and Eq. (5.36) can be used to find the initial and final conditions. Initially, the system is at steady-state and we can set $dN(0) = dN_p(0) = 0$. At $t = 0^+$, the current has abruptly increased, but no time has elapsed so no change in carrier density is observed. Because the photon density can only increase as a result of an increase in gain, which in turn follows the carrier density increase, it also must still be at its initial value. From this we conclude that $dN(0^+) = dN_p(0^+) = 0$. Equation (5.36) at $t = 0^+$ therefore becomes

$$\frac{d}{dt}\begin{bmatrix} dN(t) \\ dN_p(t) \end{bmatrix} = \frac{\eta_i}{qV}\begin{bmatrix} dI \\ 0 \end{bmatrix}. \qquad (t = 0^+) \tag{5.62}$$

This gives us our set of initial conditions: the initial slope of $dN(t)$ is finite and proportional to the current step, whereas the initial slope of $dN_p(t)$ is zero (the carrier density must rise before the gain can begin to make a change in the photon density).

Due to the clamping of the carrier density above threshold, we know that the final value of $dN(t)$ must be approximately zero after the transient has damped

[6]In some treatments, ω_R is *by definition* set equal to ω_{osc}, and the damping factor is set equal to $\gamma/2$ to comply with the natural roots of the transfer function. The problem with setting $\omega_R = \omega_{osc}$ is that ω_{osc} goes to zero at high output powers. The more standard definition of ω_R (as defined in this chapter) allows us to use Eq. (5.51) for all but very small output powers. As for the damping factor, the alternate definition causes no harm, and in fact would perhaps be the better way to define it. However, in the literature, the damping factor is usually defined such that the decay rate is $\gamma/2$, as we have it here. A word of warning: although either set of definitions can be used, one must be careful not to mix and match results found using either set (an error commonly made in the literature).

out. This fact, together with the initial condition on the slope suggests that $dN(t)$ is a damped *sine* wave of the form:

$$dN(t) = dN_0 e^{-\gamma t/2} \sin \omega_{osc} t. \tag{5.63}$$

From Eq. (5.62), we find $dN_0 = \eta_i dI/(qV\omega_{osc})$.

For the photon density, we know the initial slope is zero. This suggests a damped *cosine* solution. However, because the current has increased, the steady-state output power must also increase. In other words, the final value of $dN_p(t)$ must have a finite positive value. Hence, the solution is more likely equal to the final value less a damped cosine of the same initial value:

$$dN_p(t) = dN_p(\infty)[1 - e^{-\gamma t/2} \cos \omega_{osc} t]. \tag{5.64}$$

Upon closer examination, we find that the initial slope of the term in brackets is not zero but is equal to $\gamma/2$. For small damping we can neglect this. However for the more general case, we need to add another term which is zero initially but has an initial slope $-\gamma/2$. A damped sine wave with the appropriate prefactor can be used to give

$$dN_p(t) = dN_p(\infty)\left[1 - e^{-\gamma t/2} \cos \omega_{osc} t - \frac{\gamma}{2\omega_{osc}} e^{-\gamma t/2} \sin \omega_{osc} t\right]. \tag{5.65}$$

This function now has zero initial slope and reaches $dN_p(\infty)$ in the steady state.

To determine $dN_p(\infty)$, we know that as $t \to \infty$, the time derivatives must go to zero in Eq. (5.36) allowing us to write

$$\begin{bmatrix} \gamma_{NN} & \gamma_{NP} \\ -\gamma_{PN} & \gamma_{PP} \end{bmatrix} \begin{bmatrix} dN(\infty) \\ dN_p(\infty) \end{bmatrix} = \frac{\eta_i}{qV} \begin{bmatrix} dI \\ 0 \end{bmatrix}. \quad (t = \infty) \tag{5.66}$$

Using Cramer's rule to solve for the steady-state values, we obtain

$$dN(\infty) = \frac{\eta_i dI}{qV} \frac{\gamma_{PP}}{\omega_R^2} \approx \frac{\eta_i\ dI}{qV} \Gamma \tau_p \cdot \frac{a_p}{a}, \tag{5.67}$$

$$dN_p(\infty) = \frac{\eta_i dI}{qV} \frac{\gamma_{PN}}{\omega_R^2} \approx \frac{\eta_i\ dI}{qV} \Gamma \tau_p. \tag{5.68}$$

The photon density transient is now completely specified. However, we find that our initial assumption that $dN(t)$ must return to zero in the steady state is not quite right. Two factors affect the final carrier density. First of all, to increase the photon density, the gain must be brought closer to threshold as suggested in Fig. 5.3. As a result, the carrier density must increase slightly. In addition, gain compression reduces the gain at a given carrier density as the photon density is increased. Thus to maintain a given gain, the carrier density must again increase with increasing photon density. Both of these factors are included in γ_{PP}; however,

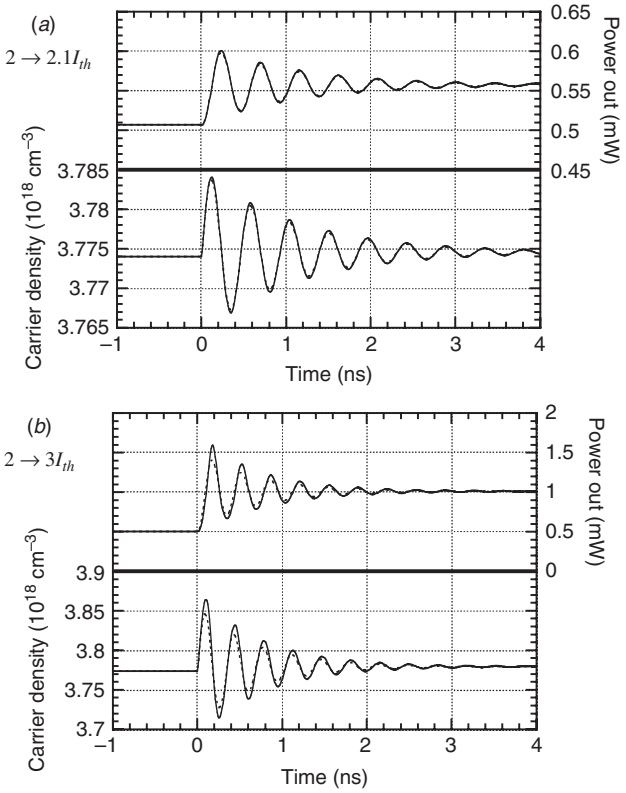


FIGURE 5.6: Comparison of exact numerical calculation (−) with analytical approximations (−−) for a small-signal step transient. An in-plane laser with parameters given in Table 5.1 is biased at twice threshold initially. Threshold gain, $\Gamma g_{th} = 50.6$ cm^{-1}; initial carrier density, $N = 3.77 \times 10^{18}$ cm^{-3}; initial photon density, $N_p = 1.23 \times 10^{14}$ cm^{-3}. (a) Transient from 2.0 to 2.1 I_{th}; (b) transient from 2.0 to 3.0 I_{th}.

for reasonable photon densities, the gain compression term dominates, leading to the approximation given in Eq. (5.67).

The question now is how to include a finite final value for $dN(t)$. We need a function which has zero slope initially and a finite final value, just like the photon density. In fact, noting that $dN(\infty) = (\gamma_{PP}/\gamma_{PN})dN_p(\infty)$ motivates us to write the new improved version of the carrier density transient as

$$dN(t) = dN_0 e^{-\gamma t/2} \sin \omega_{osc} t + \frac{\gamma_{PP}}{\gamma_{PN}} dN_p(t), \tag{5.69}$$

where again $dN_0 = \eta_i dI/(qV\omega_{osc})$ and $\gamma_{PP}/\gamma_{PN} \approx a_p/a$.

Although we have used rather ad hoc methods, one can easily verify that Eqs. (5.65) and (5.69) satisfy the differential rate Eqs. (5.36), which themselves are valid as long as the transients are small compared to the steady-state

carrier and photon densities. We could have alternatively derived these solutions through a more rigorous linear systems approach (Laplace transform of $H(\omega) \rightarrow$ impulse response $\rightarrow$ step function response).

Figure 5.6 gives two examples of responses to step function current transients, using both Eqs. (5.65) and (5.69) and an exact numerical solution of the rate equations. The first pair is for a small current step, which gives excellent agreement, and the second is for an order of magnitude larger step, both starting from a bias point of twice threshold. For larger bias points proportionally larger transients can be approximated by Eqs. (5.65) and (5.69). However for the best fit, the analytic solutions should use the *final* photon and carrier density values to estimate the rate coefficients (note that both ω_R and γ are larger in Fig. 5.6b).

For small-signal data that involve a square-wave-like current waveform, time-shifted versions of Eqs. (5.65) and (5.69) can be superimposed to give the net time response of the photon and carrier density, respectively. From a linear systems point of view, a square wave input can be described as one positive step function plus a second time-delayed negative step function. The linearity of the system then guarantees that the overall solution is simply the sum of the responses created by each individual step function input.

5.3.3 Small-Signal FM Response or Frequency Chirping

As we saw in the previous sections, current modulation of the active region results in a modulation of both the photon density and the carrier density. The modulation of the carrier density modulates the gain; however, it also *modulates the index* of the active region n_a. As a result, the optical length of the cavity is modulated by the current, causing the resonant mode to shift back and forth in frequency. This frequency modulation (FM) of the laser may be desirable if we wish to dynamically tune the laser. However for intensity modulation (IM) applications, FM or *frequency chirping* broadens the modulated spectrum of the laser, hindering its effectiveness in optical fiber communications.

To derive the general relationship between $\Delta\nu$ and Δn_a, we write the frequency of mode m including a passive section as

$$\nu[\bar{n}_a L_a + \bar{n}_p L_p] = mc/2. \tag{5.70}$$

To find the frequency deviation, we take the differential of this equation and include the possibility of the active index changing with carrier density. The first differential term is

$$\Delta(\nu\bar{n}_a) = \Delta\nu\bar{n}_a + \nu\left[\frac{d\bar{n}_a}{d\nu}\Delta\nu + \frac{d\bar{n}_a}{dN}\Delta N\right] = \Delta\nu\bar{n}_{ga} + \nu\frac{d\bar{n}_a}{dN}\Delta N. \tag{5.71}$$

For the passive section, we simply have $\Delta(\nu\bar{n}_p) = \Delta\nu\bar{n}_{gp}$. In both cases, the g subscript denotes group index: $\bar{n}_g \equiv \bar{n} + \nu d\bar{n}/d\nu$. Setting the differential of the

RHS of Eq. (5.70) to zero and solving for $\Delta\nu$, we obtain

$$\Delta\nu = -\frac{\Gamma_z v_g}{\lambda}\frac{d\bar{n}_a}{dN}\Delta N, \tag{5.72}$$

where $\Gamma_z = \bar{n}_{ga}L_a(\bar{n}_{ga}L_a + \bar{n}_{gp}L_p)$ and v_g is the group velocity in the active section. In cases where $L_a \ll \lambda$, it can be shown that standing wave effects as considered in Appendix 5 would also appear in Γ_z (to derive this, one would need to take into account the index discontinuities and resulting reflections between the active and passive sections).

In Chapter 6, we will find that the change in effective propagation index can be related to the change in active material index through $\Delta\bar{n} = \Gamma_{xy}\Delta n$ (see Eqs. (6.23) and (A5.13)), allowing us to set $d\bar{n}/dN = \Gamma_{xy}dn/dN$. With this substitution and with $\Gamma_{xy}\Gamma_z = \Gamma(= V/V_p)$, we obtain

$$\Delta\nu = -\frac{\Gamma v_g}{\lambda}\frac{dn_a}{dN}\Delta N. \tag{5.73}$$

The complex index is given by $\tilde{n} = n + jn_i$, and the power gain is related to the imaginary index through: $g = 2k_o n_i = 4\pi n_i/\lambda$. The relationship between how the real and imaginary indices are affected by the carrier density is described using what is referred to as the *linewidth enhancement factor*:

$$\alpha \equiv -\frac{dn/dN}{dn_i/dN} = -\frac{4\pi}{\lambda}\frac{dn/dN}{dg/dN} = -\frac{4\pi}{\lambda a}\frac{dn}{dN}, \tag{5.74}$$

where a is again used to define the differential gain. With this definition, the frequency shift in response to changes in carrier density becomes

$$\Delta\nu = \frac{\alpha}{4\pi}\gamma v_g a\Delta N. \tag{5.75}$$

The amount of frequency modulation is proportional to the linewidth enhancement factor, which is typically between 4 and 6 but can be as low as 2 in some active materials.

Using Eq. (5.44) for the AC carrier density modulation response, we immediately obtain

$$\frac{\nu_1}{I_1} = \frac{\alpha}{4\pi}\Gamma v_g a\frac{\eta_i}{qV}\cdot\frac{\gamma_{pp} + j\omega}{\omega_R^2}H(\omega). \tag{5.76}$$

Figure 5.7 compares the qualitative frequency response of this FM and the IM derived in Section 5.3.1. Both FM and IM exhibit a resonance peak; however, the FM response falls off at low frequencies at 20 dB/decade before leveling out below $\gamma_{PP}/2\pi$. Also, the peak of $\omega^2|H(\omega)|^2$ occurs directly at ω_R.

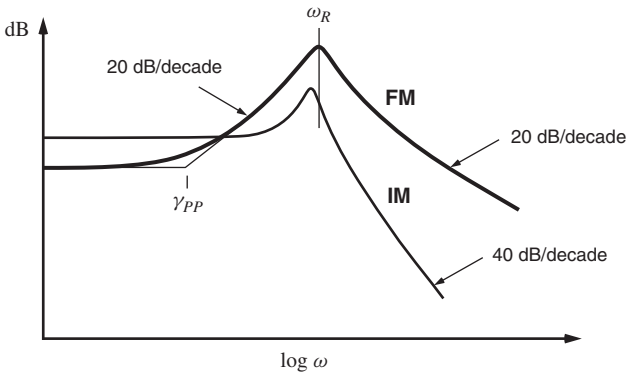


FIGURE 5.7: Qualitative comparison of the FM and IM response of a semiconductor laser.

Dividing Eq. (5.76) by Eq. (5.56) and setting $\omega_R^2 = v_g a N_p/\tau_p$, we obtain the simple result:

$$\nu_1 = \frac{\alpha}{4\pi}(\gamma_{PP} + j\omega)\frac{N_{p1}}{N_p}. \tag{5.77}$$

So we discover that the frequency chirping of the lasing spectrum increases linearly with the intensity modulation depth (because $N_{p1}/N_p = P_1/P_0$). This effect is dramatically illustrated in Fig. 5.8, which displays the lasing spectrum under different degrees of intensity modulation for a constant modulation frequency. The peaks at both extremes of the modulated spectrum result from the time averaging of the sinusoidal modulation signal (i.e., the lasing frequency spends more time on average at the extremes of the sine wave).

If we define the FM modulation index as $M = \nu_1/f$ and the IM modulation index as $m = P_1/P_0$, then using the absolute magnitude of Eq. (5.77), the ratio of the FM-to-IM modulation index becomes

$$\frac{M}{m} = \frac{\alpha}{2}\sqrt{\left(\frac{\gamma_{PP}}{\omega}\right)^2 + 1}. \tag{5.78}$$

From Eq. (5.36), $\gamma_{PP} \approx \Gamma v_g a_p N_p$. For typical numbers, this term can be in the hundreds of MHz to few GHz range, depending on the output power level. For modulation frequencies beyond this (i.e., $\omega \gg \gamma_{PP}$), $M/m \to \alpha/2$, providing us with a simple and very direct method of measuring α. For lower modulation frequencies, M/m becomes inversely proportional to the modulation frequency. By measuring M/m as a function of ω, one can then determine γ_{PP}, and ultimately determine the gain compression factor ε (via a_p) if curves are taken at different output power levels. Such a measurement is shown in Fig. 5.9. From these curves it is clear that for this laser, γ_{PP} varies from 1–3 GHz, and α is just over 6.

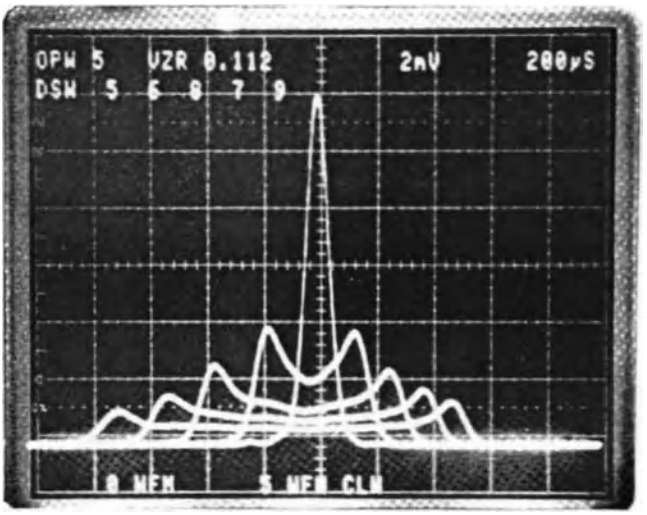


FIGURE 5.8: Time-averaged power spectra of a 1.3 μm InGaAsP laser under sinusoidal modulation at 100 MHz. The spectrum broadens with an increase in the modulation current due to frequency chirping. The horizontal scale is 0.5 Å per division. After [2]. (*Reprinted, by permission, from Journal of Applied Physics*).

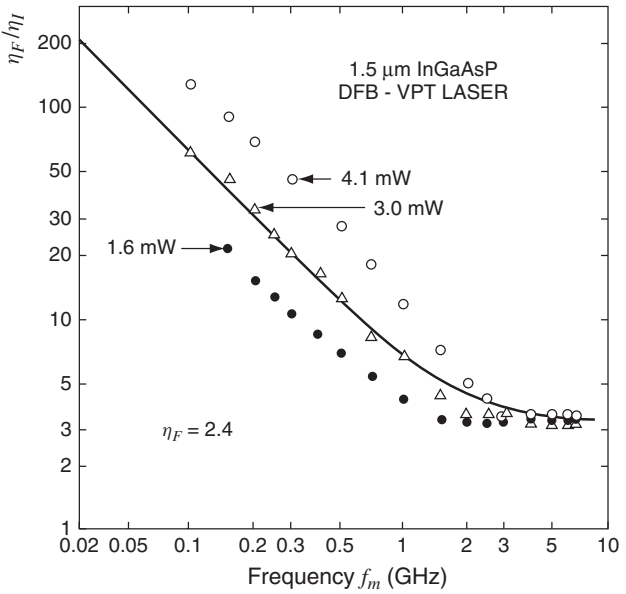


FIGURE 5.9: FM-to-IM modulation index ratio as a function of frequency for different output power levels [3], where $\eta_F \equiv M$ and $\eta_I \equiv m$. (Reproduced, by permission, from *Electronics Letters*).

Example 5.2 For the VCSEL from Example 5.1, the plot of M/m as a function of $1/f_m$ is found to decrease for low frequencies, and then saturate at a value of 2.5 for high modulation frequencies.

Problem: What is the linewidth enhancement factor for this laser?

Solution: (a) The relationship between the FM-to-IM modulation index and the linewidth enhancement factor is given by Eq. (5.78),

$$\frac{M}{m} = \frac{\alpha}{2}\sqrt{\left(\frac{\gamma_{PP}}{\omega}\right)^2 + 1}.$$

For high modulation frequencies, the expression under the square root reduces to 1, and the value from the plot corresponds to $\frac{1}{2}$ of the enhancement factor α, therefore,

$$\alpha = 5.$$

So far we have only considered changes in the index created by the modulation of the carrier density. However, at low modulation frequencies the temperature of the laser is also modulated when we apply current modulation. Because the index varies with temperature, we should expect the frequency modulation to be affected by thermal effects as well. The total FM of the laser may therefore be written as the sum of two contributions:

$$\frac{\nu_1}{I_1} = \left(\frac{\Delta\nu}{\Delta I}\right)_{carrier} + \left(\frac{\Delta\nu}{\Delta I}\right)_{thermal} \tag{5.79}$$

where

$$\left(\frac{\Delta\nu}{\Delta I}\right)_{carrier} = \frac{\alpha}{4\pi}\frac{\eta_i}{qV_p}\frac{\varepsilon}{1+\varepsilon N_p}\cdot(1+j\omega/\gamma_{PP})H(\omega),$$

$$\left(\frac{\Delta\nu}{\Delta I}\right)_{thermal} = \frac{(1-\eta_{wp})V_{th}Z_T\ d\nu/dT}{1+j\omega\tau_T}.$$

In the first equation we have again set $\gamma_{PP} \approx \Gamma v_g a_p N_p$, expanded a_p using Eq. (5.31), and set $1/\tau_p = \Gamma v_g g$, which is a good approximation above threshold. In the second equation, η_{wp} is the well-plug efficiency of the output power P_{out}/P_{in}, Z_T is the thermal impedance of the laser structure discussed in Chapter 2, and V_{th} is the threshold voltage of the laser that is assumed roughly constant above threshold. Also τ_T is the thermal time constant, which is typically in the few microseconds range, yielding thermal cutoff frequencies in the few hundred kilohertz range. Finally, $d\nu/dT = -(c/\lambda^2)d\lambda/dT$ is the shift in the mode frequency with temperature. For 1 μm emission lasers, $d\lambda/dT \sim 0.06{-}0.08$ nm/K, which translates into $d\nu/dT \sim -20$ GHz/K.

For a thermal impedance of 0.1 K/mW, the *temperature* tuning below the thermal cutoff frequency is ~ -2 GHz/mA (assuming $(1-\eta_{wp})V_{th} \approx 1V$). For a VCSEL

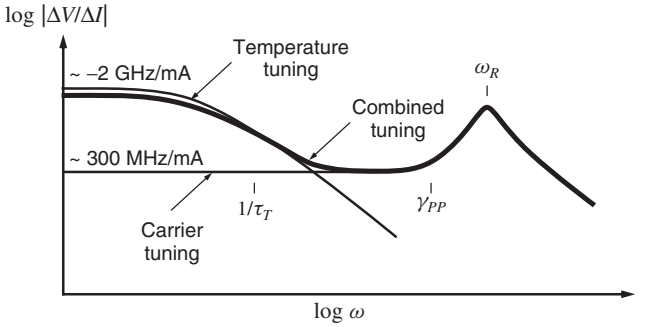


FIGURE 5.10: Sketch of the frequency tuning characteristics as a function of modulation frequency including both temperature and carrier effects (numbers indicated are for a typical in-plane laser). Because the temperature and carrier effects are opposite in sign at low modulation frequencies, one might expect a null in the combined tuning curve at the crossing point. This does not occur because the modulation phase of the temperature tuning shifts by 90° beyond the thermal cutoff frequency.

the thermal impedance is closer to 1 K/mW implying an order of magnitude larger temperature tuning than with in-plane lasers. The *carrier* tuning below $\gamma_{PP}/2\pi$ is closer to ~300 MHz/mA for numbers given in Table 5.1, and is opposite in sign to temperature tuning. Figure 5.10 sketches the important characteristics of the combined frequency tuning as a function of modulation frequency. For modulation frequencies between $1/\tau_T < \omega < \gamma_{PP}$ (i.e., ~1 MHz $< f <$ ~1 GHz), the response is flat with a value typically in the range of a few hundred MHz/mA.

Example 5.3 For the VCSEL from Example 5.1, the injection efficiency for the material is $\eta_i = 0.8$, the confinement factor is $\Gamma = 3.15\%$, and the differential efficiency is $\eta_d = 0.291$. The active region consists of 5 7-nm-wide quantum wells. Assume the mirror transmission $T_m = 0.00212$ and $L_{DBR} = 2$ μm.

Problem: (a) What is the differential gain a? (b) What is the gain compression a_p?

Solution: (a) The value of the differential gain can be computed from the known bias above threshold and resonance frequency, using the expression given by Eq. (5.58),

$$\omega_R^2 = \frac{v_g a}{qV_p}\eta_i(I - I_{th}) = \frac{\Gamma v_g a}{qV}\eta_i(I - I_{th}).$$

From here, the differential gain can be calculated for any bias point—we choose to do it for $I - I_{th} = 0.5$ mA,

$$a = \frac{(2\pi f_R)^2 qV}{\Gamma v_g \eta_i (I - I_{th})}$$

$$= \frac{(2\pi \cdot 6 \cdot 10^9 1/\text{s})^2 \cdot 1.6 \cdot 10^{-19}\text{C}(\pi \cdot 4 \cdot 10^{-8}\ \text{cm}^2)(0.035 \cdot 10^{-4}\ \text{m})}{(0.0315)\frac{3}{4}10^{10}\ \text{cm/s}(0.8)(0.5 \cdot 10^{-3})\text{C/s}}$$

$$= 1.06 \cdot 10^{-15}\ \text{cm}^2.$$

(b) Gain compression can be easily extracted from the relationship between the FM-to-IM modulation index and the linewidth enhancement factor, Eq. (5.78), using the expression for $\gamma_{pp} = \Gamma v_g a_p N_p$, Eq. (5.36). Because the value for $\frac{M}{m} = 30$, we can simplify the computation by neglecting 1 from the Eq. (5.78), thus expressing the gain compression as

$$a_p = \frac{2}{\alpha}\frac{M}{m}\frac{\omega}{\Gamma v_g N_p}.$$

We now need to compute the photon concentration. From the expression for the output power, Eq. (2.37), the photon concentration in the laser cavity is

$$N_p = \frac{\eta_i(I - I_{th})}{q v_g g_{th} V} = \frac{\eta_i(I - I_{th})\Gamma}{q v_g(\alpha_i + \alpha_m)V} = \frac{\eta_d \Gamma(I - I_{th})}{q v_g \alpha_m V}.$$

Substituting into the expression for gain compression, we now have

$$a_p = \frac{2}{\alpha}\frac{M}{m}\frac{\omega V q \frac{T_m}{L_{DBR}}}{\Gamma^2 \eta_d (I - I_{th})} = 1.95 \cdot 10^{-14}\ \text{cm}^2$$

5.4 LARGE-SIGNAL ANALYSIS

For deviations from the steady-state, which are comparable to the steady-state values themselves, our previous differential analysis of the rate equations fails to provide accurate solutions as evidenced by Fig. 5.6(b). To determine the dynamic response of the laser for large-signal inputs, we must therefore return to the general rate Eqs. (5.3) and (5.4). These equations are valid for large signal modulation provided the nonlinear changes in gain with the carrier and photon densities are included. Furthermore, they hold continuously below threshold and above threshold. Of course if the laser is always kept above threshold, the carrier density does not change by a large amount due to carrier clamping, even during large transients. The problem we face here is that these equations cannot be solved analytically. Therefore, to proceed we must use numerical techniques. Such numerical solutions to the rate equations are in principle found by iterating from one rate equation to the other using a small increment of time Δt in place of dt. As always, the power out can be calculated from N_p, and the frequency chirping can be obtained from the deviations of N from some reference value using Eq. (5.75).

5.4.1 Large-Signal Modulation: Numerical Analysis of the Multimode Rate Equations

In this section we consider numerical solutions to the multimode rate equations to give a feeling for the large-signal dynamic properties of multimode lasers. As mentioned in Section 5.2, the single-mode rate equations as given by Eqs. (5.3) and (5.4) give a good description of the dynamics of operation in single-frequency lasers. Generally, however, several different modes with different resonant wavelengths exist. In well-designed in-plane lasers these are generally due to different axial modes, whereas in VCSELs these are usually associated with different lateral modes. Each mode will have a different cavity loss and a different gain because of their different wavelength. These differences will be accentuated in lasers with frequency-dependent losses. Because all modes interact with a common reservoir of carriers, they are indirectly coupled even though they are orthogonal solutions to the electromagnetic wave equation.

Writing a separate photon density for each mode indexed by the integer m, Eqs. (5.3) and (5.4) become

$$\frac{dN}{dt} = \frac{\eta_i I}{qV} - (R_{sp} + R_{nr}) - \sum_m v_{gm} g_m N_{pm}, \tag{5.80}$$

$$\frac{dN_{pm}}{dt} = \left[\Gamma_m v_{gm} g_m - \frac{1}{\tau_{pm}} \right] N_{pm} + \Gamma_m R'_{spm}, \tag{5.81}$$

and g_m is given by Eq. (5.29) with a photon density N_{pm}. For numerical analysis, we replace dN, dN_{pm}, and dt by $\Delta N, \Delta N_{pm}$, and Δt, respectively, then multiply through by Δt. This time increment is set to some sufficiently small value so that the derivatives are accurately estimated. Because we can expect the response will involve oscillations near ω_R, we should set $\Delta t \ll 1/\ \omega_R$. The equations are then successively iterated to increment the carrier density and photon density for each mode from the initial values. Thus, after the ith iteration ($t = i\,\Delta t$), the carrier density is given by $N(i) = N(i-1) + \Delta N$, etc. In practice, numerical techniques such as the fourth-order Runge–Kutta method can be used to reduce the calculation time because much larger time steps can be used for the same accuracy in the solutions (the simple iteration scheme requires extremely small time steps to successfully determine the solution without introducing large errors).

To model the spectral roll-off of the gain near the gain peak, we can often approximate the gain spectrum by a Lorentzian, as illustrated in Fig. 5.11. The gain experienced by each mode m can then be described by

$$g(N, N_p, m) = \frac{1}{(1 + (\Delta M/m^2))} \cdot \frac{g_0}{1 + \sum_n \varepsilon_{nm} N_{pn}} \ln\left(\frac{N + N_s}{N_{tr} + N_s} \right), \tag{5.82}$$

where Δm is the mode number measured away from the central mode ($\Delta m = 0$), which is assumed to be aligned with the gain peak, and M is the mode number

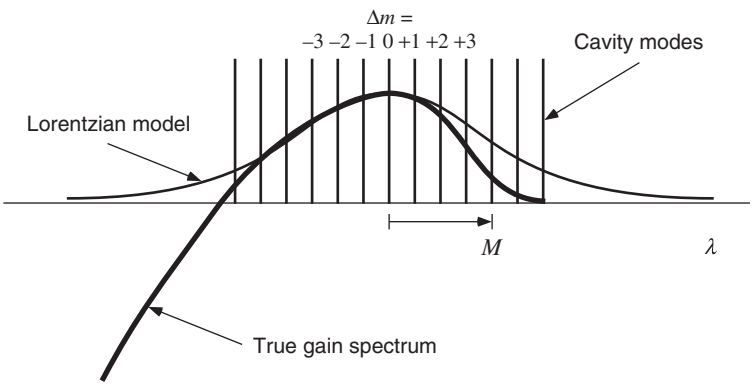


FIGURE 5.11: Simple Lorentzian model of the gain spectrum convenient for analyzing multimode lasers.

where the gain has fallen to half of its peak value. In practice, M is adjusted to match the curvature of the gain specturm near the gain peak, where the details of the spectrum are most relevant. The sum in the denominator accounts for intermodal gain compression.[7]

Figure 5.12 gives numerical plots of the carrier density and photon density for the various modes of an InGaAs/GaAs Fabry–Perot laser. No wavelength-dependent losses are assumed. The current at $t = 0$ increases from zero to $I^+ = 2I_{th}$. As shown, the carrier density initially increases as the active-region reservoir is filled. Little photon density exists until the carrier density reaches its threshold value. At this point, stimulated recombination begins to limit a further increase in carrier density as the photon density increases. This delay before the photon density "turns on" is called the *turn-on delay* of the laser.

The Fabry–Perot laser considered in Fig. 5.12 is obviously not a very good single-mode laser. Initially many modes turn on, and in the steady-state two strong side modes ($\Delta m = \pm 1$) persist. So we discover that the *dynamic* mode suppression ratio (MSR) is quite a bit different (worse) than the steady-state MSR considered at the end of Chapter 3. The challenge for state-of-the-art single-frequency lasers is to maintain a *dynamic* MSR of greater than 30 dB throughout such a turn-on transient.

Figure 5.13 illustrates the predicted turn-on characteristics of a VCSEL in which the axial mode spacing is much larger. Everything else is assumed to be the same. Here we observe the desired single-frequency behavior. In practice, such single-*axial* mode operation in VCSELs is compensated for by multi-*lateral* mode operation not considered in this calculation. One challenge in VCSEL design is to eliminate such lateral modes from appearing in the lasing spectrum.

[7]If neighboring modes have no effect on each other's gain then only diagonal terms of ε_{nm} are nonzero ($\varepsilon_{mm} = \varepsilon, \varepsilon_{nm} = 0$ for $n \neq m$) and the sum can be replaced by εN_{pm}. On the other extreme, if all modes are affected equally by any photons present, then all terms of ε_{nm} are nonzero ($\varepsilon_{nm} = \varepsilon$), and the sum can be replaced by εN_p. Spectral and spatial hole burning of the carrier population will typically lead to gain compression, which is somewhere in between these extremes.

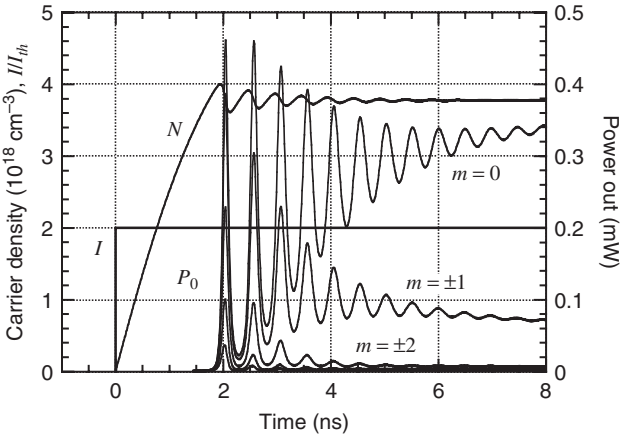


FIGURE 5.12: Large-signal modulation of an SQW InGaAs/GaAs in-plane laser using parameters given in Table 5.1. The relative gain width is set to $M = 25$ as might be appropriate for an in-plane laser. Twenty-five modes ($\Delta m = \pm 12$) are included in the calculation. The sum of gain compression terms is assumed to be εN_{pm}.

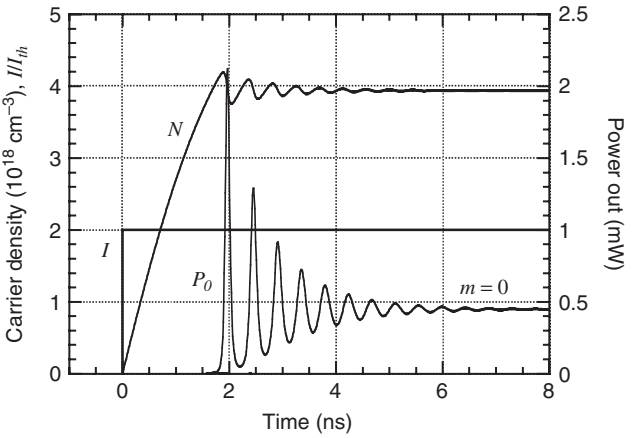


FIGURE 5.13: Large-signal modulation of a short-cavity laser using the VCSEL parameters given in Table 5.1. $M = 3$ is assumed.

Such single axial mode operation can also be achieved with in-plane lasers by using frequency-dependent losses as discussed in Chapter 3 (for example, a DFB or DBR can be employed).

5.4.2 Mode Locking

Because the axial modes of a laser cavity satisfy the same axial boundary conditions, giving regularly spaced modal frequencies according to Eq. (2.29), it is also

possible for them to be aligned in time, such that they will all add in phase at some point along the cavity and be generally out of phase elsewhere. The result as time proceeds is a pulsed output power intensity, with pulses whose width and repetition rate will be determined by the length and other design aspects of the laser cavity and the active region. Figure 5.14 illustrates this phenomenon at several different instants in time, for a given location along the laser cavity. This type of laser, called a mode-locked laser, may be of interest for various applications that require optical pulses, such as optical sampling, different medical applications, laser-based distance and speed measurements (LADAR and range finding), as well as optical regeneration. To illustrate how a number of continuously lasing modes can produce output light in the form of pulses, we will look at a simple case of

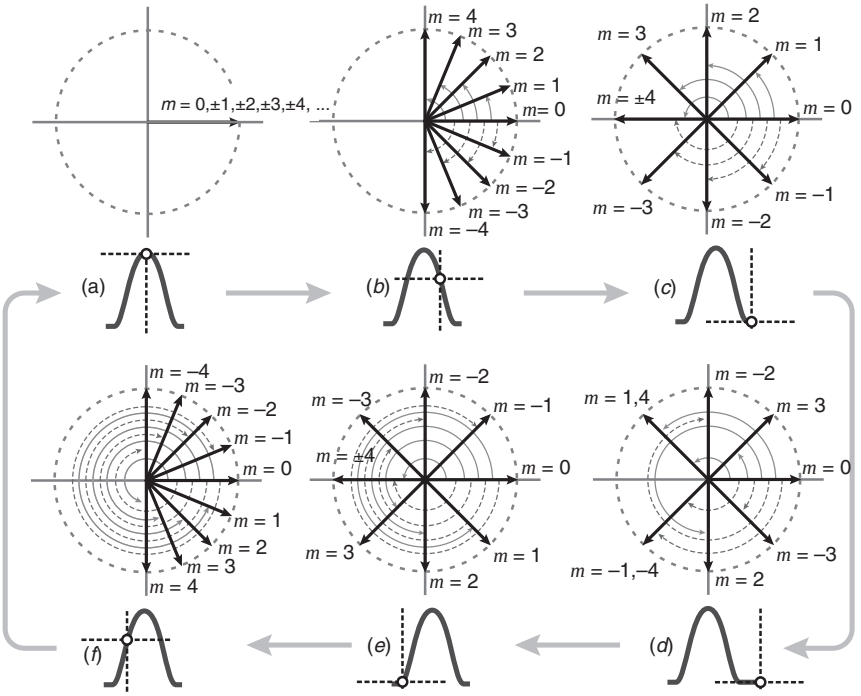


FIGURE 5.14: Phasor diagrams of the longitudinal modes in a mode-locked laser, at six distinct time instants, in one cross section of the laser. Note that the difference in angular velocity between each set of phasors is $\Delta\omega$. (a) All phasors adding in phase, corresponding to the maximum in the pulse amplitude; (b) phasors still adding to a significant positive amplitude; (c) all phasors are evenly spread and cancel out, and the pulse power is equal to zero; (d) the pulse power remains zero, as the phasors cancel out while rotating; (e) the phasors still cancel out, but a resulting power is about to grow after this time instant, as they are moving into alignment; and (f) the pulse power is growing, as more and more of the phasor components add in phase. The power will peak at when they are all in phase again, as shown in subplot (a).

a cavity with two equal amplitude adjacent longitudinal modes. Using the time and space-varying electric field representation from Eq. (2.20), and looking at the total electric field variation in time for constant z (using $z = 0$ without the loss of generality), the two mode fields can be expressed as

$$\mathscr{E}_1 = \hat{\mathbf{e}}_y E_0 U(x,y) e^{j(\omega t)} \tag{5.83}$$

$$\mathscr{E}_2 = \hat{\mathbf{e}}_y E_0 U(x,y) e^{j((\omega-\Delta\omega)t)} \tag{5.84}$$

where $\Delta\omega$ is the angular frequency difference between the two longitudinal cavity modes, per Eq. (2.32). In diode lasers, $\Delta\omega \ll \omega$, and adding the two modes to obtain the total electric field yields

$$\mathscr{E} = \mathscr{E}_1 + \mathscr{E}_2 = 2\hat{\mathbf{e}}_y E_0 U(x,y) e^{j\left(\frac{(2\omega+\Delta\omega)t}{2}\right)} e^{j\left(\frac{(\Delta\omega)t)}{2}\right)}. \tag{5.85}$$

This result corresponds to a fast oscillation at a frequency very close to ω, modulated by a slowly varying envelope, whose frequency of oscillation is $\frac{\Delta\omega}{2}$. Because the optical power is proportional the amplitude squared of the electric field, two simultaneously lasing modes give rise to pulselike optical intensity as function of time, at a given cross section of the laser.

In a practical laser cavity, the number N of cavity modes that can lase simultaneously will be determined by the gain bandwidth of the laser active region. If we now extend the two-mode approach to this case, the total electric field will be given by

$$\mathscr{E}(t) = \sum_N E_m e^{j(\omega_m t + \phi_m(t))} \tag{5.86}$$

where E_m is the amplitude, ω_m is the angular frequency, and $\phi_m(t)$ is the time-dependent phase of the mth lasing mode. The difference in angular frequency between adjacent cavity modes, per Eq. (2.32), is given by

$$\Delta\omega = \omega_m - \omega_{m-1} = \frac{2\pi c}{2\bar{n}_g L}, \tag{5.87}$$

where L is the cavity length, and $\bar{n}_g$ is the group velocity. We note that the resultant electric field is periodic in time, with a period defined by $T = \frac{2\pi}{\Delta\omega}$. We will use this information in the next subsection, when calculating the pulse repetition rate.

The intensity of the light is proportional to $|\mathscr{E}(t)|^2 = \mathscr{E}(t)\mathscr{E}^*(t)$, where $\mathscr{E}^*(t)$ is the complex conjugate of $\mathscr{E}(t)$, thus yielding

$$|\mathscr{E}(t)|^2 = \sum_N \sum_N E_m E_n e^{j(\omega_m - \omega_n)t + (\phi_m(t) - \phi_n(t))}. \tag{5.88}$$

If the phase difference term, $\phi_m(t) - \phi_n(t)$, is constant in time, we say that the modes are locked in phase. The practical implication of this case is that the temporal intensity of the light at any point in the laser can be described as a sum of a number

of phasor vectors, all rotating at angular velocities, which are integer multiples of $\Delta\omega$. Again, without the loss of generality, by fixing the phase differences at zero, assuming an even number N of modes supported, the laser gain peak at the angular frequency of ω_0, and for simplicity, assuming equal amplitudes of the modes, the resultant electric field can be described as

$$\mathscr{E}(t) = E_0 e^{j\omega_0 t} \sum_{m=-N/2}^{N/2} e^{jm\Delta\omega t} = E_0 e^{j\omega_0 t} \frac{\sin(\frac{N\Delta\omega t}{2})}{\frac{\Delta\omega t}{2}}. \tag{5.89}$$

Similarly to the two mode case, Eq. (5.85), we have the light intensity that is a product of two factors, one varying rapidly with the average frequency ω_0, and the other varying much more slowly, with the frequency of $\Delta\omega$. The slow varying term has a familiar dependence of a sinc function, encountered in Chapter 3, and responsible for the pulse shape. A graphic respresentation of the Eq. (5.89) is shown in Figure 5.14. At time $t = 0$, Fig. 5.14a, all the electric field phasor vectors are aligned along the real axis, and the optical pulse is at its peak power. As the time increases, the optical phasors rotate in the complex plane, reducing the pulse power (b), until the vector with the highest value of angular velocity, $\frac{N}{2}\Delta\omega$ reaches the negative real axis, when the pulse is no longer discernible, and its power goes to zero, as seen in (c). From this analysis, we can determine the width of the pulse by noting that

$$\frac{N}{2}\Delta\omega\Delta t_p = \pi \tag{5.90}$$

and thus

$$\Delta t_p = \frac{2\pi}{N\Delta\omega} = \frac{1}{\Delta\nu}, \tag{5.91a}$$

where $\Delta\nu$ is the gain bandwidth of the laser active region, encompassing all the active modes in the mode locked laser. This result is expected, as it is consistent with the Fourier transform relationship between the bandwidth and the pulse width.

Going back to the analysis of the mode-locked laser operation, Figure 5.14d, after the end of the pulse, the phasors will continue to rotate, spanning the whole complex plane, and the resulting output field will be zero. Another pulse will start appearing after the phasors have rotated to the positions shown in Fig. 5.14e. As they continue the rotation toward the real positive axis, the field resultant and the pulse intesity will grow (f). Finally, when the vector corresponding to the $m = 1$ (slowest) mode has rotated over the full circle and it coincides with the real positive axis, the other vectors will have rotated an integer number of full circles, and the pulse power will peak again, as shown in Fig. 5.14a. Therefore, the period corresponding to the pulse creation can be calculated based on the angular velocity of the $m = 1$ mode and the fact that this mode travels over the angle of 2π from

$$T = \frac{2\pi}{\Delta\omega} = \frac{2L\bar{n}_g}{c}. \tag{5.91b}$$

This is the same period that was established previously, from examining the case of two modes and Eq. (5.87). Intuitively, it illustrates that the pulse repetition rate is determined by the cavity parameters of the laser.

The output power from a multimode laser is equal to the sum of powers of each individual mode. The power contained in each of these modes is the same, whether or not the modes are locked in phase. If for illustrative purpuses, we assume that each mode has the same electric field amplitude E_0, the total output power when the modes are not locked will be proportional to $N \cdot |E_0|^2$. On the other hand, in the mode locked configuration, at the pulse peak, all the mode amplitudes add in phase, leading to the peak power proportional to $N^2 \cdot |E_0|^2$. Because the peak power in the mode locked laser is related to the average power $\langle P \rangle$ as

$$P = N \langle P \rangle.$$

a large number of oscillating modes can greatly enhance the pulse power in a mode locked laser. Mode locking does not change the average power of the laser beam, but rather redistributes this energy in time.

Mode Locking Techniques One simple technique that can be used for mode locking a laser is to introduce a time-variable loss inside the cavity. This type of mode locking is called active mode locking. If this loss is modulated in time at a frequency equal to the mode separation, $\Delta\omega$, it will lead to coupling of the two adjacent cavity modes. If the coupling is sufficiently strong, the phase of the different modes will be locked, resulting in mode locking.

In the time domain, with a sinusoidal loss modulation applied, the time between loss the minima is $T = 2\bar{n}_g L/c$, the round-trip time for a pulse in the cavity. Therefore, the light that arrives at the variable loss medium when the loss is at a minimum has a lower round-trip loss than light arriving at other times. This leads to the process of self-selection of the modes, where only the particular combination of modes that are locked in phase will prevail, having the lowest lasing threshold.

Another approach to achieving a mode-locked combination of modes spontaneously in the cavity, without the need for a user-supplied modulation, is called passive mode locking. This is generally accomplished by inserting a saturable absorber inside the cavity. The absorption coefficient of the saturable absorber decreases with increasing light intensity, which favors the development of high intensity pulses. The gain for the passively mode locked laser is set just below threshold, so that CW lasing will only occur if the light intensity is high enough to decrease the absorption loss in the saturable absorber. This, in turn, can occur if the modes lock together in phase to create short pulses because the peak power in each pulse is very high compared with the equivalent CW power. The pulse length is determined by the recovery speed of the saturable absorber material.

5.4.3 Turn-On Delay

The numerical simulations in Figs. 5.12 and 5.13 clearly illustrate that time is required for the carrier density to build up to the threshold value before light

is emitted. This *turn-on delay*, defined as t_d, can be detrimental in high-speed data links, however it can be useful for measuring the carrier lifetime. As a first approximation to estimating t_d, we can simply calculate the initial slope of the carrier density and draw a straight line up to the threshold value N_{th}. To determine the initial slope, we use the carrier density rate equation, assuming no appreciable photon buildup:

$$\frac{dN}{dt} = \frac{\eta_i I}{qV} - (R_{sp} + R_{nr}). \tag{5.92}$$

If we start with some initial current I_i then the term in parentheses under steady-state conditions can be set equal to $\eta_i I_i / qV$. Now we change the current instantaneously to I_f. The initial slope then becomes

$$\left.\frac{\Delta N}{\Delta t}\right|_{t=0} = \frac{\eta_i}{qV}(I_f - I_i) \approx \frac{N_{th} - N_i}{t_d} \tag{5.93}$$

Solving for the turn-on delay, we find

$$t_d \approx \frac{qV}{\eta_i}\frac{N_{th} - N_i}{I_f - I_i} \tag{5.94}$$

Replacing N with I using the carrier lifetime or differential carrier lifetime (both evaluated at threshold), the turn-on delay for the two extreme initial bias limits becomes

$$t_d \approx \tau_{th}\frac{I_{th}}{I_f}, \qquad (I_i \approx 0) \tag{5.95}$$

$$t_d \approx \tau_{\Delta Nth}\frac{I_{th} - I_i}{I_f - I_i}, \quad (I_i \approx I_{th}) \tag{5.96}$$

where $1/\tau_{th} \approx A + BN_{th} + CN_{th}^2$ and $1/\tau_{\Delta Nth} \approx A + 2BN_{th} + 3CN_{th}^2$. For small initial biases, the turn-on delay is proportional to the total carrier lifetime ($\sim 3 - 4$ ns), whereas for initial biases approaching threshold, the turn-on delay becomes proportional to the differential carrier lifetime ($\sim 1 - 2$ ns). If we know all currents involved, it is a simple matter to estimate both total and differential carrier lifetimes from measured turn-on delays. In practical applications we want to avoid a large turn-on delay. This is accomplished by (1) increasing I_f as much as possible above threshold, or (2) adjusting the bias level I_i close to threshold. In fact for I_i larger than threshold, the turn-on delay becomes very small and can be estimated using the methods discussed in Section 5.3.2.

The first-order estimates of t_d in Eqs. (5.95) and (5.96) neglect the fact that as the carrier density increases, so does the recombination term in Eq. (5.92). This increase in recombination reduces the rate of increase of N from linear to sublinear as illustrated in Fig. 5.15. Thus, although Eqs. (5.95) and (5.96) yield a reasonable first approximation to t_d, they tend to underestimate the value somewhat.

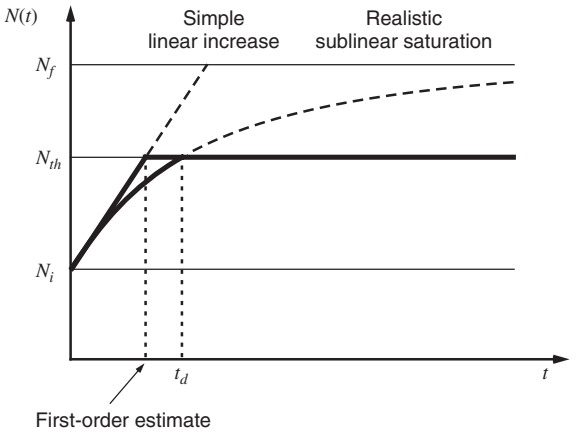


FIGURE 5.15: Carrier density as a function of time. The dashed curves show the carrier density increase without carrier clamping at threshold.

For a constant carrier lifetime (i.e., $R_{sp} + R_{nr} = N/\tau_0$), Eq. (5.92) yields an exponential saturation of the carrier density at N_f:

$$N(t) = N_i + (N_f - N_i)(1 - e^{-t/\tau_0}). \tag{5.97}$$

Of course, N will never reach N_f due to stimulated emission, which clamps N at N_{th} as Fig. 5.15 suggests. The time required for N to reach N_{th} defines the turn-on delay, and we can set $N(t_d) = N_{th}$. Solving for the turn-on delay we obtain

$$t_d = \tau_0 \ln \frac{I_f - I_i}{I_f - I_{th}}, \quad (\tau = \tau_0 \neq \tau(N)) \tag{5.98}$$

where the constant carrier lifetime has been multiplied through to convert all $N \rightarrow \tau_0 I$. Note from Fig. 5.15 that our initial linear approximation (5.85) is expected to work well when $N_f \gg N_i$ (or equivalently, when $I_f \gg I_i$), as a Taylor expansion of Eq. (5.98) for large $(I_f - I_i)$ verifies.

For short-wavelength lasers, the carrier lifetime is not a constant but is dominated by bimolecular recombination (i.e., $BN^2 \gg AN + CN^3$). For this case, the exponential saturation at N_f becomes a tanh function saturation, and the solution of Eq. (5.92) is given by

$$N(t) = N_f \tanh[BN_f t + \tanh^{-1}(N_i/N_f)]. \tag{5.99}$$

Again setting $N(t_d) = N_{th}$ and solving for t_d, we find

$$t_d = \tau_f \cdot \left[\tanh^{-1} \sqrt{\frac{I_{th}}{I_f}} - \tanh^{-1} \sqrt{\frac{I_i}{I_f}} \right], \quad (BN^2 \gg AN + CN^3) \tag{5.100}$$

where $\tau_f = 1/BN_f = \sqrt{qV/\eta_i I_f B}$, and we have set all $N \rightarrow \sqrt{\eta_i I/qVB}$.

In long-wavelength lasers, Auger recombination dominates, and we can assume $CN^3 \gg AN + BN^2$. Unfortunately, no closed-form expression for $N(t)$ can be obtained for this case. However, a solution can still be obtained from the more general definition of t_d, which is found by directly integrating Eq. (5.92) and solving for t_d:

$$t_d = \int_{N_i}^{N_{th}} \frac{dN}{N_f/\tau_f - N/\tau(N)}, \tag{5.101}$$

where $\tau_f = \tau(N_f)$. It is easily shown that both Eqs. (5.98) and (5.100) can be found from Eq. (5.92) by setting $\tau(N) = \tau_0$ and $1/\tau(N) = BN$, respectively.

With $1/\tau(N) = CN^2$, the resulting expression for the turn-on delay is rather lengthy, but is important to include for long-wavelength laser applications. Expressing t_d as two components (where the first is typically much larger than the second), we have

$$t_d = t_{d1} + t_{d2}, \qquad (CN^2 \gg AN + BN^2) \tag{5.102}$$

where

$$t_{d1} = \tau_f \cdot \frac{1}{6} \ln\left[\left(\frac{r_f - r_i}{r_f - 1}\right)^3 \frac{r_f^3 - 1}{r_f^3 - r_i^3}\right],$$

$$t_{d2} = \tau_f \cdot \frac{1}{\sqrt{3}}\left(\tan^{-1}\frac{1 + 2/r_f}{\sqrt{3}} - \tan^{-1}\frac{1 + 2r_i/r_f}{\sqrt{3}}\right).$$

The ratios are $r_f = N_f/N_{th}$ and $r_i = N_i/N_{th}$, or equivalently $r_f^3 = I_f/I_{th}$ and $r_i^3 = I_i/I_{th}$. In addition, $\tau_f = 1/CN_f^2$ where N_f is defined by $CN_f^3 = (\eta_i/qV)I_f$.

The turn-on delay for the case where $AN + BN^2 \gg CN^3$ can also be found.

Figure 5.16 shows t_d/τ_f for various initial-to-threshold and final-to-threshold current ratios using the three different expressions for the turn-on delay Eqs. (5.89), (5.91), and (5.93). With no prebias ($I_i = 0$), the turn-on delay is generally on the order of the would-be carrier lifetime. In contrast, with a prebias and a final current level that is well above threshold, the turn-on delay can be reduced to just a fraction of the final carrier lifetime.

5.4.4 Large-Signal Frequency Chirping

Here we shall briefly introduce a useful analytic formula to calculate the frequency chirp from the known intensity modulation waveform, which holds even under large-signal modulation. Taking the photon density rate equation and solving for the difference between the gain and loss, we obtain

$$\Gamma v_g g - \frac{1}{\tau_p} = \frac{1}{N_p}\frac{dN_p}{dt} - \frac{\Gamma R'_{sp}}{N_p}. \tag{5.103}$$

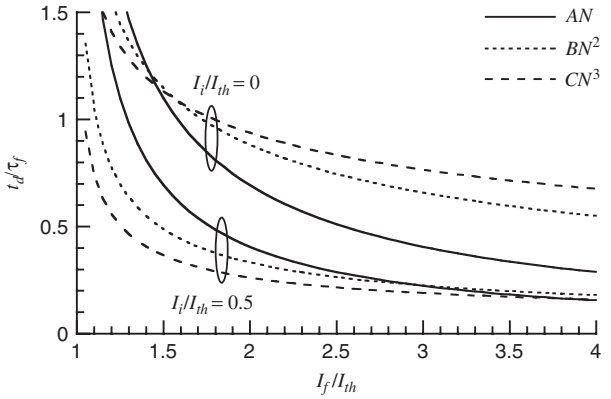


FIGURE 5.16: Normalized turn-on delay for different initial and final currents relative to threshold for the three types of recombination.

Because of carrier and gain clamping above threshold, a first-order expansion of the gain around its threshold value remains a good approximation even for large signals, allowing us to set

$$g = g_{th} + a(N - N_{th}) - a_p N_p. \tag{5.104}$$

Substituting this into Eq. (5.94), recognizing that $\Gamma v_g g_{th} = 1/\tau_p$, and using Eq. (5.75) to express the carrier density deviation as a frequency deviation, we obtain

$$\nu(t) - \nu_{th} = \frac{\alpha}{4\pi}\left[\frac{1}{N_p}\frac{dN_p}{dt} - \frac{\Gamma R'_{sp}}{N_p} + \Gamma v_g a_p N_p\right]. \tag{5.105}$$

With $N_p = N_{p0} + N_{p1}e^{j\omega t}$, the time-varying portion of the frequency chirp formula reduces to the small-signal result Eq. (5.77) derived earlier for $N_{p1} \ll N_{p0}$. If the modulated output power varies rapidly (on the order of a nanosecond or less), then the first term in brackets will usually dominate. Under these conditions and using the fact that $N_p \propto P_0$, the frequency chirp $\Delta\nu(t) = \nu(t) - \nu_{th}$ reduces to the simple result

$$\Delta\nu(t) = \frac{\alpha}{4\pi} \cdot \frac{1}{P_0(t)}\frac{dP_0(t)}{dt}. \tag{5.106}$$

With this equation, the frequency chirp for large-signal modulation can be determined directly from the shape of the modulated signal (e.g., square-wave, sine-wave, Gaussian). If we are interested in minimizing the frequency chirp, Eq. (5.106) can be used to predict which output waveforms produce the lowest chirp. We can then tailor the current input signal accordingly.

5.5 RELATIVE INTENSITY NOISE AND LINEWIDTH

Up until now we have only considered intensity and frequency responses due to the deliberate modulation of the current or some other cavity parameter. In the steady state, it has been assumed that the carrier and photon densities are constant. However, in reality random carrier and photon recombination and generation events produce instantaneous time variations in the carrier and photon densities, even with no applied current modulation. The variations in photon density lead to variations in the magnitude of the output power, which provides a noise floor, and the variations in carrier density result in variations in the output wavelength, which creates a finite spectral linewidth for the lasing mode. Before launching into the specifics of how to deal with these random noise sources, it is worth considering the implications of intensity and frequency noise in practical laser applications for motivational purposes.

5.5.1 General Definition of RIN and the Spectral Density Function

Figure 5.17 illustrates a noisy laser output for both analog and digital signal transmission. For analog applications, the noise is quantified using the electrical signal-to-noise ratio (SNR). For the laser output defined in Fig. 5.17, the SNR can be written as

$$\text{SNR} = \frac{\langle i_S^2 \rangle}{\langle i_N^2 \rangle} = \frac{\langle (P_1 \sin \omega t)^2 \rangle}{\langle \delta P(t)^2 \rangle} = \frac{m^2}{2} \frac{P_0^2}{\langle \delta P(t)^2 \rangle}, \tag{5.107}$$

where the IM modulation index is given by $m = P_1/P_0$. The $\langle\rangle$ denote the time average.

For digital applications, a decision level at the midpoint defines whether a "0" or "1" is recorded. If the noise happens to exceed $P_0/2$ in Fig. 5.17, then a false recording might be made. If the noise has a Gaussian distribution around the mean power level, then to reduce the probability of finding $|\delta P(t)| > P_0/2$ to less than 1 in 10^9 (i.e., a *bit-error rate* $< 10^{-9}$), we require that [4]

$$\frac{P_0^2}{\langle \delta P(t)^2 \rangle} > (11.89)^2, \quad (\text{for BER} < 10^{-9}) \tag{5.108}$$

where $\langle \delta P(t)^2 \rangle$ is the mean-square of the assumed Gaussian noise distribution. Bit error rate of less than 10^{-9} is sufficient for data rates lower than 10 Gigabits/s. For data rates in the range of 10–100 Gigabits/s, BERs better than 10^{-10} are generally required. However, with implementation and deployment of advanced forward error correction coding techniques, this type data BERs can be achieved through error correction from signals whose BERs are in the range of 10^{-5}.

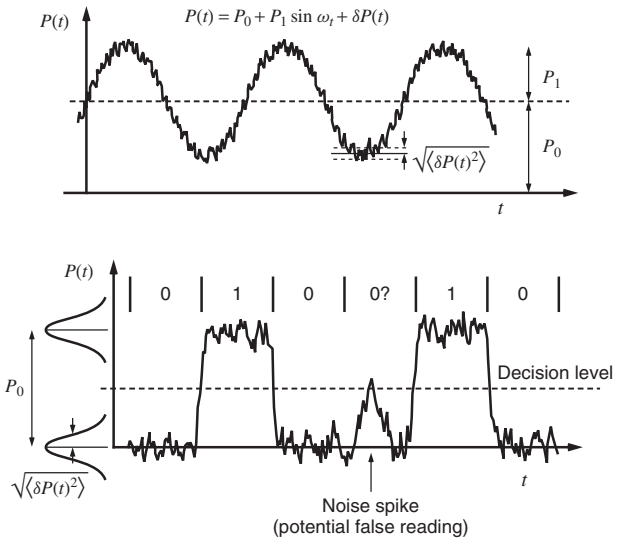


FIGURE 5.17: Noise in modulated laser signals for both analog and digital applications.

For either of these two applications, we find it useful to quantify the *relative intensity noise* (RIN) of the laser:

$$\text{RIN} \equiv \frac{\langle \delta P(t)^2 \rangle}{P_0^2}. \tag{5.109}$$

The RIN is often described in decibels, or $10\log_{10}(\text{RIN})$. For analog applications, if a given electrical SNR is required, then Eq. (5.107) can be used to determine the maximum allowable RIN. For example, if we require the SNR > 50 dB with $m = 1$, then the laser must have a RIN < -53 dB. Alternatively, for a bit-error rate (BER) $< 10^{-9}$ in digital applications, Eq. (5.108) suggests that the laser must have a RIN < -21.5 dB.

To quantify the output power fluctuations (and hence the RIN), it is more convenient to work in the frequency domain, making use of the Fourier transform pairs:

$$\delta P(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \delta P(\omega) e^{j\omega t} d\omega, \tag{5.110}$$

$$\delta P(\omega) = \int_{-\infty}^{+\infty} \delta P(t) e^{-j\omega t} dt, \tag{5.111}$$

where $\delta P(\omega)$ is the component of the noise, which fluctuates at the frequency, ω. Now suppose we were to use a spectrum analyzer to measure the *electrical* power

(i.e., the square of the optical power) associated with the noise. If the spectrum analyzer applies a narrowband filter to the signal with a passband described by $F(\omega)$, then the measured mean-square time-averaged signal would be given by

$$\langle \delta P(t)^2 \rangle = \frac{1}{(2\pi)^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \langle \delta P(\omega) \delta P(\omega')^* \rangle F(\omega) F(\omega')^* e^{j(\omega-\omega')t} \, d\omega \, d\omega'. \tag{5.112}$$

For completely random noise, the magnitude of the noise at any given frequency is completely uncorrelated with the magnitude of the noise at any other frequency. As a result, when the product of two frequency components is averaged over time, there is a delta function correlation between them (see Appendix 13). The strength of the delta function correlation is defined as the *spectral density*, $S_{\delta P}(\omega)$, of $\delta P(t)$ at ω, and we can write

$$\langle \delta P(\omega) \delta P(\omega')^* \rangle = S_{\delta P}(\omega) \cdot 2\pi \delta(\omega - \omega'). \tag{5.113}$$

With this substitution, the *measured* mean-square power fluctuation reduces to

$$\langle \delta P(t)^2 \rangle = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S_{\delta P}(\omega) |F(\omega)|^2 d\omega. \tag{5.114}$$

If the measurement filter is centered at ω_0, and is narrowband relative to variations in the spectral density, then with $F(\omega_0) = 1$ we obtain

$$\langle \delta P(t)^2 \rangle \approx S_{\delta P}(\omega_0) \int_{-\infty}^{\infty} |F(\omega)|^2 \, df = S_{\delta P}(\omega_0) \cdot 2\Delta f. \tag{5.115}$$

This relation is graphically illustrated in Fig. 5.18 for an arbitrary noise spectrum. Note that the effective measurement bandwidth is $2\Delta f$ because we

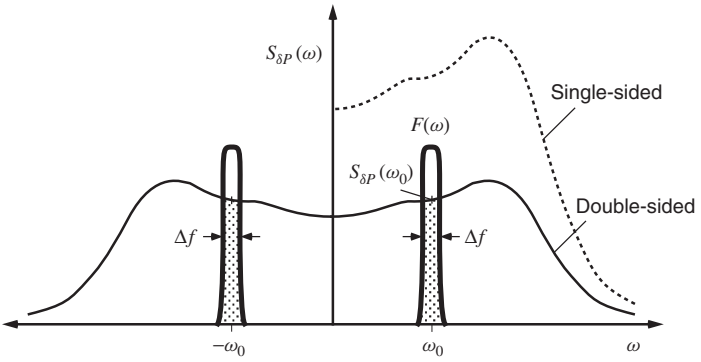


FIGURE 5.18: Measured noise using a narrowband filter. Because the noise spectral density is always an even function of frequency ($S_{\delta P}(-\omega) = S_{\delta P}(\omega)$), we can fold the spectrum in half, if desired, and define a *single-sided* spectrum existing only in the positive frequency domain that is a factor of two larger than the *double-sided* spectrum.

must include both positive and negative frequencies. We could alternatively define the spectral density as "single-sided" existing only in the positive frequency domain as indicated by the dashed line in Fig. 5.18. In this case, the measurement bandwidth would simply be Δf, and the factor of two would be lumped into the single-sided spectral density. The choice of using a single-sided or double sided spectral density is academic as long as we are consistent. In this chapter, the spectral density will always be defined as double sided. Finally note that the spectral density has units of (seconds) $\times$ (fluctuating variable units)^2.

In terms of the spectral density of the noise accompanying the signal, we can redefine the relative intensity noise as

$$\frac{\text{RIN}}{\Delta f} = \frac{2S_{\delta P}(\omega)}{P_0^2}, \tag{5.116}$$

where Δf is the filter bandwidth of the measurement apparatus (if the spectral density is defined as single sided, then the factor of 2 should be removed in Eq. (5.107)). Because the measurement bandwidth can vary from application to application, it is common to specify the quantity on the left as RIN in dB/Hz or RIN per unit bandwidth. The full RIN is then found by integrating the RIN per unit bandwidth over the (single-sided) detection bandwidth of the system of practical interest. In designing a communications system, the desired SNR or BER sets a maximum limit on the total RIN of the laser. If the RIN spectrum is flat, then the required RIN per unit bandwidth of the laser is found from

$$\text{RIN(dB/Hz)} = \text{RIN(dB)} - 10\log_{10}(\Delta f[\text{in Hz}]). \tag{5.117}$$

Example 5.4 A new DFB laser has been fabricated for use in digital transmission links. For a digital link, assume that the required bandwidth is 0.75 of the link bit rate.

Problem: (a) What is the minimum acceptable laser RIN per unit bandwidth for a link operating at 2.5 GB/s? (b) If we wanted to have an option to use the same laser in an upgraded link, operating at 10 GB/s, what is the new RIN per unit bandwidth requirement?

Solution: In a digital link, a maximum allowable bit error rate is $BER = 1 \cdot 10^{-9}$. From Eq. (5.108), the required laser RIN needs to be RIN ≤ -21.5 dB. The RIN per unit bandwidth can be computed from Eq. (5.117).

(a) for a 2.5 GB/s link, we have

$$\begin{aligned}\text{RIN(dB/Hz)} &= \text{RIN(dB)} - 10\log_{10}(\Delta f) = -21.5 - 10\log_{10}(1.875 \cdot 10^9) \\ &= -114.23 \text{ dB/Hz}.\end{aligned}$$

(b) for a 10 GB/s link, we have

$$\begin{aligned}\text{RIN(dB/Hz)} &= \text{RIN(dB)} - 10\log_{10}(\Delta f) = -21.5 - 10\log_{10}(0.75 \cdot 10^{10}) \\ &= -120.25 \text{ dB/Hz}.\end{aligned}$$

5.5.2 The Schawlow–Townes Linewidth

In addition to intensity noise, the laser also produces frequency noise, which can adversely affect the lasing spectrum. In single-mode lasers, such as the DFB and DBR types introduced in Chapter 3, the spectral width of the laser's output is reduced to that of a single mode; however, the linewidth of this single mode is still finite due to laser noise. Unless great care is taken, experiments typically show diode laser linewidths much greater than a megahertz. For many applications, such as sensor or communication systems using coherent detection, it is desirable to have linewidths much less than a megahertz. Thus, an understanding of the inherent linewidth of diode lasers is of great practical importance.

The linewidth of a diode laser results from phase fluctuations in its output. These arise from two basic sources: (1) spontaneous emission and (2) carrier density fluctuations. The first is inherent in all lasers, resulting simply from the random addition of spontaneously emitted photons to the quasi-coherent resonant cavity mode. The second is of significance only in diode lasers, and it results from the proportionality between ΔN and $\Delta\nu$ characterized by Eq. (5.75); the constant of proportionality contains the *linewidth enhancement factor*, α. This factor exists because both the gain and the index of refraction depend directly on the carrier density.

We will develop the full expression for the laser linewidth a little later, after the treatment of frequency noise. For now, we wish to consider a simplified derivation of the spontaneous emission component of the laser linewidth, which, although not entirely correct for lasers above threshold, does provide the reader with an intuitive feel for the origin of a finite laser linewidth.

In the derivation of the rate equations in Chapter 2, the cavity lifetime was introduced as the natural decay rate of photons in the resonant cavity in the absence of stimulated or spontaneous emission sources. Therefore, in the absence of sources, Eq. (5.4) has the solution

$$N_p(t) = N_{p0}e^{-t/\tau_p}. \tag{5.118}$$

The corresponding time dependence of the field is

$$\mathscr{E}(t) = E_0 e^{j\omega_0 t} e^{-t/2\tau_p} u(t), \tag{5.119}$$

where $u(t)$ is a unit step function that turns on at time zero to indicate that the field is instantaneously created at $t = 0$ by a stimulated emission event, for example. The Fourier transform of this time domain response gives the frequency domain

response of the laser cavity. The Fourier transform of an exponential is a Lorentzian. This undriven result, sometimes called the cold cavity response, is given by

$$|\mathscr{E}(\omega)|^2 = \frac{|\mathscr{E}(\omega_0)|^2}{1 + (\omega - \omega_0)^2(2\tau_p)^2}. \tag{5.120}$$

From this we can see that the full-width half-maximum (FWHM) linewidth of the cold cavity is $\Delta\omega = 1/\tau_p$. This spectral width corresponds to the filter bandwidth of the Fabry–Perot resonator mode with no active material present (hence cold cavity response). The key point here is that the resonance width is linked with the photon decay rate. Now, if we add back the stimulated term that is responsible for gain in the cavity, we see from Eq. (5.4) that the same exponential solution in time is obtained, but it is now characterized by a new *effective* cavity lifetime:

$$\frac{1}{\tau_p'} = \frac{1}{\tau_p} - \Gamma v_g g. \tag{5.121}$$

The effective cavity lifetime *increases* as the gain in the cavity compensates for cavity losses. Thus with gain, the FWHM linewidth becomes $\Delta\omega = 1/\tau_p'$, and so as τ_p' increases, the resonance width decreases. As illustrated in Section 5.2.2, Eq. (5.4) can be solved in the steady state for N_p:

$$N_p = \frac{\Gamma R_{sp}'}{1/\tau_p - \Gamma v_g g}. \tag{5.122}$$

Using Eq. (5.113) to replace $[1/\tau_p - \Gamma v_g g]$ in Eq. (5.112), we can express the driven FWHM linewidth as

$$\Delta\nu_{spon} = \frac{1}{2\pi\tau_p'} = \frac{\Gamma R_{sp}'}{2\pi N_p}. \quad \text{(spontaneous only)} \tag{5.123}$$

Equation (5.123) is equivalent to the famous Schawlow–Townes linewidth formula [6]. One central conclusion of this formula is that the linewidth varies inversely with photon density (or output power). And because the photon density in a laser can grow very large, the linewidth can collapse into a very narrow spectral line—one of the defining characteristics of lasers.

Unfortunately, this intuitive derivation has some shortcomings. Equation (5.123) does correctly give the below-threshold linewidth and is therefore accurate for amplified spontaneous emission problems. However, above threshold, the nonlinear coupling between the rate equations suppresses one of the two quadrature components of the noise (the field amplitude fluctuations are stabilized above threshold), resulting in a factor of 2 reduction in the linewidth predicted here [7, 8]. With a correction factor of 1/2, Eq. (5.123) becomes the *modified* Schawlow–Townes linewidth formula:

$$(\Delta\nu)_{ST} = \frac{\Gamma R_{sp}'}{4\pi N_p}. \tag{5.124}$$

The modified Schawlow–Townes linewidth, however, still only considers spontaneous emission noise and does not include carrier noise. To describe carrier noise, we will need to develop the Langevin rate equation approach to laser noise.

5.5.3 The Langevin Approach

To determine the laser RIN and the carrier noise, we must find the spectral density of the output power and carrier noise fluctuations. For this purpose, we introduce *Langevin* noise sources $F_N(t)$ and $F_P(t)$ as the AC driving sources for the carrier and photon densities, respectively. These sources are assumed to be white noise (see Appendix 13) and are assumed small enough that we can make use of the *differential* rate equations. For a constant drive current ($dI = 0$), Eqs. (5.32) and (5.33) become

$$\frac{d}{dt}(dN) = -\gamma_{NN}\, dN - \gamma_{NP} dN_p + F_N(t), \tag{5.125}$$

$$\frac{d}{dt}(dN_p) = \gamma_{PN}\, dN - \gamma_{PP} dN_p + F_P(t). \tag{5.126}$$

The rate coefficients are defined by Eqs. (5.34) as always. To solve these equations it is again convenient to place them in matrix form:

$$\frac{d}{dt}\begin{bmatrix} dN \\ dN_p \end{bmatrix} = \begin{bmatrix} -\gamma_{NN} & -\gamma_{NP} \\ \gamma_{PN} & -\gamma_{PP} \end{bmatrix}\begin{bmatrix} dN \\ dN_p \end{bmatrix} + \begin{bmatrix} F_N(t) \\ F_P(t) \end{bmatrix}. \tag{5.127}$$

To determine the spectral densities we must first transform to the frequency domain. Replacing all time-dependent variables with equivalent versions of Eq. (5.110), we obtain for each frequency component:

$$\begin{bmatrix} \gamma_{NN} + j\omega & \gamma_{NP} \\ -\gamma_{PN} & \gamma_{PP} + j\omega \end{bmatrix}\begin{bmatrix} N_1(\omega) \\ N_{p1}(\omega) \end{bmatrix} = \begin{bmatrix} F_N(\omega) \\ F_P(\omega) \end{bmatrix}, \tag{5.128}$$

where N_1, N_{p1}, F_N, and F_P represent the components of the noise that fluctuate at frequency ω. This result is analogous to the small-signal result obtained in Section 5.3.1. Using Cramer's rule and the definitions in Section 5.3.1, we immediately obtain

$$N_1(\omega) = \frac{H(\omega)}{\omega_R^2}\begin{vmatrix} F_N(\omega) & \gamma_{NP} \\ F_P(\omega) & \gamma_{PP} + j\omega \end{vmatrix}, \tag{5.129}$$

$$N_{p1}(\omega) = \frac{H(\omega)}{\omega_R^2}\begin{vmatrix} \gamma_{NN} + j\omega & F_N(\omega) \\ -\gamma_{PN} & F_P(\omega) \end{vmatrix}. \tag{5.130}$$

Using Eq. (5.113) we can define the carrier and photon density spectral densities as

$$S_N(\omega) = \frac{1}{2\pi}\int \langle N_1(\omega)N_1(\omega')^*\rangle d\omega', \tag{5.131}$$

$$S_{N_p}(\omega) = \frac{1}{2\pi}\int \langle N_{p1}(\omega)N_{p1}(\omega')^*\rangle d\omega'. \tag{5.132}$$

Multiplying both sides of Eq. (5.129) by $N_1(\omega')^*$, and both sides of Eq. (5.130) by $N_{p1}(\omega')^*$, taking the time average, and integrating over ω', we finally obtain

$$S_N(\omega) = \frac{|H(\omega)|^2}{\omega_R^4}[\gamma_{NP}^2\langle F_P F_P\rangle - 2\gamma_{PP}\gamma_{NP}\langle F_P F_N\rangle + (\gamma_{PP}^2+\omega^2)\langle F_N F_N\rangle], \tag{5.133}$$

$$S_{N_p}(\omega) = \frac{|H(\omega)|^2}{\omega_R^4}[(\gamma_{NN}^2+\omega^2)\langle F_P F_P\rangle + 2\gamma_{NN}\gamma_{PN}\langle F_P F_N\rangle + \gamma_{PN}^2\langle F_N F_N\rangle]. \tag{5.134}$$

where the Langevin noise source spectral densities are defined by

$$\langle F_i F_j\rangle = \frac{1}{2\pi}\int \langle F_i(\omega)F_j(\omega')^*\rangle\, d\omega'. \tag{5.135}$$

Because F_N and F_P are white noise sources, their noise spectral densities $\langle F_i F_j\rangle$ are uniformly distributed over all frequencies. Hence, the various $\langle F_i F_j\rangle$ can be regarded as constants in the frequency domain (see Appendix 13 for further details). Equations (A13.28) and (A13.26) therefore reveal that the carrier and photon density fluctuations follow a $(a_1 + a_2\,\omega^2)|H(\omega)|^2$ spectral dependence, peaking at the relaxation resonance frequency (we will quantify a_1 and a_2 later). This behavior is not surprising because the natural resonance of the carrier–photon system would be expected to accentuate and amplify any noise existing near that resonance. Furthermore, the inverse dependence on ω_R^4 suggests that the fluctuations and corresponding noise decrease with increasing output power. To fully quantify these relationships, we need to expand the terms within square brackets.

5.5.4 Langevin Noise Spectral Densities and RIN

Equations A13.28 and A13.26 reveal that with the Langevin method, the evaluation of the laser noise boils down to evaluating the spectral densities or *noise correlation strengths* $\langle F_i F_j\rangle$ between the various Langevin noise sources. Assuming $N_p V_p \gg 1$ above threshold, Appendix 13 shows that these terms reduce to the following:

$$\langle F_P F_P\rangle = 2\Gamma R'_{sp} N_p, \tag{5.136}$$

$$\langle F_N F_N\rangle = 2R'_{sp}N_p/\Gamma - v_g g N_p/V + \eta_i(I+I_{th})/qV^2, \tag{5.137}$$

$$\langle F_P F_N\rangle = -2R'_{sp}N_p + v_g g N_p/V_p. \tag{5.138}$$

The various factors appearing in these equations are essentially the summation of the shot noise associated with the random generation and recombination/escape of both photons and carriers in both reservoirs. The shot noise associated with the injected current (which gives rise to the $\eta_i I/qV^2$ term appearing in Eq. (5.137)) can in principle be eliminated or at least reduced by careful design of the current drive circuitry [9]—a point we will return to later in this section.

With Eqs. (5.136) through (5.138) the carrier Eq. (5.124) and photon Eq. (5.125) spectral density functions can be expanded. The desired output power spectral density function, $S_{\delta P}(\omega)$ needed in Eq. (5.116), can then be calculated. However, this calculation is complicated by an additional noise term in the output power due to the negative correlation between photons reflected and transmitted at the output mirror [10]. Thus, as shown in Appendix 13, $S_{\delta P}(\omega)$ is not simply related to $S_{N_P}(\omega)$ by the expected factor $(h\nu V_p v_g \alpha_m F)^2$ or $(\eta_0 h\nu V_p/\tau_p)^2$. After first deriving $S_{N_P}(\omega)$, Appendix 13 shows that for above- threshold conditions, the output power spectral density can be written as

$$S_{\delta P}(\omega) = h\nu P_0 \cdot \left[\frac{a_1 + a_2\ \omega^2}{\omega_R^4}|H(\omega)|^2 + 1\right], \tag{5.139}$$

where

$$a_1 = \frac{8\pi(\Delta\nu)_{ST}P_0}{h\nu}\frac{1}{\tau_{\Delta N}^2} + \eta_0\ \omega_R^4\left[\frac{\eta_i(I + I_{th})}{I_{st}} - 1\right],$$

$$a_2 = \frac{8\pi(\Delta\nu)_{ST}P_0}{h\nu} - 2\eta_0\ \omega_R^2\frac{\Gamma a_p}{a},$$

and $(\Delta\nu)_{ST} = \Gamma R'_{sp}/4\pi N_p, I_{st} = qP_0/\eta_0 h\nu = \eta_i(I - I_{th})$. Also, P_0 defines the power out of the desired facet.

If the emitted field is in a perfectly coherent state, the output power noise (double-sided) spectral density is limited to a minimum value of $h\nu P_0$. This quantum noise floor of the coherent field is often referred to as the *standard quantum limit*, or the shot noise floor. The two contributions within the square brackets of Eq. (5.130) can therefore be thought of as the excess intensity noise and the inherent quantum noise of the laser. At low output powers, the excess noise dominates and is dramatically enhanced near the relaxation resonance frequency of the laser. At high powers, the laser generally quiets down to the standard quantum limit except near the resonance, where the excess noise can persist even at very high output powers. Figure 5.19 illustrates these characteristics for the in-plane laser characterized in Table 5.1.

Characteristics of the RIN Spectrum The laser RIN plotted in Fig. 5.19 is related to $S_{\delta P}(\omega)$ through Eq. (5.116) and can be written as

$$\frac{\text{RIN}}{\Delta f} = \frac{2h\nu}{P_0}\left[\frac{a_1 + a_2\ \omega^2}{\omega_R^4}|H(\omega)|^2 + 1\right]. \tag{5.140}$$

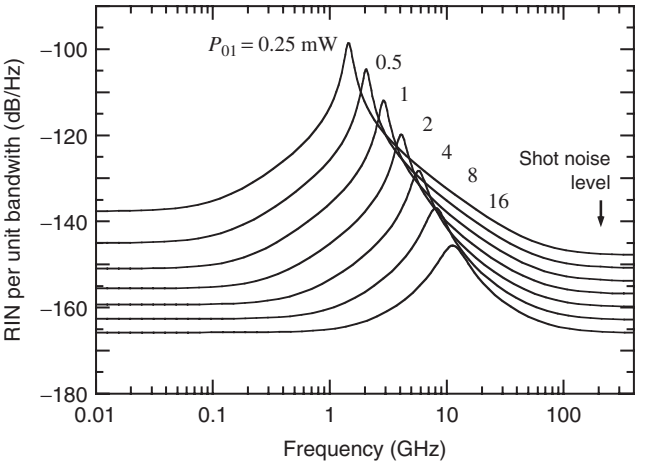


FIGURE 5.19: Calculated relative intensity noise at different output power levels for an InGaAs/GaAs in-plane laser with parameters given in Table 5.1.

The prefactor $2h\nu/P_0$ is the standard quantum limit for the minimum RIN of the laser. The frequency coefficients, a_1 and a_2, are as defined in Eq. (5.139). The first terms of both a_1 and a_2 are independent of power, whereas the latter terms depend on P_0^2 and P_0, respectively. At low powers, these latter terms can be neglected and the RIN reduces to

$$\frac{\text{RIN}}{\Delta f} = 16\pi(\Delta\nu)_{ST}\frac{1/\tau_{\Delta N}^2 + \omega^2}{\omega_R^4}|H(\omega)|^2 + \frac{2h\nu}{P_0}. \qquad \text{(low power)} \qquad (5.141)$$

This form of the RIN is commonly used in laser applications and in fact is a reasonable approximation at higher powers as well, as long as the laser current source is shot noise-limited. With Eq. (5.141), the measured RIN spectrum can be modeled using four fitting parameters: $\tau_{\Delta N}, \gamma, \omega_R$, and $(\Delta\nu)_{ST}$ [11].

When the excess noise term in Eq. (5.141) dominates, the RIN spectrum levels off to a constant value for $\omega < 1/\tau_{\Delta N}$, which is typically in the range of 100 MHz, as shown in Fig. 5.19 for the 0.25 and 0.5 mW curves. At higher powers, this range of constant RIN increases up to a few GHz, as the low frequency RIN saturates at the shot noise floor. Setting $\omega = 0$ in Eq. (5.140) and using a_1 as defined in Eq. (5.139), the low frequency RIN becomes

$$\frac{\text{RIN}}{\Delta f} = \frac{16\pi(\Delta\nu)_{ST}}{\omega_R^4\tau_{\Delta N}^2} + \frac{2h\nu}{P_0}\left[\eta_0\frac{\eta_i(I+I_{th})}{I_{st}} + (1-\eta_0)\right]. \qquad (\omega \ll \omega_R) \quad (5.142)$$

The first term decreases as $1/P_0^3$ (since $(\Delta\nu)_{ST} \propto 1/P_0$ and $\omega_R^4 \propto P_0^2$) and hence quickly drops below the shot noise floor with increasing power. The second term converges toward the shot noise $1/P_0$ power dependence at high current levels

where $\eta_i(I + I_{th})/I_{st} \rightarrow 1$. In Fig. 5.19, we can see this trend by examining the intervals between every doubling of power. At low frequencies, the steps change from 9 dB to 3 dB with increasing power as the power dependence changes from $1/P_0^3$ to $1/P_0$.

At the RIN peak, $a_2\ \omega^2 \gg a_1$ in Eq. (5.140). Hence, the frequency dependence in this range is characterized by $\omega^2|H(\omega)|^2$. Using Eq. (5.46), it can be shown that the peak of this function occurs at ω_R regardless of the extent of damping, in contrast to $|H(\omega)|^2$ which peaks right at ω_R only when $\gamma/\omega_R \ll 1$ (see Fig. 5.4). Setting $|H(\omega_R)|^2 = (\omega_R/\gamma)^2$ in Eq. (5.140), and neglecting both the noise floor limit and the second term of a_2 in Eq. (5.139), we obtain the simple result

$$\frac{\text{RIN}}{\Delta f} = \frac{16\pi(\Delta\nu)_{ST}}{\gamma^2}. \qquad (\omega = \omega_R) \tag{5.143}$$

In general, the damping factor $\gamma = Kf_R^2 + \gamma_0$ from Eq. (5.52). However, at higher powers, $\gamma \approx Kf_R^2 \propto P_0$. Hence, the RIN peak converges toward a $1/P_0^3$ power dependence because $(\Delta\nu)_{ST} \propto 1/P_0$. In other words, the RIN peak drops by 9 dB with every doubling of power, as Fig. 5.19 indicates at the higher power levels. Beyond the resonance frequency, the RIN reduces to the standard quantum limit, which drops by 3 dB with every doubling of power.

Example 5.5

Problem: For the in-plane laser whose parameters are described in Table 5.1, determine the peak RIN value at 1 mW output power.

Solution: (in-plane) At 1 mW of output power, the in-plane laser has a Schawlow–Townes linewidth of $(\Delta\nu)_{ST} = 1.07$ MHz and a damping factor of $\gamma = 2.88 \cdot 10^9$/s. The peak RIN value is given by

$$\frac{\text{RIN}}{\Delta f} = \frac{16\pi(\Delta\nu)_{ST}}{\gamma^2} = -111.88 \text{ dB/Hz}$$

This is illustrated in Fig. 5.19. This peak level of noise barely satisfies the criterion for a 2 Gbit/s (1 GHz) digital transmission link (see the previous example, following Eq. (5.117)). However, the peak occurs beyond the system bandwidth and will not add to the detected noise. When the 1 mW RIN spectrum in Fig. 5.19 is integrated from 0 to 1 GHz, the average RIN per unit bandwidth is only −140 dB/Hz, easily satisfying the transmission link requirements.

Noise-Free Operation The component of the laser noise, which derives from the noise of the current source appears in the second term of a_1 in Eq. (5.139). This is one component of the noise we can modify externally, and ideally eliminate. In Eq. (5.140), $a_2\ \omega^2|H(\omega)|^2$ is responsible for the RIN peak at ω_R, and $2h\nu/P_0$ dominates the spectrum beyond the RIN peak. Thus, $a_1|H(\omega)|^2$ only significantly affects the low frequency RIN away from the RIN peak, or at $\omega \ll \omega_R$. Within the denominator of $a_1|H(\omega)|^2$ then, we can set $(\omega_R^2 - \omega^2)^2 \approx \omega_R^4$ without introducing much

error in the total RIN spectrum (this works well as long as $(\gamma/\omega_R)^2 g 2/(1+2\sqrt{r_p})$ where r_p is the RIN peak relative to the shot noise floor). With this modification, and also concentrating on moderate-to-high powers such that we can neglect the $1/P_0^3$ term of a_1, Eq. (5.140) reduces to

$$\frac{\text{RIN}}{\Delta f} \approx \frac{2h\nu}{P_0}\left[\frac{\eta_0\eta_i(I+I_{th})/I_{st}+(1-\eta_0)+\omega^2\tau_p''^2}{1+\omega^2\tau_p''^2}\right]+\frac{2h\nu}{P_0}\frac{a_2}{\omega_R^4}\omega^2|H(\omega)|^2, \tag{5.144}$$

where $\tau_p'' \equiv \gamma/\omega_R^2 \approx \tau_p[1+\Gamma a_p/a]$. The first term provides the RIN noise floor, while the second term provides the RIN peak, which rises above the noise floor near ω_R. The various contributions to the noise floor originate from (in order of appearance in the first term): (1) injected current (I) and carrier recombination (I_{th}), (2) the random selection of output photons (partition noise), and (3) the random delay associated with photons escaping the cavity (only observable at frequencies comparable to or greater than $1/\tau_p''$).

The injected current contribution to the RIN noise floor can be reduced if a current source with subshot noise characteristics is used. Yamamoto has suggested and demonstrated along with others that if a high impedance source is used in driving the laser, then subshot noise current injection is possible [9, 12]. To accommodate such cases, we can generalize Eq. (5.144) by setting $I \to S_I/q$, where S_I is the double-sided spectral density of the injection current. For a perfectly quiet current source, $S_I = 0$. For a shot noise-limited current source, $S_I = qI$. If we have a very efficient laser with $\eta_0 \to 1$, and if we are far above threshold such that $I \gg I_{th}$ and $I_{st} \approx \eta_i I$, then Eq. (5.144) (excluding the RIN peak contribution) reduces to

$$\frac{\text{RIN}}{\Delta f} \approx \frac{2h\nu}{P_0}\cdot\frac{S_I/qI+\omega^2\tau_p''^2}{1+\omega^2\tau_p''^2}. \qquad \text{(noise floor)} \tag{5.145}$$

This function is sketched in Fig. 5.20. For a shot noise-limited current source, $S_I/qI = 1$, and the RIN noise floor reduces to a constant at the standard quantum limit, $2h\nu/P_0$, for all frequencies. At frequencies $\ll 1/2\pi\tau_p''$, the noise of the laser comes from the current pumping source. Thus, if a noiseless pumping source is used ($S_I = 0$), the output power noise can be reduced or potentially eliminated below $f \ll 1/2\pi\tau_p''$. However, Eq. (5.144) shows that the current noise contribution is replaced by partition noise as $\eta_0 \to 0$. Hence it is essential to use an optically efficient laser.

Yamamoto and Machida [9] have shown that the current injection noise into a forward-biased PN junction is dominated by the thermal noise of the series resistor such that the double-sided spectral density is $S_I = 2kT/R_S$. By using a large series resistance the pump noise can be reduced below the shot noise limit of qI. In other words, we require that $R_S \gg 2kT/qI$. For a 1 mA drive current at room temperature, this translates into $R_S \gg 50\ \Omega$. By using a constant current source with very large series resistance, the current input noise can be reduced substantially below the shot noise limit. This is the strategy some researchers have employed to

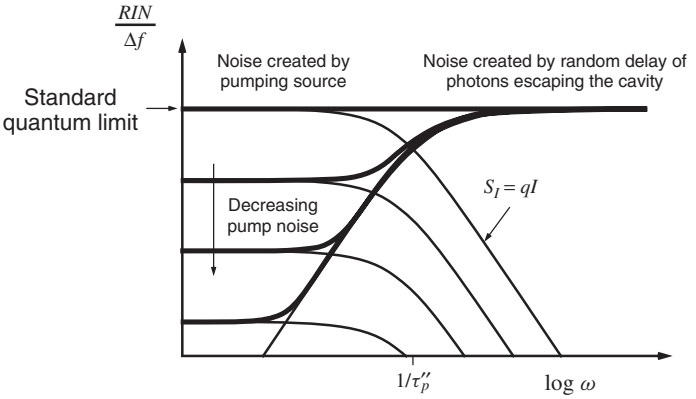


FIGURE 5.20: The two components of the relative intensity noise floor, assuming a high optical efficiency laser. Low frequency noise can be reduced below the standard quantum limit if the pumping source is quiet.

attain noise-free laser operation [9, 10, 12]. To be successful, such lasers must have a very high optical efficiency ($\eta_0 \to 1$) and must be operated well above threshold.

Another caveat we must consider, particularly with noiseless lasers is the transmission and detection process. Imagine an ideally quiet laser that emits a perfectly regular stream of photons. In the process of transmission and detection, some of these photons will be lost, randomly selected out of the uniform stream. These random vacancies in the otherwise perfectly regular photon stream produce *partition noise* in the detected signal. For high losses, the magnitude of this partition noise approaches the shot noise limit. Using the Langevin method detailed in Appendix 13 to include this partition noise (see Problem A13.1), we can derive the detected noise spectral density and corresponding RIN [12]:

$$S_{det}(\omega) = \eta_{det}^2 (q/h\nu)^2 S_{\delta P}(\omega) + (1 - \eta_{det}) q I_{det}, \tag{5.146a}$$

$$(\text{RIN})_{det} = (\text{RIN})_{laser} + (1 - \eta_{det})(\text{RIN})_{shot}, \tag{5.146b}$$

where $I_{det} = \eta_{det}(q/h\nu)P_0$ and $(\text{RIN})_{shot} = 2qI_{det}/I_{det}^2$. Here, η_{det} is the photon collection efficiency of the photodetector including coupling and transmission losses in getting from the laser to the detector (current leakage in the detector is not included in η_{det} because parallel current paths do not necessarily lead to partition noise—the same goes for η_i at the laser end [12]).

In Eq. (5.146a), the coefficient of the first term is the normal transfer function one would expect in converting the power noise to detector current noise. The second term is the standard (double-sided) shot noise term commonly attached to the detector current noise, which more correctly here includes a $(1 - \eta_{det})$ factor. Equation (5.146b) gives the corresponding RIN of the detector current. Both of these equations reveal that to achieve subshot noise performance in an optical communications system, it is essential to keep the losses in going from the laser

to the detector to a minimum. For example, if half of the light is lost before being converted to current, then the best one can hope to do is drop 3 dB below the standard quantum limit, even if the laser is perfectly quiet.

Note that the argument for high η_{det} is no different from the argument demanding a high η_0 in the laser. Essentially, a perfectly regular stream of electrons injected into the laser must have a one-to-one correspondence with electrons generated in the detector to replicate the perfect regularity. Any random division or loss of photons along the way introduces irregularities that show up as shot noise in the receiver current.

RIN in Multimode Lasers Because the derivation leading up to Eq. (5.140) used the single-mode rate equations, it is not surprising that Fig. 5.19 gives a good representation of what is experimentally observed in single-frequency lasers. However, it also is roughly valid for multimode lasers provided all modes are included in the received power. On the other hand, if only one mode is filtered out from a multimode spectrum, it is typically found to contain a much larger noise level, especially at the lower frequencies. This is because of *mode partitioning*. The energy tends to switch back and forth randomly between the various modes observed in the time-averaged spectrum causing large power fluctuations in any one mode. If all modes are included, the net power tends to average out these fluctuations. There are numerous ways in which an optical link can provide the unwanted spectral filtering of a multimode laser's output. For example, in multimode fiber the spatial modes will interfere differently for each frequency in the laser's spectrum, providing a different transmission fraction for each. This is especially accentuated if there is some incidental spatial filtering in the optical link, so that different spatial modes are coupled differently.

It is also somewhat surprising how large the mode suppression ratio (MSR) for the unwanted modes must be before the laser behaves like a single-frequency laser. Experiments have verified that significant mode partitioning can occur even for MSRs $\sim$30 dB, although this is roughly the level at which mode partitioning tends to disappear. Such single-frequency lasers also tend to be very sensitive to spurious feedback from external reflections in the optical path. This sensitivity will be the subject of later discussions in this chapter.

5.5.5 Frequency Noise

For frequency modulation applications, it is useful to determine the frequency jitter or noise. By an extension of Eq. (5.75), we can write

$$\frac{d\phi}{dt} = 2\pi\,\Delta\nu(t) = \frac{\alpha}{2}\Gamma v_g a[dN(t)] + F_\phi(t). \tag{5.147}$$

The first equality reminds us that the frequency deviation can be considered a rate equation for the field phase. The latter equation introduces a Langevin noise source for this rate equation that can be associated with the phase noise of the

laser. Converting to the frequency domain, we obtain

$$\nu_1(\omega) = \frac{\alpha}{4\pi}\Gamma v_g a N_1(\omega) + \frac{1}{2\pi}F_\phi(\omega). \tag{5.148}$$

Appendix 13 shows that the correlation strengths for the phase noise source reduce to the following:

$$\langle F_\phi F_\phi \rangle = \frac{\Gamma R'_{sp}}{2N_p}, \quad \langle F_\phi F_P \rangle = \langle F_\phi F_N \rangle = 0. \tag{5.149}$$

Note that the phase noise source is uncorrelated with photon and carrier density noise sources. To determine the frequency noise spectral density, as before we multiply both sides of Eq. (5.148) by $\nu_1(\omega')^*$, take the time average, and integrate over ω', to obtain

$$S_\nu(\omega) = \left(\frac{\alpha}{4\pi}\Gamma v_g a\right)^2 S_N(\omega) + \frac{1}{(2\pi)^2}\langle F_\phi F_\phi \rangle. \tag{5.150}$$

The cross-term does not appear because there is no correlation between the carrier and phase noise. Thus, the frequency noise has two contributions: (1) the carrier noise, which induces refractive index changes causing the lasing frequency to fluctuate, and (2) the inherent phase noise of the laser, originating from photons, which are spontaneously emitted into the mode. For semiconductor lasers, carrier noise dominates by typically more than an order of magnitude.

We can use Eqs. (5.136) through (5.138) to evaluate the carrier noise spectral density in Eq. (5.124). Appendix 13 carries out this task revealing that the carrier density noise spectral density can be well approximated by

$$S_N(\omega) = \frac{8\pi}{(\Gamma v_g a)^2}\left[\frac{\Gamma R'_{sp}}{4\pi N_p}\right] \cdot |H(\omega)|^2. \tag{5.151}$$

With Eq. (5.151) for the carrier noise and Eq. (5.140) for the phase noise, the frequency noise spectral density (double-sided) becomes

$$S_\nu(\omega) = \frac{1}{2\pi}(\Delta\nu)_{ST} \cdot (1 + \alpha^2|H(\omega)|^2), \tag{5.152}$$

where again

$$(\Delta\nu)_{ST} = \frac{\Gamma R'_{sp}}{4\pi N_p}.$$

The measured FM noise can be modeled using four fitting parameters: α, γ, ω_R, and $(\Delta\nu)_{ST}$ [11]. A typical FM noise spectrum is sketched in Fig. 5.21. Beyond the relaxation resonance frequency, the carrier noise becomes negligible, reducing the FM noise to the white noise background provided by spontaneous emission phase noise. In addition thermal contributions from the noise on the pumping source can also contribute to the FM noise at low frequencies, as discussed in Section 5.3.3.

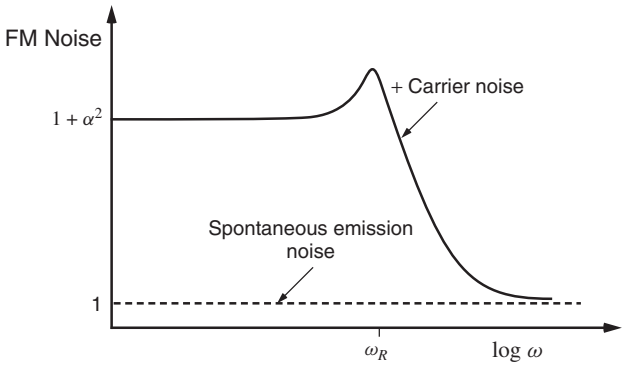


FIGURE 5.21: Frequency noise spectrum illustrating contributions from both carrier and spontaneous emission noise.

5.5.6 Linewidth

To relate the FM noise to the frequency spectrum of the laser, it is useful to introduce the *coherence time* of the laser light. Consider a measurement that mixes the emitted electric field, $\mathscr{E}(t)$, with a time-delayed version of itself, $\mathscr{E}(t-\tau)$. As long as the phases of the two fields are well correlated, the fields will add coherently. We can write the cross-term of the coherent addition, or the *autocorrelation* function of the field as

$$\langle \mathscr{E}(t)\mathscr{E}(t-\tau)^* \rangle \propto e^{j\omega\tau} e^{-|\tau|/\tau_{coh}}. \tag{5.153}$$

The first factor gives the expected interference fringes created by the coherent mixing. However, because the laser is not emitting a pure single frequency, with increasing time delay, τ, the phases of the two fields become less and less correlated and the interference fringes gradually disappear to the point where the two fields add incoherently. The envelope of the fringe pattern given by the second factor characterizes this coherence decay. As the functionality suggests, the decay can often be described by an exponential with a decay constant defined as the coherence time of the laser, τ_{coh}. However, more generally, τ_{coh} may be a function of the time delay itself, $\tau_{coh}(\tau)$, in which case the autocorrelation does not yield a simple decaying exponential.

Figure 5.22a suggests one method of directly measuring the autocorrelation function, allowing us to experimentally extract $\tau_{coh}(\tau)$. Using the measured fringe pattern in Fig. 5.22a, we can write the envelope of the autocorrelation function as

$$e^{-|\tau|/\tau_{coh}} = \frac{P_1 + P_2}{2\sqrt{P_1 P_2}} \frac{P_{\max} - P_{\min}}{P_{\max} + P_{\min}}, \tag{5.154}$$

where $P_1 \propto \langle |\mathscr{E}_1(t)|^2 \rangle$ and $P_2 \propto \langle |\mathscr{E}_2(t-\tau)|^2 \rangle$ are the individual powers in the two legs of the interferometer, and $P_{\max} \propto \langle |\mathscr{E}_1(t) + \mathscr{E}_2(t-\tau)|^2 \rangle$ and $P_{\min} \propto \langle |\mathscr{E}_1(t) + \mathscr{E}_2(t-\tau)|^2 \rangle$ are the maximum and minimum powers of the interference fringes

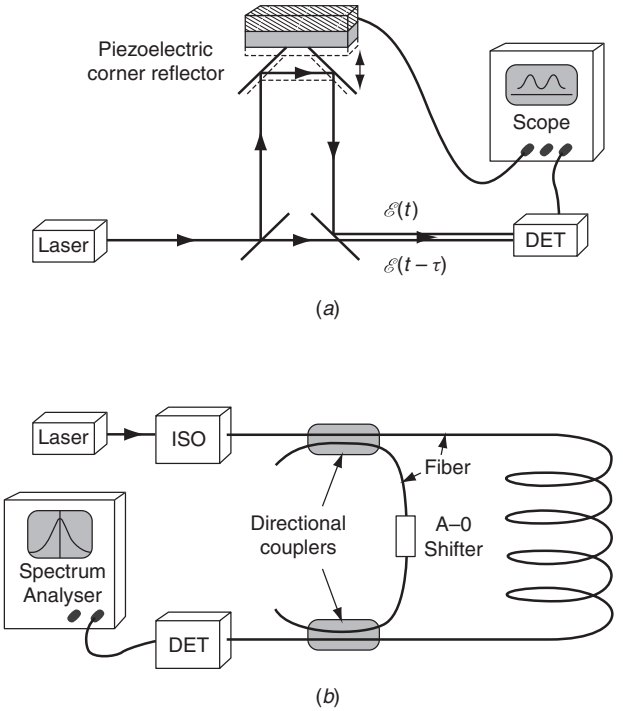


FIGURE 5.22: Schematics of two techniques to measure linewidth. (a) Coherence length measurement in which one branch of a Mach-Zehnder interferometer is increased in length until fringe visibility is reduced. A piezoelectrically driven mirror provides a small length variation to display interference fringes. (b) Self-heterodyne technique in which a long length of fiber ($\gg$ coherence length) is used to mix the laser emission with an incoherent version of itself. The difference frequency signal replicates the combined lineshape of both signals at low frequencies and is easily observed using an RF spectrum analyzer [13].

measured by the detector. The second ratio is often referred to as the *fringe visibility*. The first ratio normalizes the fringe visibility when $P_1 \neq P_2$. Measuring the normalized fringe visibility as a function of the time delay between the two arms yields the envelope of the autocorrelation function.

Once the autocorrelation is known, it can be related to the frequency spectrum of the mode, $P_0(\omega)$, through a Fourier transform (see Appendix 13):

$$P_0(\omega) \propto \mathscr{F}[\langle \mathscr{E}(t)\mathscr{E}(t-\tau)^* \rangle]. \tag{5.155}$$

For example, if τ_{coh} is a constant (i.e., not a function of τ) then the autocorrelation is a decaying exponential, and the corresponding lasing spectrum is a Lorentzian.

Experimentally, the lasing spectrum can be measured conveniently using the experimental setup illustrated in Fig. 5.22b. The laser is coupled through an optical isolator into a fiber (the isolator is necessary to remove unwanted

feedback, which can affect the laser linewidth (see Section 5.7)). One leg of fiber is fed through an acousto-optic modulator, which shifts the light frequency by typically 40 MHz. The other leg of fiber is much longer than the coherence length ($L_{coh} \equiv v\tau_{coh}$) of the laser. When the two incoherent fields are recombined in the detector, they generate a difference-frequency signal current at 40 MHz that contains the combined FM *field* noise of both light sources. The square of this current is therefore equivalent to the combined optical power spectrum. In other words, this technique effectively converts the combined optical power spectrum to an *electrical power* spectrum centered at 40 MHz, which can easily be measured using a radio-frequency spectrum analyzer. However, because it contains the noise of both signals, the electrical power 3 dB-down full-width of the spectrum is *twice* as wide as the original laser linewidth (the factor of 2 is specific to the combination of two *Lorentzian* lineshapes [13]).

Theoretically, the lasing spectrum is found by determining $\tau_{coh}(\tau)$ and applying Eq. (5.155). We can relate the coherence time to the FM noise and the measurement time delay as follows [14]:

$$\frac{1}{\tau_{coh}} = \pi\tau \int_{-\infty}^{+\infty} S_v(\omega) \frac{\sin^2(\omega\tau/2)}{(\omega\tau/2)^2}\, d\omega. \tag{5.156}$$

The coherence time is related to the *integrated* frequency noise of the laser times a $\text{sinc}^2(\omega\tau/2)$ function.[8] Because the first zero of $\text{sinc}^2(\omega\tau/2)$ occurs at $\omega = 2\pi/\tau$, the entire function scales toward zero with increasing time delay. This effect is illustrated in Fig. 5.23a. As the peaks and nulls of $\text{sinc}^2(\omega\tau/2)$ sweep by the resonance peak of the noise with increasing time delay, they produce periodic undulations in the overlap (because the resonance peak is either included or excluded from the overlap). The effect of this on $1/\tau_{coh}$ is shown in Fig. 5.23(b). The decaying oscillations that appear in the otherwise linear τ/τ_{coh} are the direct result of the resonance peak in the noise. The oscillation frequency is equal to the peak frequency of the resonance ($\sim \omega_R$), whereas the magnitude of the oscillations is governed by how sharp and strong the resonance is ($\sim 1/\gamma$).

Eventually, τ/τ_{coh} relaxes back to a linear relationship. In fact, for $\tau g 2\pi/\omega_R$, the $\text{sinc}^2(\omega\tau/2)$ function is peaked in the low-frequency portion of the FM noise spectrum, and we can approximate $S_v(\omega) \approx S_v(0)$ inside the integral. This leads to the simple result:

$$\frac{1}{\tau_{coh}} \approx \pi\tau S_v(0) \int_{-\infty}^{+\infty} \frac{\sin^2(\omega\tau/2)}{(\omega\tau/2)^2}\, d\omega = 2\pi^2 S_v(0). \qquad (\tau \gg 2\pi/\omega_R) \tag{5.157}$$

[8]This result comes about as follows. Assume the field is $\propto e^{j\omega t} e^{j\phi(t)}$ where $\phi(t)$ is a fluctuating phase created by frequency noise. The autocorrelation is then $\propto \langle e^{j\,\Delta\phi(\tau)}\rangle = \exp[-\langle\Delta\phi(\tau)^2\rangle/2]$, and we conclude that $\tau/\tau_{coh} \equiv \langle\Delta\phi(\tau)^2\rangle/2$. Using the Fourier transform pair $\Delta\phi(\tau) = \phi(\tau) - \phi(0) \Leftrightarrow \phi(\omega)(e^{j\omega\tau} - 1)$ and Eq. (5.113), we can derive another Fourier transform pair $\langle\Delta\phi(\tau)^2\rangle \Leftrightarrow 2S_\phi(\omega)(1 - \cos\omega\tau)$ where $S_\phi(\omega)$ is the phase noise spectral density. Because phase and frequency are related by a time derivative we can set $S_\phi(\omega) \to (2\pi)^2 S_v(\omega)/\omega^2$. This leads to $\langle\Delta\phi(\tau)^2\rangle = 2\tau/\tau_{coh} \Leftrightarrow (4\pi)^2 S_v(\omega)\sin^2(\omega\tau/2)/\omega^2$. Expressing the Fourier transform explicitly using Eq. (5.110) and rearranging we obtain Eq. (5.156).

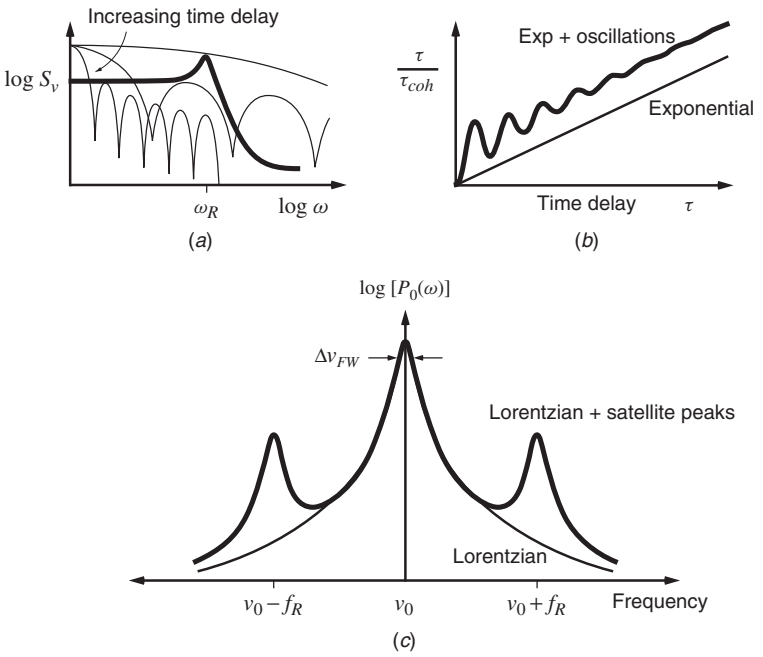


FIGURE 5.23: (a) Overlap between FM noise spectrum and $\text{sinc}^2(\omega\tau/2)$ function for increasing time delays. (b) Resultant functionality of τ/τ_{coh} vs. time delay. (c) Lasing spectrum with and without satellite peaks.

The integral is equal to $2\pi/\tau$, canceling out the dependence on time delay. Thus, for time delays much greater than $1/f_R, \tau_{coh}$ is a constant, τ/τ_{coh} increases linearly, and the autocorrelation reduces to a decaying exponential. For shorter time delays, the decaying exponential contains some ripples.

From a physical point of view, it is the relaxation oscillations occurring in response to random absorption/emission events that enhance the FM noise near the resonance frequency. And because these oscillations are created by noise events that occur earlier in time, we could argue that some component of the noise at any given time mirrors the noise that existed $\sim 1/f_R$ seconds ago. Thus, we might expect a higher correlation or coherence between fields that are delayed by time intervals of $\sim 1/f_R$. This is exactly what Fig. 5.23b reveals.

Taking the Fourier transform of $e^{-|\tau|/\tau_{coh}}$ produces the power spectrum of the laser as shown in Fig. 5.23c. Without ripples in τ/τ_{coh}, the spectrum is a Lorentzian. With ripples, the Lorentzian acquires a series of satellite peaks spaced by the resonance frequency. The magnitude of the satellite peaks depends on the damping factor of the laser (less damping $\Rightarrow$ stronger peaks). However, the FWHM linewidth of the spectrum is essentially the same with or without these satellite peaks.

Using the long time delay expression for τ_{coh} Eq. (5.148) to isolate the exponential component of the autocorrelation and then Fourier transforming, we obtain

the dominant Lorentzian lineshape. In general, if the exponential decay constant is τ_0, then the transformed Lorentzian FWHM is $\Delta\omega = 2/\tau_0$. In the simpler treatment of linewidth leading to Eq. (5.123), τ_0 corresponds to the field decay or $2\tau_p$, yielding $\Delta\omega = 1/\tau_p$. Here, τ_0 corresponds to the coherence time such that

$$\Delta\nu_{FW} = \frac{1}{\pi\,\tau_{coh}} = 2\pi S_\nu(0). \tag{5.158}$$

Inserting Eq. (5.152) with $|H(0)|^2 = 1$, we obtain

$$\Delta\nu_{FW} = (\Delta\nu)_{ST}(1+\alpha^2) = \frac{\Gamma R'_{sp}}{4\pi N_p}(1+\alpha^2). \tag{5.159}$$

Thus, the modified Schawlow–Townes expression for linewidth is *enhanced* by $1+\alpha^2$ in semiconductor lasers (this is where α gets its name as the linewidth enhancement factor) [5, 14]. The 1 represents the spontaneous emission noise contribution, and the α^2 represents the carrier noise contribution.

The above-threshold linewidth can be rewritten in terms of external parameters by setting $R'_{sp} = \Gamma v_g g_{th} n_{sp}/V, N_p = P_0/h\nu V_p v_g \alpha_m F$, and $\alpha_m F = \eta_0 \Gamma g_{th}$:

$$\begin{aligned}\Delta\nu_{FW} &= \frac{(\Gamma v_g g_{th})^2 \eta_0}{4\pi P_0} n_{sp} h\nu (1+\alpha^2) \\ &= 38\ \text{MHz} \times \left(\frac{4.2}{\bar{n}_g}\right)^2 \left(\frac{\Gamma g_{th}}{50\ \text{cm}^{-1}}\right)^2 \left(\frac{\eta_0}{0.4}\right)\left(\frac{n_{sp}}{1.5}\right)\left(\frac{h\nu}{1.5\ \text{eV}}\right) \\ &\quad \left(\frac{1\ \text{mW}}{P_0}\right)\left(\frac{1+\alpha^2}{26}\right).\end{aligned} \tag{5.160}$$

Typical numbers have been used to evaluate the linewidth (the value for η_0 refers to the single-facet efficiency, and hence P_0 refers to single-facet power). Generally speaking, powers in the milliwatts range in a typical semiconductor laser produce linewidths in the tens-of-megahertz range. Thus, to achieve sub-megahertz linewidths it is necessary to increase the power well above 10 mW. Another approach to reducing the linewidth involves the use of external cavities, which we explore in Section 5.7.

Example 5.6 For the VCSEL from Examples 5.1, 5.2 and 5.3, we are interested in calculating the linewidth. Assume $n_{sp} = 1.25$.

Problem: What unmodulated linewidth would be expected at the three biases, assuming no external feedback?

Solution: To compute the linewidth, we need to use the expression for the modified Schawlow–Townes linewidth, given by Eq. (5.159),

$$\Delta\nu = \frac{\Gamma R'_{sp}}{4\pi N_p}\,(1+\alpha^2).$$

Plugging in the expression for spontaneous emission into a mode, Eq. (4.67), we have

$$\Delta\nu = \Gamma(1+\alpha^2)\frac{\Gamma v_g g_{th} n_{sp}}{4\pi V N_p} = (1+\alpha^2)\frac{v_g^2 \frac{\eta_i}{\eta_d}\alpha_m^2 n_{sp} q}{4\pi \eta_d (I - I_{th})} = 49.49 \text{ MHz}$$

for $I - I_{th} = 0.5$ mA. For $I - I_{th} = 1.0$ mA and $I - I_{th} = 2.0$ mA, the linewidths are 24.475 MHz and 12.37 MHz, respectively.

5.6 CARRIER TRANSPORT EFFECTS

In the sections so far we have neglected any transport time for carriers to reach the active region. For cases where the intrinsic region of the *pin* diode and the active region are one and the same, such as for simple bulk DH structures, this assumption is generally good. However, for separate-confinement heterostructures (SCHs) used with quantum-well active regions, it has been found that transport effects must be considered [15]. This structure was introduced in Fig. 1.8. Figure 5.23 gives a slightly more detailed SCH schematic for the present discussion.

For a number of years during the 1980s there was a dilemma as to why quantum-well lasers were not providing the modulation bandwidths predicted by the calculations given earlier. From Eq. (5.51) and earlier versions, it was clear that the resonant frequency was directly proportional to the differential gain, which is much higher with quantum-well active regions. Thus, it was surprising when experiments showed quantum-well laser bandwidths about the same as for bulk DH structures. Once it was realized that carrier injection delays due to transport effects were significant, laser designs were modified, and bandwidths increased dramatically. In this section we shall review the relevant theory.

With reference to Fig. 5.24 we construct a set of three rate equations for the SCH design. The change is that the carrier equation has been replaced by two new equations. The first is for the carrier density in the barrier regions, N_B, and the second is for the carrier density in the active region, N. Multiple quantum-well active regions can also be treated by lumping all the barrier regions together. The new rate equations are

$$\frac{dN_B}{dt} = \frac{\Gamma_q \eta_i I}{qV} - \frac{n_B}{\tau_s} + \frac{\Gamma_q N}{\tau_e}, \tag{5.161}$$

$$\frac{dN}{dt} = \frac{N_B}{\Gamma_q \tau_s} - N\left[\frac{1}{\tau} + \frac{1}{\tau_e}\right] - v_g g N_p, \tag{5.162}$$

$$\frac{dN_p}{dt} = \left[\Gamma v_g g - \frac{1}{\tau_p}\right] N_p + \Gamma R'_{sp}, \tag{5.163}$$

where $\Gamma_q = V/V_{SCH}$ is the fraction of the SCH region filled by the quantum-well active region, N_B/τ_s is the loss rate of carriers from the SCH region to the quantum-well active region, N/τ_e is the loss rate of carriers from the quantum wells

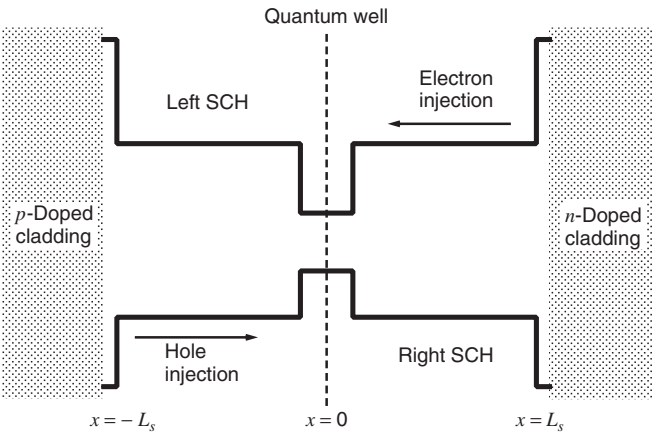


FIGURE 5.24: Schematic diagram of a single quantum-well laser with a separate-confinement heterostructure (SCH) used in the carrier transport model.

to the SCH region, and the other symbols are as defined before. In this simplified treatment, we ignore the fact that the holes come from one side of the junction and the electrons from the other. Fortunately, other more in-depth treatments, which include holes and electrons separately yield essentially the same results as this simplified analysis.

Limiting our attention to small-signal perturbations, we can linearize the rate equations as we did in Section 5.3 to obtain the differential rate equations. In matrix form, these become

$$\frac{d}{dt}\begin{bmatrix} dN_B \\ dN \\ dN_p \end{bmatrix} = \begin{bmatrix} -\gamma_{BB} & \gamma_{BN} & 0 \\ \gamma_{NB} & -\gamma_{tNN} & -\gamma_{NP} \\ 0 & \gamma_{PN} & -\gamma_{PP} \end{bmatrix} \begin{bmatrix} dN_B \\ dN \\ dN_p \end{bmatrix} + \frac{\Gamma_q \eta_i}{qV} \begin{bmatrix} dI \\ 0 \\ 0 \end{bmatrix}. \tag{5.164}$$

The rate coefficients of the upper left 2×2 submatrix are given by

$$\begin{aligned} \gamma_{BB} &= 1/\tau_s, \qquad \gamma_{BN} = \Gamma_q/\tau_e, \\ \gamma_{NB} &= 1/\Gamma_q\tau_s, \qquad \gamma_{tNN} = \Gamma_{NN} + 1/\tau_e. \end{aligned} \tag{5.165}$$

The other rate coefficients are as defined before in Eq. (5.34). Converting to the frequency domain and applying Cramer's rule, we obtain

$$N_{p1}(\omega) = \frac{\Gamma_q \eta_i I_1(\omega)}{qV} \cdot \frac{1}{\Delta_t} \begin{vmatrix} \gamma_{BB} + j\omega & -\gamma_{BN} & 1 \\ -\gamma_{NB} & \gamma_{tNN} + j\omega & 0 \\ 0 & -\gamma_{PN} & 0 \end{vmatrix}, \tag{5.166}$$

where

$$\Delta_t \equiv \begin{vmatrix} \gamma_{BB} + j\omega & -\gamma_{BN} & 0 \\ -\gamma_{NB} & \gamma_{tNN} + j\omega & \gamma_{NP} \\ 0 & -\gamma_{PN} & \gamma_{PP} + j\omega \end{vmatrix}. \tag{5.167}$$

The only nonzero portions of Δ_t are (1) the product of the upper-left element with the determinant of the lower right 2×2 submatrix and (2) the product $\gamma_{BN}\gamma_{NB}(\gamma_{PP} + j\omega)$. Pulling out the factor $(\gamma_{BB} + j\omega)$ in the denominator, setting $\gamma_{BN}\gamma_{NB}/(\gamma_{BB} + j\omega) = \gamma_{BN}\gamma_{NB}/\gamma_{BB} - j\omega(\chi - 1)$, and recognizing that $\gamma_{NN} = \gamma_{tNN} - \gamma_{BN}\gamma_{NB}/\gamma_{BB}$, we obtain

$$N_{p1}(\omega) = \frac{\Gamma_q \eta_i I_1(\omega)}{qV} \cdot \frac{\gamma_{NB}\gamma_{PN}}{\gamma_{BB} + j\omega} \cdot \frac{1/\chi}{\omega_{tR}^2 - \omega^2 + j\omega\gamma_t}, \tag{5.168}$$

where

$$\omega_{tR}^2 = (\gamma_{NP}\gamma_{PN} + \gamma_{NN}\gamma_{PP})/\chi, \tag{5.169}$$

$$\gamma_t = \gamma_{NN}/\chi + \gamma_{PP}, \tag{5.170}$$

$$\chi \equiv 1 + \frac{\gamma_{BN}\gamma_{NB}/\gamma_{BB}}{\gamma_{BB} + j\omega}. \tag{5.171}$$

The new term χ is referred to as the *transport* factor. Using the definitions of the rate coefficients (Eqs. (5.35) and (5.156)), and additionally setting $\gamma_{PN} \approx \Gamma v_g a N_p \approx \Gamma_\chi \omega_{tR}^2 \tau_p$ and $\tau_p = \eta_0 h\nu N_{p1} V_p / P_1$, we obtain the small-signal output power modulation response:

$$\frac{P_1(\omega)}{I_1(\omega)} = \eta_i \eta_0 \frac{h\nu}{q} \cdot \frac{1}{1 + j\omega\tau_s} \cdot \frac{\omega_{tR}^2}{\omega_{tR}^2 - \omega^2 + j\omega\gamma_t}, \tag{5.172}$$

with

$$\omega_{tR}^2 = \omega_R^2/\chi \approx v_g(a/\chi)N_p/\tau_p, \tag{5.173}$$

$$\gamma_t = v_g(a/\chi)N_p\left[1 + \frac{\Gamma a_p}{a/\chi}\right] + \frac{1}{\tau_{\Delta N}\chi} + \frac{\Gamma R'_{sp}}{N_p}, \tag{5.174}$$

$$\chi = 1 + \frac{\tau_s/\tau_e}{(1 + j\omega\tau_s)} \approx 1 + \frac{\tau_s}{\tau_e}. \tag{5.175}$$

The approximate expression for the relaxation resonance frequency ignores the last two terms in Eq. (5.49), and makes use of the fact that in most lasers, $a/\tau_p \gg \Gamma a_p/\tau_{\Delta N}$. The approximate expression for the transport factor neglects the frequency roll-off of the second term. For typical SCH-SQW lasers at room temperature, $\tau_e \sim 100 - 500$ ps, $\tau_s \sim 20 - 100$ ps, and the transport factor $\chi \sim 1.2$.

Using the approximate expression for ω_{tR}^2, the damping factor can be rewritten in terms of the K-factor:

$$\gamma_t = K_t f_{tR}^2 + \gamma_{t0}, \tag{5.176}$$

where

$$K_t = 4\pi^2 \tau_p \left[1 + \frac{\Gamma a_p}{a/\chi}\right] \quad \text{and} \quad \gamma_{t0} = \frac{1}{\tau_{\Delta N}\chi} + \frac{\Gamma R'_{sp}}{N_p}. \tag{5.177}$$

In examining Eqs. (5.173) and (5.174), we conclude that transport across the SCH region *effectively* reduces the differential gain from a to a/χ, and *effectively* increases the differential carrier lifetime from $\tau_{\Delta N}$ to $\tau_{\Delta N}\chi$. The gain compression factor remains unmodified in this model. The reduced differential gain decreases the relaxation resonance frequency and increases the K-factor. The increase in differential carrier lifetime reduces the damping factor offset, γ_{t0}. In addition (and perhaps most significantly), there is a new prefactor that provides a low-pass filtering effect with a cutoff at $\omega = 1/\tau_s$. This prefactor provides an insidious parasitic-like roll-off that is indistinguishable from an RC roll-off.

Figure 5.25 gives examples of the effects of transport in SCH structures. Note that with the narrower SCH region, the low-pass filter effect is eliminated allowing for a much larger modulation bandwidth.

5.7 FEEDBACK EFFECTS AND INJECTION LOCKING

Signals returning to a laser cavity, or being injected into the cavity from the outside can have important effects on lasing. Although optical feedback effects in lasers are unavoidable, and generally we try to minimize them due to their negative impact on laser's linewidth, deliberate, controlled optical feedback or injection locking can improve the laser's performance through noise suppression, reduced nonlinear distortion and modulation bandwidth enhancement [22]. In this section, we will analyze both effects, starting with optical feedback, and modifying the results from this case to correspond to the injection locking case as well.

5.7.1 Optical Feedback Effects—Static Characteristics

In typical applications of diode lasers some light is unintentionally reflected back into the laser cavity. For example, the reflection from the front surface of a fiber is about 4%, and perhaps one-tenth of this may be coupled back into the laser cavity in a pigtailing arrangement. Surprisingly, even such small amounts of feedback can have dramatic effects on the laser's linewidth and noise properties. This is especially true for single-frequency lasers. In fact, 0.4% (−24 dB) is considered to be a large feedback. Experiments have shown that even −60 dB of feedback can be unacceptable with single-frequency lasers.

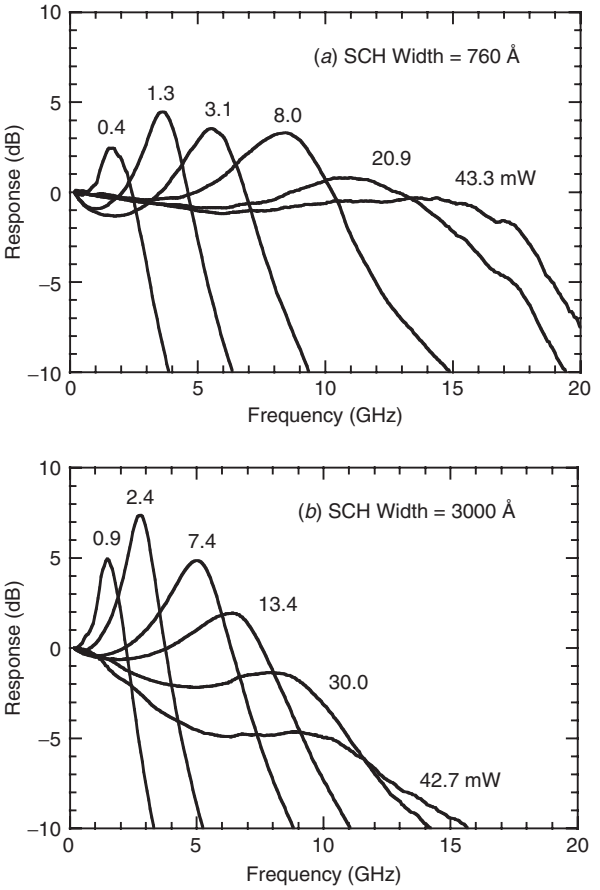


FIGURE 5.25: Modulation response for (a) a narrow SCH and (b) a wide SCH-SQW laser (the lasers are otherwise identical). The modulation response for the wide SCH laser, at comparable power levels, shows the detrimental effects of the low frequency roll-off due to carrier transport. After [15]. © 1992 IEEE.

To analyze the effects of feedback, we consider the three-mirror cavity results of Chapter 3. Figure 3.9 and Eq. (3.35) are our starting point. They describe an effective mirror concept that folds the effects of the external cavity back into the active laser diode section. From this starting point, we explore the effects of feedback on the laser's output power and spectrum. After the static characteristics are quantified, we consider the dynamic effects on the laser's linewidth and noise spectrum.

Equation (3.35) describes the vector addition of the primary reflection at the laser facet, r_2, with the feedback term. This is shown schematically in Fig. 5.26. For relatively weak feedback, the external cavity resonance can be neglected (i.e., we can ignore the denominator in Eq. (3.35), so that the in-phase (*ip*) and quadrature

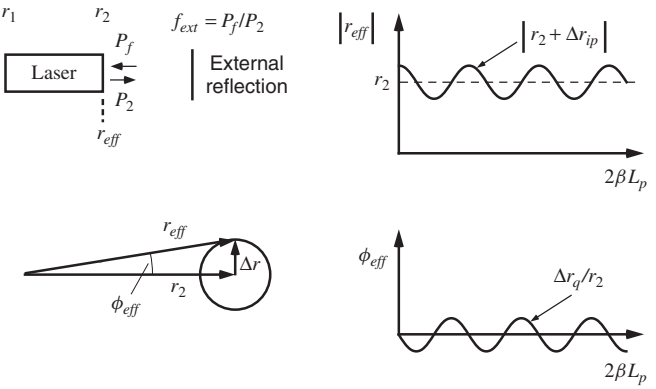


FIGURE 5.26: Polar plot (left) of effective reflection from output side of laser; Δr represents feedback that can change both the amplitude and phase of r_{eff}. Plots of the magnitude and phase of r_{eff} illustrate the quadrature relationship.

(q) components of the feedback term, $r_{eff} - r_2 = \Delta r$, become

$$\begin{aligned} \Delta r_{ip} &= t_2^2\sqrt{f_{ext}}\cos(2\beta L_p), \\ \Delta r_q &= -t_2^2\sqrt{f_{ext}}\sin(2\beta L_p), \end{aligned} \tag{5.178}$$

where $-2\beta L_p$ is the round-trip phase of the external cavity. Also, the external feedback level, f_{ext} represents the fraction of emitted *power* coupled back to the laser. It replaces r_3 in Eq. (3.35) to account for additional coupling and propagation losses encountered through one round-trip path of the external cavity.

The in-phase component of the feedback, Δr_{ip}, affects the laser by changing the magnitude of the effective reflectivity at mirror #2. This modifies the photon lifetime of the cavity, which in turn modifies the threshold gain, threshold carrier density, and threshold current of the laser. The perturbed photon lifetime is given by

$$\frac{1}{\tau_p'} = \frac{1}{\tau_p} + v_g\Delta\alpha_m = \frac{1}{\tau_p} + \frac{v_g}{L}\left[\frac{-\Delta R}{R}\right], \tag{5.179}$$

where $R = r_1 r_2$ and $\Delta R = r_1\Delta r_{ip}$. Expanding these terms and defining the round-trip time of the laser cavity as $\tau_L = 2L/v_g$, we can express the change in mirror loss as

$$v_g\Delta\alpha_m = -2\frac{t_2^2}{r_2}\frac{\sqrt{f_{ext}}}{\tau_L}\cos(2\beta L_p) = -2\kappa_f\cos(2\beta L_p). \tag{5.180}$$

The feedback rate $\kappa_f = (t_2^2\sqrt{f_{ext}}/r_2)/\tau_L$ introduced here represents the normalized reflected field injection rate per round-trip time of the cavity (i.e., per bounce).

The change in mirror loss gives rise to a shift in the threshold modal gain, or $\Delta(\Gamma g_{th}) = \Delta\alpha_m$. For small changes, we can use the differential gain $a = \Delta g/\Delta N$ to determine the corresponding shift in threshold carrier density, and we can write

$$\Delta g_{th} = -\frac{2\kappa_f}{\Gamma v_g}\cos(2\beta L_p), \tag{5.181}$$

$$\Delta N_{th} = -\frac{2\kappa_f}{\Gamma v_g a}\cos(2\beta L_p), \tag{5.182}$$

The perturbed threshold current can also be described as follows:

$$I'_{th} = I_{th} - |\Delta I_{th}|\cos(2\beta L_p), \tag{5.183}$$

where ΔI_{th} can be found by defining a differential gain per unit current density, a_J, similar to Eq. (5.182). When the external reflection is in phase with the reflection at mirror #2 (i.e., when $2\beta L_p = 2\pi m$), the threshold level decreases due to an increase in r_{eff}. When the external reflection is out of phase (i.e., when $2\beta L_p = 2\pi m + \pi$), the threshold level increases due to a decrease in r_{eff}.

The change in mirror loss also affects the differential efficiency and output power. Expanding Eqs. (3.31) through (3.33) to first-order in $\Delta r_{ip}/r_2$, we can obtain new perturbed values for the total differential efficiency, η'_d, the fraction of light emitted through each mirror, $F'_{1,2}$, and the corresponding power out, $P'_{01,02}$. Using Eq. (5.183) and assuming both laser facets are lossless (i.e., $t^2 = 1 - r^2$), the perturbed power out of mirrors #1 and #2 becomes

$$P'_{01,02} = P_{01,02}[1 - \mu_{1,2}\cos(2\beta L_p)], \tag{5.184}$$

where

$$\mu_{1,2} = \left[2\kappa_f\tau_p\left(\frac{\eta_i}{\eta_d} - 1\right)\right] \pm \left[\kappa_f\tau_L(1 - F_{1,2})\left(\frac{1 + r_2^2}{t_2^2}\right)\right] - \left[\frac{|\Delta I_{th}|}{I - I_{th}}\right].$$

The upper sign is for μ_2, whereas the lower sign is for μ_1. As a reminder, $|\Delta I_{th}|$ represents *half* of the peak-to-peak change in threshold current.

Equation (5.184) shows that if we vary the optical path length between the laser and the reflection causing the feedback, the output power will oscillate sinusoidally as the external reflection goes in and out of phase with the reflection at mirror #2. The oscillation is the combined result of (in order of appearance in $\mu_{1,2}$): (1) the change in overall differential efficiency, (2) the change in the fraction of power emitted out of either facet, and (3) the shift in threshold current. The first two effects reinforce each other (upper sign) for light emitted out mirror #2, but work against each other (lower sign) for light emitted out mirror #1. Also, the first effect is enhanced for small η_d, whereas the second effect is enhanced when the fraction of light emitted from the measured facet is small. The third effect is enhanced near threshold. At certain operating points all three effects can cancel each other out.

However, for bias currents well above threshold, the third effect is negligible. By recording the peak-to-peak power variation as the feedback phase is scanned, we can directly measure $\mu_{1,2}$. This provides a method of estimating both κ_f and the feedback level f_{ext}. However, the technique is limited to relatively large feedback levels because the change in output power for very weak feedback is hard to detect.

While Δr_{ip} affects the threshold level and power out, the quadrature component of the feedback, Δr_q modifies the phase angle of the net reflection at mirror #2 from zero to $\phi_{eff} = \Delta r_q / r_2$, affecting the cavity's resonant wavelength. With no feedback, the resonance condition is met when $-2\beta_0 L = 2\pi m$. With feedback, the accumulated round-trip phase must now satisfy: $-2\beta L + \phi_{eff} = 2\pi m \equiv -2\beta_0 L$. In other words, β must adjust itself such that $2(\beta - \beta_0)L = \phi_{eff}$. In a dispersive medium, $\Delta\beta = \Delta\omega / v_g$ allowing us to write the frequency shift due to the external feedback as

$$\Delta\omega_\phi = \frac{\phi_{eff}}{\tau_L} = -\kappa_f \sin(2\beta L_p). \tag{5.185}$$

If we specify the round-trip phase change as $\Delta\phi_{r-t} = -\Delta\omega\tau_L$, we find that this equation is equivalent to $-\Delta\phi_{r-t}/\tau_L = \phi_{eff}/\tau_L$, which reduces to $\Delta\phi_{r-t} + \phi_{eff} = 0$. In other words, the round-trip phase change compensates for the altered phase of mirror #2.

An additional round-trip phase change is created by the shift in threshold carrier density via the linewidth enhancement factor, α. This leads to an additional shift in the lasing frequency. Placing the threshold carrier density shift Eq. (5.173) into Eq. (5.75), the resulting carrier-induced frequency shift is given by

$$\Delta\omega_N = -\alpha\kappa_f \cos(2\beta L_p). \tag{5.186}$$

Adding Eqs. (5.185) and (5.186), and using the trigonometric identity

$$a \sin\phi + b \cos\phi = \sqrt{a^2 + b^2} \sin[\phi + \tan^{-1}(b/a)],$$

the total frequency shift becomes

$$\Delta\omega = -\kappa_f \sqrt{1 + \alpha^2} \sin(2\beta L_p + \phi_\alpha) \equiv \frac{\phi'_{eff}}{\tau_L}, \tag{5.187}$$

where $\phi_\alpha = \tan^{-1}\alpha$ (which varies from 0 to $\pi/2$ as α increases). By defining ϕ'_{eff} as the effective phase of mirror #2 factoring in the carrier-induced round-trip phase shift, Eq. (5.187) becomes equivalent to $-\Delta\phi_{r-t}/\tau_L = \phi'_{eff}/\tau_L$, which reduces to $\Delta\phi_{r-t} + \phi'_{eff} = 0$. Note that α enhances ϕ'_{eff} and pulls the quadrature $\sin x$ dependence toward a $\cos x$ dependence more in sync with the threshold gain variation (5.172).

Our goal is to determine the mode frequency shift, $\Delta\omega$. We start by writing the feedback phase directly in terms of the frequency shift and the external cavity

round-trip time $\tau_{ext} = 2L_p/v_{gp}$:

$$2\beta L_p = \Delta\omega\tau_{ext} + 2\beta_0 L_p. \tag{5.188}$$

$2\beta_0 L_p$ is the feedback phase that would exist at the laser's unperturbed frequency. Substituting this into Eq. (5.187) leads to a transcendental equation that cannot be solved for $\Delta\omega$ analytically. However, we can get a feel for the solutions by plotting the left- and right-hand sides individually and looking for frequencies that satisfy the equation. Such a graphical solution is shown in Fig. 5.27. The negative of the round-trip phase shift, $-\Delta\phi_{r-t} = \Delta\omega\tau_L$ for three internal cavity modes and the effective phase of mirror #2, ϕ'_{eff} are all plotted as a function of frequency. A cavity mode appears wherever the phases cancel, $-\Delta\phi_{r-t} = \phi'_{eff}$, which means wherever the two curves intersect.

For weak feedback, only one mode per original internal cavity mode exists. However, for strong feedback we find that multiple modes clustered near each original internal cavity mode become possible. The spacing between these *external cavity modes* is somewhat uneven for the case shown in the figure. However, as the feedback and oscillation amplitude increase, the external mode spacing evens out approaching $\Delta\omega_m = \pi/\tau_{ext}$ (one for every zero crossing). This is half the mode spacing one might expect. For example, the unperturbed internal cavity modes are spaced by $\Delta\omega_m = 2\pi/\tau_L$. The reason for the π phase interval instead of the normal 2π phase interval is that both in-phase and out-of-phase feedback reflection cases (which are spaced by π) represent valid solutions (for $\alpha = 0, 2\beta_0 L_p = \pi m$)

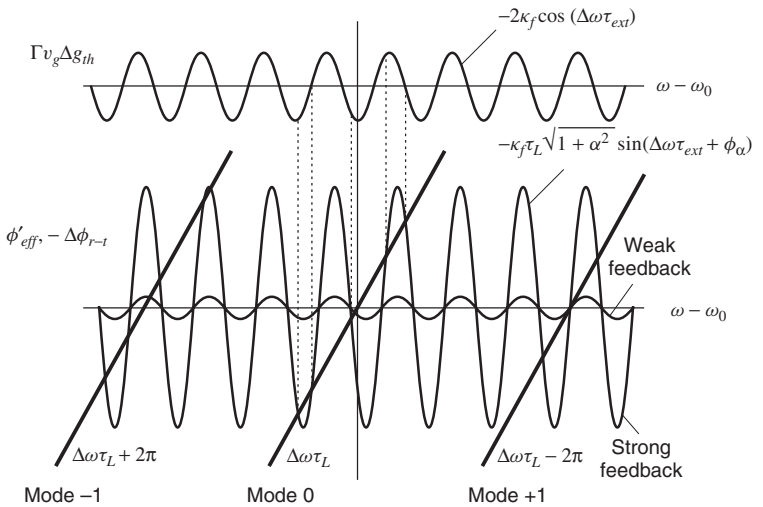


FIGURE 5.27: Graphical solution of the laser cavity modes with external feedback. Cavity modes for both weak and strong feedback are found at the intersection of the straight and sinusoidal phase curves. The relative threshold gains of the external cavity modes clustered near mode 0 in the strong feedback case are indicated by the dashed lines (the curves assume $2\beta_0 L_p = 2\pi m$).

because both cases leave the reflection phase of mirror #2 unperturbed. The catch is that the out-of-phase solutions have a higher threshold gain (see Eq. (5.181)), and hence are not as prominent in a multimode lasing spectrum.

In general, the solutions do not usually yield either pure in-phase or out-of-phase feedback modes (identified by alignment with the extremes in Δg_{th}). However, modes do show tendencies toward one or the other. For example, in Fig. 5.27, the first, third, and perhaps fifth external cavity modes originating from mode 0 might be classified as relatively in phase, whereas the second and clearly the fourth are closer to out-of-phase modes. Of the three relatively in-phase feedback modes, the one with the lowest threshold gain (i.e., the one with the most in-phase reflection) becomes the dominant mode in the spectrum, replacing what was the internal mode of the laser with no feedback. Note from the figure that the dominant cavity mode is not necessarily the external mode closest to the original internal mode.

The transition from single to multiple mode solutions signifies a distinct change in the mode spectrum of the laser and is useful to quantify. From Fig. 5.27, it is evident that multiple solutions cannot exist when the slope of the straight curve ($-\Delta\phi_{r-t}$) exceeds the maximum slope of the oscillatory curve (ϕ'_{eff}). Using Eq. (5.187), the maximum slope of ϕ'_{eff}/τ_L is $\kappa_f \tau_{ext}\sqrt{1+\alpha^2}$, whereas the slope of $-\Delta\phi_{r-t}/\tau_L$ is 1. Thus, multiple solutions cannot exist for

$$C = \kappa_f \tau_{ext}\sqrt{1+\alpha^2} < 1. \tag{5.189}$$

The feedback coefficient, C, characterizes the level of feedback in relation to how it affects the mode structure of the laser ($C = |d\phi'_{eff}/d(\omega\tau_L)|_{\max}$). From Fig. 5.27, it is obvious that rapid and/or large oscillations in ϕ'_{eff} can lead to multiple solutions. Equation (5.189) reflects this observation showing that both long external cavities and/or strong feedback can produce multiple external cavity modes.

A good example of how feedback can affect the *LI* curve of a laser is shown in Fig. 5.28. As shown in the inset, the substrate-air interface of the substrate-emitting VCSEL provides a strong external reflection. The oscillations in output power in this case are not caused by changes in the external path length, but rather by a steady increase in the lasing wavelength with increasing current (which is equally effective at scanning the feedback phase, $-4\pi n_p L_p/\lambda$, through multiples of 2π). We can observe a cancellation between the differential efficiency change and the threshold shift effects at about 3 mA. We can also observe that the oscillations are not entirely sinusoidal as Eq. (5.184) might suggest. This is due to *mode pulling*, where the feedback itself modifies β in often unpredictable ways (we will consider these effects next). To prevent such oscillations in output power, an antireflection coating is usually applied to the substrate of the VCSEL.

5.7.2 Injection Locking—Static Characteristics

The case of injection locking is analogous to the case of optical feedback, with a difference that the external (master) signal coupled into the cavity is now uncorrelated to the original (slave) laser signal, and generally on a different wavelength.

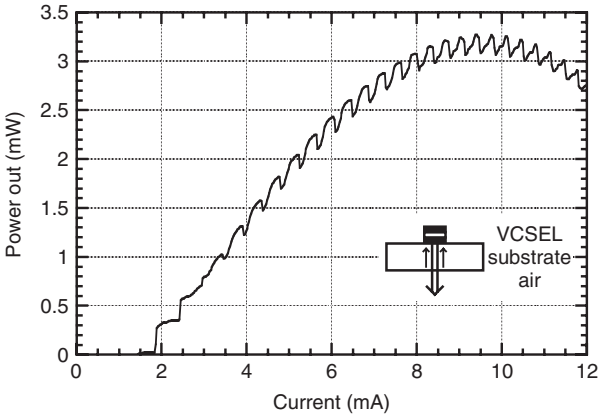


FIGURE 5.28: Continuous-wave *LI* curve of a substrate-emitting VCSEL with strong feedback from the substrate-air interface. As the current increases, the temperature of the VCSEL increases. This increases both the index and optical path of the resonator cavity, leading to a steady increase in lasing wavelength (and shift in feedback phase) with increasing current. The peaks (valleys) in the output power occur when the feedback field subtracts from (adds to) the recirculating field in the VCSEL cavity (the opposite of what one might think). The rollover of the *LI* curve beyond 8 mA is due to a variety of thermal effects.

Still, like in the case of optical feedback, this injected signal will cause the frequency shift that can be described by two distinct effects—gain and phase condition changes inside of the slave laser. The interplay between the injected field and the local field will cause the output signal from the slave laser to appear at the wavelength of the master laser, even though the slave laser itself will continue lasing at a frequency corresponding to one of its cavity modes. However, because the carrier density in the cavity will be perturbed by this external stimulus, the cavity modes of the slave laser will be shifted relative to their unperturbed state. Similar to the case of optical feedback, we will use the injected electric field $\mathscr{E}_i$ over the cavity round-trip time τ_L to define the normalized master laser field injection rate as

$$\kappa_i = \frac{\mathscr{E}_i}{\mathscr{E}\tau_L}, \tag{5.190}$$

where $\mathscr{E}$ is the electric field of the slave laser. A phasor model describing the operation of the injection locked slave laser is illustrated in Fig. 5.29. The change in value of the complex electric field $\mathscr{E}$ of the slave laser is caused by the injected master signal $\mathscr{E}_i$, represented by a real vector.

Over a time period of Δt, the injected signal introduces an out-of-phase electric field component into the slave laser's electric field, $\Delta\mathscr{E} = \mathscr{E}\kappa_i \Delta t$, as shown in Fig. 5.29. This addition will alter the phase of the slave vector by $\Delta\phi$. From Fig. 5.29, we can observe that

$$\Delta\phi \approx \sin(\Delta\phi) = \frac{\Delta\mathscr{E}_s}{\mathscr{E}} = \frac{\Delta\mathscr{E}\sin\phi}{\mathscr{E}} = \kappa_i \Delta t \sin\phi, \tag{5.191}$$

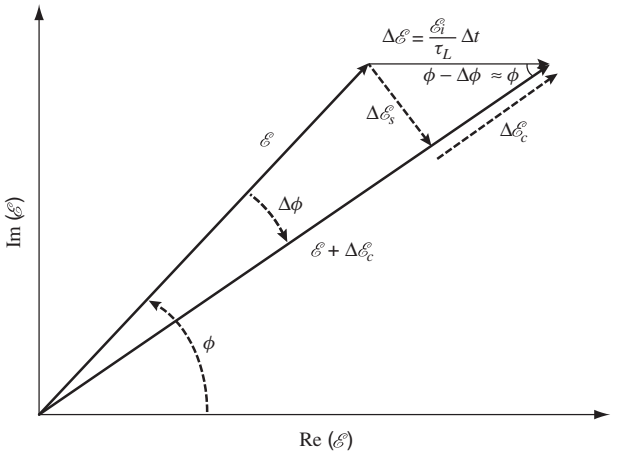


FIGURE 5.29: Injection locking mechanism that keeps a slave laser locked in a time Δt, the injected electric field adds a real component of the electric field to the slave laser signal, causing a phase shift equal to $\Delta\phi$.

which is valid for small values of Δt and consequently $\Delta\phi$. From here, the angular frequency change of the slave laser is

$$\Delta\omega = -\frac{\Delta\phi}{\Delta t} = -\kappa_i \sin\phi. \tag{5.192}$$

The second mechanism that alters the frequency of the slave laser is its gain perturbation, caused by the injected signal. The condition imposed by the slave laser's steady-state operation is that its output power be constant in time. Thus, over the same time period of observation Δt, the gain in the slave laser needs to change to maintain that constant output. Optical power P is proportional to the electric field squared, $P = c^2|\mathscr{E}|^2$, where c is a constant. The change in output power from the slave laser is given by

$$\Delta P = P(\Delta t) - P(0) = c^2(|\mathscr{E}| + \Delta\mathscr{E}_c)^2 - c^2|\mathscr{E}|^2 \approx 2c^2|\mathscr{E}|\Delta\mathscr{E}_c. \tag{5.193}$$

Again, referring to Figure 5.29, we have that

$$\Delta\mathscr{E}_c \approx \Delta\mathscr{E}\cos\phi, \tag{5.194}$$

leading to

$$\Delta P = 2c\sqrt{P}\Delta\mathscr{E}\cos\phi. \tag{5.195}$$

Another way to determine the change in the output power is from the rate equation for photons, Eq. (2.17). In a time period Δt, we can calculate the change

in the photon density, which is proportional to the change of the light power,

$$\Delta P = \Gamma v_g \cdot \Delta g \cdot P \cdot \Delta t. \tag{5.196}$$

We can now equate the expressions for the power change ΔP, from Eqs. (5.193) and (5.196), to get

$$\Delta g = 2\frac{\mathscr{E}_i}{\mathscr{E} v_g \Gamma \tau_L} \cos\phi = 2\frac{\kappa_i}{\Gamma v_g} \cos\phi. \tag{5.197}$$

Analogous to the case of optical feedback, the change in optical gain causes an additional round-trip phase change through the refractive index change, and can be expressed via the linewidth enhancement factor, α. Using Eq. (5.75), and the relationship between the angular velocity and frequency, $\Delta\omega = 2\pi\Delta\nu$, the resulting carrier induced frequency shift is

$$\Delta\omega = -\frac{\alpha}{2}\Gamma v_g \cdot \Delta g = -\alpha\kappa_i \cos\phi. \tag{5.198}$$

Similarly, the change in carrier density can be expressed from the change in gain, obtaining the result identical to Eq. (5.182), with κ_i replacing κ_f. The total frequency shift, using expression 5.192, becomes

$$\Delta\omega = -(\kappa_i \sin\phi + \alpha\kappa_i \cos\phi) = -\kappa_i\sqrt{1+\alpha^2}\sin(\phi + \phi_\alpha), \tag{5.199}$$

where $\phi_\alpha = \tan^{-1}\alpha$. This expression also is identical to the expression obtained for the frequency shift due to optical feedback, Eq. (5.187). This is no coincidence, as the case of optical feedback can be treated as a special case of optical injection, by realizing that $P_1 = \frac{t_2^2}{r_2}P_f$, where the injected master power is given by $P_1 = c^2|\mathscr{E}_i|^2$.

5.7.3 Injection and Feedback Dynamic Characteristics and Stability

The static behavior of the laser discussed in the previous section considered the basic changes to the threshold, output power, and mode spectrum induced feedback or external injection into the laser. However, such steady-state solutions may not always be stable against random fluctuations in the carrier and photon density. In other words, the inherent noise of the laser can induce dynamic instabilities, which can literally run wild under certain external signal conditions.

To solve for the dynamic behavior of a laser under feedback or optical injection, we need to generalize our rate equations, by replacing the photon density rate equation with two separate rate equations describing the electric field amplitude and phase changes in time. We can then apply the same technique of differential analysis to these three rate equations to observe how the laser will behave dynamically in response to the perturbations due to feedback/optical injection. The response will in large part be directed by the determinant of the matrix of the system of equations,

$$\frac{d}{dt}\begin{bmatrix} 0 \\ 0 \\ \Delta J \end{bmatrix} = \begin{bmatrix} \gamma_{EE} + j\omega & \gamma_{E\phi} & \gamma_{EN} \\ \gamma_{\phi E} & \gamma_{\phi\phi} & \gamma_{\phi N} \\ \gamma_{NE} & 0 & \gamma_{NN} + j\omega \end{bmatrix} \begin{bmatrix} dE \\ d\phi \\ dN_p \end{bmatrix}. \tag{5.200}$$

Based on the solutions for laser behavior under both static and dynamic conditions, we can define different ranges of laser operation. We are particularly interested in the laser's locking range, under which the laser will operate with stable, narrow linewidth and be locked to the perturbation signal.

The first condition for stability comes from Eqs. (5.187) and (5.198), and the fact that the phase of the (slave) laser has to be real. This condition will be fullfilled if the value of the sine function is less than unity, which will be true for angle ϕ in the range of

$$-\frac{\pi}{2} - \tan^{-1}\alpha \leq \phi \leq \frac{\pi}{2} - \tan^{-1}\alpha. \tag{5.201}$$

The second condition comes from the requirement that the change in carrier density must be negative, so that the total carrier density can never be higher than the threshold carrier density, N_{th}. From this, we note that $\cos\phi > 0$, and $-\frac{\pi}{2} \leq \phi$. Combining the two requirements for the (slave) laser phase, and using Eqs. (5.187) and (5.198), we finally arrive at the Morgensen's locking range [22],

$$-\kappa_i\sqrt{1+\alpha^2} \leq \Delta\omega \leq \kappa_i. \tag{5.202}$$

The final constraint comes from the dynamic equation solution for the frequency response, and from the knowledge of the stability theory that states that real parts of roots of the system determinant must remain negative for a system to remain stable. Unfortunatelly, the locking range based on this constraint needs to be determined numerically.

A diagram showing the locking range on a detuning frequency versus injection ratio plot is shown in Figure 5.30. The solid curves show the boundaries between optical injection locking and nonlocking regions. The boundary of the unstable and stable injection locking areas is denoted by a dotted curve. The asymmetric feature of stable injection locking again originated from the fact that the α linewidth enhancement factor has a nonzero value in semiconductor lasers. The case of optical feedback is represented by the line obtained for detuning frequency equal to zero.

5.7.4 Feedback Effects on Laser Linewidth

In this section, we analyze qualitatively random fluctuations in the carrier and photon density caused by feedback, and their impact on laser linewidth. Two dominant factors are responsible for these instabilities.

The first factor is related to the mode solutions depicted in Fig. 5.27. Under certain conditions, the two intersecting curves can run tangent to each other as demonstrated by mode $+1$ with weak feedback (at the zero crossing), or mode -1 with strong feedback (at the extreme of the sine wave). In such cases, the resonance

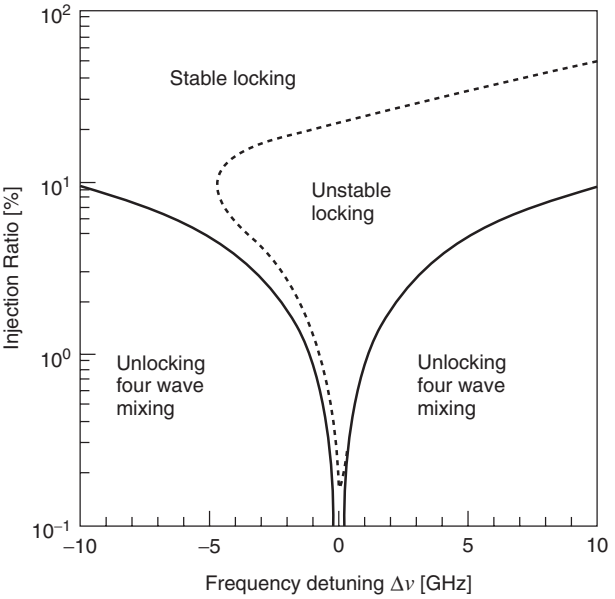


FIGURE 5.30: Locking and unlocking regions in phase space of frequency detuning and injection field. *Reprinted with permission from Springer-Verlag*, [22].

condition is not well defined making the exact position of the mode extremely sensitive to carrier-induced fluctuations in the round-trip phase. The random positioning of the mode also affects the threshold gain, which can lead to sporadic mode hopping if other solutions have similar threshold gains. The second factor involves the external round-trip delay of the feedback field relative to the recirculating mode. This time-delayed reinjection of power and phase fluctuations can accentuate the instability of the laser considerably. The net result is that external feedback can affect the noise, linewidth, and dynamic properties of single-frequency lasers very dramatically. Certain levels of feedback and delay can result in self-pulsations, very large linewidths, or large enhancements in the low-frequency noise.

Experimentally, five distinct regimes of laser performance under feedback have been identified [19]. These are indicated in Fig. 5.31, for reflection distances of 5 cm and higher. For zero distances, it is interesting to note that there is a correspondence with the regions indicated in Figure 5.31—that is, regions II and III disappear, region V becomes the stable injection locking region, and region IV, coherence collapse, becomes the unstable injection locking region. Finally, region I remains a single mode region in both cases.

To understand the origins of these five regions under feedback, let's begin at the bottom of the plot in regime I. For weak feedback and relatively short external cavities, $C < 1$ and only one mode for each original internal laser mode exists. In this regime it has been observed that the laser linewidth is narrowed for certain feedback phases and broadened for others. A similar decrease and increase occurs

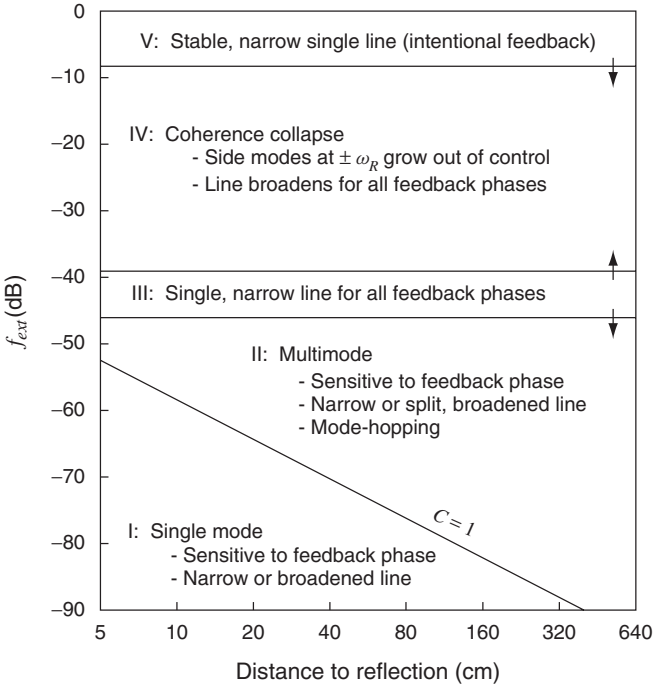


FIGURE 5.31: Sketch of the five feedback regimes. The regime boundaries plotted here were measured on one particular laser and will vary somewhat, depending on the laser structure and bias level. For example, the arrows indicate the movement of the boundaries with increasing output power. After Ref. [16].

in the RIN level. Figure 5.27 can be used to explain this behavior. As mentioned earlier, when the two intersecting curves run tangent to each other, the mode position can fluctuate leading to an increase in linewidth. For mode +1 with weak feedback, the phase slopes of $-\Delta\phi_{r-t}$ and ϕ'_{eff} are parallel and an instability is expected. However, when the phase slopes are crossed like at mode 0 with weak feedback, the mode position is actually stabilized by the feedback, narrowing the linewidth and quieting the laser. A more rigorous analysis [17] reveals that the laser linewidth varies inversely with the difference between the phase slopes squared,[9] or $\propto [d(\Delta\phi_{r-t} + \phi'_{eff})/d(\omega\tau_L)]^{-2}$. Using Eq. (5.187) to evaluate the derivatives, we can write the linewidth with feedback relative to the unperturbed linewidth as

$$\Delta\nu = \frac{\Delta\nu_0}{[1 + C\cos(2\beta L_p + \phi_\alpha)]^2}, \tag{5.203}$$

where again $\phi_\alpha = \tan^{-1}\alpha$ and C is given by Eq. (5.189).

[9]This relation holds as long as $\tau_{ext} < \tau_{coh}$, such that the feedback field interferes coherently with the recirculating field in the cavity.

When the feedback phase is adjusted to $2\beta L_p = 2\pi m - \phi_\alpha$, the phase slopes are crossed and the linewidth is narrowed to $\Delta\nu_0/(1+C)^2$. When $2\beta L_p = 2\pi m + \pi - \phi_\alpha$, the phase slopes are parallel and the linewidth is broadened to $\Delta\nu_0/(1-C)^2$. In between these extremes (for example at mode -1), the feedback phase approaches zero slope and the linewidth $\rightarrow \Delta\nu_0$. The crossed phase slopes configuration of mode 0 is often referred to as having "in-phase" feedback even though this is only true for $\phi_\alpha = 0$. Likewise, the parallel phase slopes configuration of mode $+1$ is referred to as having "out-of-phase" feedback. For large α, both configurations are closer to having a quadrature feedback phase. We can tune the dominant mode of the laser (e.g., mode 0) to either of these configurations by adjusting $2\beta_0 L_p$, which has the effect of shifting all sinusoidal curves in Fig. 5.27 to the left or right by as much as we desire. Every π change in $2\beta_0 L_p$ alternates the phase slope alignment from crossed to parallel. In any case, regardless of the feedback phase, when $C < 0.05$, the feedback modifies the linewidth by $< 10\%$ and the laser can be considered sufficiently isolated for many applications. As $C \rightarrow 1$, the narrowing and broadening increase until eventually Eq. (5.203) predicts an infinite linewidth for the out-of-phase feedback configuration (i.e., when the two phase slopes are perfectly tangent to each other). In reality, the laser mode splits into two modes [16, 18] and this defines the boundary between regimes I and II.

For long external cavities, even relatively weak feedback can lead to $C > 1$ and we enter regime II as illustrated in Fig. 5.31. Regime II is characterized by the same feedback phase sensitivity described for regime I. However, in this regime multiple solutions are allowed. Experimentally, the in-phase feedback configuration of mode 0 continues to provide a narrowing single line as C increases. Apparently when the linewidth is stabilized by the crossed phase slopes, the laser favors this mode even in the presence of other mode solutions. For the out-of-phase feedback configuration of mode $+1$, the one mode splits into three mode solutions as the extremes of the sine wave penetrate across the straight phase curve with increasing feedback. The three modes of mode $+1$ can be observed in Fig. 5.27 for strong feedback. Of the three modes, the zero-crossing mode generally has a higher threshold gain and hence does not appear in the spectrum. The outer two mode solutions, however, can be tuned slightly via $2\beta_0 L_p$ until they have identical threshold gains. For this feedback phase, the laser line splits into two closely spaced modes. Thus, the out-of-phase feedback configuration in regime II is characterized by a double peaked lineshape.

As the feedback increases, the separation between the two modes increases and eventually reaches $2\pi/\tau_{ext}$ (one full cycle of the sine wave). Thus, the overall effective laser linewidth can increase to $\Delta\nu \sim 1/\tau_{ext}$ [18]. Furthermore, the laser does not exist in both modes at the same time. Rather, it jumps between the two modes at a rate in the few MHz range. This mode hopping creates low-frequency noise that can be undesirable in some applications. However, for both regimes I and II, one might argue that intentional feedback with the proper feedback phase can improve the laser performance. In practice, maintaining the proper feedback phase can prove challenging and it is generally best to either keep $C < 0.05$ or stay away from regimes I and II entirely.

As the two modes of the out-of-phase feedback configuration (mode +1) reach a separation close to $2\pi/\tau_{ext}$ with increasing feedback, the rate of jumping decreases and eventually stops altogether marking the transition to regime III. The laser then settles down to one of the two modes. It is interesting to point out that the two outer modes originating from mode +1 have now shifted on the sine wave over to a crossed phase slopes, and hence stable, narrow linewidth configuration. In fact they are approaching this level in Fig. 5.27. As a result, the linewidth is narrowed for both in-phase and out-of-phase feedback configurations, making regime III independent of feedback phase. The boundary between regimes II and III has been found experimentally to be independent of external cavity length (as have all other higher feedback regime boundaries indicated in Fig. 5.31). Although regime III offers stable, low noise, narrow single-line laser operation, it unfortunately spans only a small range of feedback levels. The feedback regime opens up for higher output powers [16], but is still relatively small. Designing an intentional feedback system with tight tolerances to maintain the feedback level in regime III is possible but not very practical.

As the feedback level is increased still further, the laser enters a new regime that is characterized by complex dynamic interactions between the many relatively strong multiple external mode solutions. Rate equation models of the field and phase including the time delay of the feedback field reveal unstable solutions and chaotic behavior at these high feedback levels [19]. Indeed, experiments reveal linewidths so large that the coherence length of the laser drops below 1 cm (as opposed to tens of meters). For this reason, regime IV is referred to as the coherence collapse regime. The collapse begins at the transition to regime IV, where the normally small satellite peaks created by noise-induced relaxation oscillations (depicted in Fig. 5.23) grow larger and larger, eventually becoming comparable to the central peak and broadening the linewidth dramatically. It has been suggested that the transition between regimes III and IV occurs when the feedback rate $\kappa_f \sim \omega_R$ [16]. A more refined theoretical estimate of the critical level of feedback is given by [20]

$$f_{ext}|_{crit} = \frac{\tau_L^2}{16C_e^2}(Kf_R^2 + \gamma_0)^2 \left[\frac{1+\alpha^2}{\alpha^4}\right], \tag{5.204}$$

where we define a laser coupling factor, $C_e = t_2^2/2r_2$ (for Fabry–Perot cavities). The K-factor and γ_0 are given by Eq. (5.53). The critical feedback level is increased for larger output power (via f_R^2), smaller α, longer cavity lengths, and smaller laser coupling factors. The maximum linewidth and RIN level within the coherence collapse regime have also been theoretically estimated and experimentally verified. They are given by [19]

$$\Delta\nu|_{\max} = f_R\sqrt{1+\alpha^2}\ \sqrt{\ln 4}, \tag{5.205}$$

$$\left.\frac{\text{RIN}}{\Delta f}\right|_{\max} = \frac{1}{\gamma} = \frac{1}{Kf_r^2 + \gamma_0}. \tag{5.206}$$

For a resonance frequency of 5 GHz and $\alpha = 5$, we obtain a maximum linewidth of 30 GHz! However, for a typical damping factor of 10^{10} s^{-1} at a few milliwatts of output power, we obtain a RIN of -100 dB/Hz. This level of RIN is large but not extreme. Thus, simple systems may be able to function adequately in regime IV as long as the linewidth is not of critical importance. Figure 5.32 summarizes the impact on laser linewidth due to feedback as it is scanned through regimes I through IV [21]. The envelope superimposed on the measurements reveals the linewidth variation created by the feedback phase.

Finally, at very high feedback levels when the external mirror becomes comparable to the laser's mirror feedback, a stable regime can be observed. Regime V is usually achieved only when the feedback is deliberate and well

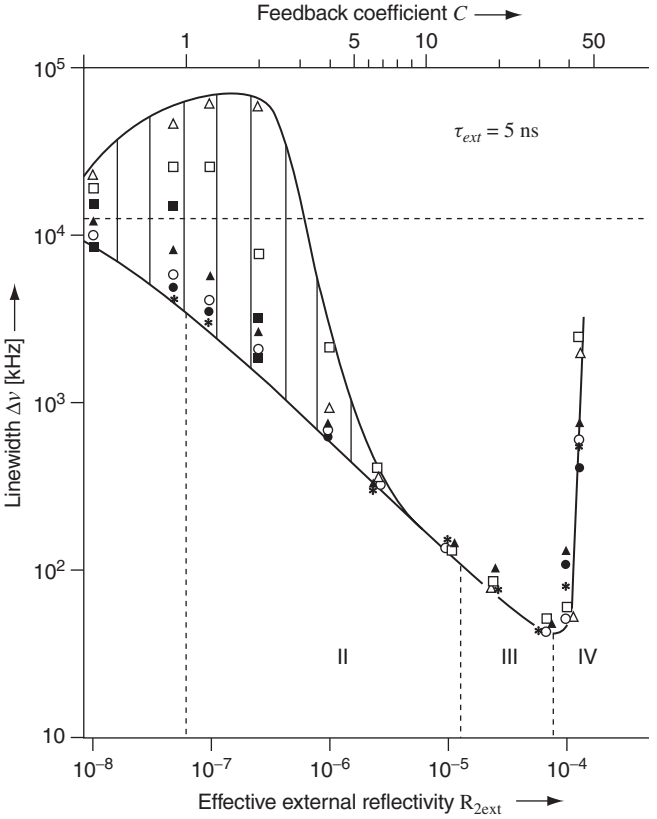


FIGURE 5.32: Spectral linewidth versus the amount of external optical feedback [18]. The roman numerals I–IV denote the regimes given in Fig. 5.31. In this example the assumed parameters are: $\tau_{ext} = 5$ ns, $\tau_L = 9$ *ps*, $R_2 = 0.32$, $\alpha = 6$, and 5 mW/facet out for an InGaAsP/InP laser ($R_{2ext} \equiv t_2^4 f_{ext}$). (*Reprinted, by permission, from IEEE Journal of Quantum Electronics © IEEE* 1990).

designed. For example, monolithic complex cavities considered in Chapter 3 fall into this category. An external mirror with good coupling to the laser cavity via an antireflection coating on mirror # 2 can also reach feedback levels in regime V. The mode solutions in this feedback regime are stable with very narrow linewidths. If the mirror # 2 reflection is small, it is sufficient to treat the entire laser cavity as a solitary laser with mode spacing given by $2\pi/(\tau_L + \tau_{ext})$. The effect on linewidth is then simply calculated by including a fill factor in the modal gain. The expression for linewidth derived earlier Eq. (5.151) depends on Γ^2. Thus, all else being equal, we can describe the line narrowing in regime V by

$$\Delta\nu = \Delta\nu_0 \left[\frac{\bar{n}_{ga} L_a}{\bar{n}_{ga} L_a + \bar{n}_{gp} L_p} \right]^2 . \tag{5.207}$$

This result is not entirely different from Eq. (5.203) evaluated at the in-phase condition because $C \propto \bar{n}_{gp} L_p / \bar{n}_{ga} L_a$. With typical numbers such as $L_a = 300\ \mu\text{m}, L_p = 5\ \text{cm}, \bar{n}_{ga} = 4.2$, and $\bar{n}_{gp} = 1$, the line narrowing is $\sim (1/40)^2$. The external cavity can therefore reduce a 100 MHz linewidth down to 62.5 kHz! Combined with a frequency-selective filter to attain single-mode operation, such deliberate external cavity feedback can be quite useful for applications demanding highly coherent light.

As for unintentional feedback, it is unfortunately very easy to fall into regime IV with small reflections. For on-chip optical feedback reduction, the main strategy is to reduce the number of optical interfaces with large discontinuities, as well as to angle these interfaces, to minimize the coupling back of the reflected light into the fundamental mode—through phase front missalignment [26]. For

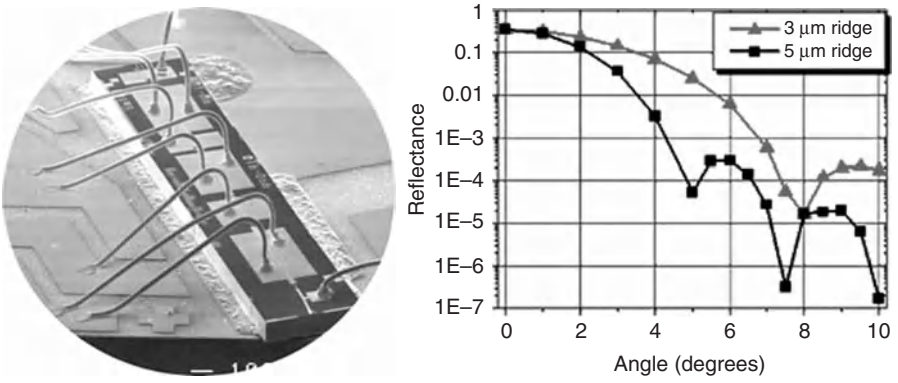


FIGURE 5.33: (left) A multisection tunable DBR laser, with an angled, tapered output waveguide (right) Calculated reflectance for a particular output waveguide geometry, as a function of angle of incidence, and for two different ridge waveguide widths.

facet reflection suppression, a combination of angled interface and optical beam expansion (that amplifies the effect of the angled interface) is most often used. The simplest implementation is a curved tapered output waveguide, as shown in Figure 5.33 on the left. The combination of these approaches can reduce the reflectance by several orders of magnitude, as shown in Figure 5.33 (right), before an antireflection coating is applied.

For off-chip reflections, the main sources of reflection are micro-optic components used to couple the light into the fiber, as well as fiber connectors at moderate distances. Optical isolators that provide greater than 60 dB of isolation are therefore often required.

REFERENCES

1. Scott JW, Thibeault BJ, Mahon CJ, Coldren LA, Peters FJ. High modulation efficiency of intracavity contacted vertical cavity lasers. *Applied Phys Lett* 1994;**65**: 1483.
2. Dutta NK, Olsson NA, Koszi LA, Besomi P, Wilson RB. Frequency chirp under current modulaton in InGaAsP injection lasers. *J Appl Phys* 1984;**56**: 2167.
3. Schimpe R, Bowers JE, Koch TL. Characterisation of frequency response of 1.5 μm InGaAsP DFB laser diode and InGaAs PIN photodiode by heterodyne measurement technique. *Electron Lett* 1986;**22**: 453.
4. Yariv A. *Optical electronics*. 4th ed. Philadelphia: Saunders Collete Publishing; 1991. Chapter 10.
5. Henry CH. Theory of the linewidth of semiconductor lasers. *IEEE J Quantum Electron* 1982;**QE-18**: 259.
6. Schawlow AL, Townes CH. Infrared and Optical Masers. *Phys Rev* 1958;**112**: 1940.
7. Lax M. Classical noise. V. Noise in self-sustained oscillators. *Phys Rev* 1967;**160**: 290.
8. Gerhardt H, Welling H, Güttner A. Measurements of the laser linewidth due to quantum phase and quantum amplitude noise above and below threshold. *I. Z Physik* 1972;**253**: 113.
9. Yamamoto Y, Machida S. High-impedance suppression of pump fluctuation and amplitude squeezing in semiconductor lasers. *Phys Rev A* 1987;**35**: 5114.
10. Yamamoto Y, Imoto N. Internal and external field fluctuations of a laser oscillator: Part I—Quantum mechanical Langevin treatment. *IEEE J Quantum Electron* 1986;**QE-22**: 2032.
11. Kikuchi K, Okoshi T. Measurement of FM noise, AM noise, and field spectra of 1.3 μm InGaAsP DFB lasers and determination of the linewidth enhancement factor. *IEEE J Quantum Electron* 1985;**QE-21**: 1814.
12. Goobar E, Karlsson A, Björk G, Rigole P-J. Quantum correlated light beams and sub-Poissonian electrical partition noise in parallel and series coupled semiconductor light emitters. *Phys Rev Lett* 1993;**70**: 437.
13. Richardson WH, Machida S, Yamamoto Y. Squeezed photon-number noise and sub-Poissonian electrical partition noise in a semiconductor laser. *Phys Rev Lett* 1991;**66**: 2867.
14. Bachor H-A, Rottengarter P, Savage CM *Appl Phys B* 1992;**55**: 258.

15. Machida S, Yamamoto Y, Itaya A. Observation of amplitude squeezing in a constant-current–driven semiconductor laser. *Phys Rev Lett* 1987;**58**: 1000.
16. Okoshi T, Kikuchi K, Nakayama A. Novel method for high resolution measurement of laser output spectrum. *Electron Lett* 1980;**16**: 630.
17. Henry CH. Theory of the phase noise and power spectrum of a single mode injection laser. *IEEE J Quantum Electron* 1983;**QE-19**: 1391.
18. Nagaragan R, Ishikawa M, Fukushima R, Geels, RS, Bowers JE. High speed quantum-well lasers and carrier transport effects. *IEEE J Quantum Electron* 1992;**28**: 1990.
19. Tkach RW, Chraplyvy AR. Regimes of feedback effects in 1.5-μm distributed feedback lasers. *IEEE J Lightwave Technol* 1986;**LT-4**: 1655.
20. Agrawal GP. Line narrowing in a single-mode injection laser due to external optical feedback. *IEEEJ Quantum Electron* 1984;**QE-20**: 468.
21. Helms J, Kurzke C, Petermann K. External feedback requirements for coherent optical communication systems. *IEEE J Lightwave Technol* 1992;**LT-10**: 1137.
22. Ohtsubo J. *Semiconductor laser: stability, instatability and chaos*. New York: Springer-Verlag; 2008.
23. Wang J, Petermann K. Noise analysis of semiconductor lasers within the coherence collapse regime. *IEEE J Quantum Electron* 1991;**QE-27**: 3.
24. Helms J, Petermann K. A simple analytic expression for the stable operation range of laser diodes with optical feedback. *IEEE J Quantum Electron* 1990;**QE-26**: 833.
25. Petermann K. *Laser diode modulation and noise*. New York: Kluwer Academic Publishers; 1991.
26. Marcuse D. Reflection loss of laser mode from tilted end mirror. *J Lightwave Technology* 1989;**7**(2): 336–339.

READING LIST

Argawal GP, Dutta NK. *Semiconductor lasers*. New York: Van Nostrand Reinhold; 1993. Chapter 6.

Ebeling KJ. *Integrated opto-electronics*. Berlin: Springer-Verlag; 1993. Chapter 10.

Henry CH. Line broadening of semiconductor lasers. In: Yamamoto Y, editor. *Coherence, amplification, and quantum effects in semiconductor lasers*. New York: Wiley-Interscience; 1991. Chapter 2.

Ohtsubo J. *Semiconductor lasers: stability, instability and chaos*. Berlin: Springer-Verlag; 2008.

Petermann K. *Laser diode modulation and noise*. New York: Kluwer Academic Publishers; 1991. Chapters 4 & 9.

Yariv A. *Optical electronics*. 4th ed. Philadelphia: Saunders College Publishing; 1991. Chapters 10 & 15.

PROBLEMS

1. What is the current relative to threshold in Fig. 5.1?

2. Derive N_1 for P_1 for a mirror-loss-modulated laser instread of a current-modulated laser (i.e., assume $dI(t) = 0$ and $d\alpha_m(t) = \alpha_{m1}e^{j\omega t}$). What is the major distinction between the two types of modulation?

3. Derive $N_{p1} + N_{p2}$ assuming there are two competing modes in the cavity and that intermodal gain compression can be neglected. Under what conditions does the modulation response of the total photon density resemble the modulation response of either of the two modes separately (in other words, under what conditions can we ignore the division of photons into separate modes)?

4. Derive P_1/I_1 for a laser that has a series resistance, R, and a parallel shunt capacitance, C, across the diode such that not all of the terminal current, I_1, makes it to the active region. Express the result in terms of an overall transfer function, $H_{RC}(\omega)H(\omega)$, which includes RC parasitics. For a series resistance of 10 Ω and a laser with negligible damping and a relaxation resonance frequency of 20 GHz, determine the capacitance that reduces the 3 dB bandwidth of the laser to 90% of its value with no parasitics.

5. By solving for the poles of Eq. (5.59), show that Eq. (5.60) is correct.

6. Using Table 5.1, evaluate all terms in Eq. (5.49), and thus, verify Eq. (5.51).

7. Using the analytic approximations for a step function in current, plot the power out of one end of a 300-μm-long, 3-μm-wide cleaved-facet InGaAs/GaAs 3-QW buried-heterostructure laser versus time in response to a current that abruptly rises by 10% from an initial value of twice threshold. The laser emits at 0.98 μm and has an internal loss of 5 cm^{-1}, an internal efficiency of 80%, gain and recombination parameters as in Table 5.1 for an 80 Å quantum-well, and a confinement factor of 6%.

8. For the transient response given in the example of Fig. 5.6a, plot the frequency chirping versus time assuming a linewidth enhancement factor of 5.

9. For the in-plane device described in Table 5.1 and Fig. 5.6, a small signal sinusoidal current of varying frequency is applied at a bias current of twice threshold. The frequency response peaks at some frequency and then falls off rapidly. What is this peak frequency? By how much is it different from the relaxation resonance frequency?

10. An edge-emitting, mode-locked laser in InP, operating at 1.55 μm, consists of an active region, a 50-μm-long saturable absorber, and 2 DBR mirrors, whose lengths are $L_1 = 100$ μm and $L_2 = 200$ μm. The coupling constant for the gratings in the DBR mirrors is $\kappa = 0.0131/\mu m$. Assume that the group index in all sections is equal to 3.8, and that the laser waveguide supports only one lateral and one transverse mode. AR coatings are applied to the laser facets.

What is the length of the active region if this laser generates pulses with the repetition rate of 40 GHz?

11. For the mode locked laser from the previous example, assuming that the gain bandwidth is 3.2 nm, and that the cavity modes are evenly distributed over this range, what is the number of supported cavity modes and what is the width of the pulses emitted by this laser?

12. A 1.55-μm InGaAsP/InP laser with a linewidth enhancement factor of 5 is modulated from 1.5 to $4.5I_{th}$, creating an output power variation of 1–7 mW. Assuming a trapezoidal output at 1 Gb/s with 0.5 ns linear transitions and 0.5 ns constant power plateaus at each extrema, plot the chirping for two periods of a 10101... sequence.

13. A 1.3-μm InGaAsP/InP laser with a threshold carrier lifetime of 3 ns is biased at $0.9I_{th}$. At time $t = 0$ the drive current is suddenly increased to $2I_{th}$. What is the turn-on delay? Make whatever assumptions are necessary.

14. Consider the noise emitted from an InGaAs/GaAs VCSEL parameterized in Table 5.1, at 100 MHz.

(a) What is the RIN level for $I = 2I_{th}$? (Can P_0 be assumed small?)

(b) What is the RIN level at $I = 5I_{th}$? (Can P_0 be assumed large?)

15. What is the expected linewidth for the two cases of Problem 5.14, assuming a linewidth enhancement factor of 5?

16. Estimate the 3 dB modulation bandwidth for the two cases of Problem 5.14.

(a) Neglect transport effects.

(b) Include transport effects wherein the lifetime of carriers in the SCH region due to relaxation into the quantum wells is 50 ps and the leakage out of the well can be neglected.

(c) Include transport wherein leakage out of the well to the SCH region is additionally characterized by a lifetime of 200 ps.

17. A cleaved facet, 400-μm-long edge emitting GaAs multi quantum-well laser lases at 860 nm and its threshold current is 10 mA. The laser injection efficiency is $\eta_i = 0.78$, the average internal cavity losses are $< \alpha_i >= 14\ \text{cm}^{-1}$, and the active and passive region volumes are $V = 5\ \mu\text{m}^3$, and $V_p = 200\ \mu\text{m}^3$, respectively. The laser's differential gain is $a = 5 \cdot 10^{-16}\ \text{cm}^2$ and $\beta_{sp} = 1.3$. The laser is initially biased with a current source set to 5 mA. Then, the laser is biased above threshold, by applying a 35 mA bias current.

(a) What is the turn-on delay of this laser?

(b) What is the laser output power?

(c) What is linewidth of this laser after its output stabilizes, with a bias current (I = 35 mA), if its linewidth enhancement factor is equal to 3?

18. For the laser from previous example, under the same operating conditions,
(a) What is the 3 dB modulation frequency f, assuming that damping is negligible?
(b) What is the maximum the linewidth of this laser under coherence collapse?

19. For the laser from the previous two examples, if we now bias it at 1 mA above threshold and apply a large signal current modulation, the output power waveform can be approximated by the signal sketched out in Fig. 5.34. Sketch the large signal chirp of this laser as a function of time, labeling the time, chirp amplitude and slope.

20. Verify Eq. (5.154). If the normalized fringe visibility decays exponentially and is equal to 0.2 at a time delay of 3 ns, what is the linewidth of the laser?

21. In Fig. 5.28, estimate μ_2 near 2.5 mW of output power. From this determine the feedback level f_{ext} using the VCSEL parameters in Table 5.1, assuming that all of the light is coupled out of mirror #2, and that we are well above threshold. Use a VCSEL cavity length and group index of 1 μm and 4.2, respectively. Also express the feedback level in decibels.

22. Estimate the substrate thickness of the VCSEL in Fig. 5.28, assuming a constant voltage of 3 volts above threshold, a thermal impedance of 3 C/mW, a wavelength shift of 0.08 nm/°C, a substrate–air reflection of 32%, and a substrate group index of 4.2.

23. The intensity feedback from the end of a fiber is 4% and about one-quarter of this is coupled back into the lasing mode (excluding transmission through mirror # 2) in a 300-μm-long cleaved-facet 1.55 μm in-plane laser.

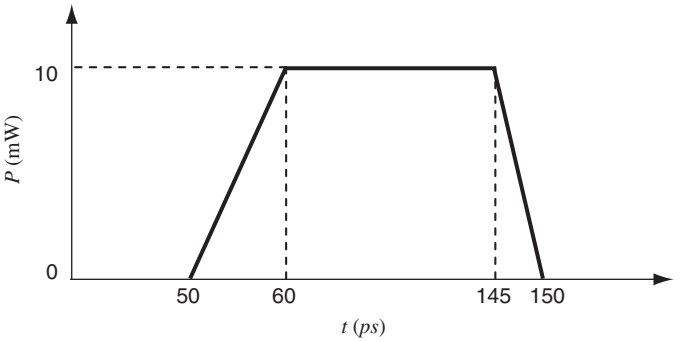


FIGURE 5.34: Idealized output power waveform for large signal modulation of a laser in Problem 5.19.

(a) Plot the feedback parameter, C vs. L_p, the fiber length ($n_{fiber} = 1.45$), for $\alpha = 0, 2, 4, 6$, up to $L_p = 1$ km.

(b) Plot the linewidth narrowing and broadening factor applicable in regimes I through III vs. C for $0 < C < 10$. Discuss what really happens near $C \sim 1$.

CHAPTER SIX

Perturbation, Coupled-Mode Theory, Modal Excitation, and Applications

6.1 INTRODUCTION

This chapter is the first of two that focus more on the electromagnetic aspects of lightwave propagation, particularly as applied to diode lasers and related photonic integrated circuits. In this chapter, after examining the guided mode power and effective width, a concept we will be using through the remainer of the book, we introduce the powerful perturbation and coupled-mode approaches to approximately solve very complex problems, which otherwise might only be addressed numerically.

To use these approaches we generally must know at least some of the eigenmodes of a relatively simple waveguide configuration. Then, the trick is to express the solution to some perturbed or more complex configuration in terms of this original basis set of eigenmodes. As we shall see, by using the orthogonality relationships among the original basis set, it is possible to derive some general formulae that are helpful in solving specific problems. The usefulness of these formulae derives from the fact that they only involve physical dimensions and some of the original basis functions.

To get started, we refer back to Chapter 2 to recall a convenient form for the electric field, $\mathscr{E}$, of some waveguide eigenmode. For mode m, then, the electric field as a function of space and time, Eq. (2.19), can be written as

$$\mathscr{E}_m(x,y,z,t) = \hat{\mathbf{e}}_i E_{0m} U_m(x,y) e^{j(\omega t - \beta_m z)}, \tag{6.1}$$

Diode Lasers and Photonic Integrated Circuits, Second Edition.
Larry A. Coldren, Scott W. Corzine, and Milan L. Mašanović.
© 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

where $\hat{\mathbf{e}}_i$ is the unit vector along the ith coordinate (giving the polarization direction), E_{0m} gives the magnitude of the field, and U_m is the normalized transverse mode shape for mode m. For convenience, we can also combine the polarization into U_m, so that $\mathbf{U}_m = \hat{\mathbf{e}}_i U_m$. Thus, making use of the orthogonality between eigenmodes, we have

$$\int \mathbf{U}_m^* \cdot \mathbf{U}_n \, dA = \delta_{mn}, \tag{6.2}$$

where δ_{mn} is the Kronecker delta function, which equals unity for $m = n$ and zero otherwise. (For orthogonal polarizations the dot product also gives zero, even for $m = n$.) Orthogonal modes of a uniform waveguide do not interact. It is also worth mentioning that in the normal course of solving a waveguide problem, one finds that the eigenfunctions U_m provide a complete set.

Given that the U_m form a complete set, we can express an arbitrary field in the vicinity of a waveguide by a normal-mode expansion of all of the waveguide eigenmodes (including unguided radiation modes). Thus,

$$\mathscr{E}(x, y, z, t) = \sum_m \mathscr{E}_m(x, y, z, t), \tag{6.3}$$

where the amplitudes of the various terms in the summation are given by E_{0m} in Eq. (6.1). Both Eqs. (6.1) and (6.3) are solutions to the wave equation,

$$\nabla^2 \mathscr{E} + \varepsilon(x, y, z) k_0^2 \mathscr{E} = 0, \tag{6.4}$$

where $\varepsilon(x, y, z)$ is the relative dielectric constant, and k_0 is the free-space propagation constant for the medium of interest. For a single mode of a waveguide, we can also use Eq. (6.1) in (6.4) to obtain a wave equation for the transverse mode profile, U:

$$\nabla_T^2 U + [\varepsilon(x, y, z) k_0^2 - \beta^2] U = 0, \tag{6.5}$$

where we have used $\nabla^2 \mathscr{E} = \partial^2 \mathscr{E} / \partial z^2 + \nabla_T^2 \mathscr{E}$.

Although the higher order modes of even a simple waveguide may be complicated, it is fortunate that we need only know the details of at most two modes for all of the discussion to follow in this chapter. In fact, only one mode must be characterized for many problems.

6.2 GUIDED-MODE POWER AND EFFECTIVE WIDTH

In this and the subsequent chapter, many calculations require a normalization of the eigenmode or some integration over a part of it. In this section we calculate the time-averaged Poynting vector power, $\boldsymbol{P}$, carried by a guided mode along the direction of propagation, which can be useful in such calculations. For a simple three-layer slab guide, illustrated in Figure 11.1, we find that the value of the Poynting vector

can be expressed in terms of the magnitude squared of the electric field in the center of the guide and an effective width, d_{eff}, which accounts for the energy stored in the evanescent fields. Also, for two-dimensional guides in which the lateral width is much wider than the transverse width, we get a similar result.

The time-averaged power propagating in the z-direction is given by

$$P_z = \frac{1}{2}\mathrm{Re}\iint (\boldsymbol{\mathscr{E}} \times \boldsymbol{\mathscr{H}}^*) \cdot \hat{\mathbf{e}}_z \, dx \, dy. \tag{6.6}$$

The electric field, $\boldsymbol{\mathscr{E}}$, is given by Eq. (6.1), and the magnetic field, $\boldsymbol{\mathscr{H}}$, is obtained from the curl of $\boldsymbol{\mathscr{E}}$. We initially consider the TE modes with the electric fields aligned along the x-direction. Because only $\mathscr{E}_x$ and $\mathscr{H}_y$ survive the vector products, and $\mathscr{H}_y = (j/\omega\mu)d\mathscr{E}_x/dz$, we have

$$P_z = \frac{\beta}{2\omega\mu}\iint |\mathscr{E}_x|^2 \, dx \, dy = \frac{1}{2\eta_g}\iint |\mathscr{E}_x|^2 \, dx \, dy, \tag{6.7}$$

where $\eta_g = \omega\mu/\beta = 377 \ \Omega/\bar{n}$ is the waveguide impedance, and

$$\mathscr{E}_x = E_{0m} U_{xm}(x, y) e^{j(\omega t - \beta z)}. \tag{6.8}$$

The form of the transverse mode shape, $U(x, y)$, for the even (symmetric) modes is given by Eqs. (A3.17), (A3.21), and (A3.22) for a rectangular guide cross section.

For simplicity we wish to look at a guide with a lateral width much wider than the transverse width or $w \gg d$. For the lateral dimension we consider only two extreme cases: (a) a fundamental lateral mode or (b) a uniform field. In both cases because w is large, we neglect any fields for $|y| > w$. Thus, the lateral integration gives in case (a),

$$\int_{-\infty}^{\infty} |U(x, y)|^2 dy = |U(x)|^2 \int_{-w/2}^{w/2} \cos^2\left(\frac{\pi}{2w} y\right) dy = |U(x)|^2 \frac{w}{2}, \tag{6.9}$$

(fundamental lateral mode)

and in case (b),

$$\int_{-\infty}^{\infty} |U(x, y)|^2 dy = |U(x)|^2 \int_{-w/2}^{w/2} dy = |U(x)|^2 w. \tag{6.10}$$

(uniform lateral field)

Thus, the lateral integration adds only a multiplication by the width in the case of uniform fields or the half-width for the cosinusoidal lateral mode.

Now, we can plug *U(x)* from Eqs. (A3.17) and (A3.21) into Eq. (6.7) and perform the integration. For the even modes of a symmetric waveguide,

$$P_z = \frac{w/\rho}{2\eta_g} |E_0 U_0|^2 \left[2\int_0^{d/2} \cos^2(k_{1x} x) \, dx + 2\int_{d/2}^{\infty} \cos^2(k_{1x} d/2) e^{-2\gamma_x (x - d/2)} \, dx \right], \tag{6.11}$$

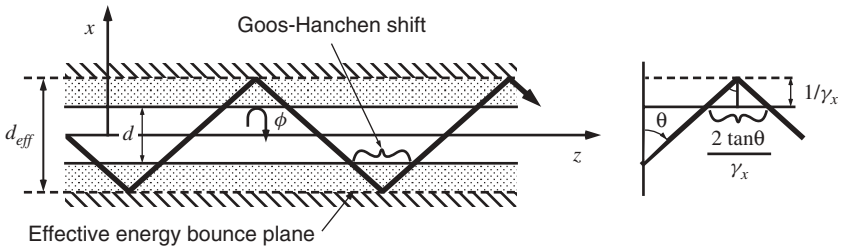


FIGURE 6.1: Zigzag ray picture of energy propagation in a three layer waveguide.

where $\rho = 2$ or 1 for lateral waveguide cases (a) or (b), respectively. Using the waveguide characteristic equation, Eq. (7.17) or (A3.9), to simplify, it can be shown that

$$P_z = \frac{w/\rho}{2\eta_g}|E_0U_0|^2\frac{d_{eff}}{2}, \tag{6.12}$$

where

$$d_{eff}^{TE} = d + \frac{2}{\gamma_x}. \quad \text{(symmetric guide)} \tag{6.13}$$

For the odd modes, it can be shown that the result is the same. Also, if we had considered the more general asymmetric waveguide, we would have obtained

$$d_{eff}^{TE} = d + \frac{1}{\gamma_{2x}} + \frac{1}{\gamma_{3x}}. \quad \text{(asymmetric guide)} \tag{6.14}$$

In all cases, we see that the power propagating down the waveguide is proportional to the magnitude squared of the peak field amplitude in the guide, $|E_0U_0|^2$, and an effective width, d_{eff}, which equals the slab thickness plus the decay lengths of the evanescent fields in the cladding regions. This is very useful in trying to approximate integrals involving the transverse mode, such as the perturbation integrals of the previous chapter. It is also interesting to note that the integrals in Eq. (6.11) contribute a factor $d_{eff}/2$, just as if the transverse field were a half-period of a cosine that goes to zero at $|x| = d_{eff}/2$. However, such a cosine would have a slightly longer transverse period than the actual cosine of the guide.

If we repeat the above for the TM modes we would find that

$$d_{eff}^{TM} = d + \frac{\vartheta_2}{\gamma_{2x}} + \frac{\vartheta_3}{\gamma_{3x}}, \tag{6.15}$$

where

$$\vartheta_i = \left[\frac{k_{1x}^2 + \gamma_{ix}^2}{k_{1x}^2 + (\varepsilon_1/\varepsilon_i)^2\gamma_{ix}^2}\right]\frac{\varepsilon_1}{\varepsilon_i}.$$

In addition to the total power, it is useful to know the relative power contained in the guide and cladding layers, represented by the two integrals in Eq. (6.11). We can write the terms corresponding to each integral conveniently for the fundamental TE mode of a symmetric guide by using the normalized frequency, $V = d\sqrt{k_x^2 + \gamma_x^2}$, introduced in Appendix 3. With Eq. (7.17), we can set $\cos(k_x d/2) = k_x d/V$ and $\sin(k_x d/2) = \gamma_x d/V$ in the evaluated integrals to obtain

$$d_{eff} = \left[d + \frac{2\gamma_x d^2}{V^2}\right] + \left[\frac{2}{\gamma_x} - \frac{2\gamma_x d^2}{V^2}\right]. \tag{6.16}$$

The first set of brackets gives the relative power in the guide and the second set gives the relative power in the cladding layers. Eq. (A3.14) gives the resulting fundamental TE mode confinement factor (aside from an $n_1/\bar{n}$ factor).

For weak guiding either through a small index difference or a narrow guide width (characterized by a small V), the tangent function in Eq. (A3.9) reduces to its argument from which we can derive $\gamma_x d \to \sqrt{1 + V^2} - 1 \approx V^2/2$. In this weak-guiding limit, we can use Eq. (6.13) to show that $\Gamma_x \to 2d/d_{eff}$. This result can also be obtained by considering a thin region of width d_A, which only samples the peak field of the guide. The power contained in this region is $\propto |E_0 U_0|^2 d_A$. In view of Eq. (6.12), the fraction of power is then $\Gamma_x = d_A/(d_{eff}/2) = 2d_A/d_{eff}$. This result applies equally to a quantum well of width d_A placed at the peak of a mode in an SCH waveguide or to a weakly guided mode in which the guide itself only samples the peak field, for which we can set $d_A = d$. For the latter case, as the guiding becomes stronger, the prefactor of 2 approaches 1 as both Γ_x and d/d_{eff} converge toward unity. In general, for strong guiding, $d/d_{eff} < \Gamma_x < 1$, and $\Gamma_x \neq d/d_{eff}$ until both are unity. Thus, the d_{eff} defined in Chapter 2 (see Section 2.4) does not refer to this effective width, but the equivalent width of a mode with constant cross-section that is by definition equal to d_A/Γ_x (where d_A is the active region width).

6.3 PERTURBATION THEORY

Many real waveguide structures involve a slight perturbation from a mathematically more simple structure, for which the eigenmode shapes, U_m, and propagation constants, β_m, are known. The perturbation can usually be expressed in terms of a change in relative dielectric constant, $\Delta\varepsilon$, which for generality can be complex and/or periodic along z.

If we replace ε by $\varepsilon + \Delta\varepsilon$ in Eq. (6.4), we most generally must assume that Eq. (6.3) must be used to represent the perturbed field. However, first we consider sufficiently weak perturbations where scattering to other modes can be neglected. This might occur most easily in a single-mode guide. Also, we assume $\Delta\varepsilon$ is uniform along z for this first example. Thus, in response to $\varepsilon \to \varepsilon + \Delta\varepsilon$, let $\beta \to \beta + \Delta\beta$, and $U \to U + \Delta U$ in Eq. (6.5). That is,

$$\nabla_T^2(U + \Delta U) + [(\varepsilon + \Delta\varepsilon)k_0^2 - (\beta + \Delta\beta)^2](U + \Delta U) = 0. \tag{6.17}$$

Multiplying this out, and dropping the unperturbed transverse wave equation, which equals zero, and the second-order perturbation terms, we are left with

$$\nabla_T^2 \Delta U + \varepsilon k_0^2 \Delta U + \Delta\varepsilon k_0^2 U - 2\beta\Delta\beta U - \beta^2 \Delta U = 0. \tag{6.18}$$

Now, we introduce a technique that we shall use again and again in this chapter to simplify complex expressions:

Multiply by the complex conjugate of the transverse mode and integrate over the cross section.

In later cases, modal orthogonality will help to remove many of the unwanted terms. In this case, we obtain

$$2\beta\Delta\beta \int |U|^2 \, dA = \int \Delta\varepsilon k_0^2 |U|^2 \, dA + \int [(\nabla_T^2 \Delta U) U^* + \varepsilon k_0^2 \Delta U U^* - \beta^2 \Delta U U^*] \, dA. \tag{6.19}$$

As shown in Appendix 14, the second term on the right is negligible, provided that both ΔU and U vanish at infinity and that the unperturbed ε and β are mostly real. This indicates that the small change in the transverse mode shape, ΔU, due to $\Delta\varepsilon$ will have *no effect* on the propagation constant to first order. Thus, we fortunately do not have to worry about how the changing mode shape will change the averaging, at least for this first-order approximation, and we only need know the original unperturbed transverse mode.

Then, solving for $\Delta\beta$, we have the desired perturbation formula

$$\Delta\beta = \frac{\int \Delta\varepsilon k_0^2 |U|^2 \, dA}{2\beta \int |U|^2 \, dA}. \tag{6.20}$$

If U is normalized according to Eq. (6.2), the denominator integral is just unity. Usually, the index perturbation is limited in lateral extent, and it may even be constant over some range. In these cases the integration is easily performed. It is important to keep in mind that all quantities, except for the actual perturbation, $\Delta\varepsilon$, are for the original unperturbed problem. Next we consider the example of adding a quantum-well active region to a simple three-layer slab waveguide to form a more complex SCH-QW laser.

Example 6.1 The addition of a quantum well to a simple three-layer slab guide is an example of a relatively easy problem to solve using perturbation theory. This approach gives a quick and familiar result for a relatively complex, and certainly technologically important, problem.

Problem: Applying the perturbation theory, calculate the expressions for the modal gain and effective index change caused by insertion of a quantum well layer of thickness d_A.

Solution: Figure 6.2 schematically shows the problem at hand. The solution using Eq. (6.20) requires that we know only the unperturbed transverse eigenmode, $U(x, y)$, and propagation constant, β, in addition to the actual perturbation, $\Delta\tilde{\varepsilon} = \Delta\varepsilon_r + j\,\Delta\varepsilon_i$, which in this case is complex. The integration is limited to the area of the active region, A, since $\Delta\tilde{\varepsilon}$ is zero elsewhere. Starting from the definition of $\varepsilon = (n + jgc/2\omega)^2$, we can expand $\Delta\tilde{\varepsilon}$ in terms of its real index and gain components in the active region. Then, expanding $\Delta\beta$ in terms of the real effective index and modal gain perturbations, we have

$$\Delta\beta = \frac{2\pi\,\Delta\bar{n}}{\lambda} + j\,\frac{\langle g\rangle_{xy}}{2} = \frac{k_0^2}{2\beta}\,\frac{2n_{av}\displaystyle\int_A \left(\Delta n_A + j\,\frac{\lambda g_A}{4\pi}\right)|U|^2\,dA}{\displaystyle\int |U|^2\,dA}, \tag{6.21}$$

where $n_{av} = (2n_A + \Delta n_A)/2$. Solving for the modal gain and effective index perturbations separately, we have

$$\langle g\rangle_{xy} = \frac{\dfrac{n_{av}g_A}{\bar{n}}\displaystyle\int_A |U|^2\,dA}{\displaystyle\int |U|^2\,dA} = \Gamma_{xy}g_A, \tag{6.22}$$

and

$$\Delta\bar{n} = \frac{\dfrac{n_{av}\Delta n_A}{\bar{n}}\displaystyle\int_A |U|^2\,dA}{\displaystyle\int |U|^2\,dA} = \Gamma_{xy}\Delta n_A, \tag{6.23}$$

where we have assumed that the gain and index perturbations are constant over the active region at g_A and Δn_A, respectively. For U properly normalized, the integral in the denominator equals one.

With reference to Appendix 5 and Chapter 2, we see that Eqs. (6.22) and (6.23) are the same expressions as derived there for the modal gain and effective index increase for such a waveguide configuration.

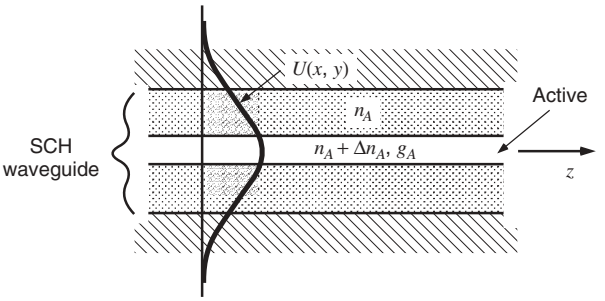


FIGURE 6.2: Quantum-well laser schematic wherein the active layer is considered to be the perturbation.

6.4 COUPLED-MODE THEORY: TWO-MODE COUPLING

By introducing index and loss perturbations along a waveguide, it is possible to facilitate a coupling between basis modes, which, of course, are orthogonal without any such perturbation. Considering the basis set of an isolated uniform waveguide, the perturbation usually has to have some periodicity along the propagation direction to cause such a coupling, since the propagation constants of the two uncoupled modes are different. An important practical example is a grating, which has a period such that scattering from one waveguide mode couples coherently to another counterpropagating mode.

On the other hand, if two identical uniform guides are brought together so that their unperturbed degenerate modes overlap, the existence of the index of one guide will perturb the mode of the other such that coherent coupling results. This is called a directional coupler, since only modes traveling in the same direction with about the same velocity can couple. More generally, periodic index perturbations can be added to coupled waveguides to provide more complex couplers. In what follows, we shall first consider simple gratings in a single waveguide, then the coupling of uniform waveguides, and finally, the so-called grating-assisted coupling between two waveguides.

6.4.1 Contradirectional Coupling: Gratings

General Theory Initially, we consider the coupling of two identical modes of the same waveguide propagating in opposite directions. Such coupling is obviously important in reflecting a particular mode within a laser cavity, for example. With this assumption and limiting the discussion to index-guided structures where the phase fronts are perpendicular to the axis of the guide, we can express the net field as the sum of the two identical counterpropagating modes. That is, using Eqs. (6.1) and (6.3)

$$\mathscr{E}(x, y, z, t) = \hat{\mathbf{e}}_i U(x, y)[E_f e^{j(\omega t - \beta z)} + E_b e^{j(\omega t + \beta z)}], \tag{6.24}$$

where E_f and E_b give the amplitudes of the two counterpropagating modes.

Now, we plug Eq. (6.24) into the wave Eq. (6.4) with an assumed perturbation, $\varepsilon \to \varepsilon + \Delta\varepsilon(z)$. We let $\Delta\varepsilon$ be z-dependent, since we suspect that this will be necessary to provide the desired coupling. If there is coupling, then the amplitudes E_f and E_b will vary with z, so the full wave Eq. (6.4) is needed. However, if we assume that the coupling will be weak, the amplitudes should vary relatively slowly. That is, we will be able to drop terms containing their second derivatives in the expansion of the wave equation if the coefficients are not large. Also, for generality we allow a small ΔU, although we suspect that it will not show up in the final result, in analogy with the perturbation result of Section 14.2.

Observing that the unperturbed ε and U will generate terms that satisfy the unperturbed wave equation and dropping the second-order terms in the perturbations as well as the second derivatives, we are left with

$$\begin{aligned} -2j\beta(U+\Delta U)\frac{dE_f}{dz}e^{-j\beta z} + 2j\beta(U+\Delta U)\frac{dE_b}{dz}e^{j\beta z} \\ = -[\nabla_T^2\Delta U + (\varepsilon k_0^2 - \beta^2)\Delta U][E_f e^{-j\beta z} + E_b e^{j\beta z}] \\ - \Delta\varepsilon(z)k_0^2 U[E_f e^{-j\beta z} + E_b e^{j\beta z}]. \end{aligned} \tag{6.25}$$

We now dot multiply by U^* and integrate over the cross section. Then, we see that the first term on the right is zero according to Appendix 14, as was the case in Eq. (6.19) above. Next, we recognize that the terms containing $\Delta U\,(dE_i/dz)$ are really second-order terms and are negligible in comparison to the others. Finally, we divide by $-2j\beta\int|U|^2\,dA$, to get

$$\frac{dE_f}{dz}e^{-j\beta z} - \frac{dE_b}{dz}e^{j\beta z} = -j\frac{k_0^2}{2\beta}[E_f e^{-j\beta z} + E_b e^{j\beta z}]\frac{\int\Delta\varepsilon(z)|U|^2\,dA}{\int|U|^2\,dA}. \tag{6.26}$$

Note that if $\Delta\varepsilon(z)$ were uniform along z, we would be left simply with the perturbation formulas for both forward and backward traveling waves independent of each other, since $dE_f/dz = -j\Delta\beta E_f$, and $dE_b/dz = j\Delta\beta E_b$, in that case. Any DC component in $\Delta\varepsilon(z)$ has exactly the predicted effect of uniformly changing the propagation constants by $\Delta\beta$. However, to couple forward to backward waves and vice versa, it can be seen that $\Delta\varepsilon(z)$ must contain factors exp $(\pm 2j\beta z)$.

More generally, we allow $\Delta\varepsilon(z)$ be a complex spatially periodic function with possibly a nonuniform cross section. Thus, Fourier analysis can be applied to give

$$\Delta\varepsilon(x,y,z) = \sum_{l\neq 0}\Delta\varepsilon_l(x,y)e^{-jl(2\pi/\Lambda)z}, \tag{6.27}$$

where $2\pi l/\Lambda$ are the various space harmonics of the arbitrary periodic perturbation of fundamental period Λ. Using this in Eq. (6.26), and looking for possible solutions that couple forward to backward waves and vice versa, we identify terms that could provide similar exponential factors on each side of the equation for all z. That is,

$$\frac{dE_f}{dz}e^{-j\beta z} = -j\frac{k_0^2}{2\beta}E_b e^{j\beta z}e^{-jl(2\pi/\Lambda)z}\frac{\int\Delta\varepsilon_{+l}(x,y)|U|^2\,dA}{\int|U|^2\,dA}, \tag{6.28}$$

and the comparable equation for dE_b/dz from the remaining terms,

$$\frac{dE_b}{dz}e^{j\beta z} = j\frac{k_0^2}{2\beta}E_f e^{-j\beta z}e^{jl(2\pi/\Lambda)z}\frac{\int\Delta\varepsilon_{-l}(x,y)|U|^2\,dA}{\int|U|^2\,dA}. \tag{6.29}$$

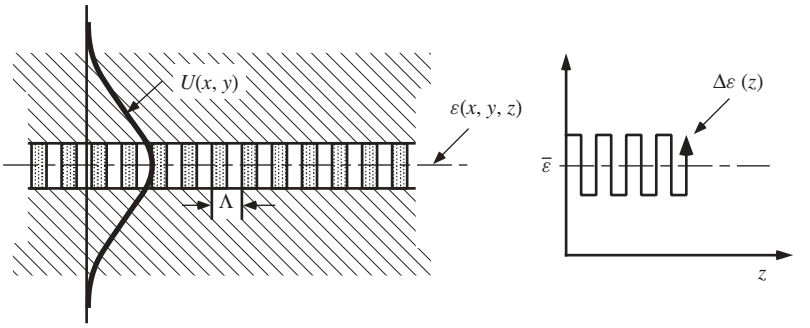


FIGURE 6.3: Waveguide containing index grating for contradirectional coupling. The relative dielectric constant (square of index of refraction) varies periodically along z about some average value, $\bar{\varepsilon}$, so that reflected components add in phase.

From both Eqs. (6.28) and (6.29) we see that the condition,

$$l\frac{2\pi}{\Lambda} \approx 2\beta = \frac{4\pi\bar{n}}{\lambda}, \tag{6.30}$$

must be met for dE_f/dz or dE_b/dz to have the same sign over some distance, or put another way, for coherent phasing of the coupling. All other terms can be neglected, since we are only interested in the net effect on E_f and E_b over a distance of several wavelengths or more. That is, the differential Eqs. (6.28) and (6.29) would not provide for any net change in E_f and E_b versus z unless Eq. (6.30) is satisfied. Eq. (6.30) indicates that for a good reflective grating, the period of the index perturbation, Λ, should be some multiple, l, of $\lambda/2\bar{n}$. This is called the Bragg condition. Figure 6.3 illustrates a waveguide containing an index grating for forward to backward wave coupling.

Before continuing to find solutions to Eqs. (6.28) and (6.29), it is convenient to make some substitutions. As already suggested, the propagation constant at the Bragg condition is given by Eq. (6.30) with $\equiv$ replacing $\approx$. More specifically, we let $\beta_0 \equiv l\pi/\Lambda$, which also defines the Bragg wavelength, $\lambda_0/\bar{n} \equiv 2\Lambda/l$ (note that higher orders of l allow us to use a longer grating period for the same λ_0). In addition, we let

$$\kappa_{\pm l} = \frac{k_0^2}{2\beta}\frac{\int \Delta\varepsilon_{\pm l}(x,y)|U|^2\,dA}{\int |U|^2\,dA}. \tag{6.31}$$

As before, if U is normalized according to Eq. (6.2), the denominator drops out. It is also important to remember that $\Delta\varepsilon_{\pm l}$ are Fourier coefficients as given by Eq. (6.27), not the actual amplitude of the index perturbation.

Example 6.2 A cosinusoidal grating perturbation, whose index varies as $\delta n_{gd}\cos(2\pi z/\Lambda - \phi)$, was introduced into a uniform slab waveguide, whose refractive index is n_{gd}.

Problem: Determine the coupling coefficients $\kappa_{\pm 1}$ for this type of perturbation.

Solution: With a given perturbation, two terms will result from Eq. (6.27) for $l = \pm 1$, each with a magnitude, $\Delta\varepsilon_{l=\pm 1} = \delta\varepsilon/2$. That is, we can expand the cosine to find

$$\begin{aligned}\Delta\varepsilon(x,y,z) &= \frac{\delta\varepsilon(x,y)}{2}e^{-j\phi}e^{j(2\pi/\Lambda)z} + \frac{\delta\varepsilon(x,y)}{2}e^{j\phi}e^{-j(2\pi/\Lambda)z} \\ &= \Delta\varepsilon_{-1}(x,y)e^{j(2\pi/\Lambda)z} + \Delta\varepsilon_{+1}(x,y)e^{-j(2\pi/\Lambda)z}. \end{aligned} \tag{6.32}$$

To calculate the magnitude of $\delta\varepsilon$, we start from the relationship between the permitivity and the index of refraction, and compute $\delta\varepsilon = (n_{gd} + \delta n_{gd})^2 - n_{gd}^2 \approx 2 \cdot n_{gd} \cdot \delta n_{gd}$. Thus, we can use the Fourier components from Eq. (6.32) in Eq. (6.31) to write the coupling coefficient for the cosinusoidal dielectric perturbation with an arbitrary phase relative to $z = 0$. That is,

$$\kappa_{\pm 1} = \kappa e^{\pm j\phi} = \frac{n_{gd}}{\bar{n}}\frac{\pi}{\lambda}\Gamma_{xy_g}(\delta n_{gd})e^{\pm j\phi}, \tag{6.33}$$

where for the second equality, we have assumed a uniform grating perturbation over some lateral portion of the waveguide, for which Γ_{xy_g} is the lateral grating region confinement factor.

Finally, if we define the *detuning parameter* as $\delta \equiv \beta - \beta_0$, Eqs. (6.28) and (6.29) become

$$\frac{dE_f(z)}{dz} = -j\kappa_l E_b(z)e^{2j\delta z}, \tag{6.34}$$

and

$$\frac{dE_b(z)}{dz} = j\kappa_{-l}E_f(z)e^{-2j\delta z}. \tag{6.35}$$

Equations 6.34 and 6.35 are relatively clear statements of the coupling between the forward- and backward-going lightwaves. They both indicate that the change in amplitude of one wave is directly proportional to the amplitude of the other, and that the proportionality involves a coupling constant, $\kappa_{\pm l}$, which depends on the overlap of a periodic index perturbation and the modes' energy density, as well as a dephasing factor, which deviates from unity if the Bragg condition is not met. It is also worth noting from Eq. (6.31) or Eq. (6.33) that $\kappa_{-l} = \kappa_l^*$ for a real $\Delta\varepsilon(z)$ and β, and that this condition is still approximately true for waveguides with small gains or losses (i.e., β predominately real), as is usually the case.

So now all we need to do is solve the coupled Eqs. (6.34) and (6.35). Unfortunately, these equations are still not in a very convenient form for mathematical solution. The solution for the net electric field along the guide is given by Eq. (6.24) as a linear superposition of the forward and backward fields. We also

can express this net field in terms of two new z-varying factors, $A(z) = E_f(z)e^{-j\delta z}$, and $B(z) = E_b(z)e^{j\delta z}$, so that we have,

$$\mathscr{E}(x, y, z) = \hat{\mathbf{e}}_i U[E_f(z)e^{-j\beta z} + E_b(z)e^{j\beta z}] = \hat{\mathbf{e}}_i U[A(z)e^{-j\beta_0 z} + B(z)e^{j\beta_0 z}]. \tag{6.36}$$

The latter form combines both coupling and dephasing effects within the z-varying coefficients, A and B, since $\beta_0 = \pi l/\Lambda$ is a constant for a given grating.

Expressing $E_f(z)$ and $E_b(z)$ in Eqs. (6.34) and (6.35) in terms of $A(z)$ and $B(z)$, we obtain

$$\frac{dA}{dz} = -j\kappa_l B - j\delta A, \tag{6.37}$$

and

$$\frac{dB}{dz} = j\kappa_{-l} A + j\delta B. \tag{6.38}$$

These coupled differential equations can now be straightforwardly solved in the usual way by letting

$$A = A_1 e^{-\sigma z} + A_2 e^{\sigma z}, \tag{6.39}$$

and

$$B = B_1 e^{-\sigma z} + B_2 e^{\sigma z}. \tag{6.40}$$

Plugging Eqs. (6.39) and (6.40) into Eqs. (6.37) and (6.38), and separately equating the coefficients of $e^{\sigma z}$ and $e^{-\sigma z}$ (since the intermediate expressions must be true for all z) leads to relationships between A_1 and B_1, and A_2 and B_2, e.g.,

$$B_1 = \frac{j\kappa_{-l} A_1}{-\sigma - j\delta} \quad \text{and} \quad B_2 = \frac{j\kappa_{-l} A_2}{\sigma - j\delta}, \tag{6.41}$$

as well as a solution for σ given by

$$\sigma^2 = \kappa_l \kappa_{-l} - \delta^2 = \kappa^2 - \delta^2, \tag{6.42}$$

where for the last equality we have made use of Eq. (6.33).

For waveguides that contain gain or loss, β is complex, and therefore, so is the detuning parameter, δ. Thus, following previous nomenclature, we also have

$$\tilde{\sigma}^2 = \kappa^2 - \tilde{\delta}^2, \tag{6.43}$$

where

$$\tilde{\delta} = \tilde{\beta} - \beta_0 = \frac{2\pi\bar{n}}{\lambda} + j\frac{\langle g\rangle_{xy} - \alpha_i}{2} - \frac{\pi l}{\Lambda}. \tag{6.44}$$

In cases where $\Delta\varepsilon(z)$ is complex (i.e., where there is a significant periodic gain or loss), κ is complex, and $\kappa_{-l} \neq \kappa_l^*$. However, Eq. (6.43) is still valid since this condition has not been used in its derivation. For simultaneous cosinusoidal index and gain perturbations over some region of the waveguide cross section, the derivation of Eq. (6.33) yields

$$\kappa = \frac{n_{gd}}{\bar{n}}\frac{\pi}{\lambda}\Gamma_{xy_g}\left(\delta n_{gd} + j\frac{\delta g_g}{2k_0}\right), \tag{6.45}$$

where δg_g is the magnitude of the cosinusoidal gain perturbation in the grating region. Thus, Eq. (6.45) can be used in Eq. (6.43) and elsewhere for this case.

Using Eqs. (6.39) and (6.40) with $\sigma = j\sqrt{\delta^2 - \kappa^2}$ in the expression for the propagating electric field, Eq. (6.36), we observe four combinations of phase factors, one for each of the A and B coefficients:

$$j\beta_g = \pm j(\beta_0 \pm \sqrt{\delta^2 - \kappa^2}). \tag{6.46}$$

The propagation constant of the grating, β_g, therefore, has four possible solutions at any given frequency, ω, or equivalently, $\delta = \beta - \beta_0 = (\omega - \omega_0)/v_g$. Far away

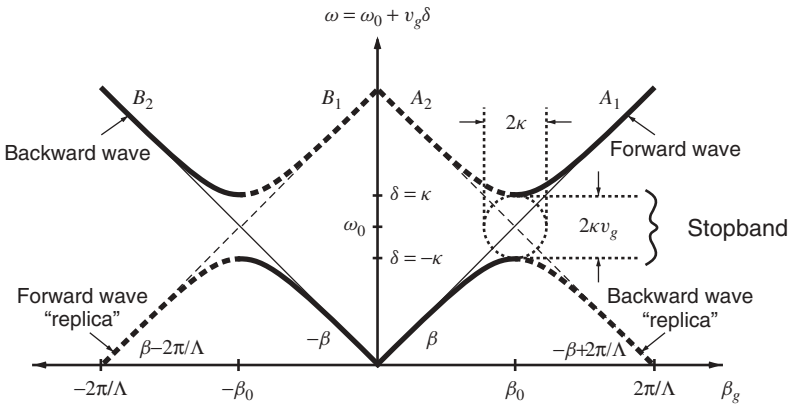


FIGURE 6.4: An $\omega - \beta$ diagram for the coupled β_g-solutions to contradirectional coupling in a grating, with the uncoupled solutions denoted by thinner straight lines. Each of the grating-generated replica solutions (dashed) and ordinary forward and backward wave solutions (solid) correspond to one of the A or B coefficients indicated. The extent of the stopband in both ω and β directions is also shown (where v_g is the group velocity of the unperturbed mode). However, the complex $\tilde{\beta}_g$-solutions which exist throughout the stopband are not shown. Finally, the scale of the stopband has been exaggerated somewhat, considering that $\kappa \ll \pi/\Lambda$ to satisfy the weak coupling criterion.

from the Bragg frequency ($|\delta| \gg \kappa$), these solutions are $\pm\beta$ and $\pm(\beta - 2\beta_0)$, which as shown in Fig. 6.4, are (a) the unperturbed forward and backward wave propagation constants, $\pm\beta$, and (b) grating-induced replicas of these propagation constants displaced by $\mp 2\pi/\Lambda$ (for $l = 1$). Away from the Bragg condition, the two sets of solutions do not interact and the latter solutions never contain much energy ($A_2, B_1 \approx 0$).

At the Bragg frequency ($\delta = 0, \omega_0 = (c/\bar{n})(\pi/\Lambda)$), the solutions are $\pm(\beta_0 \pm j\kappa)$. That is, all field solutions propagate at $\pm\beta_0$, but grow or decay at the rate κ as they travel due to the coherent transfer of energy from one solution to the other. In this case, the solutions interact heavily, and all four A and B coefficients can have significant amplitude. Detuning to $|\delta| = \kappa$, or $\omega = \omega_0 \pm v_g\kappa$, the propagation constants become purely real and equal to $\pm\beta_0$, or $\pm\pi/\Lambda$. Detuning further, the solution split and approach their uncoupled values asymptotically. Thus, as illustrated in the figure, the coupling between the solutions creates a "stopband" for $|\delta| < \kappa$. Within this stopband, the solutions have complex propagation constants, the real parts remaining fixed at $\pm\beta_0$. It is in this regime where an incident field can decay in the grating and eventually be reflected back out (hence the name *stop*band).

Finite-Length Gratings with no Back Reflections Now, we are finally ready to look at some practical characteristics of a real physical problem. For example, consider the grating of finite length, L_g, in Fig. 6.5. We assume a forward ($+z$-traveling) wave incident from the left that enters the grating at $z = 0$ and divides into a net reflected backward ($-z$-traveling) wave emerging at $z = 0$, as well as a transmitted forward wave at $z = L_g$. Of course, if the grating contains gain or loss the sum of the transmitted and reflected powers will not equal the incident

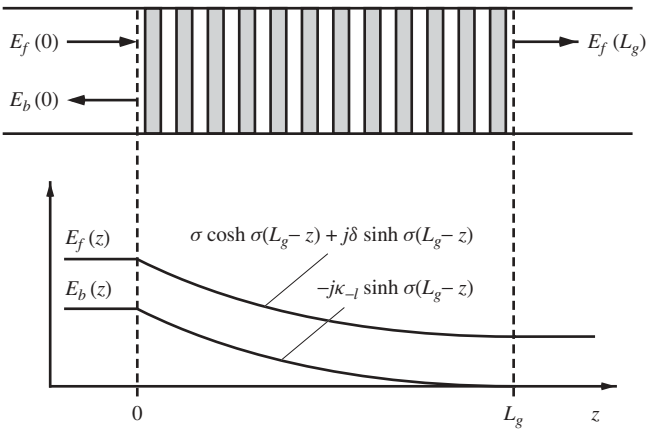


FIGURE 6.5: Grating of finite length. The forward wave decays as the backward wave grows. The backward wave is zero at $z = L_g$, and a portion of the forward wave is transmitted. The functionality of $A(z)$ and $B(z)$ within the grating is also shown. The shaded areas indicate regions of lower index.

power. Within the grating the forward wave decays as more and more energy is coupled to the backward wave. The backward wave grows as it propagates to the left. We also assume no reflections beyond the grating, so the backward wave has no amplitude at $z = L_g$.

For this case of no reflection from the right side of the grating, we have the boundary condition, $E_b(z = L_g) = 0$, which requires that $B(z = L_g) = 0$. The ratio of the amplitudes of the net reflected wave to the incident wave is defined as the grating reflection coefficient,

$$r_g = \frac{E_b(0)}{E_f(0)} = \frac{B(0)}{A(0)} = \frac{B_1 + B_2}{A_1 + A_2}. \tag{6.47}$$

From the right-hand boundary condition at $z = L_g$ and Eq. (6.40), $B_2 = -B_1 e^{-2\sigma L_g}$. Using this and Eq. (6.41), we obtain

$$\tilde{r}_g = -j \frac{\tilde{\kappa}_{-l} \tanh \tilde{\sigma} L_g}{\tilde{\sigma} + j\tilde{\delta} \tanh \tilde{\sigma} L_g}. \tag{6.48}$$

The tildes have been added to explicitly indicate that the coupling constant, κ, detuning parameter, δ, and decay constant, σ, can be complex. The reflectivity from the other side of the grating is given by Eq. (6.48) times $e^{-2j\beta_0 L_g}$, with $\tilde{\kappa}_l$ replacing $\tilde{\kappa}_{-l}$ (the phase factor arises from the definitions of A and B in Eq. (6.36)).

Implicit in the derivation is the assumption of a reference plane at $z = 0$. If the grating is unshifted (i.e., $\phi = 0$ in Eq. (6.33)) and the index perturbation is real then, κ_{-l} is real, and we have a maximum positive deviation at $z = 0$ (i.e., a cosinusoidal space harmonic in $\Delta\varepsilon$.) (In this case, the grating has real, positive,

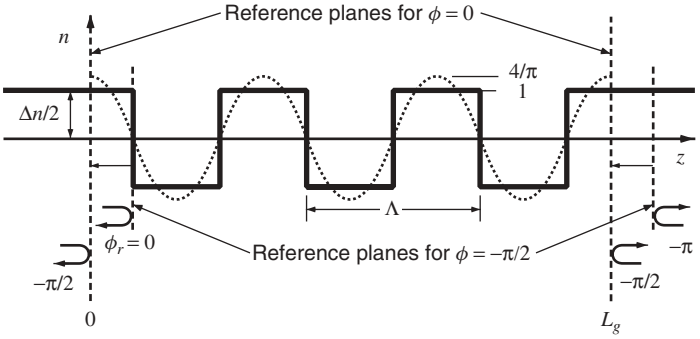


FIGURE 6.6: Illustration of reference planes used to define $z = 0$ and $z = L_g$ in a finite-length grating. The coupled-mode equations assume the index profile has a Fourier component that varies as $\cos(2\beta_0 z - \phi)$. The first set of reference planes are drawn for $\phi = -\pi/2$, which aligns the $z = 0$ plane with an index down step of a square wave grating (these are the reference planes assumed in Chapter 3). The second set shifts the planes to the left for a more symmetric placement by setting $\phi = 0$.

and symmetrical Fourier coefficients.) Equation (6.48) shows that at the Bragg frequency, the reflection coefficient has a phase of $-\pi/2$ referenced to this plane. Figure 6.6 illustrates this choice of reference planes and its correspondence to a square wave index grating, for which the reflection phase is more obvious. That is, at the Bragg frequency, where all reflection components add in phase, we know that the reflection phase from a square wave grating must be zero referenced to an index down step. The first Fourier component of this grating has a zero crossing at this point and is a maximum one quarter-wave back. Thus, for a reference plane placed at this maximum of the first Fourier component, the grating phase would be $-\beta\Lambda/2 = -\pi/2$.

More generally, the reflection phase, ϕ_r, of waves incident from the left is determined by $-j\kappa_{-l} \propto -je^{-j\phi}$ at the $z = 0$ plane, for a lossless grating at the Bragg frequency. For waves incident from the right, ϕ_r is determined by $-j\kappa_l e^{-2j\beta_0 L_g} \propto -je^{j\phi}e^{-2j\beta_0 L_g}$ at the $z = L_g$ plane. These reflection phases are shown in Fig. 6.6 for two different choices of the grating phase, ϕ, and a grating with $L_g = m\Lambda$. When combining a grating with other elements, the symmetric reference planes are a convenient choice. To use them, we adjust the input plane to a cosine peak by setting $\phi = 0$. However if $L_g \neq m\Lambda$, the output plane will not coincide with a cosine peak if it is placed at L_g. To retain the symmetry, we must shift the output plane slightly to create a reference plane separation of $L_r = m\Lambda$. This is accomplished by adding a phase delay, $e^{-2j\beta(L_r - L_g)}$, to the reflection phase from the right. The total reflection phase from the right then becomes $-je^{j\phi}e^{-2j\beta_0 L_r}e^{-2j\delta(L_r - L_g)} \approx -je^{j\phi}$, where the approximation holds for $L_r \sim L_g$, or $\delta = 0$. In practice then, for any grating length, the reflection phase at the Bragg frequency can be set to $-\pi/2$ from both sides of the grating. Away from the Bragg frequency and/or with gain or loss, the reflection phases remain symmetric (for $L_r \sim L_g$), but are no longer exactly equal to $-\pi/2$.

Although our choice of reference planes is arbitrary, it is important to realize that our choice of L_g is not, since it determines the reflectivity magnitude. Typically with in-plane lasers, L_g corresponds to the physical length of the grating (i.e., where the uniform index stops to where the uniform index begins again). However, this may not always be the case. For example, in Figs. 6.5 and 6.6 (and the gratings analyzed in Chapter 3), L_g is $\Lambda/2$ longer than the physical length of the grating. The reason for this has to do with how the index outside the grating compares with the *average* index inside the grating. If there is a mismatch, additional reflections are created at the boundaries that can either enhance or reduce the overall grating reflectivity. In coupled-mode theory, this is approximately taken into account by modifying L_g. In Fig. 6.6, each $\Lambda/4$ segment contributes a $\Delta n/2$ step. Thus, adding a mismatch of $\Delta n/2$ at a physical boundary should be roughly equivalent to increasing L_g by $\Lambda/4$, as long as the mismatch reflection adds in phase with the grating reflections. In Figs. 6.5 and 6.6, the index on either side of the grating is $\Delta n/2$ larger than the average grating index. The two $\Delta n/2$ mismatches therefore effectively increase L_g by $\Lambda/2$. In the case of long gratings and small mismatches typically encountered with in-plane lasers, such subtle modifications to L_g are not important to consider.

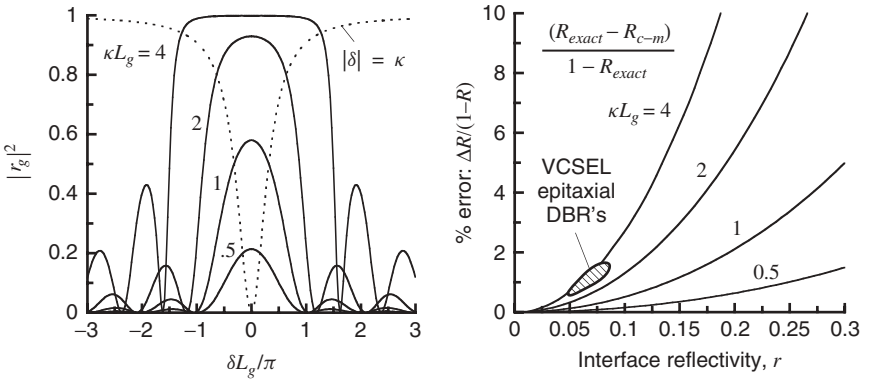


FIGURE 6.7: The left plot shows the power reflection spectrum of the grating for four different values of κL_g versus the detuning parameter, $\delta \equiv \beta - \beta_0 = (\omega - \omega_0)/\omega_0 \cdot (\pi/\Lambda)(\bar{n}_g/\bar{n})$, where $\Lambda = \lambda_0/2\bar{n}$. Exact and coupled-mode results are indistinguishable on this scale. The right plot shows the percent error in transmission at resonance calculated using the coupled-mode formula. For example, if the exact result predicts $R_g = 99\%$, then a 10% transmission error would correspond to a coupled-mode prediction of $R_g = 98.9\%$.

The power reflection spectrum of the grating, given by the absolute square of Eq. (6.48), is plotted on the left in Fig. 6.7 for different values of κL_g. The curves have similar characteristics to those of Fig. 3.14, with a sin *x*/*x* spectrum for small κL_g and the development of a well-defined transmission stopband for large κL_g. The dashed curve tracks the $|\delta| = \kappa$ transition. Outside this transition point, (i.e., for $|\delta| > \kappa$), the reflectivity drops off dramatically. Thus, the transmission stopband (or reflection bandwidth) of the grating can be characterized by $\sim 2\kappa$. From the definition of the grating penetration depth given in Chapter 3, we observe that the reflection bandwidth is roughly the reciprocal of the penetration depth ($L_{eff} \approx 1/2\kappa$).

It may seem surprising that the coupled-mode formula given by Eq. (6.48), which has little resemblence to Eq. (3.34) derived from transmission matrix theory, leads to a nearly identical reflection spectrum (as revealed by the right plot in Fig. 6.7). It turns out that there is a correspondence. It is just obscured by the math. For the interested reader, Appendix 7 reveals this hidden correspondence between the two formulas.

As in Chapter 3, the reflection parameter of interest is κL_g. This can be realized in the present case by evaluating Eq. (6.48) at the Bragg frequency by using Eq. (6.43),

$$\tilde{r}_g = -j \tanh \kappa L_g, \quad (\delta = 0) \tag{6.49}$$

and noting that for small κL_g, $|r_g| \to \kappa L_g$. That is, κ can be interpreted as the reflection per unit length, just as in Chapter 3. However, unlike the result of Chapter 3 discussed in the caption of Fig. 3.14, the coupled-mode prediction is characterized only by the product κL_g, independent of whether κ is small or large. In fact, for

large perturbations (large κ, and more significantly, large reflection per half-period, $2r$, using the jargon of Chapter 3), the coupled-mode result begins to break down even for small κL_g. However, even for interface field reflectivities (i.e., $\Delta n/2n$) as high as 18% and $\kappa L_g = 4$, Fig. 6.7 reveals that the error in transmission at the center of the stopband is still less than 10%. Figure 6.7 also shows that the range of interface reflectivities and κL_g used in typical VCSEL semiconductor DBRs would lead to transmission errors of $\approx$1% if the coupled-mode formula (6.37) were used. Thus, coupled-mode theory still works surprisingly well even in this regime. However, in practice the *terminations* used with VCSEL mirrors are far from the average index of the DBRs, and this ultimately makes the coupled-mode formula impractical for such applications. In such cases where the terminations have a very different index (a DBR terminating in air, for example), the analysis at the end of Appendix 7 can be used to determine the peak reflectivity; however, the reflectivity *spectrum* must be calculated numerically using the transmission matrix techniques introduced in Chapter 3.

Other important grating relationships useful for DBR and DFB lasers may also be found in Chapter 3. For example, the reflection for small κL_g is given by Eq. (3.58), and the grating effective reflection plane spacing (or penetration depth), L_{eff}, is given by Eq. (3.63).

The transmission through the grating in Fig. 6.5 is given by

$$t_g = \frac{E_f(L_g)}{E_f(0)} e^{-j\beta L_g} = \frac{A(L_g)}{A(0)} e^{-j\beta_0 L_g}. \tag{6.50}$$

The phase factors arise from how E_f and A are defined relative to the full electric field in Eq. (6.36). Using Eq. (6.39) for $A(L_g)$, the boundary condition $B(z = L_g) = 0$, and Eq. (6.41) to relate the A_i's to the B_i's, we obtain

$$\tilde{t}_g = \frac{\tilde{\sigma}\ \text{sech}\,\tilde{\sigma}\ L_g}{\tilde{\sigma} + j\tilde{\delta}\ \tanh\ \tilde{\sigma}\ L_g} e^{-j\beta_0 L_g}. \tag{6.51}$$

The tildes have again been added to indicate the possibility of complex factors. Note that the prefactor for transmission is $\tilde{\sigma}$ instead of $\tilde{\kappa}_{-l}$ as it was for reflection. This has to be the case because the transmission should not go to zero when the coupling goes to zero. In fact, for a lossless grating with $\kappa = 0$, we see that $\sigma \to j\delta$ and $t_g \to e^{-j\beta L_g}$, which is the simple expected transmission delay between the two reference planes. Also note that because the transmission only depends on the coupling through $\tilde{\kappa}_l \tilde{\kappa}_{-l}$, the phase of the grating relative to the reference planes, ϕ, drops out (from Eq. (6.33)). So unlike the reflection phase, the transmission phase does not depend on the specific placement of the reference planes, only on the separation between. For example, at the Bragg frequency, the transmission phase is $-\pi L_g/\Lambda$, which for $L_g = m\Lambda$, reduces to $-m\pi$, regardless of where we place the reference planes relative to the grating. However, if we separate the reference planes by $L_r \neq L_g$, then the phase factor in Eq. (6.51) should be replaced with $e^{-j\beta_0 L_r} e^{-j\delta(L_r - L_g)} \approx e^{-j\beta_0 L_r}$ (the approximation is good for $L_r \sim L_g$). Finally, it should be noted that the transmission from the other side of the grating is identical to Eq. (6.51).

6.4.2 DFB Lasers

As discussed in Chapter 3, a DFB laser consists of grating mirrors that also provide gain. For a DFB laser, at threshold, the gain is sufficient to overcome internal and transmission losses and provide an output with no input. In terms of the previous section, the characteristic equation for such a laser can be obtained from the poles of r_g or t_g, and these can be obtained by setting the denominator of the grating reflection or transmission coefficient, Eqs. (6.48) or (6.40), respectively, to zero, as long as there aren't any end reflections past the grating. Thus, the characteristic equation, which determines the threshold gain and wavelength, is

$$\tilde{\sigma}_{th} = -j\tilde{\delta}_{th} \ \tanh \tilde{\sigma}_{th} L_g . \tag{6.52}$$

The symbols have been already defined, but for clarity we repeat at threshold,

$$\begin{aligned} \tilde{\sigma}_{th}^2 &= \kappa^2 - \tilde{\delta}_{th}^2 \\ \tilde{\delta}_{th} &= \tilde{\beta}_{th} - \beta_0, \\ \tilde{\beta}_{th} &= \beta_{th} + j\,\frac{\langle g\rangle_{xy_{th}} - \langle \alpha_i \rangle_{xy}}{2}, \\ \beta_{th} &= 2\pi \bar{n}/\lambda_{th} \quad \text{and} \quad \beta_0 = \pi l/\Lambda. \end{aligned} \tag{6.53}$$

Again, the fundamental quantities we are usually after are the threshold modal gain, $\langle g\rangle_{xy_{th}}$, and wavelength, λ_{th}. Thus, solutions such as those plotted in Fig. 3.26 can be obtained numerically by plugging these definitions into Eq. (6.52).

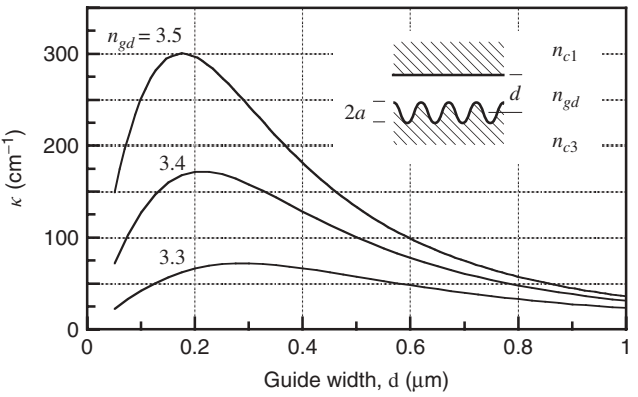


FIGURE 6.8: Coupling constant for the fundamental TE mode as a function of guide width in a three-layer waveguide for three different values of the guiding index. Parameters assumed are: $\lambda = 1.55$ μm, $n_c = 3.17$ (InP at 1.55 μm), $2a = 50$ nm, and $G = 1$ (sinusoidal first-order grating).

The coupling constant κ is calculated from Eq. (6.31) for the specific dielectric grating perturbation involved. In practical DFB lasers a number of different configurations are used. Figure 6.8 shows a popular waveguide cross section, along with calculated values for κ over a range of corrugation depths. Since the thickness of some layer is usually varied in typical semiconductor waveguide gratings rather than the index over the layer as assumed in the above calculations, it is generally most convenient to first calculate the effective index as a function of z over a period of the perturbation. This is accomplished by using Eq. (6.20) repeatedly over the perturbation period. Then, this effective index variation is used as if the index perturbation were uniform across the waveguide cross section. Finally, Fourier analysis is used to obtain the space harmonic of interest. Having obtained the magnitude of this space harmonic, $\delta\bar{n}$, we can replace δn_{gd} in Eq. (6.33) by $\delta\bar{n}$ with $\Gamma_{xy_g} = 1$ and $n_{gd} = \bar{n}$ to obtain

$$\kappa_{\pm 1} = \frac{\pi}{\lambda}\delta\bar{n} \cdot e^{\pm j\phi}. \tag{6.54}$$

For a square wave corrugation pattern, the effective index alternates abruptly between $\bar{n}_1$ and $\bar{n}_2$, implying that $\bar{n}(z)$ is also a square wave with peak-to-peak variation $\Delta\bar{n} = \bar{n}_2 - \bar{n}_1$. The fundamental cosine component of this square wave varies from peak-to-peak by $(4/\pi)\Delta\bar{n}$ which we can set equal to $2\delta\bar{n}$, yielding $\delta\bar{n} = (4/\pi)(\Delta\bar{n}/2)$ (see Fig. 6.6). Thus, for a square wave corrugation, $\kappa = 2\Delta\bar{n}/\lambda$, which is the same result we found in Chapter 3 (see discussion following Eq. (3.59)). For other corrugation patterns, $\bar{n}(z)$ and $\delta\bar{n}$ must generally be calculated numerically, making this approach to determining the coupling constant less than ideal.

The coupling constant has also been derived from a different standpoint. Instead of developing a coupling of waves, the mode is defined in terms of rays bouncing down the guide at some angle and with some effective width, which takes into account the penetration into the cladding layers (see Chapter 7). The number of bounces per unit length can easily be determined from this. Next, the grating perturbation is considered to be a diffraction grating with a plane wave incident at some angle, which at every bounce diffracts some energy backward at the same angle as the incident ray. The amount of energy diffracted per bounce is characterized by the diffraction efficiency of the grating. The advantage to this approach is that a closed-form expression for the coupling constant can be derived, which is in complete agreement with the numerical procedure outlined above. The expression for a three-layer guide, as shown in Fig. 6.8, is given by [1]

$$\kappa = Gk_0 a \frac{n_{gd}^2 - \bar{n}^2}{2\bar{n}d_{eff}}, \tag{6.55}$$

where $d_{eff} = d + \frac{1}{\gamma_{x1}} + \frac{1}{\gamma_{x3}}$, n_{gd} is the refractive index of the waveguide, and $\bar{n}$ is the effective index of the mode. The decay rates into cladding regions $i = 1$ and 3 are defined as $\gamma_{xi} = k_0\sqrt{\bar{n}^2 - n_{ci}^2}$, and d is the width of the guiding region. The

guide and cladding indices are, n_{gd} and n_{c1}, n_{c3}. The total height of the grating is $2a$, and $k_0 = 2\pi/\lambda$. Finally, G is related to the spatial Fourier coefficient of the grating pattern: (1) $G = 1$ for a sinusoidal variation, (2) $G = 4/\pi$ for a square wave pattern, (3) $G = 8/\pi^2$ for a triangular sawtooth pattern, and (4) $G = (4/\pi)$ sin x/x for a graded square wave, where $x = 0$ for no grading (perfect square wave) and $x = \pi/2$ for complete grading (perfect sawtooth pattern), and in general, $x = (\pi/2) \times$ (fraction of graded material). For $k_0 a$ approaching unity, G generally becomes more complex than listed here. If a higher-order component of the Fourier spectrum of the grating pattern is used for coupling modes, then G would have to be modified accordingly.

The coupling constant is shown in Fig. 6.8 as a function of the guide width, for a total grating height of 50 nm. We see that there is an optimum guide width, which becomes more peaked as the index step between the guide and cladding increases. By converting from a sinusoidal grating to a square wave grating, the coupling constant can be increased by $4/\pi \approx 1.27$. Also, there is a linear dependence on grating depth, implying that very deep grooves can be used to obtain large κ. In general, the curves in Fig. 6.8 can be scaled for different G and a.

The closed-form expression for the coupling constant Eq. (6.55) can only be used if the effective index of the guide is known, which in general must be found using the techniques discussed in Appendix 3. However, if the guide is symmetric ($n_c = n_{c1} = n_{c3}$), the effective index is given exactly by [2]

$$\bar{n}^2 = n_{gd}^2 b + n_c^2(1-b), \quad \text{where} \quad b \approx 1 - \frac{\ln(1 + V^2/2)}{V^2/2}. \tag{6.56}$$

In this expression, b is defined as the *field* confinement factor (i.e., the field instead of the field squared is used as the weighting function). As with Γ, we can approximate b using the normalized frequency, $V = k_0 d\sqrt{n_{gd}^2 - n_c^2}$. The approximation in Eq. (6.56) produces less than 1.2% error in the effective index over all ranges of V. In fact, using Eq. (6.56) in Eq. (6.55), we would obtain curves almost indistinguishable from those plotted in Fig. 6.8.

Example 6.3 As indicated in Chapter 3 and illustrated in Fig. 6.9, it is not uncommon to have facet reflections at the end of the grating in practical DFB lasers.

Problem: Determine the characteristic equation of a DFB laser with facet reflections.

Solution: This situation can be modeled using the scattering theory developed in Chapter 3. Since we have the reflection and transmission parameters of a symmetrical grating, we can describe the grating by a scattering matrix and include it within a Fabry–Perot resonator formed by the facet reflections.

The grating and back reflector can be combined to get an effective reflector, which together with the front reflector form a simple two-mirror resonator for which we can write a characteristic equation. First, we combine the grating and the back reflector assuming that the propagation phase is included in $r_2' = r_2 e^{-2j\beta L_2}$.

Following Fig. 6.9,

$$r_{eff} = r_g + \frac{t_g^2 r_2'}{1 - r_g r_2'}, \tag{6.57}$$

and

$$t_{eff} = \frac{t_g t_2'}{1 - r_g r_2'}, \tag{6.58}$$

where $t_2' = r_2 e^{-j\beta L_2}$ is the transmission coefficient of facet 2 with phase included. Equations (6.57) and (6.58), of course, are the same as Eqs. (3.20) and (3.21), although here we have included the phase of the reflections and transmissions relative to the grating reference planes in r_g and r_2', and t_g and t_2', respectively. Note also that with the reference planes of the previous section, the reflection from either side of the grating is r_g, unlike the case in Chapter 3, where we used physical reference planes and the reflection from opposite sides of the grating had opposite signs (see Fig. 6.6).

Now, we can use Eqs. (6.57) and (6.58) as the properties of a new effective back mirror, and write the net transmission through the entire structure. That is,

$$t_g' = \frac{t_1 t_{eff} e^{-j\beta L_1}}{1 - r_1 r_{eff} e^{-2j\beta L_1}} = \frac{t_1 t_g t_2 e^{-j\beta L_1}}{\left[1 - r_1 r_g e^{-2j\beta L_1} - \frac{r_1 r_2 t_g^2 e^{-2j\beta L_1}}{1 - r_g r_2}\right](1 - r_g r_2)}. \tag{6.59}$$

The characteristic equation is again given by the poles, or,

$$\frac{r_1 r_2 t_g^2 e^{-2j\beta L_1}}{(1 - r_1 r_g e^{-2j\beta L_1})(1 - r_2 r_g)} = 1. \tag{6.60}$$

As before, this can be solved for the threshold gain and wavelength using the constituent relationships. For the simple case of no grating, we can set $r_g = 0$, and Eq. (6.60) reduces to the Fabry–Perot threshold condition, $r_1 r_2 t_g^2 = 1$. Alternatively, for $r_1 = r_2 = 0$, it reduces to the standard DFB threshold condition, $t_g = \infty$. And for $r_1 = 0 (r_2 = 0)$, it reduces to $r_2 r_g = 1 (r_1 r_g = 1)$, which is the threshold condition for a laser with active mirrors. When solving for threshold, keep in mind that the phase delays implicit in r_1 and r_2 include the round-trip distance between facet-1 or facet-2 and the nearest cosine maximum of the grating index variation (since we have chosen $\phi = 0$ and $L_r = m\Lambda$ for our grating reference planes).

6.4.3 Codirectional Coupling: Directional Couplers

In this section, two different modes propagating in the same direction are considered. These may be (a) orthogonal modes of the same waveguide, or (b) modes of two, initially separate, waveguides. A dielectric perturbation can couple these

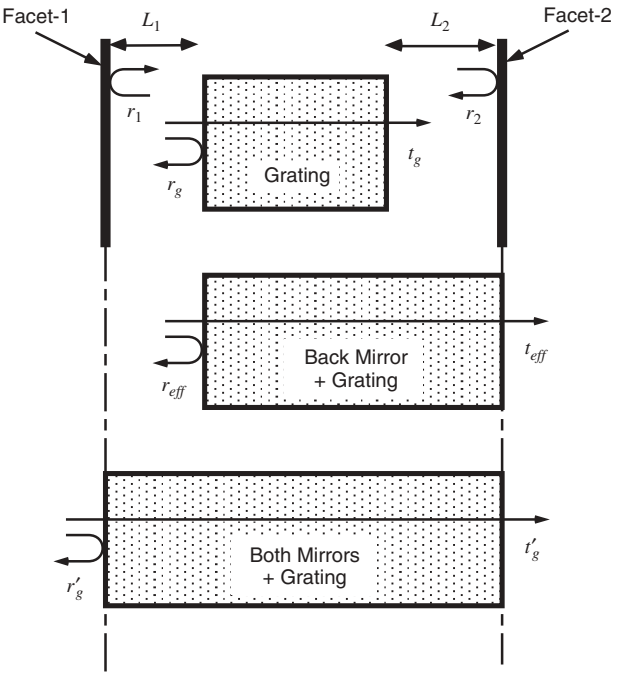


FIGURE 6.9: Progression of calculations to get net transmission and/or reflection from a grating with facet reflections. Any distances to the facet reflections are included in the phases of r_1 and r_2 for convenience. That is, reference planes are those of the grating.

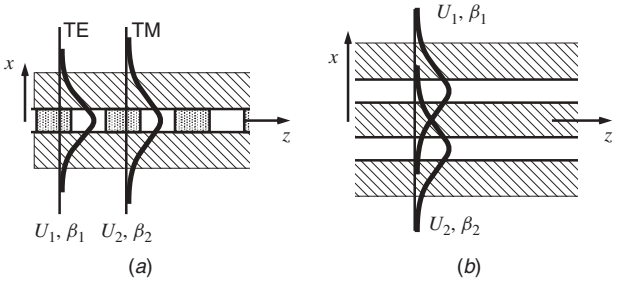


FIGURE 6.10: Codirectional coupling schemes. Uncoupled eigenmodes (a) in the same waveguide and (b) in different waveguides.

two modes in either case. In Case (a), the index of the single guide is perturbed, usually in a periodic fashion, and in Case (b) each initially separate guide acts as a dielectric perturbation for the other. As indicated in Fig. 6.10, an example of Case (a) is TE to TM coupling in one guide; and an example of Case (b) is a directional coupler for modes of separate, but closely spaced waveguides. In either case, we can express the net field as

$$\mathscr{E}(x,y,z) = E_1(z)\mathbf{U}_1(x,y)e^{-j\beta_1 z} + E_2(z)\mathbf{U}_2(x,y)e^{-j\beta_2 z}, \tag{6.61}$$

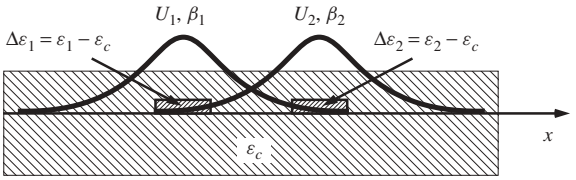


FIGURE 6.11: Cross-sectional schematic of directional coupler showing added index regions for waveguides which also serve to perturb the other waveguide. Eigenmodes and propagation constants refer to unperturbed values.

where $\mathbf{U}_1$ and $\mathbf{U}_2$, and β_1 and β_2 represent the unperturbed solutions for the two uncoupled modes, and the E_i's give the amplitude of each component in the net field.

For coupling between the modes, we need to add a dielectric perturbation. For Case (a), we assume $\varepsilon(x,y,z) = \varepsilon_{gd}(x,y) + \Delta\varepsilon(x,y,z)$, where $\varepsilon_{gd}(x,y)$ is the unperturbed waveguide, and $\Delta\varepsilon(x,y,z)$ is the (possibly periodic) dielectric perturbation seen by both $\mathbf{U}_1$ and $\mathbf{U}_2$. Referring to Fig. 6.11 for Case (b), we observe that $\Delta\varepsilon_2$ acts as the dielectric perturbation for $\mathbf{U}_1$, while $\Delta\varepsilon_1$ acts as the perturbation for $\mathbf{U}_2$. Hence, for this case, $\varepsilon(x,y) = \varepsilon_c + \Delta\varepsilon_1(x,y) + \Delta\varepsilon_2(x,y)$, where the latter two terms act as both waveguides and perturbations simultaneously. The equations to follow are derived for Case (b). However, by setting $\Delta\varepsilon_1 = \Delta\varepsilon_2 = \Delta\varepsilon$ in these equations, we obtain the results for Case (a) as well.

Proceeding as before, we add the dielectric perturbation, write the new perturbed mode profiles as $\mathbf{U}_i + \Delta\mathbf{U}_i$, and plug the assumed net field into the wave Eq. (6.4). After dropping out second-order terms (assuming $\beta_1 \sim \beta_2$) and the unperturbed solutions, this leads to

$$2j\beta_1\mathbf{U}_1\frac{dE_1}{dz}e^{-j\beta_1 z} + 2j\beta_2\mathbf{U}_2\frac{dE_2}{dz}e^{-j\beta_2 z} = \Delta\varepsilon_2(x,y)k_0^2\mathbf{U}_1E_1e^{-j\beta_1 z} + \Delta\varepsilon_1(x,y)k_0^2\mathbf{U}_2E_2e^{-j\beta_2 z}. \tag{6.62}$$

Additional terms containing $\Delta\mathbf{U}_1$ and $\Delta\mathbf{U}_2$ analogous to those appearing on the right side of Eq. (6.25) also exist. However, we have left them out of Eq. (6.62) anticipating that we will be able to set them to zero if we dot multiply by $\mathbf{U}_1^*$ or $\mathbf{U}_2^*$ and integrate over the transverse cross section, as we have done in every other perturbation problem (we go to the trouble of first defining perturbed mode profiles $\mathbf{U}_i' = \mathbf{U}_i + \Delta\mathbf{U}_i$ and then setting the $\Delta\mathbf{U}_i$ terms to zero to prove that to first order, we only need to know the unperturbed mode profiles, $\mathbf{U}_i$, to estimate the coupling between modes).

Performing the integration of Eq. (6.62) after dot multiplying by $\mathbf{U}_1^*$ (the parallel case with $\mathbf{U}_2^*$ will be the same except for permuted subscripts), we obtain

$$2j\beta_1\frac{dE_1}{dz}e^{-j\beta_1 z}\int|U_1|^2\,dA + 2j\beta_2\frac{dE_2}{dz}e^{-j\beta_2 z}\int\mathbf{U}_1^*\cdot\mathbf{U}_2\,dA = k_0^2E_1e^{-j\beta_1 z}\int\Delta\varepsilon_2|U_1|^2\,dA + k_0^2E_2e^{-j\beta_2 z}\int\Delta\varepsilon_1\mathbf{U}_1^*\cdot\mathbf{U}_2\,dA. \tag{6.63}$$

To go further, we must consider Cases (a) and (b) separately. However, in both cases we argue that the second term on the left side of the equation and the first term on the right side are negligible or unimportant for coupling. Then, we can proceed to calculate a simple expression for dE_1/dz. Following this, we can write the corresponding expression for dE_2/dz, which follows by multiplying Eq. (6.62) by $\mathbf{U}_2^*$ and integrating. The justification for dropping the unwanted terms goes as follows for the two cases.

Case (a) $\mathbf{U}_1$ *and* $\mathbf{U}_2$ *are Eigenmodes of the Same Guide:* For this case, the second term on the left side of Eq. (6.63) is identically zero, since the unperturbed eigenmodes are orthogonal. The first term on the right side of Eq. (6.63) is not necessarily negligible, but it gives only the single-mode β-perturbation calculated earlier from the DC part of $\Delta\varepsilon$, and no mode-coupling effect. In fact, we already know that $\Delta\varepsilon$ must be periodic along the z-axis to couple the two different β's in this case, so this periodic portion will integrate to zero in this first right-hand term.

Case (b) $\mathbf{U}_1$ *and* $\mathbf{U}_2$ *are Eigenmodes of Two Separate Guides:* Figure 6.11 illustrates the overlapping fields in a directional coupler and the fact that the placement of guide # 2 near guide # 1 is the perturbation to be analyzed for guide # 1. On the left side of Eq. (6.63), we argue that the first term is much larger than the second for weakly coupled guides, since there is no place in the cross section of integration where both modes have coincidental large amplitudes. The coefficients of the integrals should be comparable. Put another way, we could say that since we choose to neglect the second term on the left side, the theory is only valid for very small overlap of the two modes quantified by $\int |U_1|^2\, dA \gg \int \mathbf{U}_1^* \cdot \mathbf{U}_2\, dA$. (When this condition is not met, a superposition of the exact "supermodes" of the five-layer system is required to determine the net energy transfer from the vicinity of one dielectric stripe to the other. This will be discussed later.)

To neglect the first term on the right side for case (b), we realize that the integration is limited to the cross section of guide # 2, where according to the weak overlap assumption, $\mathbf{U}_1$ must be very small, so the first integral on the right is indeed small. The second term survives since $\mathbf{U}_1 \gg \mathbf{U}_2$ over guide # 1 where the integration exists. Also, this term is still much larger than the second term on the left side, since $k_0^2 E_2 \gg \beta_2 dE_2/dz$.

With these arguments then, we proceed with the analysis of the directional coupler using only the leftmost and rightmost terms in Eq. (6.63). Thus, for guide # 2 perturbing guide # 1:

$$2j\beta_1 \frac{dE_1}{dz} e^{-j\beta_1 z} \int |U_1|^2\, dA = k_0^2 E_2 e^{-j\beta_2 z} \int \Delta\varepsilon_1 \mathbf{U}_1^* \cdot \mathbf{U}_2\, dA. \tag{6.64}$$

For guide # 1 perturbing guide # 2, an equation similar to Eq. (6.63) can be derived by dot multiplying Eq. (6.62) by $\mathbf{U}_2^*$ and integrating. After dropping the small terms,

$$2j\beta_2 \frac{dE_2}{dz} e^{-j\beta_2 z} \int |U_2|^2\, dA = k_0^2 E_1 e^{-j\beta_1 z} \int \Delta\varepsilon_2 \mathbf{U}_2^* \cdot \mathbf{U}_1\, dA. \tag{6.65}$$

Or, after rearranging both,

$$\frac{dE_1}{dz} = -j\frac{k_0^2}{2\beta_1}E_2 e^{-j(\beta_2-\beta_1)z}\frac{\int \Delta\varepsilon_1 \mathbf{U}_1^* \cdot \mathbf{U}_2\, dA}{\int |U_1|^2\, dA}, \tag{6.66}$$

and

$$\frac{dE_2}{dz} = -j\frac{k_0^2}{2\beta_2}E_1 e^{-j(\beta_1-\beta_2)z}\frac{\int \Delta\varepsilon_2 \mathbf{U}_2^* \cdot \mathbf{U}_1\, dA}{\int |U_2|^2\, dA}. \tag{6.67}$$

These are the basic coupled mode expressions for codirectional coupling. They show that unless $\Delta\varepsilon_i$ contains some spatially periodic factor, β_1 must nearly equal β_2 for coherent coupling (monotonic growth of E_i) over some distance. This is the case for uniform directional couplers. However, if $\beta_1 \neq \beta_2$, then $\Delta\varepsilon_i$ must contribute the difference by including a periodicity. Such is generally the case in our Case (a), where $\mathbf{U}_1$ and $\mathbf{U}_2$ are unperturbed eigenmodes of the same guide, and it can be the situation in our Case (b) if the two coupled waveguides are different.

In the case of $\beta_1 \neq \beta_2, \Delta\varepsilon_i$ can be Fourier analyzed, as in Eq. (6.27), and Eqs. (6.66) and (6.67) can be modified to explicity show the z-dependence of $\Delta\varepsilon_i$, analogous to Eqs. (6.28) and (6.29). As in the case of contradirectional coupling, these modified equations give a requirement on the spatial period of the perturbation for coherent addition of the coupling. For the present codirectional coupling case,

$$l\frac{2\pi}{\Lambda} = |\beta_2 - \beta_1| = \frac{2\pi}{\lambda}|\bar{n}_2 - \bar{n}_1|. \tag{6.68}$$

As can be seen, this condition leads to a somewhat coarser period than in the reflective grating case. In fact, the period is generally tens of wavelengths long, since the difference between the mode effective indicies, $\bar{n}_2 - \bar{n}_1$, is usually not so large.

Also, in Case (a) there is no distinction between $\Delta\varepsilon_1$ and $\Delta\varepsilon_2$, since there is only one perturbation of a single waveguide. Thus, the subscripts can be dropped in Eqs. (6.66) and (6.67) in future codirectional coupling results for analyzing this case. In what follows, however, we shall explicitly treat the slighltly more complex Case (b), which generally requires the distinction of $\Delta\varepsilon_1$ from $\Delta\varepsilon_2$. Of course, it also leads to the technologically important four-port directional coupler.

Before moving on to the four-port directional coupler, we choose to rewrite the coupled-mode Eqs. (6.66) and (6.67) in terms of the "normalized amplitudes," a_i's, introduced in Chapter 3. Using Eq. (3.2) then, we make the substitutions,

$$\begin{aligned} a_1(z)\sqrt{2\eta_1} &= E_1(z)e^{-j\beta_1 z}, \\ a_2(z)\sqrt{2\eta_2} &= E_2(z)e^{-j\beta_2 z}. \end{aligned} \tag{6.69}$$

In this case, the assumed total electric field may be written as,

$$\mathscr{E} = \sqrt{2\eta_1}a_1\mathbf{U}_1 + \sqrt{2\eta_2}a_2\mathbf{U}_2. \tag{6.70}$$

From Eq. (6.70) it can be seen that the power flow in the positive-z direction is just $|a_1|^2 + |a_2|^2$, as should be the case for the normalized amplitudes. Note that the cross terms are negligible for the same reasons as given earlier.

Plugging Eq. (6.69) into Eqs. (6.67) and (6.68) and neglecting the difference between η_1 and η_2, we find

$$\begin{aligned} \frac{da_1}{dz} &= -j\beta_1 a_1 - j\kappa_{12} a_2, \\ \frac{da_2}{dz} &= -j\beta_2 a_2 - j\kappa_{21} a_1, \end{aligned} \tag{6.71}$$

where

$$\begin{aligned} k_{12} &= \frac{k_0^2}{2\beta_1} \frac{\displaystyle\int_{G1} (\varepsilon_1 - \varepsilon_c)\mathbf{U}_1^* \cdot \mathbf{U}_2 \, dA}{\displaystyle\int |U_1|^2 \, dA}, \\ k_{21} &= \frac{k_0^2}{2\beta_2} \frac{\displaystyle\int_{G2} (\varepsilon_2 - \varepsilon_c)\mathbf{U}_2^* \cdot \mathbf{U}_1 \, dA}{\displaystyle\int |U_2|^2 \, dA}, \end{aligned} \tag{6.72}$$

and the integration is shown explicitly as only over the region of perturbation where $\Delta\varepsilon_i = (\varepsilon_i - \varepsilon_c)$ is added to form the additional guide as indicated in Fig. 6.11. That is, $G2$ represents the cross-sectional area of added index material to form guide # 2, etc. Again, recall that the ε's are always the relative dielectric constants. Also, note that although the coefficient of $\kappa_{ij}, k_0^2/(2\beta_i) = \omega/(2c\bar{n}_i)$, is linearly proportional to the optical frequency, the overlap integrals can decrease rapidly as the wavelength decreases; thus, κ_{ij} actually tends to decrease slightly with increasing optical frequency in a directional coupler.

Example 6.4 Two uniform slab waveguides, 0.1 μm in width, are spaced by 0.3 μm. The index of the slabs is 3.4, and that of the surrounding regions is 3.17, for the operating wavelength of 1.55 μm. These values can be used to calculate the effective index of each waveguide, $\bar{n}_{eff} = 3.1836$, and the confinement factor $\Gamma = 0.11$, based on the material from Appendix 3.

Problem: What is the ratio of amplitudes of the fundamental mode from the second waveguide to the fundamental mode of the first waveguide, at the center of the first waveguide?

Solution: To compute the ratio of normalized fields, we need to make a couple of observations. Inside waveguide #1, the normalized field profile $U_1(x)$ will have a cosine dependence, and outside, it will exponentially decay toward zero. Since the confinement factor Γ is relatively low, we can make the following approximations to the shape of the normalized electric field profiles: $U(x)$ is approximately constant inside the waveguide, $U_1(x) = U_1(0)$, and $U(x)$ is approximately linear across

the second waveguide. Thus, to link the amplitude of the normalized field of the fundamental mode to the amplitude of the tail of the other waveguide's fundamental mode, we need to calculate the exponential decay constant for this waveguide system,

$$\gamma_x = \sqrt{\beta^2 - k_0^2 \cdot n_c^2} = \frac{2\pi}{1.55\ \mu\text{m}}\sqrt{3.1836^2 - 3.17^2} = 1.192\ \mu\text{m}^{-1}.$$

Based on our assumptions, $\frac{U_2(0)}{U_1(0)} \approx e^{-\gamma \cdot x}$, where $x = 0.3 + 0.05\ \mu$m, yielding

$$\frac{U_2(0)}{U_1(0)} \approx e^{-0.192 \cdot 0.35} = 0.659.$$

In lossless waveguides, $\Delta\varepsilon_1$ and $\Delta\varepsilon_2$ are real and the phase fronts of the modes are flat and perpendicular to the propagation direction, implying that both U_1 and U_2 are real functions. As a result, both κ_{12} and κ_{21} are real. This remains approximately true for waveguides with small gains or losses, as long as $\Delta\varepsilon_1$ and $\Delta\varepsilon_2$ are predominantly real, as is usually the case. In such cases, we can use the real part of the perturbation in each guide to estimate κ_{12} and κ_{21}. In addition, if the guides are symmetric, then $\kappa_{12} = \kappa_{21} = \kappa$ (for asymmetric guides we can still define an average $\kappa = \sqrt{\kappa_{12}\kappa_{21}}$).

Equations (6.71) are already in a convenient form for solution, so we proceed toward a general solution. To this end, we assume trial solutions, $a_1 = a_{10}e^{-j\beta_c z}$, and $a_2 = a_{20}e^{-j\beta_c z}$, where a_{10} and a_{20} are real constants independent of z and β_c is some unknown propagation constant, presumably with multiple solutions, for the coupled waveguides. The general solutions for the $a_i(z)$'s will then be a linear superposition of all of the particular solutions. Plugging the trial solutions into Eqs. (6.71), and solving for β_c, we obtain

$$\beta_c = \frac{\beta_1 + \beta_2}{2} \pm \sqrt{\left(\frac{\beta_1 - \beta_2}{2}\right)^2 + \kappa_{12}\kappa_{21}}, \tag{6.73}$$

or

$$\beta_c = \bar{\beta} \pm s, \tag{6.74}$$

where

$$\bar{\beta} = \frac{\beta_1 + \beta_2}{2} \quad \text{and} \quad s = \sqrt{\left(\frac{\beta_1 - \beta_2}{2}\right)^2 + \kappa_{12}\kappa_{21}}. \tag{6.75}$$

We can also define a codirectional detuning parameter, $\delta \equiv (\beta_2 - \beta_1)/2$, such that $s^2 = \kappa_{12}\kappa_{21} + \delta^2$, making it the codirectional analog of Eq. (6.42).

With Eq. (6.74), the general solutions are of the form:

$$\begin{aligned} a_1(z) &= e^{-j\bar{\beta}z}[A_1 e^{jsz} + A_2 e^{-jsz}], \\ a_2(z) &= e^{-j\bar{\beta}z}[B_1 e^{jsz} + B_2 e^{-jsz}], \end{aligned} \tag{6.76}$$

and we are now ready for a real problem with real boundary conditions.

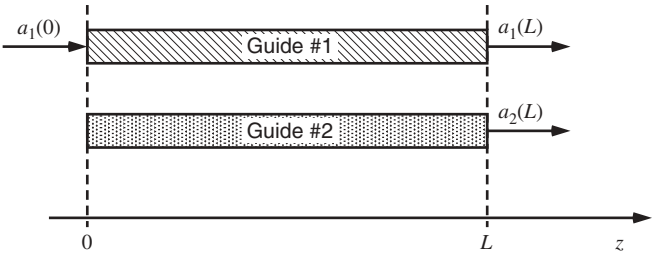


FIGURE 6.12: Four-port directional coupler showing a single input excited.

The Four-Port Directional Coupler Figure 6.12 schematically shows a four-port directional coupler formed by two coupled waveguides of finite length, L. In fact, we can derive most of the desired directional coupler expressions without being specific about the length, but we will assume that the outputs are matched so that no reflections exist. This allows us to derive the general scattering parameters for the directional coupler, and from these we can add a wide variety of different boundary conditions that can be handled from the scattering theory.

To get started, we let $a_1(0) = a_1(0)$, and $a_2(0) = 0$, and plug into Eq. (6.76) to find that $B_1 = -B_2$, and $a_1(0) = A_1 + A_2$. Next, we evaluate Eq. (6.71) at $z = 0$, and using Eqs. (6.74) and (6.76), solve for B_2. We find that

$$B_2 = \frac{\kappa_{21}}{2s} a_1(0). \tag{6.77}$$

Again using Eq. (6.76) this gives

$$a_2(z) = -j \frac{\kappa_{21}}{s} a_1(0) e^{-j\bar{\beta}z} \sin sz. \tag{6.78}$$

Next, we solve for A_1 and A_2 using these same equations. This leads to

$$a_1(z) = a_1(0) e^{-j\bar{\beta}z} \left[\cos sz + j \frac{\beta_2 - \beta_1}{2s} \sin sz \right]. \tag{6.79}$$

Therefore, generally, by linear superposition letting the input at port #2, $a_2(0) = a_2(0)$, we have

$$a_1(z) = \left[a_1(0) \left(\cos sz + j \frac{\beta_2 - \beta_1}{2s} \sin sz \right) - j \frac{\kappa_{12}}{s} a_2(0) \sin sz \right] e^{-j\bar{\beta}z}, \tag{6.80}$$

and, likewise by symmetry, we can interchange subscripts to solve for $a_2(z)$,

$$a_2(z) = \left[-j \frac{\kappa_{21}}{s} a_1(0) \sin sz + a_2(0) \left(\cos sz - j \frac{\beta_2 - \beta_1}{2s} \sin sz \right) \right] e^{-j\bar{\beta}z}. \tag{6.81}$$

Figure 6.13 illustrates these normalized amplitudes versus z for $\beta_1 = \beta_2$ and $\beta_1 \neq \beta_2$ under the initial assumption that $a_2(0) = 0$.

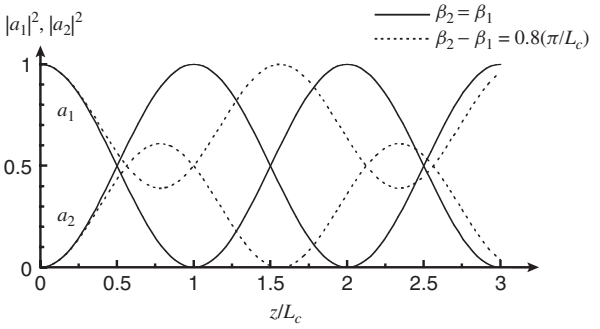


FIGURE 6.13: Energy exchange between two coupled waveguides as a function of propagation distance for matched and mismatched propagation constants.

In the simple case of identical coupled guides, $\beta_1 = \beta_2$ and $s = \sqrt{\kappa_{12}\kappa_{21}} \equiv \kappa$. For a coupler of length L, Eq. (6.78) gives

$$\left|\frac{a_2(L)}{a_1(0)}\right|^2 = \sin^2 \kappa L, \tag{6.82}$$

from which the length, L_c, for full coupling from guide #1 to guide #2 is found to be

$$L_c \equiv \frac{\pi}{2\kappa}. \tag{6.83}$$

This is generally referred to as the *coupling length*. Odd multiples of the length given by Eq. (6.83) also provide full coupling, as indicated by Fig. 6.13.

Example 6.5 A vertical directional coupler is created from the vertical waveguide structure described in Example 6.4.

Problem: (a) What is the coupling coefficient κ for this coupler? (d) What is the coupling length L_c of this coupler?

Solution: (a) Starting from the definition of κ, and keeping in mind the discussion about normalized fields from the previous example, we have

$$\kappa = \frac{k_0^2}{2 \cdot \beta} \frac{\int_{G1} (\varepsilon_1 - \varepsilon_2) \mathbf{U}_1^* \cdot \mathbf{U}_2 \, dA}{\int |U_1|^2 \, dA} = \frac{\pi}{\lambda \cdot \bar{n}_{eff}} (n_{gr}^2 - n_c^2) \frac{\int_{G1} |U_1|^2 \, dA \frac{U_2(0)}{U_1(0)}}{\int |U_1|^2 \, dA}$$

where we have approximated $U_1(x)$ and $U_2(x)$ with their average values over the range of integration, waveguide G1. Using the results from Example 6.4, the coupling coefficient is given as

$$\kappa = \frac{\pi}{\lambda \cdot \bar{n}_{eff}} (n_{gr}^2 - n_c^2) \frac{U_2(0)}{U_1(0)} \Gamma = \frac{\pi}{1.55 \ \mu\text{m} \ \cdot 3.1836} (3.4^2 - 3.17^2)(0.11)(0.659)$$
$$= 0.0681 \ \mu\text{m}^{-1}.$$

(b) From the definition of the coupling length, Eq. (6.83),

$$L_c \equiv \frac{\pi}{2\kappa} = \frac{\pi}{2 \cdot 0.0659^{-1}} = 23.82 \ \mu\text{m}.$$

Viewing the directional coupler as a four-port network, as suggested by Fig. 6.12, it is desirable to develop the four-port scattering matrix, so that the techniques of Chapter 3 can be applied. For the labeling given in Fig. 6.14, Eqs. (6.78) through (6.81) can be used to identify these coefficients by inspection. That is,

$$\mathbf{S} = \begin{bmatrix} 0 & \sqrt{1-c^2} & 0 & -jc \\ \sqrt{1-c^2} & 0 & -jc & 0 \\ 0 & -jc & -0 & \sqrt{1-c^2} \\ -jc & 0 & \sqrt{1-c^2} & 0 \end{bmatrix} e^{-j\bar{\beta}L}, \tag{6.84}$$

where for $\kappa_{12} \approx \kappa_{21} \approx \kappa$,

$$c = \frac{\kappa}{s} \sin sL. \tag{6.85}$$

Note that we have retained the use of $a_i(z)$ and $b_i(z)$ for the forward and backward waves in guide i, rather than going to the standard "input" and "output" definitions in Fig. 6.14. To conform to the standard scattering theory jaron used in Chapter 3, we must use $a_1 \equiv a_1(0), b_1 \equiv b_1(0), a_2 \equiv b_1(L), b_2 \equiv a_1(L), a_3 \equiv a_2(0), b_3 \equiv b_2(0), a_4 \equiv b_2(L)$, and $b_4 \equiv a_2(L)$.

The general scattering matrix given by Eq. (6.84) can be combined with various boundary conditions to develop scattering matrices of more complex photonic integrated circuits. For example, consider the configuration shown in Fig. 6.16 where guide #2 has reflectors r_3 and r_4 at $z = 0$ and $z = L$, respectively. These provide boundary conditions that create backward traveling waves, $b_1(z)$ and $b_2(z)$. In the present case, it can be seen that $b_2(L) = r_4 a_2(L)$ and $a_2(0) = r_3 b_2(0)$. After a little algebra, the scattering parameters for the resulting two-port network can be obtained. That is,

$$\mathbf{S} = \begin{bmatrix} -\frac{c^2 r_4 e^{-j\bar{\beta}L}}{1-r_3 r_4(1-c^2)e^{-2j\bar{\beta}L}} & \sqrt{1-c^2} - \frac{c^2 r_3 r_4 \sqrt{1-c^2} e^{-2j\bar{\beta}L}}{1-r_3 r_4(1-c^2)e^{-2j\bar{\beta}L}} \\ \sqrt{1-c^2} - \frac{c^2 r_3 r_4 \sqrt{1-c^2} e^{-2j\bar{\beta}L}}{1-r_3 r_4(1-c^2)e^{-2j\bar{\beta}L}} & -\frac{c^2 r_3 e^{-j\bar{\beta}L}}{1-r_3 r_4(1-c^2)e^{-2j\bar{\beta}L}} \end{bmatrix} e^{-j\bar{\beta}L}. \tag{6.86}$$

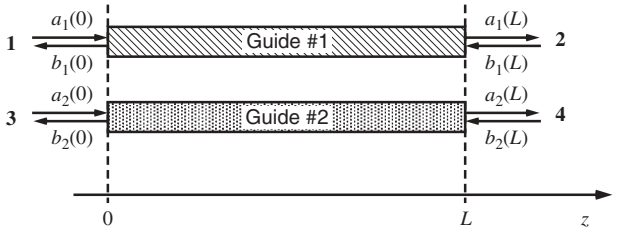


FIGURE 6.14: Schematic of four-port network formed by two coupled waveguides.

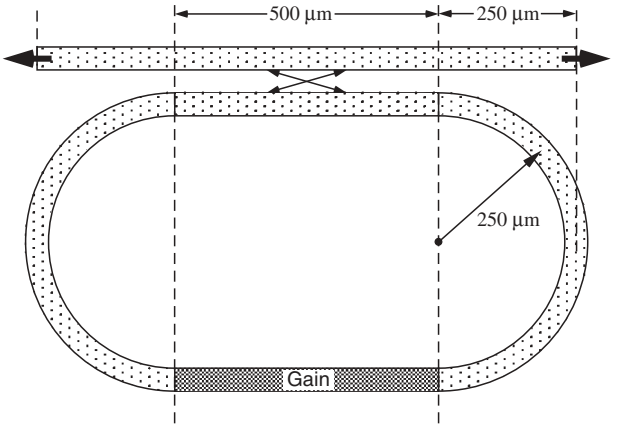


FIGURE 6.15: Ring laser with directional-coupler output tap.

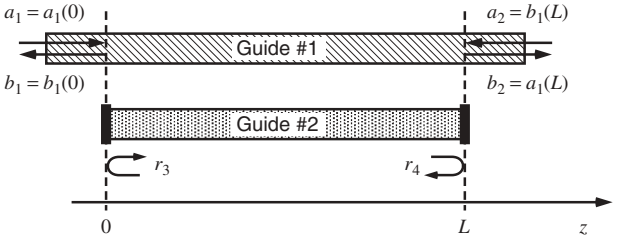


FIGURE 6.16: Two-port formed by placing reflectors at ports 3 and 4 of a directional coupler.

Applications of Directional Couplers: Ring Structures Directional couplers (and multimode interference couplers, discussed later in this Chapter) are key components in implementing waveguide ring structures, illustrated in Fig. 6.15, which can be used as optical filters or simple single-mode laser structures. As the light from the coupler is coupled into a ring waveguide, for certain wavelengths, phase matching will occur between the light coupled in, and out of the ring, leading to a resonance. In the next example, we will derive some important properties of a ring resonator, consisting of entirely passive waveguides.

Example 6.6 A planar waveguide ring resonator, similar to the one shown in Figure 6.15, is coupled to a straight waveguide. The ring length is L_r, the straight waveguide length is L_s. The one pass coupling magnitude due to the coupling length L_c between the straight guide and the ring is c.

Problem: Calculate the S_{21} parameter for this ring.

Solution: Using Mason's rule from Chapter 3, and the results on directional couplers from this chapter, we can directly calculate S_{21} by calculating the transfer

function between the normalized input field a_1 and the normalized output field b_2,

$$b_2 = S_{21} a_1$$

Converting this structure into a signal flow chart (which is left to the reader as an exercise), we have that

$$\Delta = 1 - \sqrt{1 - c^2} e^{-j\tilde{\beta} L_r}.$$

There are two independent forward paths, and one loop in this system, therefore,

$$T_1 \cdot \Delta_1 = \sqrt{1 - c^2} e^{-j\tilde{\beta} L_s} \cdot (1 - \sqrt{1 - c^2} e^{-j\tilde{\beta} L_r}),$$

and

$$T_2 \cdot \Delta_2 = (jc)^2 \cdot e^{-j\tilde{\beta} L_r} \cdot e^{-j\tilde{\beta} L_s}.$$

Finally, the value for S_{21} is

$$S_{21} = \left(\sqrt{1 - c^2} e^{-j\tilde{\beta} L_s} + \frac{(jc)^2 \cdot e^{-j\tilde{\beta} L_r} e^{-j\tilde{\beta} L_s}}{1 - \sqrt{1 - c^2} e^{-j\tilde{\beta} L_r}} \right)$$

This result is analogous to the result for a Fabri–Perot cavity, in which the reflection is replaced by $r = \sqrt{1 - c^2}$ and transmission by $t = jc$. The pole in the denominator will determine the transfer function of the ring resonator, and this structure can be used as a filter. An example application will be discussed in Chapter 8.

Ring lasers represent an attractive implementation of single mode lasers, due to their simple building blocks and fabrication processes involved. In addition, the "mirrors" in a ring laser are actually light couplers, which allow for a portion of the light to escape the cavity. Figure 6.15 illustrates the use of a directional coupler as an output coupler for a ring laser. In a ring laser, the round trip cavity length is the circumference of the cavity, and there are generally two propagating modes, which are uncoupled (clockwise and counterclockwise), and whose degeneracy must be removed through insertion of differential loss to achieve stable operation. One way to remove the degeneracy, particularly suitable for PIC ring laser implementation is through seeding of a particular mode by an on-chip light source. The priciple of designing a ring laser is illustrated in the following example.

Example 6.7 We wish to analyze a ring laser device as shown in Fig. 6.15, assuming a 4-quantum-well gain region from Fig. 4.31. The passive waveguide cross section in the transverse direction is 0.3 μm thick, and the waveguide index is 3.37. The waveguide is clad by InP, with an index of 3.17 at 1.55 μm. The lateral

width is adjusted to insure single mode operation. A BH structure is assumed, clad by InP. The waveguide internal losses are $\alpha_i = 10\ \text{cm}^{-1}$, the bend radiative losses (to be discussed in detail in Chapter 7) $\alpha_R = 16.3\ \text{cm}^{-1}$, and the waveguide separation in coupler $s = 1\ \mu\text{m}$.

Problem: (a) Determine the output coupling loss (analogous to mirror loss). (b) Calculate the threshold modal gain, threshold current, and differential efficiency.

Solution: (a) For the calculation of the coupling loss, we first need to determine κ, the guide-to-guide coupling constant from Eq. (6.72). Assuming the lowest-order symmetric lateral mode, we can plug Eqs. (6.8), (A3.17), and (A3.21) into Eq. (6.72) and solve. For identical guides coupled laterally, we find that

$$\kappa = \Gamma_x \frac{2k_0^2(\varepsilon_1 - \varepsilon_c)\gamma_y \cos^2(k_{1y} w/2)}{\beta w_{eff}(\gamma_y^2 + k_{1y}^2)} e^{-\gamma_y s},$$

where Γ_x is the transverse confinement factor, ε_1 and ε_c are the relative dielectric constants (indices squared) of the waveguide slab and surrounding cladding, γ_y is the lateral decay constant outside of the guides, w is the lateral waveguide width, w_{eff} is the effective width (given by Eq. (6.13), with y replacing x), and s is the waveguide separation in the coupler region.

First, we determine the guide width, w, and the lateral propagation and decay constants using the effective index technique. For single lateral mode, $V_y \leq 3.25$. Therefore, from Eq. (A3.12), $w \leq 1.2\ \mu\text{m}$. Then, using $\bar{n}_2 = 3.24$ and $\bar{n} = 3.216$, we get $k_y = 1.60/\mu\text{m}$, and $\gamma_y = 2.20/\mu\text{m}$. These numbers give $w_{eff} = 2.51\ \mu\text{m}$. Then, we solve for the coupling constant between the two guides in the directional coupler,

$$\kappa = 0.49 \frac{2(2\pi/1.55\ \mu\text{m})^2(3.216^2 - 3.17^2)2.2/\mu\text{m} \times \cos^2(1.6 \times 1.2/2)}{[2\pi(3.216)(1.55\ \mu\text{m}]2.51\ \mu\text{m}\ (2.2^2 + 1.6^2)/\mu\text{m}^2} e^{-2.2\times 1}$$
$$= 15.7\ \text{cm}^{-1}.$$

The fractional power coupled from one guide to the other over the 500-μm-long coupler length is

$$c^2 = \sin^2 \kappa L = \sin^2(15.7 \times 0.05) = 0.498,$$

omitting internal losses. This means that $(1 - c^2) = 0.502$ of the power is transmitted through one of the guides, again omitting losses.

(b) For the ring laser, we need to identify the various lengths so that we can use Eq. (2.23) or (2.25) for the threshold modal gain. We only consider propagation in one direction, since there is no coupling between forward and backward waves. Clearly, the active length L_a is 500 μm, and the total passive length, $L_p = (500\pi + 500)\ \mu\text{m} = 2071\ \mu\text{m}$. The equivalent of the mean mirror reflectivity, R, in Eqs. (2.23) and (2.25) is just $(1 - c^2)$, as might be verified by

reviewing the derivation leading up to Eq. (2.23). Thus, the output coupling loss (equivalent of mirror loss) is

$$\alpha_m = \frac{1}{L_{tot}} \ln \frac{1}{1 - c^2} = 2.7 \text{ cm}^{-1}.$$

The average internal modal loss consists of the internal losses and of the radiation losses in the bends. The total average internal modal loss is

$$\langle \alpha_i \rangle = \frac{(2 \times 500 \text{ μm})10 \text{ cm}^{-1} + (1571 \text{ m})16.3 \text{ cm}^{-1}}{2571 \text{ μm}} = 13.8 \text{ cm}^{-1}.$$

Now, we can calculate the threshold modal gain as

$$\Gamma g_{th} = \langle \alpha_i \rangle + \alpha_m = 16.5 \text{ cm}^{-1}.$$

The confinement factors for the transverse, lateral, and axial directions are 0.06, 0.84, and 0.194, respectively. Thus, the threshold gain, $g_{th} = 1688 \text{ cm}^{-1}$, from which we can extrapolate from Fig. 4.31 that $J_{th} = 5.2 \text{ kA/cm}^2$. Then, the threshold current, $I_{th} = J_{th} w L_a / \eta_i = 52$ mA. This is a very high threshold gain and current density. In a real case, we would try to include more quantum wells to increase the modal gain by increasing the transverse confinement factor. For example, with six quantum wells the threshold gain and current density would be reduced to 1125 cm^{-1} and 3.3 kA/cm^2, more reasonable values for 1.55 μm.

Using the above numbers, but excluding loss in the output waveguide, the differential efficiency is

$$\eta_d = (0.8) \frac{2.7}{16.5} = 13.1\%.$$

For the output power, this must be multiplied by the additional attenuation factor from the coupler to the output facet. The coupler loss has been included in the ring cavity, but not for the energy coupled over. A careful review of the theory reveals that the average propagation constant is used in the propagation delay for both the coupled and uncoupled portions of the lightwave. If it is complex, an attenuation factor will be introduced in both cases. That is, in the present case the coupled fractional power is $c^2 \exp(-\alpha_i L)$, and the uncoupled fractional power is $(1 - c^2) \exp(-\alpha_i L)$. We have already included all the loss in the uncoupled fraction by including the coupler length in L_p, but this only includes the loss in one guide for the coupler as a whole (i.e., half of the coupler loss for the coupled fraction). Thus, we must include an additional factor of $e^{-(10 \times 0.25)}$ for the energy coupled over. (Put another way, we have modeled the coupler as an effective lumped coupler at the center of the actual coupler.) Since we have an additional 250 μm in the extension guide after the coupler, the net reduced differential efficiency becomes

$$\eta'_d = 0.131 e^{-(10 \times 0.05)} = 7.9\%.$$

This ring laser is obviously not a very efficient device, primarily because of the long sections of lossy passive waveguide.

6.4.4 Codirectional Coupler Filters and Electro-optic Switches

$\beta_1 \approx \beta_2$ If two identical waveguides are brought into close proximity to form a directional coupler, the device tends to function over a broad bandwidth of input wavelengths. This is because the unperturbed eigenmodes will be the same and the dispersion properties of each component guide will be the same. Thus, phase-matched coupling will always occur. This kind of dispersion is illustrated in Fig. 6.17, where we have plotted the two β solutions of Eq. (6.73) as a function of optical frequency. The overlap between the unperturbed eigenmodes will vary with wavelength, and this results in a change in the coupling constant, κ, according to Eq. (6.72). This gives the gradual spreading of the two propagation constant solutions as κ increases for longer wavelengths (lower optical frequencies).

For a given coupling level, a length can be chosen for 100% coupling from one guide to the other at the operating wavelength. This is given by the coupling length, L_c, in Eq. (6.83) or odd multiples thereof. If the wavelength is now changed, the net transfer would be less for wavelengths either shorter or longer than this according to Eq. (6.82). In fact, to make a wavelength-selective filter, it is common to use a device many coupling lengths long, so that as the wavelength changes and κ changes, the transfer fraction, Eq. (6.83), will vary sinusoidally. Thus, at one wavelength the device can have an odd number of transfer lengths, while at another it can have an even number, yielding either 100% or 0% transfer,

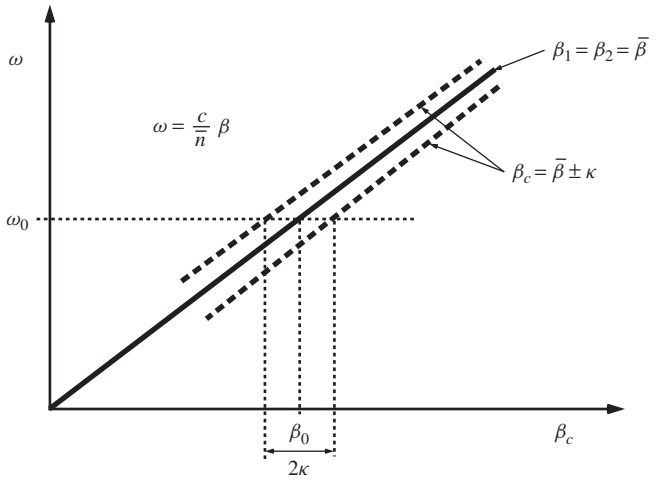


FIGURE 6.17: Dispersion diagram for identical coupled waveguides near the operating wavelength. The operating radial optical frequency, ω_0, results in two propagation constant solutions displaced $\pm\kappa$ from β_0, the value of the uncoupled β for each guide at this point.

respectively. This kind of filter is useful for separating relatively widely spaced wavelengths, such as 1.3 μm from 1.55 μm; however, because κ is a slowly varying function of wavelength, this approach is not effective in providing relatively narrow filter passbands. The quantitative details of the filtering action of such four-port directional couplers are given by Eqs. (6.80) through (6.84). These also allow for slight mismatches in the unperturbed propagation constants as discussed previously.

A voltage-controlled switch or modulator can be constructed of such nominally identical coupled waveguides, if the device is constructed in electro-optic material. The III–V compound semiconductors are electro-optic, and thus, such devices can be compatible with diode lasers. Appendix 15 gives a brief introduction to the electro-optic effect. Without going into a myriad of details, suffice it to say that for certain orientations of electro-optic materials, the application of an electric field with a frequency much lower than that of the optical wave (from DC up to at least 100 GHz) leads to a change in the index of refraction for certain polarizations of the lightwave. Thus, by applying differing electric fields to the waveguides in a directional coupler the propagation constants can be changed slightly, so that $\beta_1 \neq \beta_2$ and the coherence of the coupling can be reduced according to Eqs. (6.78) and (6.79).

As an example of electro-optic modulation, Fig. 6.18 shows a vertical directional-coupler configuration, which allows separate control of the applied low-frequency electric fields in each guide, along with a plot of the power transfer fraction $|a_2(L)/a_1(0)|^2$ from Eq. (6.78) for a general directional coupler with $L = L_c$ and $5L_c$. As can be seen, for a directional coupler that is one coupling length long, no transfer occurs between the guides when

$$\frac{\beta_2 - \beta_1}{2} = \sqrt{3}\kappa, \tag{6.87}$$

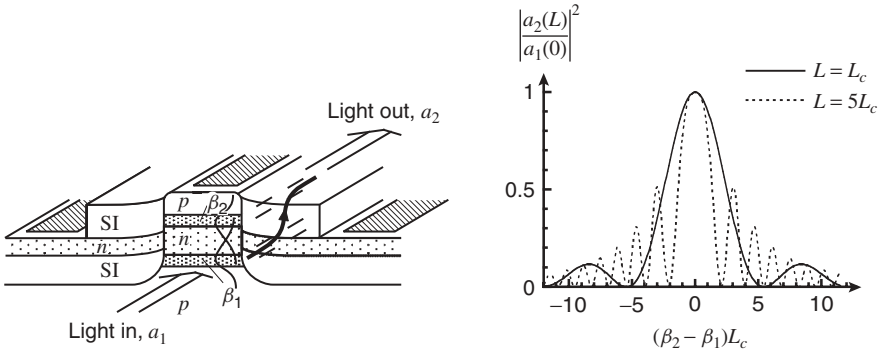


FIGURE 6.18: Schematic of vertical directional-coupler electro-optic switch and example plot of power transfer versus the normalized deviation of the index of one guide from the other.

since s has increased from κ to 2κ at this point. More generally, for a coupler of length $L = mL_c$, the first null in transmission from one guide to the other occurs when the mismatch is

$$(\beta_2 - \beta_1)L_c = \pi\sqrt{[(m+1)/m]^2 - 1}. \tag{6.88f}$$

Thus, as Fig. 6.18 illustrates, a longer coupler (larger m) requires less of a mismatch to shut off the power transfer.

$\beta_1 \neq \beta_2$ If the two waveguide constituents in a directional coupler are not identical, or if two dissimilar modes of the same waveguide are coupled, then a much sharper wavelength filtering action is possible. The conditions for coherent addition of any coupling were discussed in conjunction with Eqs. (6.66) and (6.67). There are several specific possibilities worth mentioning. Two cases have already been outlined in Fig. 6.10. In Case (a) some coarse periodicity in the perturbation was found necessary to couple modes with different propagation constants in the same waveguide. In Case (b), if the β's are different, a modulation of the index will again generally be necessary. In both cases, Eq. (6.68) gives the relationship between the periodicity and the difference in propagation constants. However, in some cases the waveguides can be different, but due to their different dispersive properties, they can have the same propagation constant at some particular wavelength, even with no index modulation. Thus, this becomes a third case, Case (c), where a codirectional waveguide filter is possible. These three cases are summarized by the $\omega - \beta$ diagrams in Fig. 6.19.

To understand the filtering action of these cases more quantitatively, we go back to Eqs. (6.66) and (6.67) and Fourier analyze $\Delta\varepsilon$ as suggested thereafter. This generates a $\pm 2\pi/\Lambda$ term in the argument of the exponential and replaces $\Delta\varepsilon$ in the integrals by the Fourier coefficients $\Delta\varepsilon_{\pm l}$. Now for the specific case illustrated in the top of Fig. 6.19, $\beta_1 > \beta_2$, and for phase matching we choose $-2\pi/\Lambda$ in Eq. (6.66) and $+2\pi/\Lambda$ in Eq. (6.67). Equivalently, we let

$$\beta_1' = \beta_1 + \frac{2\pi}{\Lambda}, \tag{6.89}$$

and replace β_1 in all of the equations to follow with β_1'. Thus, when we arrive at Eq. (6.78), s and $\bar{\beta}$ are primed, which denotes the use of Eq. (6.89) in place of β_1.

The use of Eq. (6.89) results in a relatively large crossing angle for the two effective propagation constants as shown in Fig. 6.19. Since the proximity of the two curves near ω_0 determines the degree of phase coherence for the coupling, the large crossing angle suggests a relatively narrow filter band. In fact, this is the case. Figure 6.20 gives an example of the transfer function, $|a_2(L)/a_1(0)|^2$, versus the detuning parameter, from Eq. (6.78) using Eq. (6.89). The detuning parameter, δ, is defined as

$$\delta \equiv (\beta_2 - \beta_1')/2 = (\omega - \omega_0)/\omega_0 \cdot (\pi/\Lambda)(\Delta\bar{n}_g/\Delta\bar{n}), \tag{6.90}$$

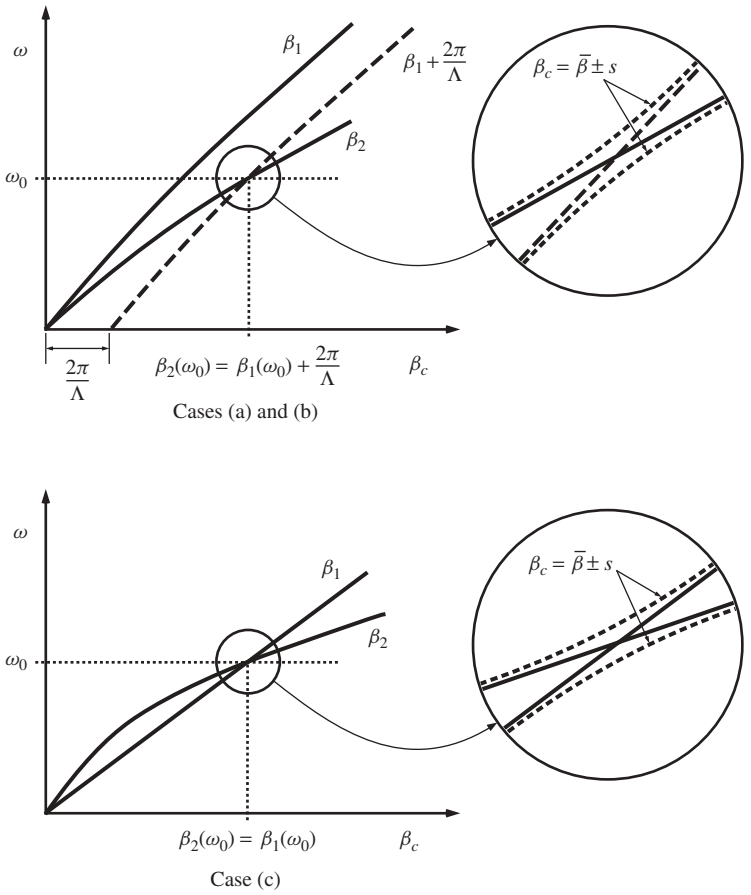


FIGURE 6.19: Dispersion characteristics for two classes of codirectional coupler filters in which two dissimilar modes are coupled. At the top, a grating perturbation is included to provide for phase matching at ω_0. Either two modes of a single guide, Case (a), or coupled waveguides, Case (b), can be used. At the bottom, Case (c), two different waveguide geometries are engineered to have identical β's at ω_0 prior to coupling. The crossing regions are expanded at the right insets to show the pair of new β_c-solutions resulting from the coupling.

where $\Lambda = \lambda_0/\Delta\bar{n}$. The ratio of the *group* effective index difference, $\bar{n}_{g2} - \bar{n}_{g1}$, to the effective index difference, $\bar{n}_2 - \bar{n}_1$, for typical semiconductor coupled waveguides is $\Delta\bar{n}_g/\Delta\bar{n} \approx 3$, which reduces the filter bandwidth accordingly.

Switchable or tunable filters result from the combination of the above filters with electro-optic material. The application of a field in one guide changes its refractive index which changes the slope of its dispersion curve on the $\omega - \beta$ diagram. Since the slopes of the two dispersion curves are similar and since the offset $2\pi/\Lambda$ does not change, the intersection point, ω_0, changes by a much larger relative

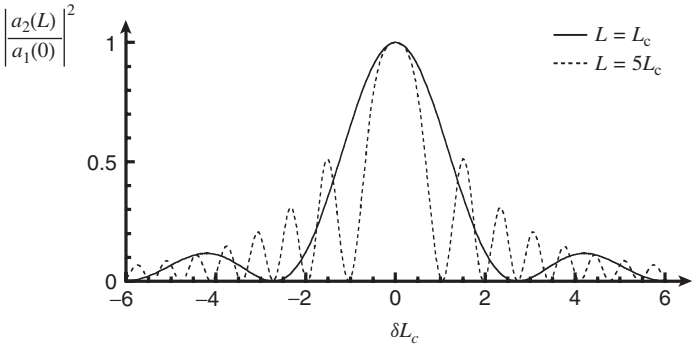


FIGURE 6.20: Power transfer for two different transfer lengths in grating-coupled waveguides versus the detuning parameter.

amount than the index (much like how the intersection point on a pair of scissors moves more rapidly than the motion of the blades themselves, particularly when the blades are near parallel). That is, the center frequency of the grating-assisted codirectional coupler filter can be tuned by a much larger relative amount than the index. A similar action takes place for Case (c) in Fig. 6.19, since here also, the two uncoupled dispersion curves cross at a small angle. This enhanced tuning rate is in stark contrast to most filters, including the contradirectional grating filter, whose center or Bragg frequency tunes by the same relative amount as the index.

To be more specific, for the grating-assisted codirectional coupler filter shown at the top of Fig. 6.19, where $\beta_2 = \beta_1 + 2\pi/\Lambda$ for phase matching, we find that the center frequency and wavelength vary as

$$\frac{\Delta\omega}{\omega} = -\frac{\Delta\lambda}{\lambda} = \frac{\Delta\bar{n}_2}{\bar{n}_{g2} - \bar{n}_{g1}}, \tag{6.91}$$

in response to an index change in guide #2. The g subscript denotes the group index, which appears due to the frequency dependence of the indices. For Case (c) in Fig. 6.19, we get the same result, but via a different derivation. Relative to simple tuning of a Bragg grating, where $\Delta\lambda/\lambda = \Delta\bar{n}/\bar{n}_g$, in a grating assisted codirectional coupler filter the tuning is enhanced, such that

$$\frac{\Delta\lambda}{\lambda} = \frac{\Delta\bar{n}}{\bar{n}_{g1} - \bar{n}_{g2}} = F\frac{\Delta\bar{n}}{\bar{n}_{g1}}, \tag{6.92}$$

where the tuning enhancement factor is given by

$$F = \bar{n}_{g1}/(\bar{n}_{g1} - \bar{n}_{g2}).$$

Also, this tuning enhancement is accompanied by a similar increase in the filter's FWHM optical bandwidth, $\Delta\lambda_{1/2}$. For a uniform interaction region of length L_C,

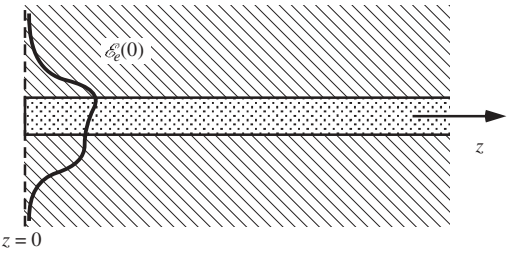


FIGURE 6.21: Illustration of waveguide excitation with some arbitrary field at $z = 0$. This can be expressed in terms of a linear superposition of all the eigenmodes of the waveguide (including radiation modes).

this bandwidth, as outlined in Fig. 6.21, is given by

$$\Delta\lambda_{1/2} = 0.8\frac{F}{\bar{n}_{g1}L_c}\lambda^2. \tag{6.93}$$

Thus, for a reasonably narrow bandwidth filter, F cannot be too large. This implies that the effective index difference between the two coupled modes must be relatively large. The tuning enhancement effect has been used for implementation of widely tunable lasers, as will be discussed in Chapter 8.

Example 6.8 A grating-assisted codirectional coupler is fabricated using a 0.1 μm-thick bottom waveguide, separated by 0.3 μm from the top waveguide, whose thickness changes from 0.28 to 0.32 μm periodically. The waveguides are clad with InP, and their core indices are $n_{gd} = 3.4$ μm. The peak filter wavelegth is 1.55 μm, and the relative group effective index dispersion is $\Delta\overline{n}_g/\Delta\overline{n} = 3$. The average effective index of the top waveguide is $\overline{n}_t = 3.247$, and of the bottom waveguide is $\overline{n}_b = 3.184$.

Problem: (a) What is the necessary grating period for coherent coupling? (b) What is the coupler's optical bandwidth for the coupler length of $L_c = 243$ μm?

Solution: (a) From the definition of the grating period, Eq. (6.68),

$$\Lambda = \frac{\lambda}{\overline{n}_t - \overline{n}_b} = \frac{1.55\ \mu\text{m}}{3.247 - 3.184} = 24.3\ \mu\text{m}.$$

(b) To determine the coupler's bandwidth, we use Figure 6.20. From there, we conclude that the value for $\delta L_c = \pm 1.25$ at half maximum of the power transfer curve, which corresponds to the filter bandwidth. From the definition of detuning, Eq. (6.90), and substituting $\Delta\lambda/\lambda = (\omega - \omega_0)/\omega_0$, we have

$$\frac{\Delta\lambda}{\lambda} = \frac{\delta L_c}{L_c}\frac{\Lambda}{\pi}\frac{\Delta\overline{n}}{\overline{n}_g} = \frac{2.5}{243\ \mu\text{m}}\frac{24.3\ \mu\text{m}}{\pi}\frac{1}{3} = 0.0265.$$

Finally, the coupler's bandwidth is

$$\Delta\lambda = 1.55 \cdot 26.5 \text{ nm} = 41 \text{ nm}.$$

6.5 MODAL EXCITATION

In the process of interconnecting various optical waveguide components together, we must constantly deal with the problem of determining how much power is transmitted and reflected at the junctions. To derive general expressions for this problem we return to the normal mode expansion, Eq. (6.3), which allows us to express an arbitrary excitation field, $\mathscr{E}_e$, in terms of a superposition of the eigenmodes, $\mathscr{E}_m$, of the new waveguide section being excited. Figure 6.21 illustrates the problem schematically. For this exercise it is convenient to work in terms of the normalized amplitudes, since their magnitude squared gives the power flow independent of the impedance of the medium. Thus, we express eigenmode m of the waveguide to be excited as

$$\mathscr{E}_m = E_{0m}\mathbf{U}_m(x,y)e^{-j\beta_m z} = \sqrt{2\eta_m}a_m\mathbf{U}_m, \tag{6.94}$$

from which the power flow in the positive z-direction is given by

$$P_{zm} = |a_m|^2 = \int \frac{\mathscr{E}_m^* \cdot \mathscr{E}_m}{2\eta_m}\, dA, \tag{6.95}$$

provided the transverse mode function $\mathbf{U}_m$ is properly normalized according to Eq. (6.2).

According to Fig. 6.21, we assume the given arbitrary field is incident on our waveguide at $z = 0$. Using Eq. (6.3) then,

$$\mathscr{E}_e(0) = \sum_m \mathscr{E}_m(0) = \sum_m \sqrt{2\eta_m}a_m(0)\mathbf{U}_m. \tag{6.96}$$

Now, we dot multiply by $\mathbf{U}_l^*$ and integrate over the cross section. That is,

$$\int \mathbf{U}_l^* \cdot \mathscr{E}_e(0)\, dA = \int \sum_m \sqrt{2\eta_m}a_m(0)\mathbf{U}_l^* \cdot \mathbf{U}_m\, dA. \tag{6.97}$$

Next, we recognize that all terms in the summation except the lth term are zero and solve for the desired eigenmode amplitude at the entrance to our waveguide, $a_l(0)$.

$$a_l(0) = \frac{1}{\sqrt{2\eta_l}} \int \mathbf{U}_l^* \cdot \mathscr{E}_e(0)\, dA. \tag{6.98}$$

We can apply Eq. (6.98) for whatever eigenmode number, l, we desire. Fortunately, we only need to know the transverse mode function of the mode(s) in which we are interested.

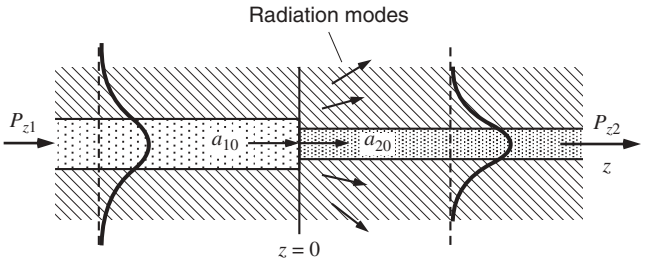


FIGURE 6.22: Illustration of butt-coupling between two waveguides and the excitation of the fundamental eigenmode of the right-hand guide by the left-hand guide.

To illustrate how to use Eq. (6.98) consider the common situation shown in Fig. 6.22, which depicts the joining of two dissimilar waveguides. As shown, power P_{z1} is incident from guide 1, and some fraction P_{z2} is transferred to the fundamental mode of guide 2. Other modes may also be excited, but we are not interested in them initially. (If the guide only supports a single guided mode, then the other modes of the summation are radiation modes, and these would generally be of little interest some distance away.)

The most difficult problem in this and many similar problems is to determine the excitation field on the *right side* of the boundary (region 2). Once it is determined, we can just plug into Eq. (6.98) for the desired a_{20}. Unfortunately, it is a rather complex problem to determine $\mathscr{E}_e(0)$ rigorously. Fortunately, for weak dielectric waveguides, it has been found that the field on the right side of the boundary is similar in shape to that on the left, but it is reduced in amplitude by some transmission coefficient. Also, in this case the transmission coefficient can be well approximated by using the waveguide effective indices in a plane wave formula. Taking all of this on faith, then,

$$P_{z2}(0^+) \approx P_{z1}(0^-)t^2, \tag{6.99}$$

where

$$t^2 \approx \frac{4\bar{n}_1\bar{n}_2}{(\bar{n}_1 + \bar{n}_2)^2}. \tag{6.100}$$

Thus,

$$\mathscr{E}_e(0^+) \approx \sqrt{2\eta_2'}ta_{10}(0^-)\mathbf{U}_{10}, \tag{6.101}$$

where η_2' is some effective impedance for the excitation field in region 2. It would be some weighted average of the impedances of the modes to be excited. Applying Eq. (6.98), we find

$$\frac{P_{20}(0^+)}{P_{10}(0^-)} = \left|\frac{a_{20}(0^+)}{a_{10}(0^-)}\right|^2 = \left|t\int \mathbf{U}_{20}^* \cdot \mathbf{U}_{10}\, dA\right|^2, \tag{6.102}$$

where we have neglected the difference between η_{20} and η_2', which is a good assumption if most of the energy goes into mode 0 or if the impedances of the primary modes excited are not so different. More generally, we could solve for P_{20}/P_{1m} to obtain the excitation of the mth mode.

6.6 TWO MODE INTERFERENCE AND MULTIMODE INTERFERENCE

The directional coupler, which we have solved exactly earlier in this chapter, can also be viewed as a five-layer waveguide. As indicated in Fig. 6.23, the first even and odd eigenmodes of the five-layer region can be superimposed to approximate the eigenmode of the individual three-layer guide, which excites the coupled region. The phase velocities of the odd and even five-layer eigenmodes are different, so that after some propagation distance, L_c, their superposition can result in most of the energy being aligned with the second high-index region, which in turn, can excite the second exit guide. Thus, most of the problem reduces to calculating the excitation of the different sets of eigenmodes at the boundaries between the uncoupled and coupled regions. Section 14.4 can be used to address this issue.

The concept of mode interference can be expanded even further, by replacing a five-layer waveguide with a single multimode waveguide, which is being excited by an input waveguide, as shown in Fig. 6.24. In this case, the input field will excite multimode waveguide's modes, all with different phase velocities. After certain propagation distance, their superposition will result in single, or multiple, regular or inverted interference images of the input field, as illustrated in Fig. 6.24. The propagation distance will depend on whether all or some of the modes in the multimode waveguide have been excited, corresponding to *g*eneral or *r*estricted interference, respectively. Restricted interference produces multiple images at shorter distances compared to the general interference case. Details about multimode interference (MMI) and different types of effects are discussed in Appendix 15. Here, we focus on the main results and their applications to photonic ICs.

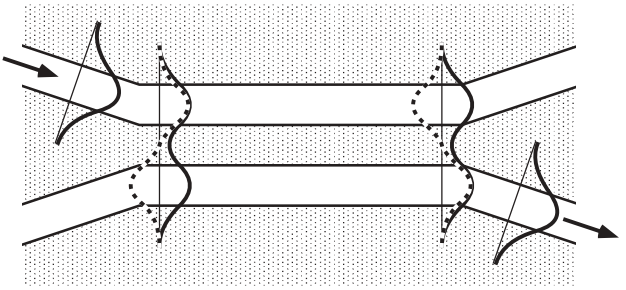


FIGURE 6.23: Directional coupler viewed in terms of the modal interference model. Input excites superposition of first even (solid) and odd (dotted) eigenmodes of the five-layer system. Over the coupling length the even mode travels some number of guided wavelengths, but the high-phase-velocity odd mode travels half a wavelength less.

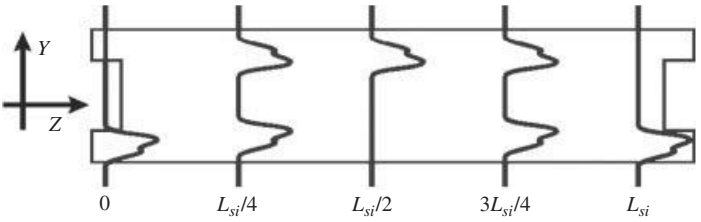


FIGURE 6.24: A schematic of a multimode waveguide, and its imaging properties—single and multiple images of the input field are reconstructed at various points.

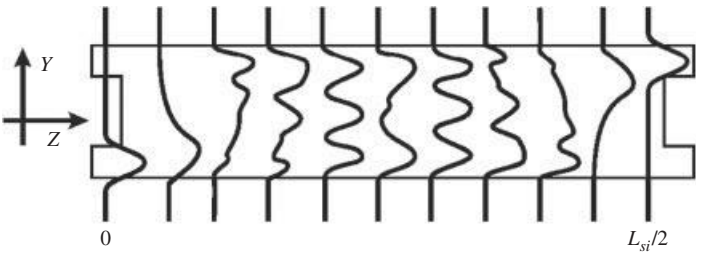


FIGURE 6.25: Calculated total electric field in a multimode waveguide, illustrating the general interference effect. Based on the final waveguide length chosen, a $1 \times N$ or $2 \times N$ splitter, $2 \times N$ coupler, or a 1×1 image reverser can be implemented.

One of the main applications of the multimode interference is for generation of multiple images of the input field. Multiple images are required for various photonic integrated circuits, such as interferometric modulators, power splitters and combiners. Of those, the most common are 1×2 and 2×2 light splitters and couplers. Recently, 2×4 MMI structures have also gained popularity as a way to realize optical hybrids used in coherent optical communications systems, discussed in Chapter 8. MMI structures are attractive and widely used due to their low inherent losses, large optical bandwidths, low polarization dependence, and simple fabrication, since MMI coupler properties, unlike directional couplers, do not depend on the reproducibility of submicrons gaps between waveguides [3]. MMI resolution depends on the number of modes supported, which depends on the waveguide width. At the same time, in Appendix 15 we have shown that the optimum MMI length depends from the MMI width squared, so it is clear that careful tradeoffs need to be made for loss and resolution.

In general, $1 \times N$ and $N \times N$ passive light splitters and couplers can be realized using the multimode interference effects; however, it is not possible to achieve arbitrary splitting ratios between different output waveguides using purely passive MMIs, in contrast to the directional couplers. The splitting ratio of the MMI couplers can be tuned through current injection induced refractive index change [4].

The most common implementation of an MMI splitter is a 1×2 symmetric interference based splitter. As explained in Appendix 15, for this component, the

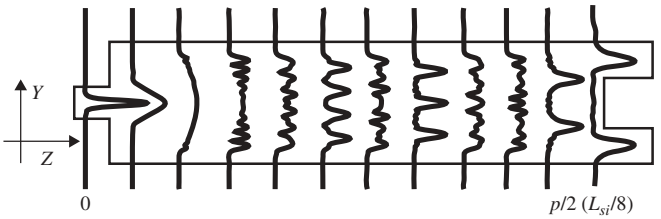


FIGURE 6.26: Response of a 1 × 2 splitter based on symmetric interference. Note the optimum length of this component, compared to the 1 × 2 splitter based on general interference case. The field response has been calculated and superimposed on the waveguide structure.

input waveguide is located at the center of the multimode waveguide, thus exciting only its even modes, and the 2-fold images are obtained at multimode waveguide distances of

$$L = \frac{p}{2}\left(\frac{L_{si}}{8}\right), \tag{6.103}$$

where p is an integer, and $L_{si} \simeq \frac{8n_l d_{eff}^2}{\lambda_0}$ is the beat length between the lateral plane wave in the structure, and the first guided mode of the MMI, as defined in Appendix 15. The output waveguides need to be located symmetrically along the x-axis, with equal spacings of d_{eff}/N starting from the bottom of the multimode waveguide. The effective width of the waveguide was defined by Eq. (6.55).

To obtain a 2 × 2 light coupler using general interference, per Appendix 15, the multimode waveguide distance needs to be

$$L = \frac{2p+1}{4}L_{si}, \tag{6.104}$$

and the output waveguide positions need to be positive and negative x position of the input waveguide, symmetric to the center of the MMI coupler.

With all MMI components, the practical limits in device implementation come from: the access waveguide width, which determines the minimum MMI width and; the component maximum length, since scattering losses will affect imaging resolution of the MMI components.

Example 6.9 Using a waveguide layer whose transverse effective index is $n_l = 3.185$, and which is clad by InP, we need to design a general interference based MMI 2 × 2 coupler operating at 1.55 μm. The MMI waveguide width is $d = 9$ μm, and the input/output access waveguide widths are 3 μm.

Problem: What is the optimum length of this MMI coupler?

Solution: From Table 6.1, the shortest general interference coupler that can be made has a length of $L = \frac{1}{4}L_{si}$, where $L_{si} = \frac{2\pi}{(k_0 n_l - \beta_0)}$, as defined in Eq. (A15.4).

Using the material from Appendix 3, we can simply calculate the propagation constant for the fundamental mode of this MMI coupler,

$$V = k_0 d\sqrt{n_{gd}^2 - n_{InP}^2} = \frac{2\pi}{1.55}9\sqrt{3.185^2 - 3.17^2} = 11.26$$

and

$$b = 1 - \frac{\ln(1 + \frac{V^2}{2})}{\frac{V^2}{2}} = 1 - \frac{\ln(1 + \frac{11.26^2}{2})}{\frac{11.26^2}{2}} = 0.934.$$

Now, we can calculate the effective index, similarly to Example 6.5,

$$\overline{n}_0 = \sqrt{b(n_{gd}^2 - n_{InP}^2) + n_{InP}^2} = \sqrt{0.934(3.185^2 - 3.17^2) + 3.17^2} = 3.184.$$

Now, $L_{si} = \frac{\lambda}{(\bar{n}_l - \bar{n}_0)} = 1550$ μm, and $L = \frac{1}{4}L_{si} = 387.5$ μm.

6.7 STAR COUPLERS

Star couplers are passive optical components that allow for the light from N input waveguides to be coupled to M output waveguides. This function is of interest in a number of applications involving photonic integrated circuits, such as novel diode laser implementations, laser array implementations, advanced integrated receiver designs, as well as arrayed waveguide grating based devices, which will be covered in the following section. The main difference between star couplers and MMI couplers, discussed in the previous section, is that star couplers rely purely on difraction of input light, whereas MMI components rely on beating of modes to form appropriate output images.

A schematic of the most common implementation of a waveguide star coupler is shown in Fig. 6.27. The power transfer between the input and output waveguides takes place in a free space region, a slab waveguide, which is placed between the two strips of input and output waveguides. Power from each of the input

TABLE 6.1: Summary of Properties of the General and Restricted (Paired and Symmetric) Interference Components

Interference Mechanism	General	Symmetric
Inputs × Outputs	$N \times N$	$1 \times N$
First single image distance	$L_{si}/2$	$L_{si}/8$
First N-fold image distance	$L_{si}/2N$	$L_{si}/8N$
Excitation restrictions	none	no excitation of modes $1, 3, 5, \ldots$
Input position(s)	any	$x = 0$
Output positions	same as inputs	at each d_{eff}/N from the MMI edge

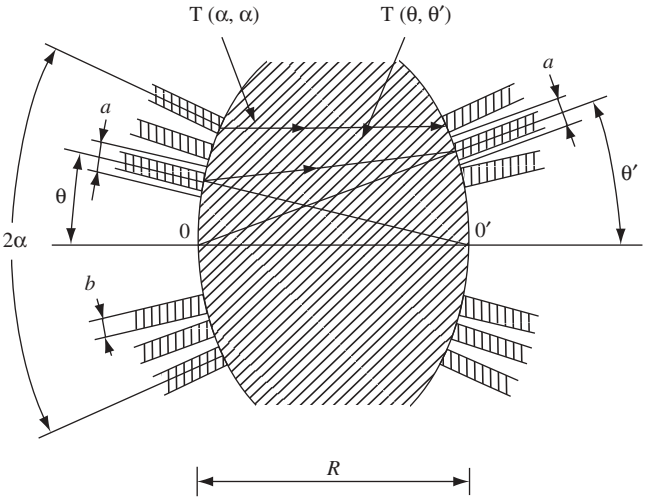


FIGURE 6.27: Illustration of a star coupler. (© IEE [5].)

waveguides is radiated in the free space region, and it is in part intercepted by the output waveguide array, and in part lost. The main causes for inherent losses of this scheme are in the finite output waveguide separation, leading to part of the light not being coupled, as well as in finite output imaging plane, causing part of the difracted light to be clipped, and thus lost. The input and the output waveguide arrays are located on two circles, which pass through the centers of each other. The key when designing a star coupler is to maximize the coupling efficiency for the output waveguides [5]. In addition, variable width output waveguides can be used to maintain the constant power splitting ratio between different output waveguides [6]. A useful method for analyzing the star coupler operation is through Gaussian approximation of the waveguide fields in input and output waveguides.

6.8 PHOTONIC MULTIPLEXERS, DEMULTIPLEXERS AND ROUTERS

Realization of photonic multiplexing, demultiplexing and wavelength routing functions is of interest for complex photonic ICs, such are wavelength division multiplexed (WDM) transmitters and receivers, or WDM wavelength selective lasers. In this section, we will introduce the key chip-scale de/multiplexing components, which utilize either waveguide, or etched difraction gratings. Their applications will be discussed in Chapter 8.

6.8.1 Arrayed Waveguide Grating De/Multiplexers and Routers

Arrayed waveguide grating (AWG) passive optical waveguide devices are the most widely deployed de/multiplexing devices in modern optical communication systems. Within the scope of our discussion, focusing on active, complex photonic ICs, AWGs can be used as building blocks to realize optical, chip-scale multiplexer, demultiplexer, cross-connect and router functions, as well as different digitally tunable diode laser cavities—all discussed in Chapter 8.

An AWG functions similar to an optical prism, by imaging the input optical field onto different spatial output locations based on the wavelength of the incoming light. Generally, an array of output waveguides is placed at desired locations on the output imaging plane in order to capture the imaged light. A 1×4 AWG demultiplexer is illustrated in Figure 6.28.

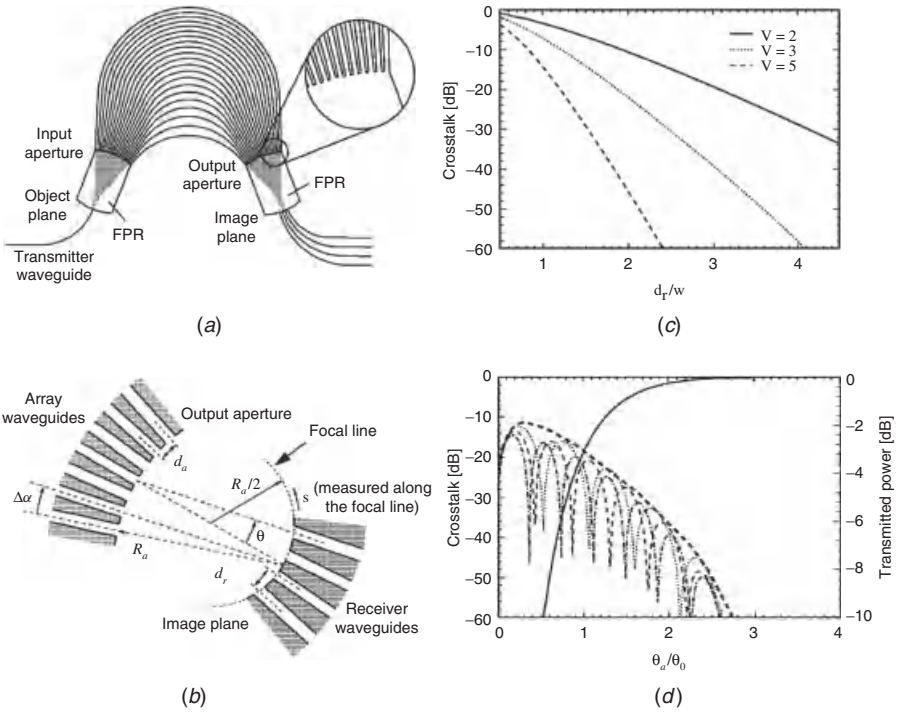


FIGURE 6.28: Illustration of an arrayed waveguide grating demultiplexer, per [7], indicating key device components and design parameters, such as the arrayed waveguide spacing d_a, and the output waveguide spacing d_r. Free propagation regions (FPR in the figure) are synonimous for star couplers. (© IEEE 1996, [7].)

AWGs consist of three main functional elements: two star couplers and a set of interconnecting arrayed waveguides, whose optical lengths, given by the product of their effective indices $\bar{n}_{wgd}$ and physical lengths, vary by a constant m number of wavelengths of light at a central wavelength λ_c from one waveguide to another,

$$\bar{n}_{wgd} \cdot \Delta L = m \cdot \lambda_c. \tag{6.105}$$

ΔL is the difference in physical lengths of adjacent waveguides in the array, which introduces a fixed optical phase difference of $\Delta\Phi$. The star couplers are generally based on the Rowland circle construction, where the radius of curvature of the input and output waveguide planes is one half of the radius of inner waveguide array planes. This arrangement ensures that the output light is focused along the circular output interface with the change in wavelength of the input signal.

In terms of device principle of operation, the light from the input waveguide radiates into the first star coupler slab waveguide and excites the modes of the arrayed waveguides at the star coupler output. After traveling through the arrayed waveguides, the light from the waveguides is difracted into the output star slab waveguide, where it constructively converges in one focal point at the output of the star coupler. This is accomplished because the path length difference between the arrayed waveguides results in a relative phase delay $\Delta\Phi$ in each waveguide, which changes with wavelength. This results in a rotation of the field phase front in the second slab and a translation of the location of the focal point as a function of the wavelength. The amount of dispersion and shift of the focal point as a function of wavelength is one of the design parameters for an AWG.

Based on the classic diffraction theory, and treating the phased array as a number of diffraction slits, we can derive the arrayed waveguide grating equation. For a signal entering the first star coupler under an angle ϕ relative to the center line, the relative phase difference between the signal paths inside the star coupler that land to two adjacent array waveguide openings will be given by $\bar{n}_{FPR} \cdot d_a \cdot \sin\phi$, where $\bar{n}_{FPR}$ is the transverse effective index of the star coupler slab, and d_a is the center-to-center spacing between the array waveguides, as illustrated in Fig. 6.28. Additional phased delays between these two paths will be introduced due to the physical length differences in the adjacent array waveguides, as well as due to the position of the output waveguide, at an angle θ relative to the centerline of the output star coupler. The Bragg's law for this type of diffraction grating is then given by

$$\bar{n}_{FPR} \cdot d_a \cdot \sin\phi + \bar{n}_{wgd}\Delta L + \bar{n}_{FPR} \cdot d_a \cdot \sin\theta = m \cdot \lambda. \tag{6.106}$$

Differentiating Eq. (6.106) with respect to the wavelength λ, keeping in mind that for realistic cases angle θ is small, and for simplicity, assuming that $\phi = 0$, we can obtain the expression for angular dispersion of an AWG(R) as

$$\frac{d\theta}{d\lambda} = \frac{\Delta L \cdot \bar{n}_g}{\lambda_c \cdot \bar{n}_{FPR} \cdot d_a}, \tag{6.107}$$

where $\bar{n}_g$ is the group index of the array waveguide. By defining the output waveguide separation in the focal plane as $ds = R_a \cdot d\theta$, the dispersion of the focal point position along the output imaging plane, s, with respect to wavelength λ is given by

$$\frac{ds}{d\lambda} = \frac{\Delta L \cdot \bar{n}_g \cdot R_a}{\lambda_c \cdot \bar{n}_{FPR} \cdot d_a}. \tag{6.108}$$

This important expression, relating the star coupler size, array waveguide spacing, and ΔL is key for design of AWG components, and its use will be illustrated later in this section.

In a configuration where we have an equal number of input and output waveguides, with proper selection of device parameters and dimensions, it is possible to realize a full optical router. In it, input light from any of the input waveguides can be imaged to any of the output waveguides, through control of the wavelength of light in input waveguides, and this particular device is called an arrayed waveguide grating router (AWGR). It is important to note that the responses of both an AWG and an AWGR are periodic, since for every change in relative optical phase difference $\Delta\Phi$ of 2π, the optical image occurs in the same physical position. The wavelength change that corresponds to the relative optical phase change of 2π determines the free spectral range (FSR) of an AWG. For an AWGR, ΔL and FSR can be related from the fact that FSR represents a frequency shift for which the phase shift in the array $\Delta\Phi = 2\pi$. From here, we have

$$FSR = \frac{c}{\bar{n}_{wgd}\Delta L}. \tag{6.109}$$

Arrayed waveguide grating devices have many degrees of freedom in their geometric parameter selection, and various design strategies are possible. Here, we outline a strategy developed by M. K. Smit [7]. This design strategy consists of: selecting the center wavelength of operation λ_c; the optical channel spacing $\Delta\lambda$, which will determine the required change of wavelength (dispersion) to steer the output light beam to one of the adjacent output waveguides; and the total number of channels N_{ch}, based on the required device function, which determines the number of input and output waveguides. Per example, a four-channel demultiplexer would require $N_{ch} = 4$.

Many important device physical features, such as the access and arrayed waveguide widths and their minimum separation, will be dictated by the maximum resolution of the device fabrication process. The arrayed waveguide spacing d_a should be as low as possible, since any light not coupled into the array will contribute to the device insertion losses. The output waveguide spacing, d_r in Figure 6.28, will directly impact the crosstalk of the AWG, since the output image at a given spatial location will have exponential tails that can couple into the adjacent output waveguides, depeding on the distance. Note that this spacing is defined in s coordinate, along the focal line of the output plane. The lower bound of this parameter can be determined, based on the desired crosstalk and receiver waveguide architecture, using the normalized crosstalk plot in Figure 6.28.

The value of key free design parameters, the length of the star coupler, R_a, and the number of array waveguides, N_a, depends on the maximum tolerable output power nonuniformity between different output waveguides, on the maximum apperturing crosstalk, and on the maximum acceptable losses in going through the star coupler. To gain an understanding of the key relationships between these different geometric and performance parameters, we approximate the input and arrayed waveguide fields as Gaussian, and then apply the known relations for Gaussian far-field to our problem.

As the input light is radiated freely into the star coupler, the size of the phased array needs to be determined based on the input beam diffraction over the length of the star coupler. The angular size of a Gaussian far-field Θ_0 emmitted from a waveguide is given by,

$$\Theta_0 \approx \frac{\lambda_c}{\bar{n}_{star}\, w_{input}\left(0.5 + \frac{1}{V-0.6}\right)}, \tag{6.110}$$

where in the case of AWG/AWGR, w_{input} is the input waveguide width, and V is its normalized V parameter, introduced in Appendix 3. The intensity of the far-field is the given by

$$I(\Theta) = I_0 e^{\frac{-2\Theta^2}{\Theta_0^2}}. \tag{6.111}$$

As the intensity from the individual waveguides will reduce by moving away from the central plane, so will their sum, which produces the focal field at an AWG output. Therefore, changing the wavelength in the input waveguide will cause the intensity of the output field to change as it physically moves along the output plane, following an intensity envelope set by the Gaussian individual array waveguides' far-field distribution. If we define the nonuniformity of the output power as the ratio (in decibels) between the intensities of the central and the outer channel, we have that

$$L_u = -10 \cdot \log(e^{\frac{2\Theta_{max}^2}{\Theta_0^2}}) \approx 8.7 \cdot \frac{\Theta_{max}^2}{\Theta_0^2}, \tag{6.112}$$

where Θ_{max} is the angular position of the outermost output waveguide, relative to the centerline. By setting the maximum acceptable nonuniformity, we can calculate the maximum dispersion angle. For a router, since the free spectral range is determined by the number of channels and channel spacing, $L_u \approx 3$ dB. Realizing, from Figure 6.28 that $\Theta_{\max} = s/R_a$, the star coupler's length will given by

$$R_a = (N_{ch} - 1) \cdot \frac{d_r}{2\Theta_{max}}, \tag{6.113}$$

where we have used the knowledge of output waveguide spacing and the number of channels to compute the maximum value for the s coordinate.

The finite apperture (star coupler size) will cause part of the input field intensity to be lost and the sidelobes of the output far field to increase, due to the input field truncation. Therefore, the maximum aperture size, Θ_a, needs to be detemined based on the crosstalk and loss specification, using Figure 6.28 (b) adapted for a particular waveguide platform used. To capture most of the light, the aperture of the input waveguide needs to be taken as approximately $\Theta_a = 3 \cdot \Theta_0$. The number of arrayed waveguides will also uniquely be determined by the choice of the aperture size. The number of the waveguides in the array that are required is then given by

$$N_a = \frac{2\Theta_a R_a}{d_a} + 1. \tag{6.114}$$

Finally, the length difference between the adjacent array waveguides can be determined from the AWG diffraction Eq. (6.108), since the output waveguide spacing d_r and the frequency channel spacing Δf_{ch} determine the amount of dispersion $ds/d\lambda = d_r/\Delta f_{ch}$,

$$\Delta L = \frac{d_r \cdot f_c \cdot \bar{n}_w \cdot d_a}{\Delta f_{ch} \cdot \bar{n}_{FPR} \cdot R_a}. \tag{6.115}$$

To be exact, the value obtained from this expression should be compared with the value obtained from Eq. (6.105), and the value for the receiver spacing d_r can be adjusted up to exactly meet the phase relationship (and thus have m be an integer).

A number of modifications to this basic design procedure can be made to achieve polarization independent operation, and a more flattened passband response. Interested readers are referred to [7] for details.

The key limitation in terms of an AWG performance comes from the introduction of random phase errors inside the arrayed waveguide grating during fabrication. In optimized device designs, these phase errors will impact the purity of the spectral response and define the minimum crosstalk between different channels. In planar lightwave circuits, phase errors can be corrected after fabrication through individual waveguide trimming; however, this type of procedure is not practical in active devices that utilize AWGs.

With all these parameters defined, the design problem is reduced to geometric layout of the AWG component, and a number of different layout architectures have been implemented [7], [8]. Due to a large number of wave guides involved, and the sensitivity of device performance on layout, automated layout tools are generally needed for lithographic mask design.

Example 6.10 A 1x8 arrayed waveguide grating–based demultiplexer needs to be designed in Indium Phosphide, to operate with the central wavelength of 1550 nm. The optical channels that need to be multiplexed are spaced by 100 GHz. The waveguide structure used is a burried ridge, having a 300 nm quarternary InGaAsP core burried in InP, with transverse effective index of $\bar{n}_{tc} = 3.2477$ in the center

of the waveguide, and $\bar{n}_{tl} = 3.2146$ in the lateral slices. Arrayed waveguides are assumed to be 2 μm wide, with the effective index of 3.2381. Input and output waveguides are 3 μm wide, with the effective index of 3.2421. Assume that the best lithographic resolution that can be achieved repeatably is 0.7 μm.

Problem: Determine the dimensions of the star coupler, arrayed waveguide length difference ΔL and the required number of waveguides in the array, N_a.

Solution: Since the loss of the multiplexer will directly depend on the spacing of the arrayed waveguides, this spacing should be chosen to be as low as possible. With our limitations of 0.7 μm resolution, we have that $d_a = 0.7\ \mu\text{m} + 2\ \mu\text{m} = 2.7\ \mu\text{m}$.

Next, we need to determine the receiver waveguide spacing, from the maximum crosstalk level specification, in Figure 6.28. The normalized frequency V for the receiver waveguide can be calculated using parameters from Example 1.1,

$$V = k_0 d \left(\bar{n}_{tc}^2 - \bar{n}_{tl}^2\right)^{\frac{1}{2}} = \frac{2\pi}{1.55\ \mu\text{m}} 3\ \mu\text{m}\ \left(3.2477^2 - 3.2146^2\right)^{\frac{1}{2}} = 5.6$$

For the crosstalk level of −40 dB, based on Figure 6.28, the ratio of the receiver spacing to waveguide width needs to be $d_r/w \approx 1.8$. From here, we get the minimum receiver spacing as $d_r = 3 \cdot 1.8 = 5.4\ \mu\text{m}$.

Next, we need to calculate the Rowland circle radius and the number of arrayed waveguides. The angular size of the input beam is, per Eq. (6.110),

$$\Theta_0 = \frac{\lambda_c}{\bar{n}_{tc} w_{input} \left(0.5 + \frac{1}{V-0.6}\right)\sqrt{2\pi}} = \frac{1.55\ \mu\text{m}}{3.2477 \cdot 3\ \mu\text{m} \cdot (0.5 + 0.20)\sqrt{2\pi}} = 0.090\ \text{rad}.$$

Using Eq. (6.112), we can determine the maximum dispersion angle assuming that we can tolerate 3 dB nonuniformity,

$$\Theta_{max} = \sqrt{\frac{L_u \cdot \Theta_0^2}{8.7}} = \sqrt{\frac{3 \cdot 0.090^2}{8.7}} = 0.053\ \text{rad}.$$

Now, we can calculate the radius of the Rowland circle using Eq. (6.113),

$$R_a = (N_{ch} - 1) \cdot \frac{d_r}{2 \cdot \Theta_{max}} = 7 \cdot \frac{5.4\ \mu\text{m}}{2 \cdot 0.053} = 357.95\ \mu\text{m}.$$

The length difference between the waveguides, ΔL, can be calculated using the expression 6.115,

$$\Delta L = \frac{d_r \cdot f_c \cdot \bar{n}_w \cdot d_a}{\Delta f_{ch} \cdot \bar{n}_{tc} \cdot R_a} = \frac{5.4\ \mu\text{m}\ \cdot 193548.387\ \text{GHz}\ 3.2381 \cdot 2.7}{100\ \text{GHz}\ 3.2477 \cdot 357.95\ \mu\text{m}} = 78.60\ \mu\text{m}.$$

Finally, we can calculate the number of arrayed waveguides, given by the Eq. (6.114),

$$N_a = \frac{2 \cdot 3\Theta_0 R_a}{d_a} + 1 = \frac{2 \cdot 3 \cdot 0.090 \cdot 357.95}{2.7} + 1 = 72.6.$$

Therefore, 73 waveguides can be chosen.

6.8.2 Echelle Grating based De/Multiplexers and Routers

In the case of an AWG-based device, the functions of optical field radiation and phase control and reconstruction are performed by three separate funtional elements: the input star coupler, the arrayed waveguides and the output star coupler. One way to reduce the footprint of the multiplexer/demultiplexer function would be to replace the arrayed waveguide grating with a real diffraction step Echelle grating. In this structure, the single star coupler in a Rowland circle configuration becomes both the diffracting and the focusing element. This is illustrated in Figure 6.29, where an Indium Phosphide demultiplexer based on a curved grating (CG) on a Rowland circle is shown. The light from the input waveguide difracts into the slab region, and it reflects from the grating teeth, which introduce appropriate phase shift to refocus it onto an output waveguide, which is also located along the

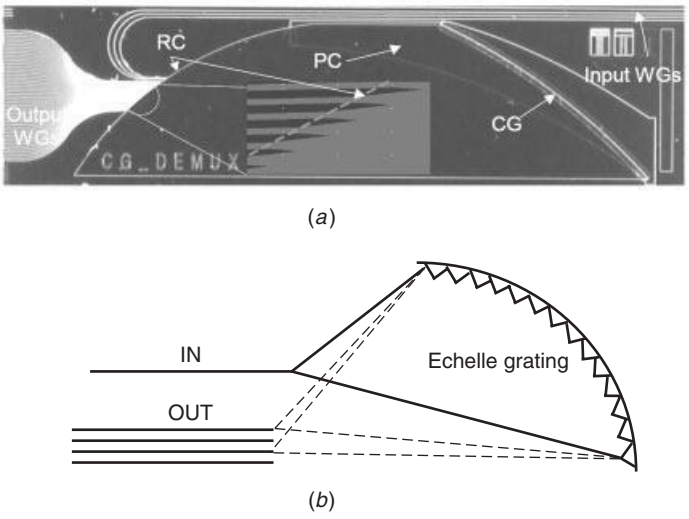


FIGURE 6.29: (a) A photograph of a waveguide Rowland circle (RC) grating demultiplexer. The device contains a curved grating (CG) of the Echelle type, and a polarization compensator (PC) birefringent waveguide. Reprinted with permission from ETRI Journal, [9] (b) Schematic illustrating the operation of this demultiplexer. This device represents a "folded" AWG, where the phase difference in the diffraction equation is introduced by the curved grating and the relative positions of the input and output waveguides.

Rowland circle. This particular device implementation also has a polarization compensation region, consisting of a birefringent waveguide, to achieve polarization independent operation. The number of grating teeth is generally a lot bigger than the number of arrayed waveguides in an AWG, thus increasing the finesse of the grating. The teeth are realized using high contrast interfaces. Application of this demux design to tunable lasers will be discussed in Chapter 8.

6.9 CONCLUSIONS

In this chapter we have introduced perturbation and coupled-mode techniques to obtain closed-form analytic solutions to relatively complex problems. These problems are limited only by the requirement that the dielectric perturbation, which creates a change in propagation constant and/or coupling between different modes, is small. Thus, the techniques are very powerful in analyzing many important practical problems. Even in cases where the perturbations are sizable, the techniques are still useful in obtaining approximate results as well as in determining the dependencies on the various parameters of the problem.

For larger perturbations, where the validity of the perturbation and coupled-mode approaches are not good, it is generally necessary to attack the entire problem, usually using some numerical technique. In many cases, however, exact solutions are possible using purely analytical techniques. For example, the case of the quantum well placed within a separate-confinement waveguide discussed above, is really a five-layer waveguide problem, for which analytic solutions exist. In fact, using the techniques to be introduced in the next chapter, waveguides of many layers can be analyzed.

In addition, we have dealt with multimode interference, a generalization of the two mode interference problem, which allows for creation of passive optical light splitters and couplers, of interest in photonic integrated circuits. Finally, we examined more passive waveguide-based components, star couplers, arrayed waveguide and Echelle gratings, from the aspect of their use in diode lasers and active PICs.

REFERENCES

1. Luo JH, Zory PS. Ray optics determination of the DFB coupling coefficient in separate confinement and multiquantum-well laser structures. *IEEE J Quantum Electron* 1994;QE-**30**: 10.
2. Chen K-L, Wang S. An approximate expression for the effective refractive index in symmetric DH lasers. *IEEE J Quantum Electron* 1983;QE-**19**: 1354.
3. Soldano LB, Pennings ECM. Optical munti-mode interference defices based on self-imaging: principles and applications. *Journal of Lightwave Technology* 1995;**13**(4): 615–627.
4. Leuthold J, Joyner CW. Multimode interference couplers with tunable power splitting ratios. *Journal of Lightwave Technology* 2001;**19**(5): 700–71-07.

5. Dragone C. Efficient $N \times N$ star coupler based on Fourier optics. *Electronics Letters* 1988;**24**(15): 942–944.
6. Okamoto K, Okazaki H, Ohmori Y, Kato K. Fabrication of large scale integrated-optic NN star couplers. *IEEE Photonics Technology Letters* 1992;**4**(9): 1032–1035.
7. Smit MK, Van Dam C. PHASAR-based WDM-devices: Principles, design and applications. *IEEE J Selected Topics in Quantum Electron* 1996;**2**: 236.
8. Kaneko A, Goh T, Yamada H, Ogawa L. Design and applications of silica-based planar lightwave circuits. *IEEE J Selected Topics in Quantum Electron* 1999;**5**: 1227.
9. Oh Kee Kwon, Chul Wook Lee, Dong Hun Lee, Eun Deok Sim, Jong Hoi Kim, Yong Soon Baek. InP-based polarization-insensitive planar waveguide concave grating demultiplexer with flattened spectral response. *ETRI Journal* 2009;**31**(2): 228–230.

READING LIST

Ebeling KJ. *Integrated opto-electronics*. Berlin: Springer-Verlag; 1993.Chapters 5 and 6.

Haus HA. *Waves and fields in optoelectronics*. Englewood Cliffs, NJ: Prentice Hall; 1984. Chapters 7 and 8.

Lee DL. *Electromagnetic principles of integrated optics*. New York: Wileu; 1986. Chapter 8.

Yariv A, Yeh P. *Optical waves in crystals*. New York: Wiley; 1984. Chapters 6 and 11.

PROBLEMS

1. A symmetric graded-index slab waveguide has a Gaussian transverse mode and an effective index of 3.5. Half of its energy is contained in the central 300 nm of the guide at 1.55 μm.

(a) What is $\Delta\beta$, if a 150-nm-wide region from the center to one side of the mode is changed so that the index changes by 0.02?

(b) If the changed region of (a) is only inserted periodically along the waveguide length with a period of $\lambda/2n$ and a 50% duty cycle, what is κ for the resulting grating?

(c) How long must the periodically perturbed waveguide section be for a power reflection coefficient of 0.5?

2. Derive an expression for the effective reflection plane separation from a grating's start, L_{eff}, using coupled-mode analysis.

3. Using the coupled-mode analysis, show that L_{eff} is also related to the energy decay length L_p in a long grating.

4. Using coupled-mode theory for the grating mirror, derive and plot the threshold modal gain versus the detuning parameter, δ, for a DBR laser that has one cleaved and one grating mirror. The grating mirror also is terminated in a cleave with a relative reflection phase of ϕ. Assume $\lambda = 1.55$ μm, no passive cavity section, an active length of approximately 250 μm, an internal loss

of 15 cm^{-1} throughout, a grating $\kappa = 50\ \text{cm}^{-1}$, and that a mode is aligned with the mirror Bragg condition at a wavelength 1.55 μm. Plot curves for $\phi = 0, 90^\circ$, and 180° with grating κL_g of both 0.5 and 1.

5. Show how a small propagation loss modifies the grating reflection and transmission coefficients. Express the results as the lossless expressions times multiplicative factors.

6. Give an expression for the threshold condition (characteristic equation) for a quarter-wave-shifted DFB laser using coupled-mode results. Neglect any facet reflections.

7. From the coupled-mode characteristic equation of a standard DFB laser with no facet reflections, calculate for the first two modes the threshold modal gain $\langle g \rangle_{xy}$ and deviation from the Bragg condition, δL, for

(a) A 1.55 μm InGaAsP/InP DH laser with a $\kappa L = 1$ and a length $L = 400$ μm.

(b) An analogous quarter-wave-shifted DFB, as considered in Problem 6.6, with the same grating and overall length.
Assume the internal modal loss is 20 cm^{-1}. Give the MSR for cases (a) and (b). Assume $\eta_r = 1, \beta_{sp} = 10^{-4}$, and $I = 2I_{th}$.

8. For a passive grating section formed of 1.3 μm bandgap waveguide material clad by InP with a symmetric triangular grating formed on one side of the 1.3 μm material so that the waveguide's width varies from 200 nm to 210 nm, calculate its coupling constant κ using coupled-mode analysis.

9. Use Eq. (6.60) to determine the normalized threshold modal gain, $(\Gamma_{g_{th}} - \alpha_i)L_g$, and wavelength deviation from the Bragg condition, δL_g, for DFB lasers with facet reflections. Plot both values versus the phase of facet reflection #2, ϕ_2, for $\kappa L_g = 1$ and mirror reflection values of

(a) $r_1 = 0$ and $|r_2| = 0.566$,

(b) $r_1 = 0.566$ and $|r_2| = 0.566$, and

(c) $r_1 = 0.566e^{j\pi/2}$ and $|r_2| = 0.566$.
If necessary, assume $\alpha_i = 15\ \text{cm}^{-1}, L_g = 300$ μm, and $\lambda = 1.55$ μm. Use the analytic equations of this chapter.

10. Two identical 3-μm-wide channel waveguides with center indexes 0.04 higher than the surrounding cladding material ($n = 3.5$) are formed with their center axes 20 μm apart. The lateral confinement factors are 40%. It is found that the lateral field of the first guide has decayed to 10% of its peak value at the center of the second guide.

(a) What is $\kappa_{21} = \kappa_{12}$? Approximate integrals, do not calculate exact mode shapes.

(b) What is the coupling length for 100% energy transfer?

11. One of the waveguides from Examples 6.4 and 6.5 is replaced by a thicker one, such that its effective index is higher and such that an axial periodic variation of

its thickness is required for coherent coupling. A rectangular grating of height 0.04 μm is used to provide a thickness that changes from 0.28 to 0.32 μm periodically. Assume a peak filter wavelength of 1.55 μm, and a relative group effective index dispersion of $\Delta\bar{n}_g/\Delta\bar{n} = 3$.

(a) What is the confinement factor and the effective index of this new guide, neglecting any interactions with the other waveguide?

(b) What is the necessary grating period for coherent coupling?

12. For the structure from the previous example,

(a) What is the grating assisted codirectional coupler optical bandwidth in nanometers?

(b) If the index of the slab region of the thick guide is now decreased by $\Delta n = 0.01$, by how much is the peak filter wavelength shifted?

(c) What are κ_{12} and κ_{21} and the coupling length?

13. What is the 3 dB bandwidth of a directional-coupler filter formed by different width and different index difference guides on InP such that the effective indexes of both are 3.5 at 1.55 μm, but at 1.50 μm the effective indexes of the two guides are 3.51 and 3.52, respectively? At 1.55 μm, $\kappa_{21} \approx \kappa_{12} = 0.01\ \mu\text{m}^{-1}$.

14. Two three-layer slab waveguides formed of 1.3 μm bandgap InGaAsP/InP are butt-coupled together such that their center axes are aligned. The left guide is 200 nm thick and the right guide is 400 nm thick. Calculate the power loss in coupling across the boundary for the fundamental transverse modes at 1.55 μm. Do the calculation twice, once for coupling left-to-right and once for coupling right-to-left.

15. Determine the length of a symmetric interference 1x2 MMI splitter operating at 1.55 μm, whose multimode waveguide width is 12 μm. The waveguide index is 3.185, and the device is clad with InP.

16. A 1×16 arrayed waveguide grating based demultiplexer needs to be designed in Indium Phosphide, to operate with the central wavelength of 1550 nm. The optical channels that need to be multiplexed are spaced by 100 GHz. The waveguide structure used is a burried ridge, having a 300 nm quarternary InGaAsP core burried in InP, with transverse effective index of $\bar{n}_{tc} = 3.2477$ in the center of the waveguide, and $\bar{n}_{tl} = 3.2146$ in lateral slices. Arrayed waveguides are assumed to be 2.2 μm wide, with the effective index of 3.2381, and input and output waveguides that are 3 μm wide, with the effective index of 3.2421. Assume that the best lithographic resolution that can be achieved repeatably is 0.5 μm.

(a) Determine the dimensions of the Star Coupler.

(b) Determine the arrayed waveguide length difference ΔL.

(c) Determine the required number of waveguides in the array N_a.

CHAPTER SEVEN

Dielectric Waveguides

7.1 INTRODUCTION

Thus far, we have managed to introduce quite a bit of material that made use of the transverse mode function, U, of some dielectric waveguide without knowing much about its actual form, aside from the brief introduction in Appendix 3. We have chosen this approach in part to emphasize that for many cases one does not need to know the details of all the possible transverse modes that might be supported by some dielectric layer structure. The primary reason has been to maintain a focus on the active device theme of this text and avoid distractions. In fact, we still do not intend to give an extremely detailed treatment of dielectric waveguides because that lies outside the theme of this text. Rather, we wish to introduce several different approaches to solving dielectric waveguide problems to both complement the field theory approach in Appendix 3 as well as provide the student with a broader, and perhaps more intuitive, understanding of the nature of waveguiding in these structures.

We begin by reviewing the reflection of plane waves that are incident at an arbitrary angle from a plane dielectric interface as illustrated in Fig. 7.1. The boundary conditions lead to general expressions for the reflection coefficients of both TE and TM polarizations. If the medium containing the incident plane wave has a higher index of refraction than that beyond the boundary, then we find that total internal reflection is possible. On the incident side of the dielectric boundary, the incident and reflected plane waves create a standing wave with a standing wave-ratio that becomes infinite for incident angles beyond the critical angle. Also, in this case we find that although the reflection coefficient is positive real (has a reflection

Diode Lasers and Photonic Integrated Circuits, Second Edition.
Larry A. Coldren, Scott W. Corzine, and Milan L. Mašanović.
© 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

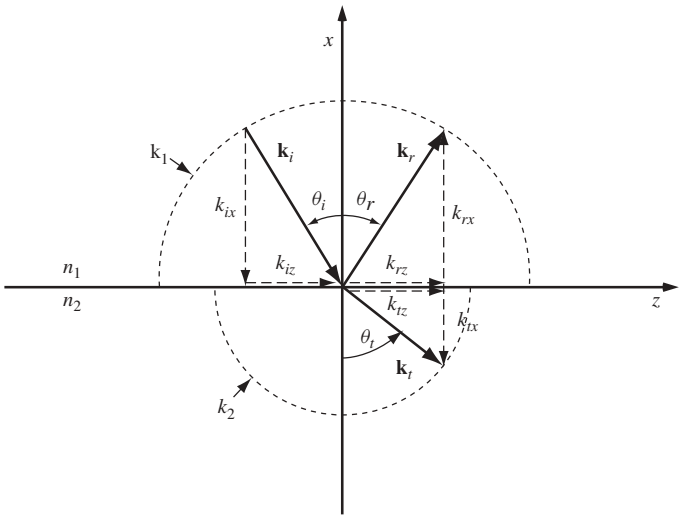


FIGURE 7.1: Illustration of a plane wave incident on a planar boundary. Angle of incidence is less than the critical angle.

phase of 0°) for incident angles smaller than the critical angle, it becomes complex beyond this angle (i.e., has a nonzero reflection phase as well as unity magnitude). Thus, for angles less than the critical angle, a standing wave maximum occurs at the boundary, but for incident angles beyond the critical angle the standing wave maximum moves back from the boundary. In what follows, we quantify the above observations.

7.2 PLANE WAVES INCIDENT ON A PLANAR DIELECTRIC BOUNDARY

Referring to the nomenclature introduced in Fig. 7.1 and earlier in Chapter 2 and Appendix 3, we express the incident, reflected, and transmitted fields for the TE and TM cases as follows:

TE

$$\begin{aligned}
\mathscr{E}_i(x,z) &= E_i(0,z)\hat{\mathbf{e}}_y e^{jk_{ix}x} e^{-jk_z z} \\
\mathscr{E}_r(x,z) &= E_r(0,z)\hat{\mathbf{e}}_y e^{-jk_{ix}x} e^{-jk_z z} \\
\mathscr{E}_t(x,z) &= E_t(0,z)\hat{\mathbf{e}}_y e^{jk_{tx}x} e^{-jk_z z}
\end{aligned} \tag{7.1}$$

TM

$$\begin{aligned}
\mathscr{H}_i(x,z) &= H_i(0,z)\hat{\mathbf{e}}_y e^{jk_{ix}x} e^{-jk_z z} \\
\mathscr{H}_r(x,z) &= H_r(0,z)\hat{\mathbf{e}}_y e^{-jk_{ix}x} e^{-jk_z z} \\
\mathscr{H}_t(x,z) &= H_t(0,z)\hat{\mathbf{e}}_y e^{jk_{tx}x} e^{-jk_z z}
\end{aligned} \tag{7.2}$$

Because the tangential electric fields, $\mathscr{E}$, and magnetic fields, $\mathscr{H}$, must be equal at $x = 0$, we have already used the fact that $k_{iz} = k_{rz} = k_{tz} = \beta$. Also, from this we have that $\theta_r = \theta_i$, and Snell's law, $k_i \sin\theta_i = k_t \sin\theta_t$. As always the propagation vector components are related by $k_{ix}^2 + k_{iz}^2 = k_i^2 = k_1^2 = k_0^2 n_1^2 = k_0^2 \varepsilon_1$, etc.

In any medium, the magnetic field of a plane wave is related to the electric field by Maxwell's curl equation. For the assumed forms of Eqs. (7.1) and (7.2) we have

$$\mathscr{H} = \frac{1}{\omega\mu}\mathbf{k} \times \mathscr{E}, \tag{7.3}$$

where μ is the magnetic permeability of the medium. Now, using Eqs. (7.1) and (7.2) in Eq. (7.3), and applying the boundary condition that the electric and magnetic fields are continuous at $x = 0$, we can solve for the ratio of the reflected to the incident electric fields [1, 2]. For the TE case with equal permeabilities,

$$\left.\frac{E_r}{E_i}\right|_{x=0} = r^{TE} = \frac{k_{ix} - k_{tx}}{k_{ix} + k_{tx}}, \tag{7.4}$$

and for the TM case,

$$\left.\frac{E_r}{E_i}\right|_{x=0} = r^{TM} = \frac{k_{ix} - \frac{\varepsilon_1}{\varepsilon_2}k_{tx}}{k_{ix} + \frac{\varepsilon_1}{\varepsilon_2}k_{tx}}. \tag{7.5}$$

For the case illustrated in Fig. 7.1, the index of refraction in region 1 is larger than that of region 2. Thus, as the incident angle is increased, at some point, k_{iz} equals k_2. This is called the critical angle, θ_c, defined explicitly using Snell's law by $\sin\theta_c = n_2/n_1$. For larger incident angles, k_{tx} must be imaginary to satisfy

$$k_{tx}^2 \equiv k_t^2 - k_{tz}^2 = k_t^2 - k_{iz}^2 = k_2^2 - \beta^2. \tag{7.6}$$

That is, for $k_{iz} > k_2, \theta_i > \theta_c$, and

$$k_{tx} = \pm j\sqrt{\beta^2 - k_2^2} = -j\gamma_{tx}, \tag{7.7}$$

where the sign of γ_{tx} is chosen for a decaying solution in region 2. This situation is shown in Fig. 7.2.

Plugging Eq. (7.7) into Eq. (7.4), we see that beyond the critical angle,

$$r^{TE} = \frac{k_{ix} + j\gamma_{tx}}{k_{ix} - j\gamma_{tx}}, \tag{7.8}$$

and ($\theta_i > \theta_c$)

$$r^{TM} = \frac{k_{ix} + j\frac{\varepsilon_1}{\varepsilon_2}\gamma_{tx}}{k_{ix} - j\frac{\varepsilon_1}{\varepsilon_2}\gamma_{tx}} \tag{7.9}$$

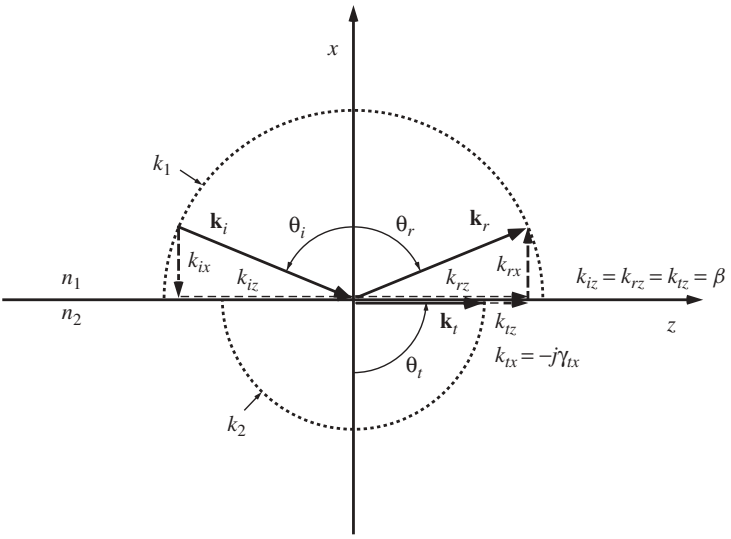


FIGURE 7.2: Plane wave incident at dielectric interface with the angle of incidence larger than the critical angle.

In both cases, we see that the magnitudes are unity, but the reflected wave has a phase angle, ϕ. That is, $r = |r|e^{j\phi}$, where from Eq. (7.8) for the TE case,

$$\phi^{TE} = \tan^{-1}\left[\frac{\gamma_{tx}}{k_{ix}}\right] - \tan^{-1}\left[\frac{-\gamma_{tx}}{k_{ix}}\right], \tag{7.10}$$

or

$$\phi^{TE} = 2\tan^{-1}\left[\frac{\gamma_{tx}}{k_{ix}}\right], \tag{7.11}$$

where

$$\frac{\gamma_{tx}}{k_{ix}} = \frac{\sqrt{1-(k_2/\beta)^2}}{\sqrt{(k_1/\beta)^2-1}} \quad \text{and} \quad \theta_i > \theta_c. \tag{7.12}$$

For the TM mode using Eq. (7.9) one obtains an equation analogous to Eq. (7.11) in which $\varepsilon_1/\varepsilon_2$ multiplies γ_{tx}.

As indicated in Fig. 7.3, the reflection phase angle is zero for incident angles up to the critical angle, where it begins to increase monotonically toward 180°. This results in standing waves with maxima that move away from the boundary for increasing incident angles, as also indicated. For total reflection there is still energy in region 2, but it decays exponentially away from the boundary, and there is no power flow in the x-direction.

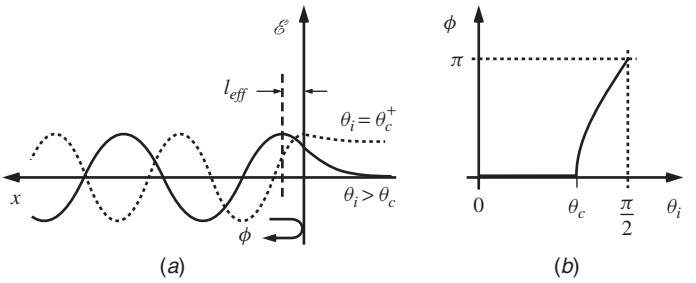


FIGURE 7.3: (a) Illustration of standing waves resulting from plane waves incident at an angle just slightly larger than (dashed) and significantly larger than (solid) the critical angle. (b) Plot of reflection phase angle versus the angle of incidence.

The standing waves can be calculated by summing the incident and reflected waves as given by Eq. (7.1). For a TE wave with $\theta_i > \theta_c, E_r(0,z) = E_i(0,z)e^{j\phi}$. Therefore,

$$\boldsymbol{\mathscr{E}}_i + \boldsymbol{\mathscr{E}}_r = E_i(0,z)\hat{\mathbf{e}}_y[e^{jk_{ix}x} + e^{j\phi}e^{-jk_{ix}x}]e^{-j\beta z}, \tag{7.13}$$

or

$$\boldsymbol{\mathscr{E}}_i + \boldsymbol{\mathscr{E}}_r = 2E_i(0,z)\hat{\mathbf{e}}_y e^{j\phi/2}\cos(k_{ix}x - \phi/2)e^{-j\beta z}. \tag{7.14}$$

Figure 7.3 gives two examples for different ϕ, illustrating how the peak of the cosine shifts away from the interface for larger θ_i.

The separation between the field maximum and the waveguide boundary in Fig. 7.3 is used to define the effective length l_{eff}, which is related to the reflection phase by

$$r = |r|e^{j\phi} = |r|e^{-2jk_{ix}(-l_{eff})} = |r|e^{2jk_{ix}l_{eff}},$$

or

$$2l_{eff} = \phi/k_{ix}. \tag{7.15}$$

Example 7.1

Problem: A 1.3Q InGaAsP dielectric layer was grown on an InP wafer. A 1.55-μm TE polarized plane wave is incident at the interface between these two materials, as illustrated in Fig. 7.2. Determine the phase angle ϕ and l_{eff} for (a) the wave angle of incidence of $\theta_i = 60^\circ$ (b) the wave angle of incidence of $\theta_i = 70^\circ$.

Solution: First, we need to compute the critical angle for this dielectric interface,

$$\theta_c = \sin^{-1}\left(\frac{n_2}{n_1}\right) = \sin^{-1}\left(\frac{3.17}{3.4}\right) = 68.8^\circ,$$

where the values for indices were obtained from Table 1.1.

(a) Since the angle of incidence in this case is lower than the critical angle, based on Fig. 7.3, the reflection angle is $\phi = 0$, and from Eq. (7.15), $l_{eff} = 0$.

(b) In this case $\theta_i > \theta_c$, and we have total internal reflection. Therefore, there will be a phase shift associated with the reflection. To calculate this phase, based on Eq. (7.11), we need to calculate the decay constant y_{tx} and the wave constant k_{ix}.

$$\begin{aligned}\gamma_{tx} &= \sqrt{\beta^2 - k_2^2} = \sqrt{(k_i \sin\theta_i)^2 - (k_0 n_2)^2} = k_0\sqrt{(n_1 \sin\theta_i)^2 - (n_2)^2} \\ &= k_0\sqrt{10.208 - 10.049} = 0.399k_0.\end{aligned}$$

We also need to calculate the value for k_{ix},

$$k_{ix} = k_i \cos\theta_i = 1.163k_0.$$

Finally, we can calculate the reflection phase shift, based on Eq. (7.11),

$$\phi^{TE} = 2\tan^{-1}\left(\frac{\gamma_{tx}}{k_{ix}}\right) = 2\tan^{-1}\left(\frac{0.399}{1.163}\right) = 0.661$$

and the effective length,

$$l_{eff} = \frac{\phi^{TE}}{2k_{ix}} = \frac{0.661}{2 \cdot 3.4 \cdot 2\pi/1.55\ \mu\text{m}} = 25.37\ \text{nm}$$

7.3 DIELECTRIC WAVEGUIDE ANALYSIS TECHNIQUES

7.3.1 Standing Wave Technique

Now, with the above preparation, we can begin to consider the construction of a waveguide that makes use of multiple total internal reflections. In Fig. 7.4 a standing wave resulting from the total internal reflection of a plane wave (such as is shown in Fig. 7.3) is illustrated. The first maxima occurs a distance l_{eff} from the boundary as discussed above. The dashed lines correspond to symmetry planes where for the given angle of incidence another boundary identical to the original one could be inserted without changing the standing wave pattern between the two boundaries.

We can see that with this construction, we have actually formed a waveguide, which traps the plane wave at the original incident angle, forcing it to zigzag

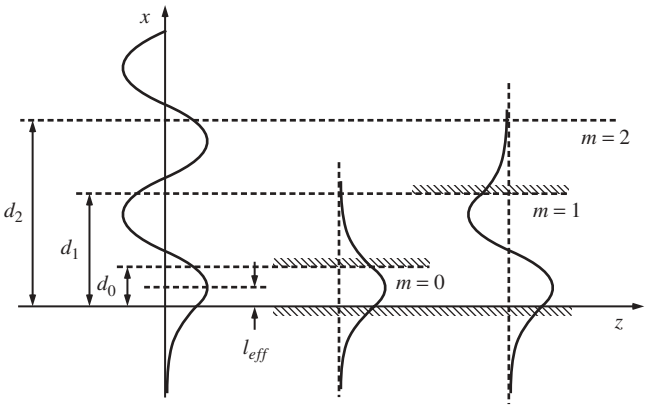


FIGURE 7.4: Construction of waveguides by inserting a second interface (top cross-hatched region) at the symmetry point on a given standing wave (i.e., given ray angle). Given standing wave at left; fundamental ($m = 0$) and first higher-order mode ($m = 1$) illustrated to the right.

indefinitely back and forth so that the net propagation of optical energy is only in the z-direction, as illustrated in Fig. 7.5. The field has the form given by Eq. (7.14) between the boundaries and is evanescent in the two outer regions according to $\mathscr{E}_t$ in Eq. (7.1), using Eq. (7.7) for k_{tx}.

From Fig. 7.4 the constructed waveguide width is seen to be

$$d = 2l_{eff} + m\frac{\lambda_x}{2}, \tag{7.16}$$

where m is the mode number, equal to zero for the lowest-order symmetric mode, one for the first odd mode, etc., and $\lambda_x = 2\pi/k_{ix}$ is the standing wave wavelength

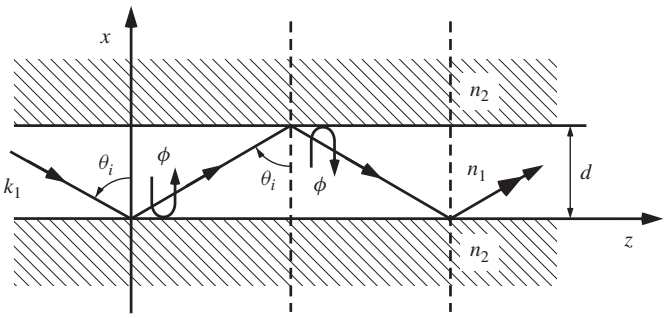


FIGURE 7.5: Illustration of zigzag ray picture of waveguiding. Each ray of the plane wave reflects off alternate boundaries with the same angle of incidence and a reflection phase that provides in-phase addition with other rays propagating in the same direction.

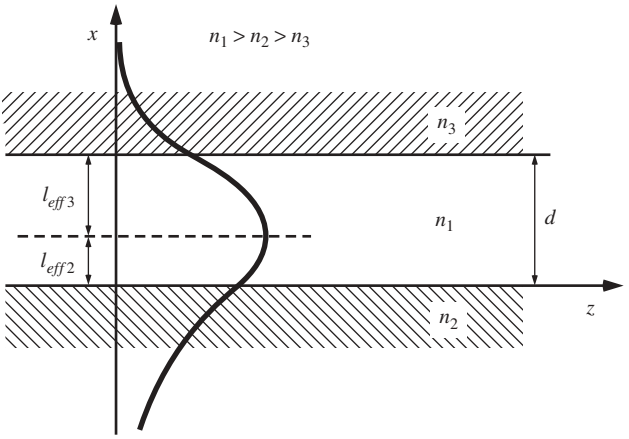


FIGURE 7.6: Fundamental mode of an asymmetric waveguide. The ray angle remains the same throughout, but the phase angle for the top reflection is larger than for the bottom. (Note the change in index definition in this chapter relative to Appendix 3.)

along the x-direction. Now, using Eqs. (7.11), (7.15), and (7.16), we obtain the waveguide dispersion relation for the TE modes,

$$d = \frac{\phi}{k_{ix}} + m\frac{\pi}{k_{ix}} = \frac{2}{k_{ix}}\tan^{-1}\left(\frac{\gamma_{tx}}{k_{ix}}\right) + m\frac{\pi}{k_{ix}}$$

or

$$k_{ix}d = 2\tan^{-1}\left(\frac{\gamma_{tx}}{k_{ix}}\right) + m\pi, \qquad m = 0, 1, 2, \ldots \tag{7.17}$$

Equation (7.17) is equivalent to Eq. (11.10), the dispersion relationship derived in Appendix 3 for the symmetric three-layer slab waveguide. Equation (7.12) expands γ_{tx}/k_{ix} in terms of the waveguide propagation constant along z, β. For the TM modes the result is the same, except $\varepsilon_1/\varepsilon_2$ multiplies γ_{tx}.

Using these same techniques the dispersion relationship for a general three-layer asymmetric guide can be derived. As shown in Fig. 7.6 the two cladding layers have different indices of refraction. Thus, the reflection phases at the bottom and top interfaces, ϕ_2 and ϕ_3, and the separations of the standing wave maxima, $l_{eff\,2}$ and $l_{eff\,3}$, respectively, are different also.

Constructing an asymmetric guide with $d = l_{eff\,2} + l_{eff\,3}$, and using Eqs. (7.15) and (7.11), we find the dispersion relationship for the TE modes to be

$$k_{1x}d = \frac{\phi_2}{2} + \frac{\phi_3}{2} + m\pi = \tan^{-1}\left(\frac{\gamma_{2x}}{k_{1x}}\right) + \tan^{-1}\left(\frac{\gamma_{3x}}{k_{1x}}\right) + m\pi, \tag{7.18}$$

where as before,

$$k_{1x} = \sqrt{k_1^2 - \beta^2}, \quad \gamma_{2x} = \sqrt{\beta^2 - k_2^2}, \quad \gamma_{3x} = \sqrt{\beta^2 - k_3^2}. \tag{7.19}$$

The electric fields are given by Eq. (7.14) between the boundaries with $\phi = \phi_2$, and Eq. (7.1) in the evanescent regions using Eqs. (7.7) and (7.19) for k_{tx}.

Example 7.2 Using the material from Example 7.1, a slab waveguide was constructed by regrowing InP on top of the InGaAsP layer whose total thickness is d.

Problem: What is the waveguide thickness d that supports the first order mode at 1.55 μm, incident at $\theta_i = 70^\circ$?

Solution: Since we are looking for the first order mode, refering to Fig. 7.4, we expect the mode to have one null, and the waveguide width will be given by Eq. (7.16), where $m = 1$. Since all the parameters for the mode (refractive indices, angle of incidence) are identical to those from Example 7.1, we can use already calculated values for k_{ix} and l_{eff},

$$k_{ix} = k_i \cos\theta_i = 1.163 k_0 = \frac{2\pi}{1.55\ \mu\text{m}} 1.163 = 4.712\ \mu\text{m}^{-1}$$

and

$$l_{eff} = 25.37\ \text{nm}.$$

Using Eq. (7.16), and the relationship for $\lambda_x = 2\pi/k_{ix} = 1.333$ μm, we have that

$$d = 2l_{eff} + m \cdot \frac{\lambda_x}{2} = 50.74\ \text{nm} + 1 \cdot \frac{1.333}{2}\mu\text{m} = 717\ \text{nm}.$$

7.3.2 Transverse Resonance

Another way of analyzing a dielectric waveguide is the transverse resonance technique. This is equivalent to the waveguide construction technique given above and the field-theory technique given in Appendix 3. As indicated in Fig. 7.7, transverse resonance means that the transverse round-trip phase must be a multiple of 2π after a complete cycle of a constituent ray. That is, the fields of an eigenmode must reproduce themselves after the plane wave components have zigzagged up and down one complete transverse cycle as the energy propagates down the guide. This is exactly the same condition as we imposed earlier in determining the modes of a Fabry–Perot resonator, only here we have generalized the situation to the case of nonnormal incidence. In other words, the component of the k-vector normal to the boundaries determines the phase progression along that direction.

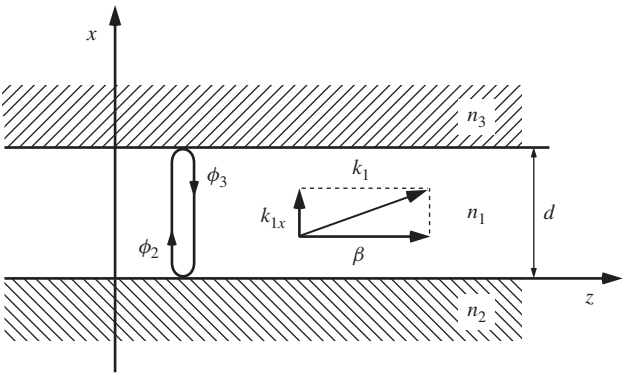


FIGURE 7.7: Waveguide cross section showing variables relevant to a transverse resonance calculation.

For a transverse mode, we must have transverse resonance, or

$$e^{-2jk_{1x}d}e^{j\phi_2}e^{j\phi_3} = e^{-2jm\pi}, \tag{7.20}$$

where the variables are as defined above in Eqs. (7.19) and (7.11). This implies that

$$2k_{1x}d - \phi_2 - \phi_3 = 2m\pi, \tag{7.21}$$

which is equivalent to Eq. (7.18).

The transverse resonance technique is a relatively simple approach to obtaining a dispersion relationship for a complex dielectric waveguiding structure. That is, regions 2 and 3 can contain a number of dielectric interfaces, and using the techniques of Chapter 3, we can calculate the net reflection coefficient from which we can obtain the reflection phases, ϕ_2 and ϕ_3. Then, Eq. (7.21) can be applied. For a lossless waveguide, the magnitude of these reflection coefficients must be unity. We shall later deal with cases where the reflection is slightly less than unity, but where the optical energy is still relatively well guided aside from a slight propagation loss.

Cutoff and "Leaky" or "Quasi Modes" The point where the ray angle becomes sufficiently large (measured from the guide axis, z) so that some energy is transmitted at one boundary or the other is generally referred to as *cutoff*. (This is the point where the angle of incidence no longer exceeds the critical angle.) However, the optical energy may still continue to propagate with only modest attenuation since the reflection magnitude at the waveguide walls is still relatively large for angles near cutoff. The light that leaks out of the waveguide also tends to radiate in the forward direction so that it may run parallel to the guide for some distance.

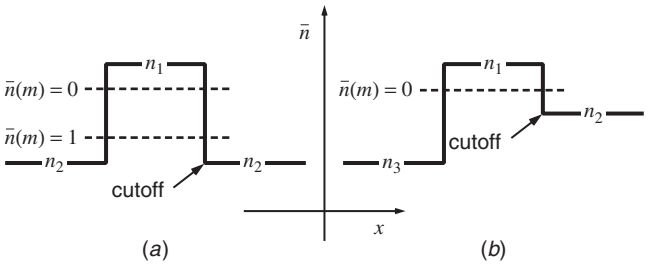


FIGURE 7.8: Illustration of effective index levels for mode numbers, m, in the (a) symmetric and (b) asymmetric cases. For decreasing frequency (increasing wavelength), $\bar{n}$ moves down toward the cutoff levels.

These are some of the key differences between dielectric and metal waveguides, in which cutoff tends to result in highly attenuated signals, and/or large reflections.

Put in more mathematical terms, cutoff is where $\phi_j \to 0$, or where $\beta \to k_j$, $j = 2, 3$. In other words, the waveguide propagation constant must always be larger than either of the cladding plane wave propagation constants. This situation is illustrated in Fig. 7.8 in terms of the effective index, $\bar{n} = \beta\lambda/2\pi$.

For the symmetric guide the dispersion relation, Eq. (7.17), becomes

$$k_{1x}d|_{cutoff} \leq m\pi, \quad (n_2 = n_3) \tag{7.22}$$

from which we can see that the fundamental mode, $m = 0$, has no cutoff except at zero frequency, or zero width, d. For the asymmetric guide, the dispersion relation, Eq. (7.18), becomes

$$k_{1x}d|_{cutoff} \leq \tan^{-1}\left(\frac{\gamma_{3x}}{k_{1x}}\right) + m\pi. \quad (n_1 > n_2 > n_3) \tag{7.23}$$

Here we see that the fundamental mode can be cut off for some finite frequency, or a small enough waveguide width. In Appendix 3 normalized curves for the dispersive properties of such waveguides are given. In Fig. 7.9, we plot the effective index versus normalized frequency for an example asymmetric guide.

Even if the reflection coefficient at one or both waveguide boundaries falls below unity for some ray angle, it is still possible to satisfy the transverse resonance condition. The dashed curves in Fig. 7.9 give the dispersion of these "leaky modes" as the ray angle within the waveguide increases toward being normal to the sidewall boundaries, where $\bar{n} \to 0$. Since the reflection phase goes to zero below cutoff, these leaky modes are basically continuations of the guided mode characteristics, and they satisfy Eqs. (7.22) or 7.23 for $k_{1x}d < k_{1x}d|_{cutoff}$, depending upon whether the guide is symmetric or not. For these cases, the optical energy still circulates in the transverse direction and builds up via constructive interference just as in any

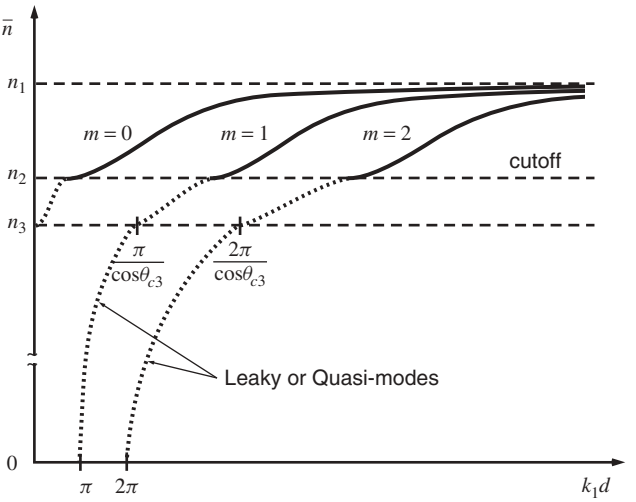


FIGURE 7.9: Schematic example of effective index dispersion curves for three lowest-order modes in an asymmetric slab. Solid curves give effective index for true guided modes. Cutoff is defined where the modes become leaky. Dotted curves indicate locus of points satisfying transverse resonance below cutoff. The angle θ_{c3} is the critical angle at the 1–3 interface. This is not the cutoff angle, which is the critical angle at the 1–2 interface.

Fabry–Perot resonator. Thus, optical energy may still be concentrated about the guide axis, and it may be transported along the z-direction with only modest loss. The "modes" of such a structure are usually referred to as *leaky* or *quasi* modes, since the propagating energy in the z-direction does change its magnitude along the waveguide. Axial modes of a Fabry–Perot laser cavity are an example of such leaky modes, since there must be mirror transmission for useful output.

Radiation Modes The leaky modes discussed above are really not true modes but quasi modes as already suggested. Strictly speaking, modes must have uniform profiles and magnitudes along the z-direction. But the guided modes, obtained earlier in this chapter or in Appendix 3, do not form a complete set because a superposition of them cannot in general synthesize an arbitary field. For such a complete set, we need to consider the "radiation modes," which can have other ray angles besides those of the guided and transverse resonant leaky modes.

To construct the radiation modes of some slab waveguide structure as true modes, we must insert some additional perfectly reflecting boundaries, which can be placed sufficiently far away in the x-direction so that any arbitrary field profile in the vicinity of the central waveguide slab can be completely described. Practically speaking, these inserted boundaries are really more of a mathematically necessary artifact than something that we actually insert in an experiment. Figure 7.10 shows the structure to be analyzed.

As can be seen in Fig. 7.10, perfect reflectors are placed on each side of the waveguide slab a distance $l/2$ away. Now, the transverse resonance technique can be

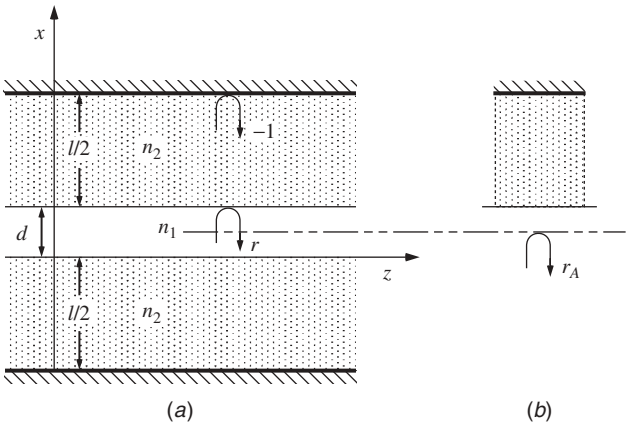


FIGURE 7.10: (a) Waveguide cross section illustrating perfectly reflecting planes inserted at $x = \pm l/2$ to provide for a set of radiation modes. (b) Net reflection from top half, r_A, calculated from the multiple interfaces using k_{ix} as the transverse propagation constant.

applied to obtain the dispersion relationship for the radiation modes. To make this a bit simpler, we calculate the net reflections looking in the positive and negative x-directions relative to a single reference plane in the center of the waveguide, r_A and r_B, respectively. Then, our transverse resonance condition, Eq. (7.20), reduces to

$$r_A r_B = 1 = e^{-2jm\pi}. \tag{7.24}$$

For the symmetric case in question, $r_A = r_B$, and we refer to the inset in Fig. 7.10 to calculate r_A,

$$r_A = re^{-jk_{1x}d} + \frac{-(1-r^2)e^{-jk_{2x}l}}{1-re^{-jk_{2x}l}}e^{-jk_{1x}d}. \tag{7.25}$$

Using Eq. (7.4) for r and inserting Eq. (7.25) into the dispersion relation Eq. (7.24), we can derive a dispersion relationship good for both even and odd radiation modes [3],

$$\cot\left(\frac{k_{2x}l}{2}\right) = \frac{k_{1x}}{k_{2x}}\tan\left(\frac{k_{1x}d}{2} - \psi\right), \qquad \psi = \begin{cases} 0 & \text{even} \\ \pi/2 & \text{odd} \end{cases} \tag{7.26}$$

Generally, we choose $l \gg d$, so that the solutions to the dispersion relationship, Eq. (7.26), are closely spaced in frequency (or ray angle). This provides a good basis set that we can use to analyze the propagation of arbitrary optical energy profiles along z. Actually, the set is only complete for describing fields contained completely within l. Thus, l must be at least as large as the extent of the field to

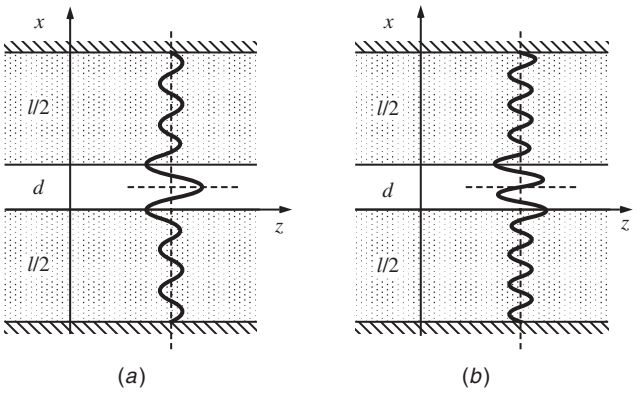


FIGURE 7.11: Schematic examples of even (a) and odd (b) radiation modes constructed by placing perfect reflectors $\pm l/2$ from either side of the waveguide slab. In these examples, transverse resonance is approximately satisfied in the waveguide slab as well.

be approximated. In practice, setting $l/d > 10$ is usually sufficient to provide the basis functions to synthesize arbitrary field profiles concentrated near the central waveguide.

The field profiles of the radiation modes are standing waves between the outer mirrors with some variation in magnitude and phase over the waveguide slab region. Some of the radiation modes may satisfy transverse resonance for the waveguide interfaces. The fields of this subset will show the expected increase in magnitude over the waveguide due to the coherent addition of multiple reflections there. As set up, we find odd and even modes with respect to the waveguide region. That is, the modes will have either a node or antinode in the center of the waveguide. Figure 7.11 gives examples.

Figure 7.12 summarizes the location of the radiation modes relative to the guided modes on an $\omega - \beta$ plot. Radiation modes that satisfy transverse resonance in the waveguide are shown by the dashed curves. These characteristics are the same as those of the leaky modes discussed earlier. For finite l, the location of the outer mirrors must be chosen properly to satisfy both inner and outer boundary conditions simultaneously.

Multilayer Waveguides The transverse resonance approach used in the last section for radiation modes can also be applied for the guided modes of multilayer waveguides. That is, Eq. (7.24) is always valid, and it can be applied at the center of a waveguide with any number of dielectric layers. The selection of the center is actually not critical, rather a convenient reference plane that facilitates the subsequent analytical or numerical evaluation of the waveguide dispersion properties is generally used.

Figure 7.14 shows an example of an m-layer waveguide. The evaluation of r_A and r_B can use the techniques developed in Chapter 3 for multilayer reflectors.

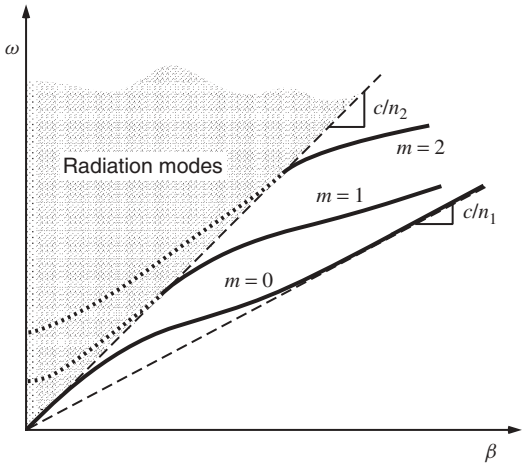


FIGURE 7.12: An $\omega - \beta$ diagram for a symmetric slab waveguide, showing relative locations of various modes. Slopes of asymptote lines equal the phase velocities of plane waves in the waveguide slab (c/n_1) and cladding (c/n_2) regions. Local slopes of modal characteristics give the group velocities. Again, dotted curves give dispersion of radiation modes that satisfy transverse resonance in the guide.

However, here the axial propagation constant, $k_z \equiv \beta$, is replaced by the transverse propagation constant, k_x, in all of the calculations. For only a few layers, it is possible to derive closed form expressions, as for the radiation modes above. For example, for a five-layer guide, Eq. (3.41) can be applied for both r_A and r_B. If many periods of a pair of layers are used, then the grating formulas can be adapted. That is, Eq. (3.54) can be used for r_A and r_B, provided again that appropriate k_xs replace the βs.

Example 7.3 Transverse resonance technique can be applied to complex multilayer waveguide claddings. To illustrate the use of the method, we are analyzing a slab waveguide, whose core index is $n_a = 3.253$, and whose cladding index is $n_1 = 3.17$. This is effectively a buried rib waveguide based on the structure from Example 1.1.

Problem: Applying generalized transverse resonance technique, calculate and plot the resonance condition ($\mathrm{Re}(r_A^2)$ and $\mathrm{Im}(r_A^2)$) versus the effective index, for the waveguide width of $d = 2.5$ μm and the light wavelength of $\lambda = 1.55$ μm.

Solution: To solve this problem, we will utilize the transverse resonance condition as given by Eq. (7.20). Using the transmission matrix theory, we will compute the reflectivities r_A and r_B from the center of the waveguide forward, illustrated in Fig. 7.14. For our case, the resonance condition reduces to

$$r_A r_B = 1.$$

Referring to Chapter 3, and assuming lossless material, we obtain the expressions for the transmission matrix T_p of a transmission line,

$$\mathbf{T_p} = \begin{bmatrix} e^{j\beta L} & 0 \\ 0 & e^{-j\beta L} \end{bmatrix}$$

and of an interface reflection T_r,

$$\mathbf{T_r} = \frac{1}{t_{12}} \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix}.$$

Since we are looking for the resonances in the transverse direction, we have to go back to the definition of reflection coefficients given by Eqs. (7.4) and (7.5) to calculate them, and then use Eq. (3.12) for the transmission coefficient. Therefore, we have that

$$r_{12} = \frac{\sqrt{n_a^2 - \bar{n}_{eff}^2} - \sqrt{n_1^2 + \bar{n}_{eff}^2}}{\sqrt{n_a^2 - \bar{n}_{eff}^2} + \sqrt{n_1^2 + \bar{n}_{eff}^2}}$$

$$t_{12} = \sqrt{1 - r_{12}^2}.$$

From Chapter 3, Table 3.1, we also have that $r_A = \frac{T_{21}}{T_{11}}$. For our problem, the composite T matrix that we need to compute is given by

$$T = T_p T_r.$$

Since the effective index is related to the propagation constant β as $\beta = k_x = \frac{2\pi}{\lambda_0}\sqrt{n_a^2 - \bar{n}^2}$, we can compute the value for T numerically for each value of β, and look for the solution of the transverse resonance condition under which $\mathrm{Im}(r_A^2) = 0$ and $\mathrm{Re}(r_A^2) = 1$. The corresponding plot is shown in Fig. 7.13. We notice that there are three different modes supported by this waveguide - their effective indices are

$$n_{eff\,1} = 3.179, \quad n_{eff\,2} = 3.221 \quad \text{and} \quad n_{eff\,3} = 3.244.$$

7.3.3 WKB Method for Arbitrary Waveguide Profiles

In all of what we have discussed so far in this chapter the index of refraction was constant over some regions, and it jumped abruptly at boundaries there between. Thus, in each region the assumptions involved in the derivation of the transverse wave equation (e.g., uniform dielectric constant) were valid, and exact overall solutions could be found by applying the appropriate boundary conditions.

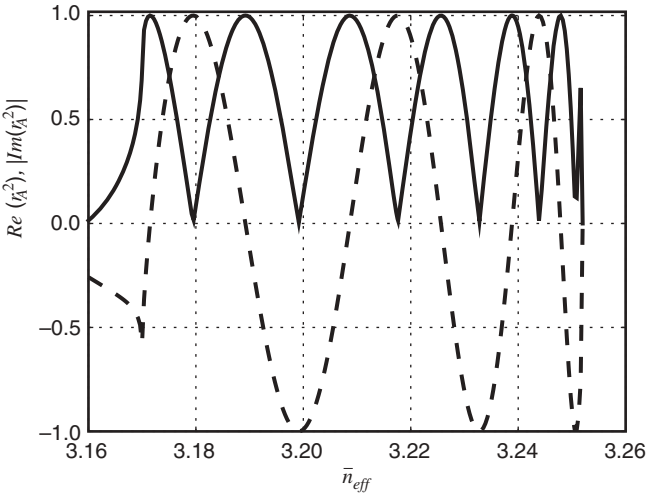


FIGURE 7.13: Solution for effective indices of modes supported by the waveguide from Example 7.3, using transverse resonance method. The plot is showing the real value of r_A^2 (dashed), as well as the absolute value of $\mathrm{Im}(r_A^2)$ (solid).

In many practical cases the index varies continuously over some waveguide region rather than discontinuously as above. Thus, strictly speaking the transverse wave Eq. (6.5) is not valid as discussed in Chapter 6. However, in cases where the dielectric constant varies slowly, so that $\nabla\varepsilon(x,y)$ is small relative to β, we can still apply it with reasonably good results. Otherwise, the wave equation retains terms involving $\nabla\varepsilon(x,y)$, and it becomes very difficult to solve. Figure 7.15 illustrates a waveguide formed by a region in which the index varies continuously.

Wentzel, Kramers, and Brillouin have found that good approximate solutions can be derived in the case of such a slowly varying index profile. Their so-called WKB approximation involves keeping the form of the uniform-medium wave equation, but using a plane wave k-vector that can vary transversely. That is, Eq. (A3.3) becomes

$$\nabla^2 U(x,y) + [k^2(x) - \beta^2]U(x,y) = 0. \tag{7.27}$$

This is the same as Eq. (6.5) where only an x-variation is permitted. Thus, we are still assuming a uniform waveguide along the z-direction.

The WKB approximation also involves neglecting any backscattering due to the slowly varying index. For waveguiding we are only interested in rays traveling roughly perpendicular to the index gradient (or parallel to z). Thus, even for relatively rapid index changes along x, the ray will only experience a slight gradient. As might be expected in such cases, the ray would bend, and as we shall see, this bending is what provides the equivalent of the zigzagging ray of the analogous three-layer slab guide.

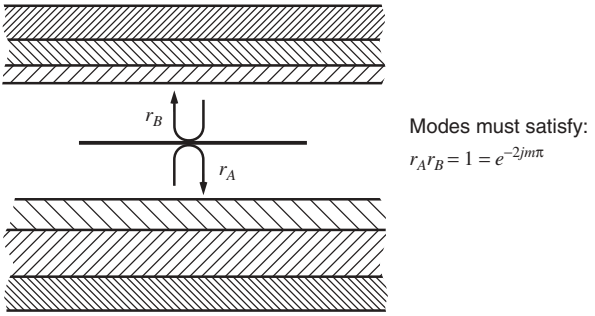


FIGURE 7.14: Schematic of generalized transverse resonance technique for the determination of modes in multilayer waveguides.

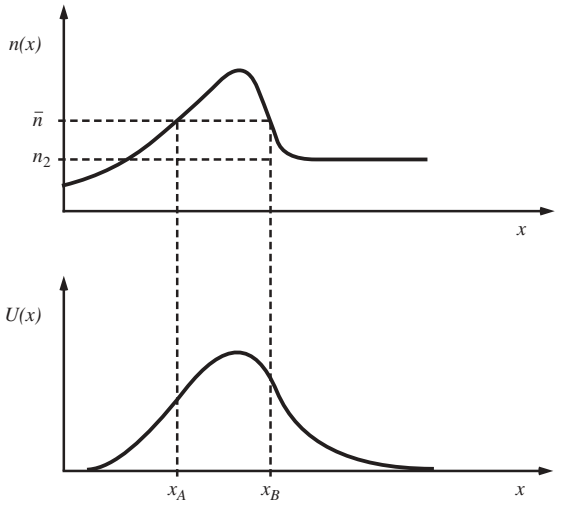


FIGURE 7.15: Transverse index and mode variation for a guide in which the index varies continuously. x_A and x_B are the ray turnaround points where the transverse mode also has points of inflection.

To get started, we express the local plane wave propagation constant, k, in terms of its vector components,

$$k_x^2(x) + k_z^2 = k^2(x), \tag{7.28}$$

where in this case, the index of refraction and the k-vector vary with x. However, for a waveguide mode, $k_z \equiv \beta \equiv k_0\bar{n}$, independent of x. Thus, we can write

$$k_x(x) = k_0\sqrt{n^2(x) - \bar{n}^2}. \tag{7.29}$$

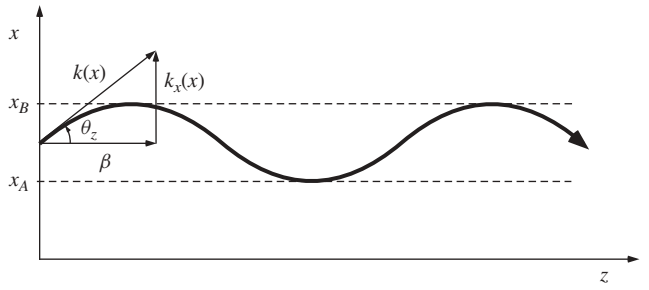


FIGURE 7.16: Sketch of meandering ray path for the case of a smoothly varying index with a maximum between the turnaround points, x_A and x_B. Ray angle with respect to the guide axis, θ_z, shown.

The cutoff condition is where $\bar{n} = n_2$, the highest adjacent cladding index.

The waveguide propagation constant can also be expressed in terms of the ray angle, $\theta_z(x)$, from which we can solve for this angle, i.e.,

$$\theta_z(x) = \cos^{-1}\left[\frac{\beta}{k(x)}\right] = \cos^{-1}\left[\frac{\bar{n}}{n(x)}\right]. \tag{7.30}$$

From Eq. (7.30) we see that the ray angle must vary with x, decreasing for decreasing $n(x)$ until $n(x) = \bar{n}$, at which point it is zero. Thus, as indicated in Fig. 7.16, the ray actually turns around at this point, meandering in a sinusoidal-like path as the lightwave propagates along z. At the turnaround points, $x = x_A, x_B$ and $k_{1x}(x) = 0$. This is consistent with Eq. (7.29). Thus we have

$$n(x_A) = n(x_B) = \bar{n} \quad \text{or} \quad k(x_A) = k(x_B) = \beta. \tag{7.31}$$

For $x > x_B$, or $x < x_A$, Eq. (7.29) shows that k_x is purely imaginary, indicating that the field decays away to provide the desired trapping of energy along the z-axis. The transverse wave Eq. (6.5) also shows that the transverse mode has inflection points (i.e., $\partial U^2(x)/\partial x^2 = 0$) at these turnaround points.

Now, we would like to develop the dispersion relationship or characteristic equation. For this, the transverse resonance technique quantified by Eq. (7.21) is used. However, in the present case the round-trip phase across the waveguide is not simply $2k_{1x}d$, but it must be found by integrating $k_x(x)$ from x_A to x_B and back. That is, Eq. (7.21) becomes

$$2\int_{x_A}^{x_B} k_x(x)\,dx = \phi_2 - \phi_3 = 2m\pi, \tag{7.32}$$

where $k_x(x)$ is given by Eq. (7.29) and the limits of integration can be expressed in terms of the effective index, $\bar{n}$, or propagation constant, β, using the actual index variation in Eq. (7.31).

The remaining problem is to figure out what the ϕ_i's are. For this we use the expressions for the reflection coefficient, either Eqs. (7.4) or (7.5) for the TE or TM modes, respectively. For k_{ix} and k_{tx}, we note that at $x = x_A$ or x_B, the transverse k_x in the present case turns from pure real to pure imaginary according to Eq. (7.29). Thus, the situation is similar to the abrupt index discontinuity case, and the reflection coefficients can be expressed as in Eqs. (7.8) and (7.9) with $k_{ix} = k_x(x_B^-)$ {or $k_x(x_A^+)$} and $\gamma_{tx} = jk_x(x_B^+)$ {or $jk_x(x_A^-)$}. That is, referring to Fig. 7.16, at the top turnaround point assuming a TE mode,

$$r^{TE}(x_B) = \frac{k_x(x_B^-) + j\gamma_x(x_B^+)}{k_x(x_B^-) - j\gamma_x(x_B^+)}. \tag{7.33}$$

Since $k_x(x \to x_B) \to 0$, and $\gamma_x(x \to x_B) \to 0$, we must expand them for $x_B^{\pm}$, letting the plane wave propagation constant, $k(x_B^-) = \beta + \delta k$, and $k(x_B^+) = \beta - \delta k$. That is, using their definitions given by Eq. (7.19), we find to first order,

$$k_x(x_B^-) = \sqrt{k^2(x_B^-) - \beta^2} = \sqrt{(\beta^2 + 2\beta\delta k) - \beta^2},$$

and

$$\gamma_x(x_B^+) = \sqrt{\beta^2 - k^2(x_B^+)} = \sqrt{\beta^2 - (\beta^2 - 2\beta\delta k)}. \tag{7.34}$$

Thus,

$$r^{TE}(x_B) = \frac{1+j}{1-j} = j, \tag{7.35}$$

and likewise for the other turnaround point. For the TM mode, γ_x is multiplied by $\varepsilon_1/\varepsilon_2$, but this approaches unity for a continuous index variation at the turn around point. Therefore, for both TE and TM modes, $\phi_2 = \phi_3 = \pi/2$, and Eq. (7.32) becomes

$$2\int_{x_A}^{x_B} k_x(x)\,dx = (2m+1)\pi, \tag{7.36}$$

or, using Eq. (7.29),

$$2k_0\int_{x_A}^{x_B} \sqrt{n^2(x) - \bar{n}^2}\,dx = (2m+1)\pi. \tag{7.37}$$

Waveguides with parabolic variation of the dielectric index are an important class of problems that can be solved using the WKB method. This is illustrated in the following example.

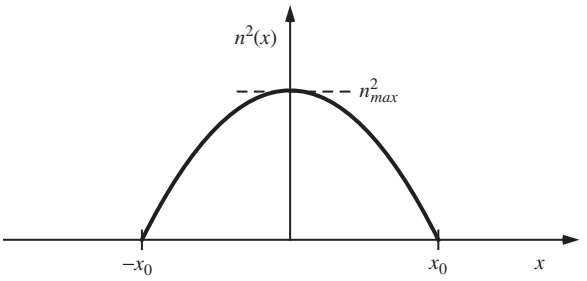


FIGURE 7.17: Parabolic dielectric constant variation given by Eq. (7.38).

Example 7.4

Problem: A waveguide consists of a core whose variation of the refractive index is given by

$$n^2(x) = n^2_{\max}\left[1 - \left(\frac{x}{x_0}\right)^2\right], \tag{7.38}$$

where we have chosen the origin of the x-axis to be at the maximum of the function. Applyng WKB theory, solve for the maximum ray excursion for mode m of this parabolic waveguide.

Solution: The parabolic function, illustrated in Fig. 7.17, is useful only for small index ranges under the WKB approximation, so that the optical energy must be well contained within $|x| \ll x_0$.

Plugging Eq. (7.38) into Eq. (7.37) gives

$$2k_0 \int_{x_A}^{x_B} \sqrt{n^2_{\max} - n^2_{\max}\left(\frac{x}{x_0}\right)^2 - \bar{n}^2}\, dx = (2m+1)\pi, \tag{7.39}$$

and $x_A = -x_B$ by symmetry. At the integration limits, the argument of the integral is zero, since $n(x_B) = \bar{n}$, or, $k_x = 0$. Therefore, we can solve for x_B from Eq. (7.38) to obtain

$$x_B = \pm x_0\sqrt{1 - (\bar{n}/n_{\max})^2} = \pm x_0\delta. \tag{7.40}$$

To make the integral more convenient to solve, we make the change of variables, $x = x_0\delta \sin\varphi$, where φ does not necessarily have any physical meaning. Then, performing the integration of Eq. (7.39) using Eq. (7.40), we find the dispersion relationship

$$\bar{n}^2 = n^2_{\max} - \frac{(2m+1)n_{\max}}{k_0 x_0}, \tag{7.41}$$

and plugging back into Eq. (7.40), the maximum ray excursion for mode m is

$$x_B = \pm x_0 \left[\frac{2m+1}{k_0 n_{\max} x_0} \right]^{1/2}. \tag{7.42}$$

The dielectric constant must continue to vary parabolically according to Eq. (7.38) over a width somewhat larger than the turnaround point separation, which is given by twice Eq. (7.42). That is, in most practical examples, it is found that the parabolic medium width must be four or five times x_B for the results to be approximately valid.

Equation (7.41) from Example 7.4 can be plugged into the transverse wave equation to obtain an expression for $U(x)$. This exercise shows that the eigenmodes of this parabolic medium are Hermite–Gaussian functions of the form,

$$U_m(x) = C_m H_m \left(\frac{\sqrt{2}x}{w_0} \right) e^{-x^2/\, w_0^2}, \tag{7.43}$$

where $H_m(\xi)$ are the Hermite polynomials and w_0 is the $1/e$ Gaussian spot size. For reference, the first three Hermite polynomials are

$$H_0(\xi) = 1, \quad H_1(\xi) = 2\xi, \quad H_2(\xi) = 4\xi^2 - 2. \tag{7.44}$$

It is also interesting to derive an expression for the ray path illustrated in Fig. 7.16. By definition,

$$\frac{dx(z)}{dz} = \frac{k_x(x)}{\beta} = \frac{k_x(x)}{k_0 \bar{n}}. \tag{7.45}$$

From Eqs. (7.29), (7.38), and (7.40)

$$k_x(x) = k_0 n_{\max} \sqrt{\delta^2 - (x/x_0)^2}, \tag{7.46}$$

and $n_{\max}/\bar{n} = (1-\delta^2)^{-1/2}$. Using Eq. (7.46) in Eq. (7.45),

$$\frac{dx(z)}{dz} = \frac{n_{\max}}{\bar{n}} \delta \sqrt{1 - \left(\frac{x}{x_0 \delta} \right)^2}, \tag{7.47}$$

and, again using the convenient change of variables, $x = x_0 \delta \sin\phi$, we can integrate to obtain

$$x(z) = x_0 \delta \sin\left[\frac{n_{\max}}{\bar{n}} \frac{z}{x_0} \right]. \tag{7.48}$$

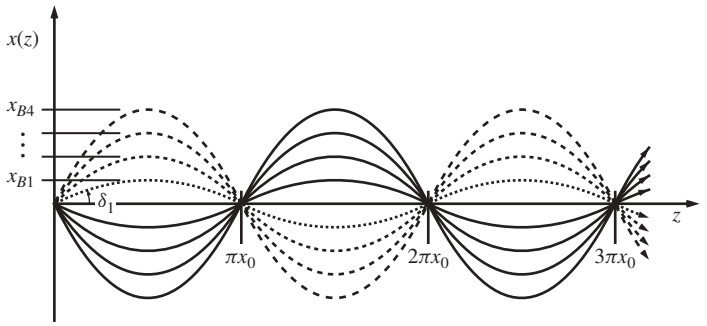


FIGURE 7.18: Illustration of ray paths in a parabolic medium that originate from a point source at the origin. Initial ray angle, δ, results in a sine wave of amplitude, $x_B = x_0\delta$, and period of approximately $2\pi x_0$ (using $n_{\max}/\bar{n} \approx 1$).

That is, in a parabolic medium, the rays oscillate around the waveguide axis as sine waves with a period of $2\pi x_0 \bar{n}/n_{\max}$. As is the case for the dispersion relationship, Eq. (7.41), this is an exact solution to the approximate wave equation. Since $\bar{n} \approx n_{\max}$, note that the ray oscillation period has only a weak dependence upon the initial ray angle, $dx(0)/dz = \delta$. However, the maximum excursion of the ray from the guide axis, $x_B = x_0\delta$, increases in direct proportion. Figure 7.18 illustrates this behavior. The fact that the ray zigzag period along the z-axis does not change much for various angles also suggests that the propagation constant does not change for the various possible modes that satisfy the transverse resonance condition. Equation (7.41) shows this small modal dispersion. Clearly, the parabolically graded waveguide has very different properties from the three-layer slab guide discussed earlier.

Now if we are interested in the eigenmodes of this parabolically graded waveguide, we can substitute the allowed values for $x_0\delta$. That is, using Eqs. (7.40) and (7.42) in Eq. (7.48), we obtain

$$x(z) = x_0\sqrt{\frac{2m+1}{k_0 x_0 n_{\max}}}\sin\left[\frac{n_{\max}}{\bar{n}}\frac{z}{x_0}\right]. \quad \text{(waveguide eigemode)} \tag{7.49}$$

Thus, only certain discrete ray angles are possible for the eigenmodes, but they still tend to have nearly the same propagation constant as mentioned above. In many cases, we may be interested in rays that do not satisfy transverse resonance, so Eq. (7.48) should continue to be used in these cases.

In the absence of birefringence, the phase fronts of traveling waves are perpendicular to the ray direction. As in the case of the three-layer slab waveguide, the phase fronts of the constituent rays superimpose to form the net phase fronts or wave crests of the eigenmode. For negligible gain or loss, the result in both cases is an eigenmode with plane phase fronts perpendicular to z. For the parabolic medium, the constituent phase fronts must bend to track the ray path, so they do

not appear to be plane waves. For the WKB approximation to be valid, however, these phase fronts must be approximately parallel over distances of a few wavelengths in the x- and z-directions. Put another way, the constituent ray angle can not change appreciably over this distance along the ray.

An example of a practical device that uses a parabolic graded-index waveguide section is the so-called graded-index rod or GRINROD lens. In this case we are not really interested in the eigenmodes of the waveguide, but rather how any given entering rays will propagate and focus after some distance. Although strictly speaking we should always express an entering field as a superposition of the eigenmodes of a waveguide to track its evolution down the guide, it is possible to use a ray-tracing approach if the waveguide is large enough. This makes particular sense, if the propagation distance is small. In the case of the GRINROD then, a pencil-like beam several wavelengths across, but much smaller than the diameter of the GRINROD will tend to propagate like a ray path given by Eq. (7.48). The eigenmode restriction given by Eq. (7.49), which provides for no change in mode cross section as a function of z, is not used because we really would have a large number of interfering eigenmodes superimposed to represent such a pencil-like beam.

GRINRODs are really cylinders with a radial index variation, but we can use our planar analysis to understand how rays will propagate along planes containing the rod axis (so-called meridional rays). For paraxial rays (small angles), Eq. (7.48) shows that a ray entering the GRINROD on axis at $z = 0$ is characterized by a slope $dx(0)/dz \approx \delta \approx \theta_z(0) \equiv \theta_0$. Thus, we write,

$$x(z) = x_0\theta_0 \sin(z/x_0), \quad \text{(GRINROD)} \tag{7.50}$$

where we already have that the inverse of x_0 is a measure of the curvature of the index-squared profile from Eq. (7.38) and θ_0 is the initial ray angle.

GRINRODs are characterized by a *pitch*, which is really just the ray oscillation period of $2\pi x_0$ for paraxial rays. Thus, a quarter-pitch GRINROD converts any and all rays entering at angles θ_{0i} at its $x - z$ *origin to rays parallel to but spaced a distance* $x_0\theta_{0i}$ from the z-axis after a quarter oscillation length of $\pi x_0/2$. That is, it acts like an ideal lens of focal length $\pi x_0/2$. Figure 7.18 illustrates this fact. GRINRODs of any length behave like lenses, but their focusing properties are not as simple as for the quarter- and half-pitch cases.

7.3.4 2-D Effective Index Technique for Buried Rib Waveguides

All of this chapter thus far has dealt only with slab waveguides in which only one-dimensional guiding was considered. For most practical applications two-dimensional guiding, which involves both the transverse (x) and lateral (y) directions, is desired. For most of these structures numerical techniques are required to derive accurate results. An exception is a cylindrical geometry for which it is possible to derive results for the circularly symmetric modes in analogy to the slab waveguides discussed above by using cylindrical coordinates in the wave equation. Here again only one variable (radial) is involved.

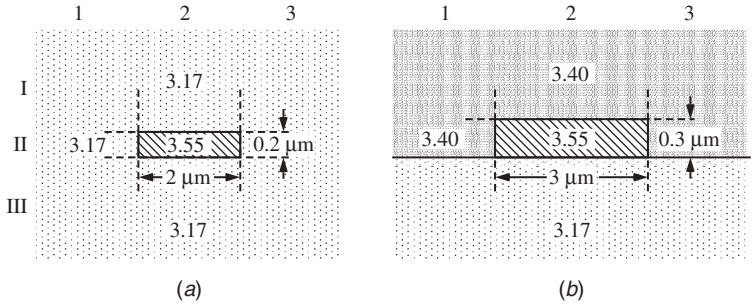


FIGURE 7.19: Example channel waveguide cross sections. (a) Waveguide channel buried in material of equal index on all sides, e.g., InP cladding over InGaAsP guide at 1.55 μm measurement wavelength. (b) Waveguide channel on substrate of one index covered with material of a different index, e.g., 1.3 μm InGaAsP covering 1.55 μm InGaAsP guide on InP substrate at 1.55 μm wavelength.

To analyze two-dimensional or channel waveguide structures involving two independent coordinates the approximate effective index technique is sometimes employed. In Appendix 3, we have already introduced a recipe for obtaining results using this technique. Here we shall briefly review this technique by means of two examples. The two waveguide cross sections that are of interest are shown in Fig. 7.19.

Example 7.5

Problem: Figure 7.19 (a) shows simple buried heterostructure laser, grown on InP and operating at 1.55 μm, in which the InGaAsP active region, $n_{wgd} = 3.55$, is buried in InP by regrowth. Determine the effective index for the fundamental mode of this waveguide using the effective index method.

Solution: To solve this problem, we shall use the normalized frequency, propagation constant, and asymmetry parameters defined in Eq. (A3.12) of Appendix 3. We explicitly consider the TE mode, but as discussed in the appendix, for small index differences we can neglect the differences between the TE and TM modes in this approximate analysis.

The first step of the effective index technique is to find the transverse effective index of each lateral region as if they were infinitely wide slab waveguides. In regions 1 and 3 the material is InP for all x. Thus, the effective index in both is clearly $\bar{n}_1 = \bar{n}_3 = 3.17$. In region 2, we use Eq. (A3.12) to obtain

$$V_2 = \frac{2\pi}{\lambda} d[n_{\text{II}}^2 - n_{\text{I}}^2]^{1/2} = \frac{2\pi}{1550 \text{ nm}} 200 \text{ nm}[3.55^2 - 3.17^2]^{1/2} = 1.30$$

and

$$a_2 = 0.$$

Using Eq. (A3.14), we can compute the value of the normalized propagation parameter b,

$$b = 1 - \frac{\ln\left(1 + \frac{V^2}{2}\right)}{\frac{V^2}{2}} = 1 - \frac{\ln\left(1 + \frac{1.3^2}{2}\right)}{\frac{1.3^2}{2}} = 0.275$$

Then, solving for the effective index in the central lateral region, $\bar{n}_2$, we find

$$\bar{n}_2 = [0.275(3.55^2 - 3.17^2) + 3.17^2]^{1/2} = 3.279.$$

Now, we have effective indices for each of the three lateral regions, and we can begin to solve the new lateral effective slab waveguide. Again using Eq. (A3.12),

$$V = \frac{2\pi}{1550 \text{ nm}} 2000 \text{ nm}[3.279^2 - 3.17^2]^{1/2} = 6.796,$$

and using Eq. (A3.14) again, we have $b = 1 - \frac{\ln\left(1 + \frac{6.796^2}{2}\right)}{\frac{6.796^2}{2}} = 0.862$. Thus, the net effective index is

$$\bar{n} = [0.862(3.279^2 - 3.17^2) + 3.17^2]^{1/2} = 3.264.$$

As discussed in Appendix 3, the net effective index can be used to calculate the lateral k_y in region 2 and the lateral decay constants, γ_y, in regions 1 and 3; however, we use $\bar{n}_2$ to obtain the transverse k_x and decay constants, γ_x, in all regions.

While Eq. (A3.14) gives us better precision for symmetric waveguides, for assymetric waveguide cases, we would have to use Fig. A3.12 to read the values for b in function of V.

In the next example, we are considering a case that cannot be solved using simple 2-D effective index technique approach, but that requires our results from the perturbation theory.

Example 7.6

Problem: Figure 7.19b shows a buried heterostructure laser operating at 1.55 μm and grown on InP, whose active region, $n_{wgd} = 3.55$, is buried by a quarternary material, $n_Q = 3.4$. Determine the effective index for the fundamental mode of this waveguide.

Solution: The waveguide structure in this example is a little more interesting, since this device is grown with a number of different index regions. Because we have only two different index layers in regions 1 and 3, we have to use a method other

than calculating the transverse effective index of a three-layer waveguide there. This would also be true if the guide in these regions were cut off. Certainly, we can go ahead and calculate the effective index in region 2, then proceed to determine the transverse mode shape, which will be used for all three lateral regions. This latter point is a good clue as to how we should calculate the effective index in the outer regions. That is, we should use the mode shape of the central region to provide the appropriate index weighting for the outer regions. (In fact, this concept even has merit in cases where there is guiding in the outer regions, however, it is generally not used because it is more difficult.)

First, we solve the transverse problem in the central region for this example waveguide (b). It should be noted that the cladding layer with the largest index should be chosen as the reference layer—denoted layer III. Otherwise, a will be negative. Evaluating Eq. (A3.12), we find that

$$V_2 = \frac{2\pi}{1550} 300 \sqrt{(3.55^2 - 3.40^2)} = 1.242,$$

and

$$a_2 = \frac{3.40^2 - 3.17^2}{3.55^2 - 3.4^2} = 1.449.$$

From Fig. A3.12, we can read off the value for $b_2 = 0.10$, thus obtaining $\bar{n}_2 = 3.415$. The small b indicates that we are approaching cutoff in this asymmetric guide.

Now, we can again look at the outer lateral regions. The best way to obtain the transverse effective index there is to use the perturbation formula from Eq. (6.23). We note that regions 1 and 3 can be constructed from region 2 by reducing the index of layer II by 0.15. The perturbation formula naturally uses the eigenmode of region 2, which is exactly what we are looking for. Thus, we proceed with $\Delta n_A = -0.15$.

After some effort in evaluating the perturbation integrals using k_x and γ_x from region 2, we find that $\overline{n}_1 \approx 3.34$.

Then, we proceed to solve the lateral problem in the same fashion and obtain $V = 8.66$, $a = 0$, $b = 0.92$, and $\bar{n} = 3.409$. Thus, we see that the fundamental mode is very near cutoff. Unfortunately, this is also where the effective index technique becomes unreliable, since the evanescent fields are expanding rapidly with slight reductions in the guide's effective index.

7.3.5 Analysis of Curved Optical Waveguides using Conformal Mapping

Curved optical waveguides are widely used in modern photonic integrated circuits. In order to precisely analyze their properties, including mode shape and excess propagation losses due to their curved nature, conformal transformation method can be used [10]. Conformal mapping allows for curved waveguides to be represented by corresponding straight waveguides with a modified lateral refractive index

distribution. The transformation is applied to the solutions of the two-dimensional scalar wave equation, given by Eq. (A3.3),

$$\nabla^2 U(x,y) + k^2(x,y)U(x,y) = 0. \tag{7.51}$$

New solutions are obtained in the coordinate system u, v, defined with respect to x and y by the relation

$$W = u + jv = f(Z) = f(x + jy) \tag{7.52}$$

where f is an analytic function. Using Cauchy–Riemann relations, $\partial u/\partial x = \partial v/\partial y$ and $\partial u/\partial y = -\partial v/\partial x$, Eq. (7.51) can be expressed as

$$\nabla^2 U(u,v) + \left|\frac{dZ}{dW}\right|^2 k^2(x(u,v), y(u,v))U(u,v) = 0. \tag{7.53}$$

The objective of the transformation is to select an $f(Z)$ that converts curved boundaries in the (x,y) plane to straight ones in the (u,v) plane. Using the values and reference planes as shown in Fig. 7.20, this can be accomplished by selecting

$$W = f(Z) = R_2 \ln \frac{Z}{R_2}, \tag{7.54}$$

for which

$$\left|\frac{dZ}{dW}\right| = e^{\frac{u}{R_2}}. \tag{7.55}$$

The result of this transformation is the waveguide shown in Fig. 7.20, whose walls are straight, and lie between $u = 0$ and $u = -R_2 \ln R_2/R_1$. The refractive index of this structure is a product of Eq. (7.55) and the refractive index in the appropriate region of the curved waveguide, Fig. 7.21. From our previous analysis of the leaky modes, it is clear that all the modes of the curved waveguide will radiate on the outer curve edge, thereby causing radiation losses. These losses can be estimated numerically or, as shown in Section 7.5, analytically by using an approximate formulation based upon the loss of the radiating tail.

One other obvious conclusion is that the fundamental mode in a curved waveguide will be asymmetric, with the mode peak located closer to the outer edge of the curve. This is illustrated in Fig. 7.20, where numerically computer fundamental modes are shown for both a straight and a curved waveguide.

Since the fundamental mode of a waveguide bend is asymmetric, the junction between a straight and a bend waveguide would normally result in the modal excitation loss, as discussed in Section 14.4. Therefore, the junction between these waveguide types requires some sort of a modal transition region. This region might be an adiabatic transition [12], or an abrupt waveguide offset [11].

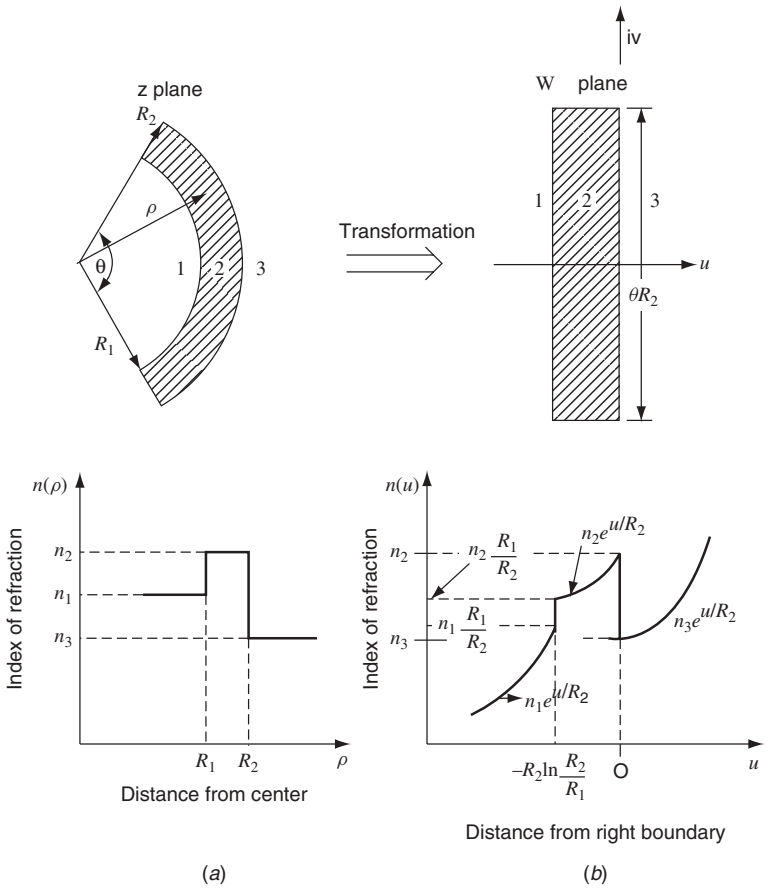


FIGURE 7.20: (a) A two-dimensional curved waveguide with radial step index distribution (b) Conformal transformation that converts the curved waveguide into a straight waveguide, with a modified lateral refractive index profile (© IEEE 1975. Figure reproduced from JSTQE, VOL. QE-11, 2, February 1975, page 75).

Example 7.7 To illustrate the use of the conformal mapping method, we are analyzing a slab waveguide whose core index is $n_a = 3.253$, and whose cladding index is $n_c = 3.17$. The radius of curvature of the center of this waveguide is $R = 600$ μm. The waveguide width is $w = 2$ μm. This waveguide is being mapped into a straight waveguide with coordinates n and u, and the right boundary of the waveguide is located at $u = 0$.

Problem: What is the refractive index profile of the straight waveguide obtained through conformal mapping?

Solution: Using Eq. (7.55), and the actual index profiles, we have that the left boundary of the waveguide will be at

$$u_{left} = -\left(R + \frac{w}{2}\ln\frac{R+\frac{w}{2}}{R-\frac{w}{2}}\right) = -\left(601\ln\frac{601}{599}\right)\mu\text{m} = -2.0033\ \mu\text{m},$$

which is really close to the original waveguide's width of 2 μm.

For $-\infty < u < u_{left}$, the refractive index profile will be given by the expression per Eq. (7.55)

$$n(u) = n_c \cdot e^{u/(R+w/2)} = 3.17 \cdot e^{u/(601\ \mu\text{m})}.$$

For $u_{left} \leq u < 0$, the refractive index profile will be given by the expression

$$n(u) = n_a \cdot e^{u/(R+w/2)} = 3.253 \cdot e^{u/(601\ \mu\text{m})}.$$

Finally, for $0 \leq u < \infty$, the refractive index profile will be given by the expression

$$n(u) = n_c \cdot e^{u/(R+w/2)} = 3.17 \cdot e^{u/(601\ \mu\text{m})}.$$

7.3.6 Numerical Mode Solving Methods for Arbitrary Waveguide Profiles

Many problems cannot be accurately solved by the effective index or WKB techniques. In these cases one generally turns to numerical techniques. Given the availability of computers and the appropriate software in today's world, such

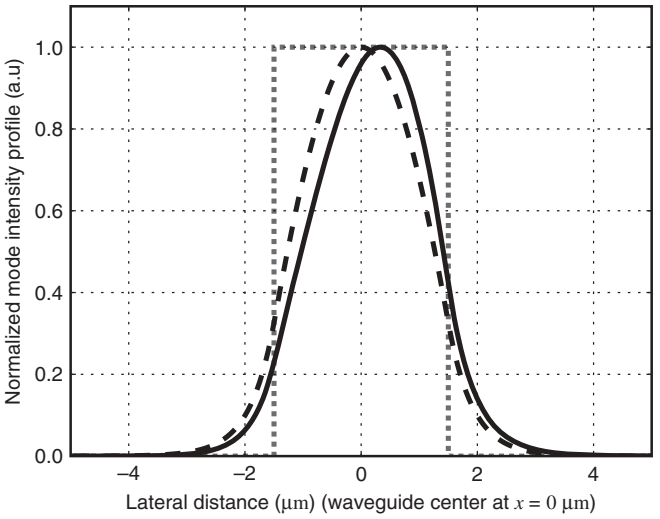


FIGURE 7.21: Fundamental mode profiles of a straight waveguide and a waveguide bend calculated numerically. The parameters used were: the cladding index $n_{cl} = 3.2291$, the core index $n_{cr} = 3.2887$, the wavelength $\lambda = 1550$ nm, the core width $w = 3$ μm, and the bend radius $R = 800$μm. The waveguide bends counterclockwise. As expected, the bend mode is shifted towards the outside edge of the curve [10].

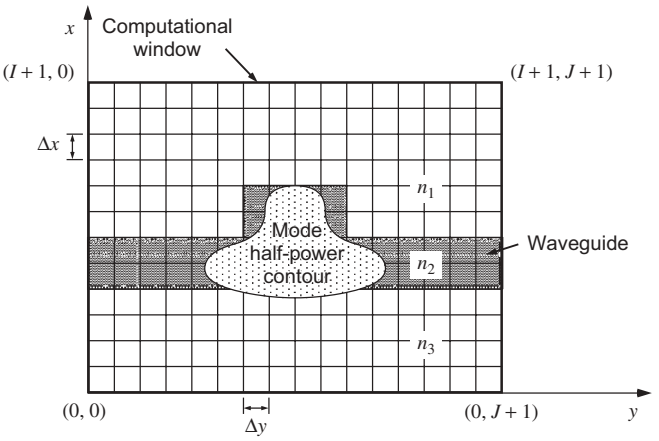


FIGURE 7.22: Illustration of finite-difference grid over a buried-rib waveguide.

numerical techniques are fairly easy to employ. The most straightforward numerical solution to the generalized scalar wave Eq. (7.27) uses what is called the *finite-difference technique*. As shown in Fig. 7.22, this involves overlaying a grid of some finite period over the lateral waveguide profile to be solved. The lateral normalized field $U(x, y)$ will be found at the nodes of the grid by converting the differential Eq. (7.27) into a set of finite-difference equations specified at the grid points. The eigenvalue of this set of linear equations will be the effective index of the mode in question. Thus, we wish to convert the partial differential Eq. (7.27) into an eigenvalue problem that can be solved by linear algebra, and linear algebra is easily done on a computer. This is the essence of the finite-difference technique.

The first step in the finite-difference procedure is to appropriately select the computational window. We wish to limit the size of the domain for computational efficiency. On the other hand, the domain should be large enough to contain the field distribution that we want to calculate. The most common approach is to set the field value to zero on the window boundary. Physical considerations should help us in determining the optimum window size. After an initial guess, one can change the window size and check to see if the result is affected. For more accurate calculations on the window boundaries, appropriate boundary conditions to make the window transparent can be used [3].

The second step is to discretize the computational window using a grid as indicated in Fig. 7.22. The grid can be uniform or nonuniform. A nonuniform grid is useful to appropriately sample strongly guided modes that may still have long evanescent tails. Variations in both the transverse and lateral dimensions must be considered. For this introductory treatment, we will assume a uniform grid from this point onward to keep the math as simple as possible. That is,

$$\begin{aligned} x &= i\,\Delta x, \quad i = 0, 1, 2, \cdots, I + 1 \\ y &= j\,\Delta y, \quad j = 0, 1, 2, \cdots, J + 1 \end{aligned} \tag{7.56}$$

Before converting the scalar wave Eq. (7.27), we expand the operator ∇^2 in rectangular coordinates and factor out the free-space plane wave propagation constant k_0 from the linear term. It then can be written in terms of the lateral index profile *n(x, y)*, and the modal effective index $\bar{n}$, as

$$\frac{\partial^2 U(x,y)}{\partial x^2} + \frac{\partial^2 U(x,y)}{\partial y^2} + k_0^2[n^2(x,y) - \bar{n}^2]U(x,y) = 0. \tag{7.57}$$

Now, we approximate the partial derivatives using the finite differences. This is accomplished by using a second-order Taylor series expansion of the fields at adjacent grid points. That is,

$$\begin{aligned} U(x+\Delta x,y) &= U(x,y) + \Delta x \frac{\partial U(x,y)}{\partial x} + \frac{(\Delta x)^2}{2} \cdot \frac{\partial^2 U(x,y)}{\partial x^2} + \cdots \\ U(x-\Delta x,y) &= U(x,y) - \Delta x \frac{\partial U(x,y)}{\partial x} + \frac{(\Delta x)^2}{2} \cdot \frac{\partial^2 U(x,y)}{\partial x^2} - \cdots \end{aligned} \tag{7.58}$$

Adding the two and solving for the second partial derivative,

$$\frac{\partial^2 U(x,y)}{\partial x^2} \approx \frac{U(x+\Delta x,y) - 2U(x,y) + U(x-\Delta x,y)}{(\Delta x)^2}, \tag{7.59}$$

where the higher-order terms can be neglected if the grid is sufficiently small. Of course, a similar expression can be derived for the second partial with respect to y.

Since we have discretized the space, we only need (or can solve for) the values of the functions at these points. Thus, using Eq. (7.56) we will denote $U(x,y) = U_j^i, U(x+\Delta x,y) = U_j^{i+1}$, and $U(x-\Delta x,y) = U_j^{i-1}$. Then,

$$\frac{\partial^2 U(x,y)}{\partial x^2} = \frac{U_j^{i+1} - 2U_j^i + U_j^{i-1}}{(\Delta x)^2},$$

and similarly

$$\frac{\partial^2 U(x,y)}{\partial y^2} = \frac{U_{j+1}^i - 2U_j^i + U_{j-1}^i}{(\Delta y)^2}. \tag{7.60}$$

The scalar wave equation can now be written in discretized form. That is,

$$\frac{U_j^{i+1} - 2U_j^i + U_j^{i-1}}{(\Delta x)^2} + \frac{U_{j+1}^i - 2U_j^i + U_{j-1}^i}{(\Delta y)^2} + k_0^2(n_j^{i^2} - \bar{n}^2)U_j^i = 0. \tag{7.61}$$

For computational ease, we divide through by k_0^2 and introduce the dimensionless quantities, $\Delta X^2 = k_0^2 \Delta x^2$ and $\Delta Y^2 = k_0^2 \Delta y^2$. Then, Eq. (7.61) can be written as the matrix eigenvalue equation,

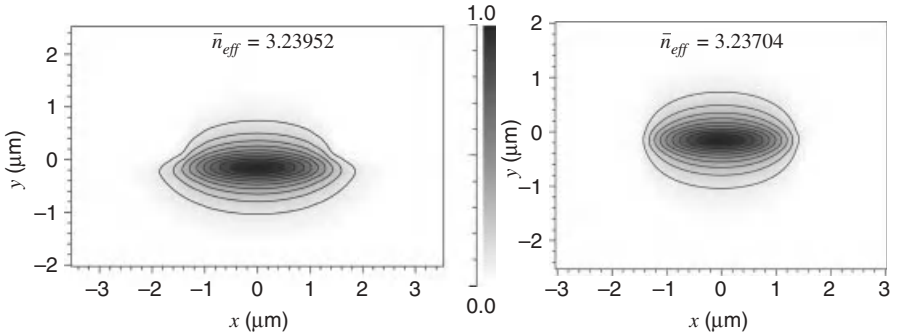


FIGURE 7.23: Mode intensity profiles for a (left) surface ridge waveguide (right) deep ridge waveguide. The details of the waveguide structures are given in the text.

$$\frac{U_j^{i-1}}{\Delta X^2} + \frac{U_{j-1}^i}{\Delta Y^2} - \left(\frac{2}{\Delta X^2} + \frac{2}{\Delta Y^2} - (n_j^i)^2\right) U_j^i + \frac{U_{j+1}^i}{\Delta Y^2} + \frac{U_j^{i+1}}{\Delta X^2} = \bar{n}^2 U_j^i, \quad (7.62)$$

for $i = 0$ to $I + 1$ and $j = 0$ to $J + 1$.

As shown in Appendix 16 this finite-difference matrix equation can be solved numerically for the eigenvalue $\bar{n}$, and the field profile U. This involves some matrix manipulation for compactness and the use of a standard linear matrix-solving algorithm, which can be found in most basic numerical analysis software packages. An example problem is also given for clarity. In many cases solutions can be generated very quickly, once the basic finite difference equations are entered. Figure 7.23 shows two mode intensity profiles, for a surface ridge waveguide and a deep ridge waveguide, calculated using the finite-difference technique. These types of structures were introduced in Chapter 1, Fig. 1.17. In both cases, the 320-nm-thick 1.4Q waveguide was grown on InP and is clad with 2 μm of InP. Both ridges are 3 μm wide. For the surface ridge, the height is 2 μm, and the ridge stops right above the waveguide layer, whereas in the deep ridge case, the waveguide is 4 μm deep.

One can improve the finite difference technique considering the vector nature of the fields [4]. Mode matching techniques that result in equivalent networks have also been developed [5, 6]. Finite-element techniques can also be applied to solve for vectorial fields [7].

7.4 NUMERICAL TECHNIQUES FOR ANALYZING PICs

7.4.1 Introduction

Numerous photonic integrated circuits cannot be accurately analyzed by simple analytic techniques. Also, even where analytical techniques are available, numerical

techniques are oftentimes employed to verify the designs before proceeding to invest serious effort in making new structures. In fact, numerical techniques are becoming increasingly available and user-friendly. In Section 7.3.6, we introduced the finite-difference technique to analyze dielectric waveguides. In this section we introduce the *beam-propagation method* (*BPM*) to analyze arbitrary waveguide propagation problems. In particular, we consider a finite-difference BPM as one of the best approaches. As we shall find it builds upon the finite-difference techniques developed in Chapter 7 and Appendix 16.

The most rigorous way of handling electromagnetic wave propagation in integrated optics is to solve Maxwell's equations with appropriate boundary conditions. However, such PIC structures have certain features that makes this approach very difficult to implement. The main reason is the large aspect ratio between the propagation distance and the transverse or lateral dimensions of the propagating energy. The cross section may be contained in a 2 μm × 5 μm window, but the propagation distance could be centimeters long. Therefore, to establish a fine enough grid for a boundary-value approach would pose a great challenge for computer memory and CPU performance. Fortunately, the guided waves in PICs have certain other properties that allow some approximations to be made. For example, in most cases the scalar wave equation is sufficient to describe the wave propagation. Second, the phase fronts of guided waves are nearly planar or their plane wave spectra are quite narrow. Therefore, they are paraxial. Third, index changes along the propagation direction tend to be small and gradual in many situations. Hence, as discussed in Chapter 6, the wave amplitudes change slowly and back reflections are negligible in these cases. Under these conditions it is possible to reduce the scalar wave Eq. (6.4) to the *paraxial wave equation*, which can be written as

$$2jk_0 n_r \frac{\partial \psi}{\partial z} = \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + k_0^2[n^2(x,y,z) - n_r^2]\psi, \tag{7.63}$$

where n_r is a reference index that describes the average phase velocity of the wave, and $\psi(x,y,z) = E(z)U(x,y)$. That is, n_r determines the rapidly varying component of the wave, and ψ includes the slowly varying amplitude along the propagation direction. Thus for the polarization of interest, $\mathscr{E}(x,y,z) = \psi(x,y,z)\exp[-jk_0 n_r z]$. The index n_r must be carefully chosen to get accurate answers. In guided-wave problems it is usually chosen as the index of the substrate.

The paraxial wave equation describes an initial value problem as opposed to a boundary value problem. As a result, one can start with an arbitrary initial wave amplitude, $\psi(x,y,0)$, which could be a Gaussian beam formed by a lens, for example. The resulting amplitude Δz away can be found by integrating the paraxial wave equation over Δz. Repeating this procedure, one can find the evolution of the initial field profile over the photonic integrated circuit. Note that one only needs the field values at $z = 0$ to calculate the field values at $z = \Delta z$. Therefore, there is no need to store or manipulate the field values at every grid point in the z-direction

as required in the solution of the boundary-value problem. Thus, BPM is much more computationally tractable and efficient. Furthermore, all parts of the wave, including the guided and radiation spectrum, are handled together. So there is no need for modal decomposition or to neglect the radiation part of the spectrum. However, since it originates from the paraxial wave equation, back reflections and wide-angle propagation are not handled.

The initial procedure for BPM involved operator techniques and FFTs [13]. But recently it has been shown that algorithms that are much more efficient and robust can be generated by using finite-difference techniques [14]. Due to its significant advantages only the finite-difference BPM (FD-BPM) will be described here. It is possible to generate FD-BPM algorithms using implicit or explicit techniques. In this basic treatment only an implicit algorithm based on the Crank–Nicolson algorithm [15] will be given. The reader is referred to the literature for discussion of the explicit FD-BPM [16].

7.4.2 Implicit Finite-Difference Beam-Propagation Method

The basic idea of the implicit finite-difference beam-propagation method (FD-BPM) is to approximate the paraxial wave Eq. (7.63) using finite-difference techniques described in Section 7.3.6. This requires choosing a computational window in the transverse dimension $(x,\ y)$ as well as choosing a grid in the z-direction. The computational window must be large enough to contain the desired field distribution all along the propagation path. However, since radiated fields can always be present, setting the field values to zero at the boundaries of the computational window can create difficulties. Since this boundary condition effectively creates a reflecting boundary, the radiated fields will reflect back and create spurious field distributions. One way to eliminate this difficulty is to use absorbing boundaries. This is achieved by introducing an artificial complex index distribution around the computational window to generate a lossy boundary. The radiated fields are then absorbed before reaching the edge of the window. In practice, less than 10 mesh points around the boundary with a complex index tends to be sufficient. Also, the imaginary part of the index is tapered so as to avoid reflections. Remember to keep the absorber outside the desired field profile. An alternative approach is to use "transparent boundary conditions" [16] to avoid the absorbers.

With the computational window and z-grid appropriately chosen, we then have $x = p\Delta x, y = q\Delta y$, and $z = l\Delta z$. The Crank–Nicolson algorithm is chosen since it is both unconditionally stable and unitary [17]. The reader is warned that a straightforward implementation of the FD-BPM can lead to an unstable algorithm, which is not power conserving, thus allowing slight round-off errors to lead to nonphysical results. With the Crank–Nicolson scheme fields will not diverge or diminish without any physical reason regardless of the mesh size, $\Delta x, \Delta y$, or Δz. But accuracy will be lost as the mesh size gets larger.

To convert Eq. (7.63) into the finite-difference form, we let $\psi(x, y, z) = \psi^l_{p,q}$, and approximate the z-partial on the left side by a forward difference

to obtain

$$\frac{\partial \psi}{\partial z} = \left[\frac{\psi_{p,q}^{l+1} - \psi_{p,q}^{l}}{\Delta z} \right]. \tag{7.64}$$

The right-hand side involves only x and y derivatives, which are approximated using the regular finite-difference approximations for second-order partial derivatives given by Eq. (7.59). This is refined by taking the average of the discretization at $z = l\Delta z$ and $z = (l+1)\Delta z$. That is,

$$\frac{\partial^2 \psi}{\partial x^2} = \frac{1}{2}\left[\frac{\psi_{p+1,q}^{l+1} - 2\psi_{p,q}^{l+1} + \psi_{p-1,q}^{l+1}}{\Delta x^2} + \frac{\psi_{p+1,q}^{1} - 2\psi_{p,q}^{l} + \psi_{p-1,q}^{l}}{\Delta x^2} \right], \tag{7.65}$$

$$\frac{\partial^2 \psi}{\partial y^2} = \frac{1}{2}\left[\frac{\psi_{p,q+1}^{l+1} - 2\psi_{p,q}^{l+1} + \psi_{p,q-1}^{l+1}}{\Delta y^2} + \frac{\psi_{p,q+1}^{l} - 2\psi_{p,q}^{l} + \psi_{p,q-1}^{l}}{\Delta y^2} \right]. \tag{7.66}$$

Also, for the final term,

$$k_0^2(n^2 - n_r^2)\psi = k_0^2 \left[\frac{(n_{p,q}^{l+1})^2 + (n_{p,q}^{l})^2}{2} - n_r^2 \right] \frac{\psi_{p,q}^{l+1} + \psi_{p,q}^{l}}{2}. \tag{7.67}$$

For normalization, we divide the whole equation by k_0^2 and call $k_0\Delta z = \Delta Z, k_0\Delta x = \Delta X$, and $k_0\Delta y = \Delta Y$. Then substituting Eqs. (7.64) through (7.67) in Eq. (7.63) and grouping terms with common q indices, we obtain

$$\begin{aligned} &- \frac{\psi_{p-1,q}^{l+1}}{2\Delta X^2} + [-b\psi_{p,q-1}^{l+1} + a_{p,q}\psi_{p,q}^{l+1} - b\psi_{p,q+1}^{l+1}] - \frac{\psi_{p+1,q}^{l+1}}{2\Delta X^2} \\ &= \frac{\psi_{p-1,q}^{l}}{2\Delta X^2} + [b\psi_{p,q-1}^{l} + c_{p,q}\psi_{p,q}^{l} + b\psi_{p,q+1}^{l}] + \frac{\psi_{p+1,q}^{l}}{2\Delta X^2}, \end{aligned} \tag{7.68}$$

where

$$\begin{aligned} b &= \frac{1}{2\Delta Y^2}, \\ a_{p,q} &= \frac{2jn_r}{\Delta Z} + \frac{1}{\Delta X^2} + \frac{1}{\Delta Y^2} - \frac{1}{2}\left[\frac{(n_{p,q}^{l+1})^2 + (n_{p,q}^{l})^2}{2} - n_r^2 \right], \\ c_{p,q} &= -a_{p,q} + \frac{4jn_r}{\Delta Z}. \end{aligned}$$

For $p = 1$ to P and $q = 1$ to Q within the computational window, we obtain $P \times Q$ coupled equations at each z step, one Eq. (7.68) for each value of p and q. The solution to these equations is best handled with matrices. The following reduction of the equations to matrix form runs parallel to Appendix 16 and assumes the same zero-field boundary conditions discussed in more detail there.

We first compact the y-direction into matrix notation by defining a vector that encompasses y for each x position, p:

$$\boldsymbol{\psi}_p^1 = \begin{bmatrix} \psi_{p,1}^l \\ \vdots \\ \psi_{p,Q}^l \end{bmatrix}. \tag{7.69}$$

Then, by vertically listing all Q Eqs. (7.68) for a given p, we can group common p indices into a matrix-difference equation along x:

$$-\mathbf{B}\boldsymbol{\psi}_{p-1}^{l+1} + \mathbf{A}_p\boldsymbol{\psi}_p^{l+1} - \mathbf{B}\boldsymbol{\psi}_{p+1}^{l+1} = \mathbf{B}\boldsymbol{\psi}_{p-1}^{l} + \mathbf{C}_p\boldsymbol{\psi}_p^{l} + \mathbf{B}\boldsymbol{\psi}_{p+1}^{l}, \tag{7.70}$$

where

$$\mathbf{B} = \frac{1}{2\Delta X^2}\mathbf{I},$$

$$\mathbf{A}_p = \begin{bmatrix} a_{p,1} & -b & 0 & \cdots & 0 & 0 \\ -b & a_{p,2} & -b & 0 & & 0 \\ 0 & -b & a_{p,3} & -b & 0 & \vdots \\ \vdots & 0 & & \ddots & & 0 \\ 0 & & 0 & -b & a_{p,Q-1} & -b \\ 0 & 0 & \cdots & 0 & -b & a_{p,Q} \end{bmatrix},$$

$$\mathbf{C}_p = \begin{bmatrix} c_{p,1} & b & 0 & \cdots & 0 & 0 \\ b & c_{p,2} & b & 0 & & 0 \\ 0 & b & c_{p,3} & b & 0 & \vdots \\ \vdots & 0 & & \ddots & & 0 \\ 0 & & 0 & b & c_{p,Q-1} & b \\ 0 & 0 & \cdots & 0 & b & c_{p,Q} \end{bmatrix}.$$

and $\mathbf{I}$ is the $Q \times Q$ identity matrix. If there is no y-direction in the problem of interest, then $\boldsymbol{\psi}_p^l \to \psi_p^l, \mathbf{A}_p \to a_p, \mathbf{C}_p \to c_p, \mathbf{B} \to 1/2\Delta X^2$, and Eq. (7.68) reduces to a simple difference equation along x.

Now writing all P Eqs. (7.70) along x in matrix form, we obtain

$$\begin{bmatrix} \mathbf{A}_1 & -\mathbf{B} & 0 & \cdots & 0 & 0 \\ -\mathbf{B} & \mathbf{A}_2 & -\mathbf{B} & 0 & & 0 \\ 0 & -\mathbf{B} & \mathbf{A}_3 & -\mathbf{B} & 0 & \vdots \\ \vdots & 0 & & \ddots & & 0 \\ 0 & & 0 & -\mathbf{B} & \mathbf{A}_{P-1} & -\mathbf{B} \\ 0 & 0 & \cdots & 0 & -\mathbf{B} & \mathbf{A}_P \end{bmatrix} \begin{bmatrix} \boldsymbol{\psi}_1^{l+1} \\ \boldsymbol{\psi}_2^{l+1} \\ \boldsymbol{\psi}_3^{l+1} \\ \vdots \\ \boldsymbol{\psi}_{P-1}^{l+1} \\ \boldsymbol{\psi}_P^{l+1} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{C}_1 & \mathbf{B} & 0 & \cdots & 0 & 0 \\ \mathbf{B} & \mathbf{C}_2 & \mathbf{B} & 0 & & 0 \\ 0 & \mathbf{B} & \mathbf{C}_3 & \mathbf{B} & 0 & \vdots \\ \vdots & 0 & & \ddots & & 0 \\ 0 & & 0 & \mathbf{B} & \mathbf{C}_{P-1} & \mathbf{B} \\ 0 & 0 & \cdots & 0 & \mathbf{B} & \mathbf{C}_P \end{bmatrix} \begin{bmatrix} \boldsymbol{\psi}_1^l \\ \boldsymbol{\psi}_2^l \\ \boldsymbol{\psi}_3^l \\ \vdots \\ \boldsymbol{\psi}_{P-1}^l \\ \boldsymbol{\psi}_P^l \end{bmatrix}, \tag{7.71}$$

which we can write symbolically as

$$\mathbf{A}\boldsymbol{\psi}^{l+1} = \mathbf{C}\boldsymbol{\psi}^l \quad \text{with} \quad \boldsymbol{\psi}^l = \begin{bmatrix} \boldsymbol{\psi}_1^l \\ \vdots \\ \boldsymbol{\psi}_P^l \end{bmatrix}. \tag{7.72}$$

The right-hand side is known since both $\mathbf{C}$ and ψ^l (the values of the field at the previous propagation step) are known. To solve for the unknown field values in the next step we have to solve this set of linear equations. As we propagate the beam we need to do this at every propagation step. If the problem is two-dimensional [i.e., if there is only one transverse dimension] $\mathbf{A}$ becomes a tridiagonal matrix. There are very efficient algorithms to solve for a tridiagonal system of equations [17]. Therefore, it is very advantageous to convert a three-dimensional problem into a two-dimensional problem using the effective index approximation, if possible. In the full three-dimensional case it is generally better to use iterative techniques rather than inversion of matrix $\mathbf{A}$ at each step.

7.4.3 Calculation of Propagation Constants in a *z*-invariant Waveguide from a Beam Propagation Solution

The field distribution $\psi(x, y, z)$ everywhere in a z-invariant waveguide can be calculated by applying the algorithm represented by Eq. (7.72). In the course of the propagating beam calculation, one calculates the correlation function, which is

$$P(z) = \iint \psi^*(x, y, 0)\psi(x, y, z)\, dx\, dy. \tag{7.73}$$

On the other hand, $\psi(x, y, z)$ can be represented by the superposition of orthogonal eigenfunctions of the z-invariant waveguide, which is

$$\psi(x, y, z) = \sum_n E_n U_n(x, y) \exp(-j\beta_{pn} z), \tag{7.74}$$

where $U_n(x, y)$ and β_{pn} are the eigenfunction and the propagation constant of the nth mode as obtained from the paraxial wave Eq. (7.63). In this expansion, it is assumed that degeneracy does not exist, which is a good approximation for dielectric waveguides. If Eq. (7.74) is substituted into Eq. (7.73), one obtains

$$P(z) = \sum_n |E_n|^2 \exp(-j\beta_{pn} z). \tag{7.75}$$

The Fourier transform of (7.75) is

$$P(\beta) = \sum_n |E_n|^2 \delta(\beta - \beta_{pn}). \tag{7.76}$$

Thus, one can find the propagation constant, β_{pn}, by numerically calculating the correlation function, $P(z)$, Fourier transforming it, and locating the peak in the Fourier domain. Ideally the accurate determination of β_{pn} can only be done by infinitely propagating the beam or when the electric field value is known over all z because only when z extends to infinity will the Fourier transform of Eq. (7.75) yield Eq. (7.76). However, in practice, one can propagate a beam only a finite length, hence field values over a certain z-range, or z-window, are known. In mathematical terms this is equivalent to multiplying Eq. (7.73) with a window function $w(z)$, which accounts for the finite length of propagation. Then the Fourier transform of the correlation function, $P_w(z)$, becomes

$$P_w(\beta) = \sum_n |E_n|^2 L(\beta - \beta_{pn}), \tag{7.77}$$

where the lineshape function for the propagation distance, D, is defined by

$$L(\beta - \beta_{pn}) = \frac{1}{D} \int_0^D \exp[j(\beta - \beta_{pn})z] w(z)\, dz. \tag{7.78}$$

Knowing this lineshape function, the propagation constant can be quite accurately determined from the spectrum, $P_w(\beta)$, using curve fitting. In the calculation, the Hanning window function, $w(z) = 1 - \cos[(2\pi z)/D]$, is used as is typically done in the literature. The eigenfunction of the paraxial wave Eq. (7.76) is identical to that of the original scalar Helmholtz equation. However, the propagation constant of the Helmholtz equation, β_h, is found from that of the paraxial equation, β_p, using the relation,

$$\beta_h = k_0 n_r (1 + 2\beta_p / k_0 n_r)^{1/2}. \tag{7.79}$$

The details of calculating the peak position from the spectrum are described in [12].

7.4.4 Calculation of Eigenmode Profile from a Beam Propagation Solution

If both sides of Eq. (7.74) are multiplied by $D^{-1}\ w(z)\exp(j\beta z)$ and integrated from 0 to D, we can obtain

$$\psi(x,y,\beta) = \frac{1}{D}\int_0^D \psi(x,y,z)\exp(j\beta z)w(z)\,dz \tag{7.80}$$

$$= \sum_n E_n U_n(x,y)L(\beta-\beta_{pn}). \tag{7.81}$$

Thus, for $\beta = \beta_{pi}$, $\psi(x,y,\beta_{pi})$ can be expressed as

$$\psi(x,y,\beta_{pi}) = E_i U_i(x,y)L(0) + \sum_{n\neq i} E_n U_n(x,y)L(\beta_{pi}-\beta_{pn}). \tag{7.82}$$

Equation (7.82) shows that the eigenmode profile $U_i(x,y)$ can be determined by evaluating the integral (7.80) with $\beta = \beta_{pi}$ provided that most of the excited power belongs to the ith mode, which is the mode of interest. In practice, such excitation can be achieved in most cases for dielectric waveguides. The detailed description of this method can be found in [19].

7.5 GOOS–HANCHEN EFFECT AND TOTAL INTERNAL REFLECTION COMPONENTS

In Section 6.2, Eq. (6.13), we have defined the effective width of a waveguide as means of relating the peak field to the total power in the mode. However, this effective width also has a physical interpretation, as suggested in Fig. 6.1. Specifically, while the phase fronts of a wave in the zigzag ray picture bounce between effective metal planes spaced apart by π/k_x, the *energy* propagating down the guide effectively bounces off of planes placed at one over the decay length into each cladding region, giving a total width of d_{eff} between bounces. This energy penetration accounts for the evanescent fields in the cladding layers and results in an axial shift and time delay of an incident beam before it is reflected off of the interface. The axial shift is known as the *Goos–Hanchen shift* and represents a very real displacement of energy that is observable in the laboratory.

To rigorously derive the Goos–Hanchen shift, L_{G-H}, we must consider the group characteristics of a spatial *wave packet* reflecting off of the interface beyond the critical angle. Fortunately, we can also guess at the result by noting the familiar relations for phase velocity, ω/β, and energy velocity, $d\omega/d\beta$. For our reflected wave, the phase fronts are advanced along z by ϕ/β, since $\exp[-j\beta z]\exp[j\phi] = \exp[-j\beta(z-\phi/\beta)]$. We therefore predict that the energy is advanced or displaced by $d\phi/d\beta$. A detailed wave packet analysis would in fact confirm this prediction. In other words, the Goos–Hanchen shift is mathematically related to the *dispersion*

of the reflection phase angle, ϕ. Using Eq. (7.11) for the reflection phase of a TE mode, we can evaluate the derivative to obtain

$$L_{G-H} = \frac{d\phi}{d\beta} = \frac{2\tan\theta}{\gamma_x}. \quad \text{(Goos–Hanchen shift)} \tag{7.83}$$

As shown on the right in Fig. 6.1, this is the same result we would obtain if the energy had to propagate (transversely) an additional $1/\gamma_x$ before reflecting back. With a Goos–Hanchen shift on both sides then, the total effective energy propagation width is just d_{eff} defined in Eq. (6.13) or (7.66), as initially assumed. Also note that the length of the guide required for the energy to make one transverse round-trip is $2d_{eff}\tan\theta$.

Goos–Hanchen effect therefore has a significant impact on the waveguide components based on total internal reflection, which will be briefly examined in the remainer of this section.

7.5.1 Total Internal Reflection Mirrors

Total internal reflection (TIR) mirrors are the most compact elements that can be used to change the direction of light in waveguides. They utilize a discrete, high-index contrast air-semiconductor interface to reflect the light off it, rather than using the deep etch to continuously guide the light in a waveguide bend. A typical TIR mirror is illustrated in Fig. 7.24a. The light is lauched from an access waveguide, into a free propagation region, where the light difracts before hitting the mirror surface. At the mirror output, the reflected light is imaged at the output waveguide aperture and coupled back into the waveguide mode. The efficiency and loss of a TIR mirror will depend on carefully tailoring the free propagation and the mirror surface positions in order to optimize this imaging into the output waveguide. This is illustrated in Fig. 7.24d. The reflection plane position will be determined by the Goos–Hanchen shift, and due to the differences in reflected phase angles (effective waveguide widths) for light waves of different polarizations, per Eq. (7.11), the optimum reflector position will generally be different for TE and TM polarized light.

The mirror surface can be planar, or concave, in which case the mirror also acts as a lens, focusing the reflected light. This is advantageous for large mirror sizes and low loss mirrors [21]. Mirror losses of less than 0.1 dB per reflection can be achieved. The surface position, verticality, and depth of the etched mirror surface are critical to achieving low loss.

Frustrated Total Internal Reflection Beam Splitters Directional coupler and multimode interference–based light splitters have already been discussed in previous chapters as very important functional components in PIC realization. Total internal reflection mirror approach can be used to produce extremely short, submicron scale slot-based light splitters. If two flat TIR mirrors are positioned back to back, with a narrow slot (index) discontinuity between their faces, light launched

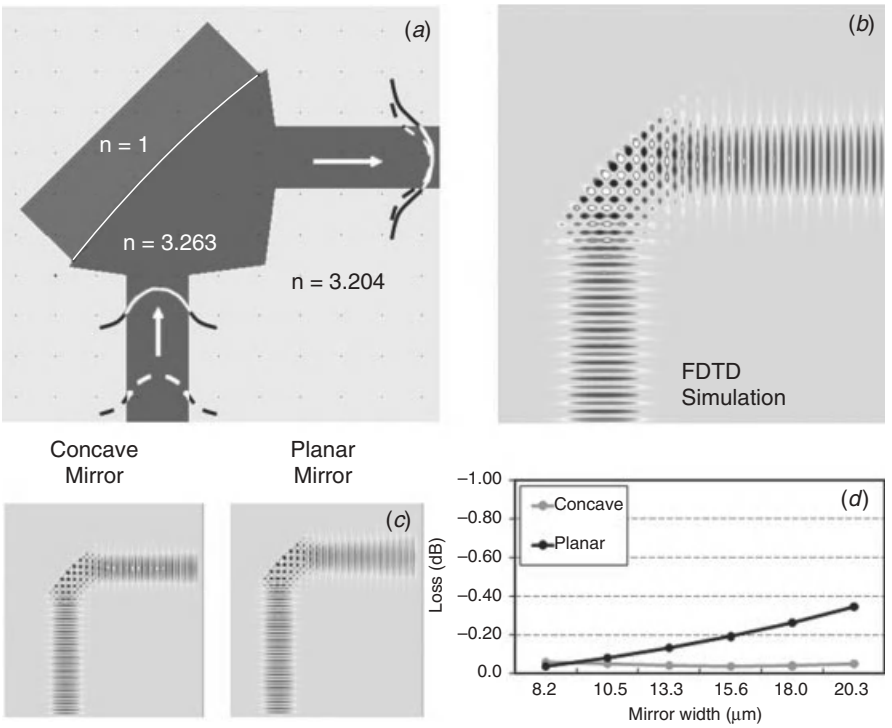


FIGURE 7.24: (a) Schematic of a total internal reflection mirror. (b) Finite-difference time domain two-dimensional simulation of the reflection for the fundamental mode at $\lambda = 1550$ nm. (c) Finite-difference time domain two-dimensional simulations for a planar and concave mirror design. (d) Loss as function of TIR mirror width for the planar and concave mirror designs. (*Reprinted with permission from Dr. Joseph Summers*).

into one access waveguide will partially continue its propagation through the slot, while the remaining light will be reflected, through frustrated total internal reflection (FTIR). This is illustrated in Fig. 7.25. FTIR can occur even when the angle of incidence onto the slot exceeds the critical angle, since, as explained in this section, there will be an evanescent field extending into the slot at the incoming TIR's interface. If the second interface of the slot is close enough that this decaying field still has a non-negligible amplitude, then FTIR occurs and a certain amount of the input power tunnels through the gap and continues to propagate into the opposite waveguide. As the gap width increases, the transmitted optical through the slot decreases exponentially. One drawback to FTIR splitters is that they have inherent losses, due to the inability to maximize the reflection and transmission at the same time, caused by a combination of Goos–Hanchen shift and diffraction effects. Similar to the TIR mirrors, the performance of these splitters will also critically depend on the verticality and smoothness of the TIR/slot faces, as well as control of the actual slot dimensions.

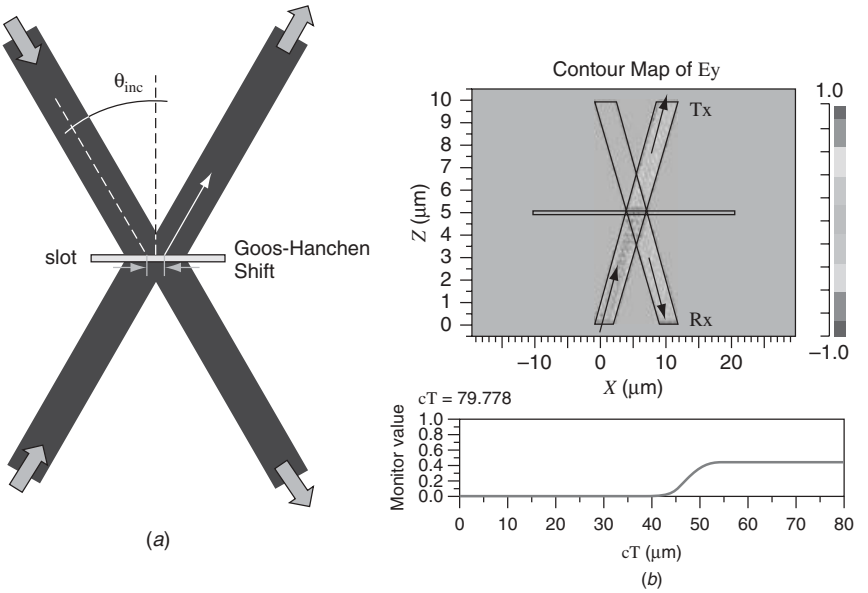


FIGURE 7.25: (a) Schematic of an FTIR slot based splitter. Θ_{inc} represents the incident angle, and the Goos-Hanchen shift is also specified. (b) FDTD simulation showing equal splitting for an FTIR air slot based splitter at $\lambda = 1550$ nm. (*Reprinted with permission from Dr. Janet Chen*).

7.6 LOSSES IN DIELECTRIC WAVEGUIDES

7.6.1 Absorption Losses in Dielectric Waveguides

Many dielectric waveguides are implemented using semiconductor alloys that have a direct bandgap, and they absorb strongly at wavelengths shorter than the corresponding band edge. These types of losses can be avoided by appropriate selection of alloy composition and operating wavelength. In applications involving photonic integrated circuits, special integration techniques for implementation of both low loss dielectric waveguides and active photonic blocks must be used.

Another important loss component in dielectric materials is free-carrier absorption, caused by the presence of free P and N carriers in the material. Free carrier absorption can be intraband (where an electron below the Fermi level absorbs a photon and moves to a free state above the Fermi level) and interband. Typical bulk losses for N-type material, doped at $1 \cdot 10^{18}$ cm^{-3}, are 22 and 4 dB/cm for GaAs and InP respectively, at 1.5 μm [13]. The greater valence band complexity, as discussed in Chapter 4, causes significantly higher absorption in P-type III-V materials, typically 56 dB/cm for $1 \cdot 10^{18}$ cm^{-3} of P doping [13]. The total absorption loss will depend on the waveguide mode overlap with the P and N doped regions. In active devices, where P doping is necessary to form a p-n junction, the losses can be mitigated by minimizing the confinement factor Γ between P regions and the mode.

Deep level traps, which cause absorption through similar mechanism to intraband absorption, can lead to losses of 3.5 and 1 dB in GaAs and InP respectively, at 1.55 μm.

7.6.2 Scattering Losses in Dielectric Waveguides

Rough interfaces on dielectric waveguide boundaries will cause additional losses through the effect of scattering. The mechanism of scattering losses can be explained by referring to the ray propagation illustration in Fig. 7.5. As a ray hits the waveguide interface at an angle that is larger than the critical angle, it reflects back from the interface. Local roughness at the interface changes the local angle of incidence, thus changing the phase of the reflected wave, and the evanescent tail of the mode, as given by Eq. (7.15). This can affect the absorption losses, as those depend on the overlap of the mode with the absorbing material. If the local roughness at the interface is such that the angle of incidence is lower than the critical angle, part of the light will be refracted, and thus lost. The reflected part of the light is now incident to the next interface with a different angle, and becomes a radiation mode.

Determining the actual scattering losses from measured roughness values is a complex problem due to the two-dimensional nature of the waveguide index distribution, dependence on both amplitude and spatial periodicity of the roughness, and the difficulty associated with measuring small amplitude roughness.

One simple approach that can be used to simplify the three dimensional problem is to use effective index method to reduce the waveguide to a two-dimensional structure. With this approximation, the scattering loss α of a symmetric planar waveguide is given by [14]

$$\alpha = \frac{\sigma^2 k_0^2 k_x}{\beta} \frac{E_S^2 \Delta n^2}{\int E^2 \, dx},$$

where σ is the interface roughness, Δn^2 is the difference in refractive indices of the core and the cladding squared, and $E_S^2 / \int E^2 \, dx$ is the normalized optical intensity at the guide-cladding interface.

A more detailed analysis of the scattering problem indicates that the above approach, while giving the correct trends and identifying the main parameters of scattering, can overestimate the scattering losses depending on the actual 3-D waveguide geometry. More accurate, and complex numerical methods, such is a volume current method [15], can be used to overcome these limitations.

7.6.3 Radiation Losses for Nominally Guided Modes

There are a few practical situations where lossless guided modes become leaky modes. One occurs when a waveguide is curved to displace the axis of the guide. This is very common in practical devices. Another occurs when some cladding is removed or replaced by higher-index material in a symmetric guide over some

device region. This creates an asymmetric guide that may be cut off (i.e., leaky). A different case is an antiguide, which is deliberately constructed to have the largest index in the cladding regions. In these latter cases, we may want to use the energy that leaks out; the axis of the guide remains the z-axis, and the optical leak rate (or propagation loss) can be calculated from the transmission that occurs for a zigzagging ray at one or both waveguide boundaries. The loss rate per unit length is the transmission loss divided by the axial length it takes a ray to cycle back and forth transversely across the guide.

Optical Antiguides and ARROW Waveguides We begin by considering a symmetric antiguide as shown in Fig. 7.26. Even though there is partial transmission at each boundary, the optical energy will be "guided" with relatively low loss provided transverse resonance is satisfied. Since there is π phase shift at each boundary for $n_1 < n_2$, the transverse resonance condition for the antiguide is

$$k_x d = (m+1)\pi. \quad \text{(antiguide)} \tag{7.84}$$

Assuming TE polarization, the transmission at each boundary is given by $(1 - r^2)$, where r is given by Eq. (7.4). A particular ray strikes a boundary once in a distance Δz. Setting $k_x \approx k_1 d/\Delta z$ in Eq. (7.84), we obtain

$$\Delta z = \frac{2n_1 d^2}{(m+1)\lambda}, \tag{7.85}$$

for paraxial rays (small angles). Assuming the power decays exponentially according to $P(z) = P(0)e^{-\alpha z}$, we can solve for the attenuation constant as the relative power change per unit length, which can be approximated by

$$\alpha = \frac{1}{\Delta z}\frac{\Delta P}{P} = \frac{(1-r^2)(m+1)\lambda}{2n_1 d^2}. \quad \text{(antiguide)} \tag{7.86}$$

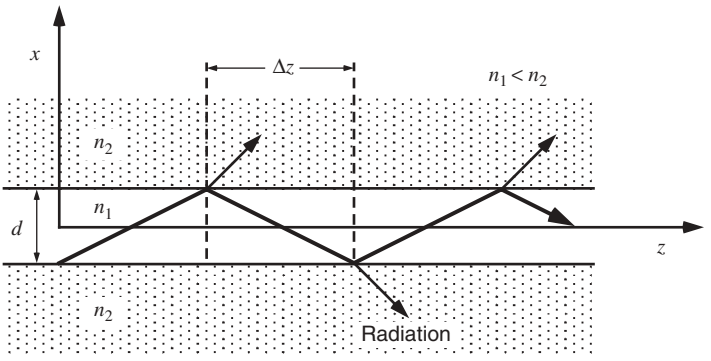


FIGURE 7.26: Illustration of radiation at boundaries in an antiguide for which $n_1 < n_2$.

Equation (7.86) gives the attenuation constant for lightwaves propagating in a symmetric antiguide. Any coherent addition of the radiated fields has been neglected. This is valid in most situations where some absorbing or scattering boundary intercepts the radiated field before it has a chance to add along any appreciable length. The attenuation can be quite high or relatively low depending upon the parameters. For example, consider a typical case where $m = 0, \lambda = 1\ \mu\text{m}, n_1 = 3.5, d = 3\ \mu\text{m}$ and the index difference gives a ray reflection coefficient, $r^2 = 0.9$. Plugging into Eq. (7.86), we find $\alpha = 1.59\ \text{cm}^{-1}$, which is not an extremely large loss for semiconductor waveguides. The d used suggests that this would be the lateral waveguide dimension rather than the transverse in the semiconductor case. For small d, e.g., one-tenth of that in the example, the loss would be 100 times larger, and this would be an extremely high loss.

If multiple boundaries are included with reflections that add in phase for some wavelength, the net propagation loss can be significantly reduced for a leaky waveguide. Also, they can operate in a quasi-single mode even for large core widths, since only a particular ray angle experiences low loss. This can lead to desirably low vertical diffraction angles at output facets in semiconductor waveguides. Such "waveguides" have been referred to as antiresonant reflecting optical waveguides or ARROW's [8]. Figure 7.27 shows an example of an ARROW guide.

For low-loss we desire a slab mode with a null at the lower interface. This is accomplished with a vertical standing wave that has a node at the interface between the slab and the first high-index reflector layer. If the top surface index difference is large, then the zigzag ray angle in the slab must be, $\theta \approx \cos^{-1}(\lambda/2n_1d_1)$. For the reflection layer of thickness t and index n_2 to be in antiresonance for high reflection,

$$t \approx \frac{\lambda}{4n_2}(2N+1)\left[1 - \frac{n_1^2}{n_2^2} + \frac{\lambda^2}{4n_2^2 d_1^2}\right]^{-1/2}, \quad N = 0, 1, 2, \ldots \tag{7.87}$$

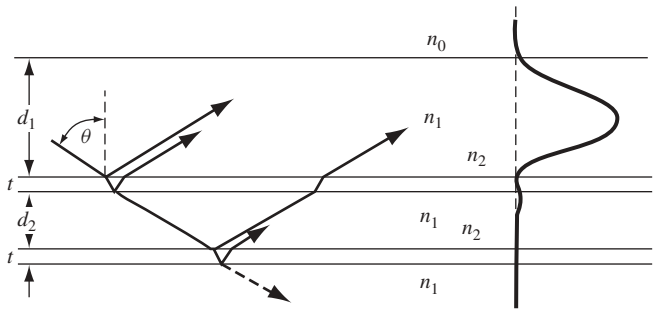


FIGURE 7.27: ARROW slab guide bounded on its upper surface by a low-index medium (n_0) and on its substrate side by one (or more) higher-index (n_2) anti-resonant reflector layers. After [8]. (*Reprinted, by permission, from Electronics Letters*).

The spacing d_2 to the second high-index reflector should also be antiresonant to maximize the net reflection. That is,

$$d_2 = \frac{d_1}{2}(2M + 1), \quad M = 0, 1, 2, \ldots \tag{7.88}$$

Using the transverse resonance recipe of Section 7.3.2 and the technique for calculating radiation loss outlined above, one can rigorously calculate the radiation loss into the substrate. Figure 7.28 gives an example result for the case outlined in Fig. 7.27. The reflector layer thicknesses are varied along with the slab thickness according to the approximate design rules given above.

In addition to having low propagation losses and large fundamental mode size, ARROW modes are also interesting because of the very low overlap of optical energy with the first reflecting layer. Thus, it is possible to have an absorbing reflector layer and still have low waveguide loss. The placement of a null at the high-loss layer provides a unique way of placing high-loss layers (e.g., gain layers) immediately adjacent to low-loss guides.

Dielectric Waveguide Bends In this section, we evaluate the radiation losses of waveguide bends. The case considered is a dielectric waveguides bend at a radius of R, as indicated in Fig. 7.29. For this problem we use a first-order perturbation approach in which the unperturbed eigenmode is used in the perturbed problem to calculate the loss. When this mode enters the bend of radius R, it is assumed to stay in the center of the guide with its phase fronts along radii from the center of curvature. From the figure we see that at some radius, x_R, the exponential tail

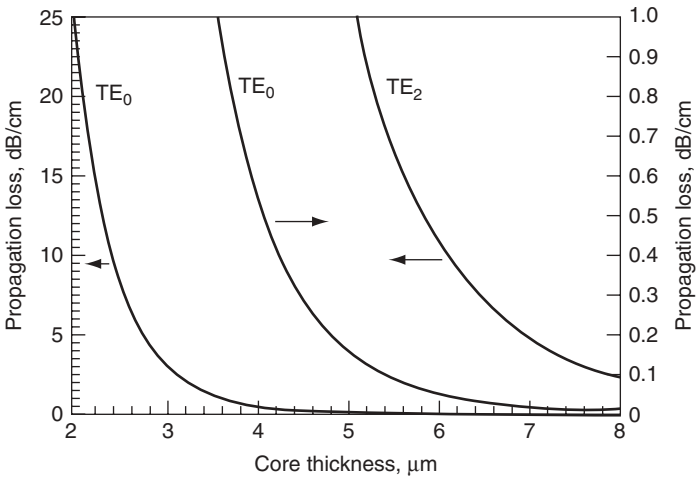


FIGURE 7.28: Computed radiation loss as a function of slab thickness for the two lowest-loss TE modes in a double-reflector-layer ARROW at 1.513 μm [8], with $n_0 = 1.0, n_1 = 3.17$, and $n_2 = 3.4$. (*Reprinted, by permission, from Electronics Letters*).

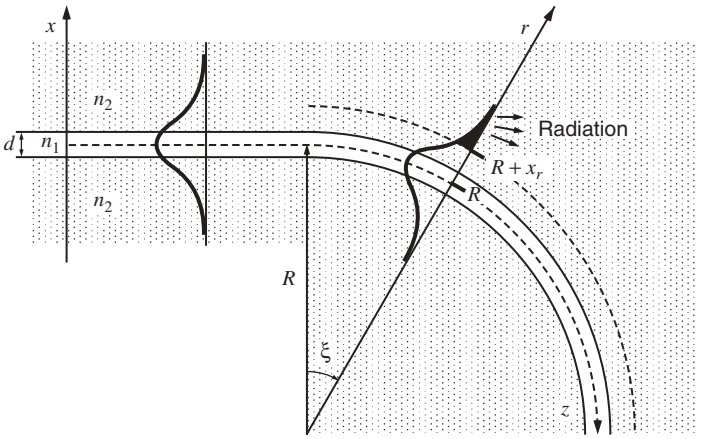


FIGURE 7.29: Illustration of an eigenmode entering a waveguide bend of radius R so that the outside tail of the mode must radiate.

of the unperturbed mode must travel faster than the speed of light in the cladding medium to keep up with the rest of the mode. Clearly, this is not allowed, and we must assume that the light beyond this radius radiates away into the cladding.

To obtain the value of x_R, we note that the local phase velocity is equal to the radial distance times the angular sweep rate, and that it is assumed to equal the unperturbed guide phase velocity, ω/β, at the center of the guide ($r = R$). That is,

$$v_p(r) = r\frac{d\xi}{dt} = r\frac{1}{R}\frac{\omega}{\beta}. \tag{7.89}$$

Now, by the criteria we have adopted, $v_p(R + x_R) = \omega/k_2$. Thus,

$$x_R = R(\beta/k_2 - 1). \tag{7.90}$$

As in the antiguide case, the problem is to determine the fraction of power lost, $\Delta P/P$, over some incremental distance, Δz, so that an attenuation constant can be estimated. Again, we shall neglect coherent addition of radiated fields, although this is generally *not* a good assumption. For the relative power lost we use the fraction of power contained in the portion of the mode that radiates (shaded region in Fig. 7.29). This is a very straightforward integration using the unperturbed mode.

Estimating the effective distance it takes this energy to be radiated out of the mode is more complex. For this we use a diffraction distance defined as the distance where the peak power density falls to $1/e$ of its initial value. This is meaningful if the diffracting energy keeps the same functional form as it diffracts, for then, the width of the one-dimensionally diffracting energy in the slab geometry has increased by e-times in this same distance. It turns out that the Hermite–Gaussian functions introduced earlier by Eq. (7.43) as the eigenmodes of a parabolic dielectric constant medium also serve as the eigenmodes of free space in the sense that

they maintain the same functional form laterally, even though their characteristic width, w, increases. Thus, they become the natural basis set in which to expand arbitrary excitation field profiles in a uniform medium. This important property comes about because the Fresnel diffraction integral generates the Fourier transform of a function, and the Hermite–Gaussian functions transform back and forth to other Hermite–Gaussian functions, reproducing the original function scaled only in width after some diffraction distance.

Therefore, our procedure for determining the diffraction distance is to expand our radiating excitation field in terms of Hermite–Gaussian basis functions and determine the distance, Δz, for their width to increase by e-times. Figure 7.30 illustrates this expansion using only the first two Hermite–Gaussian functions. In the present case the excitation field is a decaying exponential of the form,

$$\mathscr{E}_e(0) = Ae^{-\gamma_x(x-x_R)}, \quad x > x_R, \tag{7.91}$$

and the basis functions are given by Eq. (7.43). By iteration for best fit we find that the Gaussian $1/e$ spot size, w_0, must be approximately

$$w_0 \approx \frac{0.6}{\gamma_x}. \tag{7.92}$$

For a freely diffracting Hermite–Gaussian beam the characteristic width, w, increases from w_0 at $z = 0$ according to [9],

$$w^2 = w_0^2\left[1 + \left(\frac{\lambda z}{\pi n w_0^2}\right)^2\right]. \tag{7.93}$$

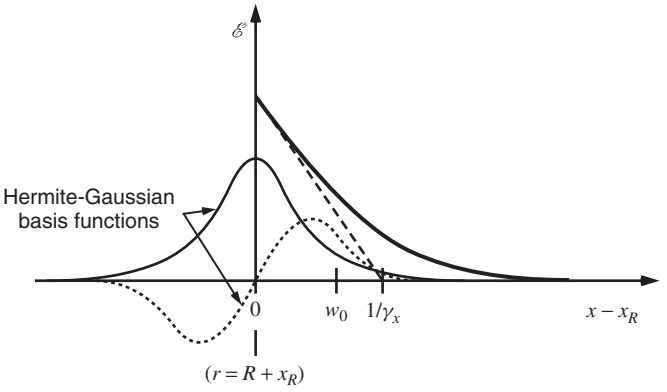


FIGURE 7.30: Radiating part of waveguide eigenmode for $x > x_R$. Superposition of the first two Hermite–Gaussian basis functions can approximate this field. Their characteristic width determines the diffraction distance.

Interestingly, this is true for all of the higher-order functions as well as the fundamental. Thus, we do not have to keep track of different characteristic widths, and the diffraction problem is completely defined once the excitation problem is solved.

Setting $w/w_0 = e$ in Eq. (7.93) for the desired diffraction length, and using the estimation for the initial width of Eq. (7.83), we obtain

$$\Delta z = \sqrt{e^2 - 1} \cdot \frac{\pi n w_0^2}{\lambda} \approx \frac{\pi n_2}{\lambda \gamma_x^2}, \tag{7.94}$$

where n_2 and γ_x are the refractive index and the unperturbed modal decay constant, respectively, in the cladding region where the radiation is occurring. It is also notable that over this distance, Δz, the wavefronts of the diffracting Hermite–Gaussian functions curve significantly, so that a coherent interaction with the planar phase fronts of the guided mode is inhibited.

Now, we can estimate the attenuation constant, α, from

$$\alpha = \frac{1}{\Delta z}\frac{\Delta P}{P} = \left(\frac{\lambda \gamma_x^2}{\pi n_2}\right) \frac{\dfrac{w}{2\eta_g}\displaystyle\int_{x_R}^{\infty} |\mathscr{E}_y|^2\, dx}{\dfrac{w}{2\eta_g}\displaystyle\int_{-\infty}^{\infty} |\mathscr{E}_y|^2\, dx}. \tag{7.95}$$

Completing the integrations similarly to the work leading up to Eq. (7.64), we obtain

$$\alpha = \frac{\lambda \gamma_x^2}{\pi n_2(\gamma_x d + 2)} \cos^2(k_{1x} d/2) e^{\gamma_x d} e^{-2\gamma_x(\beta/k_2 - 1)R}. \tag{7.96}$$

To uncover important analytical dependences and allow plotting of normalized curves, Eq. (7.96) can be approximated by

$$\alpha = C_1 e^{-C_2 R}.$$

From Eq. (7.96), it can be shown that $C_1 \approx \gamma_x^2/k_2$ and $C_2 \approx \gamma_x^3/k_2^2$, assuming weak guiding so that $\beta/k_2 \approx 1 + 0.5(\gamma_x/k_2)^2$. However, more exact calculations have shown that the second quantity is better estimated by $C_2' \approx \frac{2}{3}\gamma_x^3/k_2^2$. Using this and the fact that for weak guiding, $\gamma_x d$ can be expressed solely in terms of the normalized frequency, V, introduced in Appendix 3 (i.e., $\gamma_x d \approx \sqrt{1 + V^2} - 1$), we can generate the normalized plots similar to those given in Fig. 7.31. However, it is important to note that these calculations do not include the mode mismatch loss between the straight and curved sections, as discused in one of the previous sections.

The plot in Fig. 7.31 shows normalized radiation losses for a family of different curved waveguides, of different index contrasts. They were generated numerically using a combination of the beam propagation method and conformal mapping to represent the curved waveguides, as discussed in Section 7.3.8. By measuring the fundamental mode power as function of propagation distance, one can easily extract the radiation loss coefficient, α, without the need for any approximations.

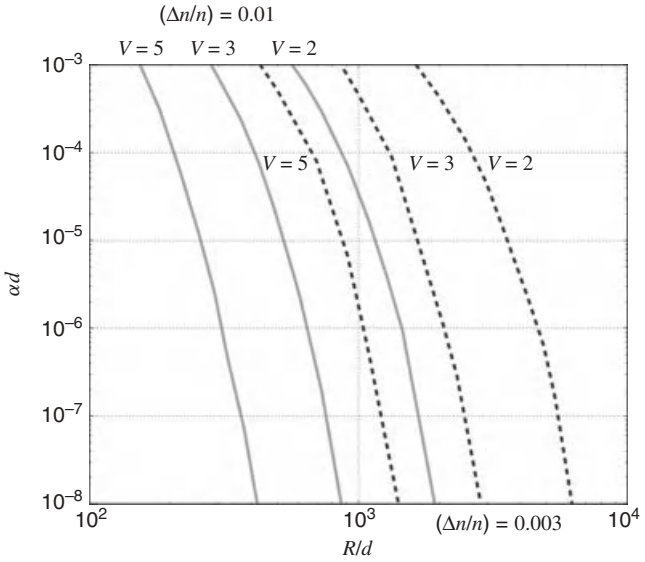


FIGURE 7.31: Bending loss as a function of bend radius for different normalized frequencies and index steps. The loss has been calculated using the conformal mapping technique in conjunction with the beam propagation technique.

REFERENCES

1. Lee DL. *Electromagnetic principles of integrated optics*. New York: Wiley; 1986. Chapter 3.
2. Haus HA. *Waves and fields in optoelectronics*. Englewood Cliffs, NJ: Prentice Hall; 1984. Chapter 2
3. Lee DL. *Electromagnetic principles of integrated optics*. New York: Wiley; 1986. Chapter 4.
4. Stern MS. Semivectorial polarised finite difference method for optical waveguides with arbitrary index profiles. *IEE Proceedings J: Optoelectronics* 1988;**135**: 56.
5. Dagli N. Equivalent circuit representation of open guided-wave structures. *IEEE J Quantum Electron* 1990;**26**: 90.
6. Peng S-T, Oliner AA. Guidance and leakage properties of a class of open dielectric waveguides: part I—mathematical formulations. *IEEE Trans Microwave Theory Tech* 1981;**29**: 843.
7. Rahman BMA, Davies JB. Finite-element analysis of optical and microwave waveguide problems. *IEEE Trans Microwave Theory Tech* 1984;**32**: 20.
8. Koch TL, Koren U, Boyd GD, Corvini PJ, Duguay MA. Antiresonant reflecting optical waveguides for III-V integrated optics. *Electron Lett* 1987;**23**: 244.
9. Haus HA. *Waves and fields in optoelectronics*. Englewood Cliffs, NJ: Prentice Hall; 1984. Chapter 5.

10. Heiblum M, Harris J. Analysis of curved optical waveguides by conformal transformation. *IEEE J Quantum Electron* 1975;**11**: 75.
11. Marcatili EAH. Bends in optical dielectric guides. *Bell Syst Tech J* 1969;**48**: 2103.
12. Heinbach M, Schienle M, Schmid A, Acklin B, Muller G. Low-loss bent connections for optical switches. *IEEE J Lightwave Technol* 1997;**15**: 833.
13. Deri RJ, Kapone E. Low-loss III-V semiconductor optical waveguides. *IEEE Journ Quantum Electronics* 1991;**27**: 626.
14. Tein PK. Light waves in thin films and integrated optics. *Appl Opt* 1971;**10**: 2395–2413.
15. Barwicz T, Haus HA. Three-dimensional analysis of scattering losses due to sidewall roughness in microphotonic waveguides. *IEEE J Lightwave Technol* 2005;**23**: 2719.
16. Chung Y, Dagli N. An assessment of finite difference beam propagation method. *IEEE J Quantum Electron* 1990;**26**: 1335.
17. Press WH, Flannery BP, Teukolsky SA, Vetterling WT. *Numerical recipes, the art of scientific computing*. New York: Cambridge University Press; 1988.
18. Smith GD. Numerical solution of partial differential equations. In: *Finite difference methods*. New York: Oxford University Perss; 1985.
19. Chung Y, Dagli N. Analysis of Z-invariant and Z-variant semiconductor rib waveguides by explicit finite difference beam propagation method with nonuniform mesh configuration. *IEEE J Quantum Electron* 1991;**27**: 2296.
20. Hadley GR. Transparent boundary condition for the beam propagation method. *IEEE J Quantum Electron* 1992;**28**: 363.
21. Summers JA, Lal V, Masanovic ML, Dagli N, Blumenthal DJ. Concave low-loss total internal reflection mirrors in indium phosphide for high fabrication tolerance. Conference on Lasers and Electro-Optics (CLEO); 2007 May 6; Baltimore.

READING LIST

Lee DL. *Electromagnetic Principles of Integrated Optics*. New York: Wiley; 1986. Chapters 3–5.

Haus HA. *Waves and Fields in optoelectronics*. Englewood Cliffs, NJ: Prentice Hall; 1984. Chapter 2.

Ebeling KJ. *Integrated Opto-electronics*. Berlin: Springer-Verlag; 1993. Chapters 2–4.

Yariv A, Yeh P. *Optical Waves in Crystals*. New York: Wiley; 1984. Chapter 11.

PROBLEMS

1. For TM plane waves incident on a planar dielectric boundary, it is possible to get total transmission as well as total reflection. This is called Brewster's angle. Derive an expression for the incident Brewster's angle in terms of the relative dielectric constants of the two media.

2. A plane wave with $\lambda = 0.85$ μm in a medium of index 3.5 is incident on another dielectric medium of index 3.0. The reflection coefficient at the planar

interface is determined to be $1\angle 60^\circ$. What is the angle of incidence for TE polarization? Repeat assuming TM polarization.

3. A symmetric three-layer slab guide has a core region 0.4 μm thick. Its index is 3.4 and that of the cladding regions is 3.2.

(a) Determine the ray angle for the reflecting plane waves that make up the fundamental TE guided mode.

(b) What guide width will provide the same ray angle for the first higher-order odd TE mode?

(c) If the cladding index were changed to 3.0 on one side, what guide width would be necessary to obtain the same ray angle for the fundamental TE mode? Assume $\lambda = 1.55$ μm.

4. For the waveguide structure as defined in Example 7.2, determine the thickness d of the waveguide that supports the fundamental mode at 1.55 μm, incident at 75°.

5. Plot the effective index versus center layer thickness, d, for $0.1 < d < 0.2$ μm with $\lambda = 1.3$ μm for the multilayer slab guide shown in Fig. 7.32. Add dispersion curves for any higher order modes as they become lossless.

6. In exciting a 0.3 μm thick GaAs slab waveguide, it seems that much of the excitation energy goes into radiation modes. In order to expand this excitation in terms of the eigenmodes of the system, perfectly reflecting planes are assumed to exist at a distance ±30 μm on each side of the slab. The cladding material is $Al_{0.2}\ Ga_{0.8}$ As and the wavelength is 1.0 μm. Plot the propagation constant versus mode number for all of the modes.

7. Using the WKB method, determine the effective index versus waveguide base width for the fundamental mode of the triangular index profile shown in Fig. 7.33 with $4.0 < d < 1.0$ μm. Assume $\lambda = 1.0$ μm. Indicate any regions where the WKB approximation may not be accurate. Also, plot the ray path for the mode at $d = 2.0$ μm.

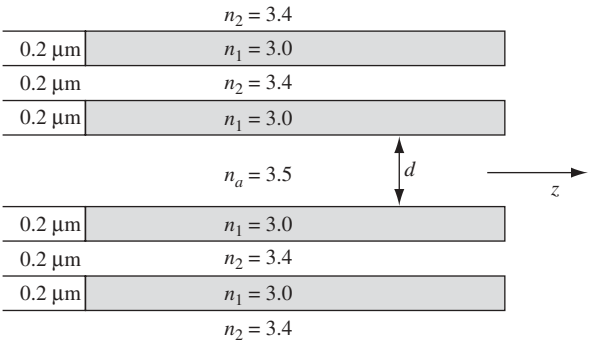


FIGURE 7.32: Schematic of multilayer slab waveguide.

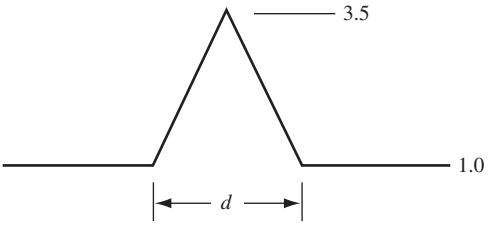


FIGURE 7.33: Triangular index profile waveguide.

8. A quarter-pitch GRINROD is 1 mm in diameter and 2 mm long. The maximum index in the center is 1.6. If an object is placed 1 mm in front of the GRINROD, what is the total distance from the object to the image, and what is the magnification?

9. A pencil-like beam of 1.3 μm light, polarized perpendicular to the plane of incidence, strikes an air interface from a medium of index of 3 at an incident angle of θ_i. Plot the shift of the reflected beam center versus θ_i for $0^\circ < \theta_i < 90^\circ$.

10. A graded index slab waveguide has an index profile given by

$$n^2(x) = 3.5^2\left[1 - \left(\frac{x}{8.5\ \mu\text{m}}\right)^2\right]$$

and it is tested at a wavelength of 0.85 μm.

(a) What are the effective indices for the first three modes?

(b) What are the eigenmode widths (measured between inflection points on the mode amplitude) for the lowest three modes?

11. For the graded index slab from the previous example,

(a) What are the ray crossing angles on axis, for the three lowest order modes?

(b) A Gaussian beam centered on its axis with a $1/e$ amplitude half width of 0.5 μm is now coupled to this guide. Assuming no reflections, what fraction of this excitation field is coupled to the fundamental mode?

12. What is the loss of the fundamental "leaky mode" that satisfies transverse resonance for an antiguide consisting of a slab 1.0 μm thick of index 3.2 sandwiched between media of index 3.4? Assume $\lambda = 1.3$ μm.

13. In order to fan out from a directional coupler to adjacent fibers, symmetric S bend waveguide sections are required as shown in Fig. 7.34. Assuming BH waveguides of 1.3 μm Q-material surrounded by InP for the 1.55 μm lightwaves, waveguide thicknesses of 0.4 μm and widths of 2.0 μm and bend radii of R_b to accomplish the total lateral shift of 70 μm, what is the total excess loss due to the four bend sections between input and output ports? $R_b = 100\ \mu\text{m}, 150\ \mu\text{m}$.

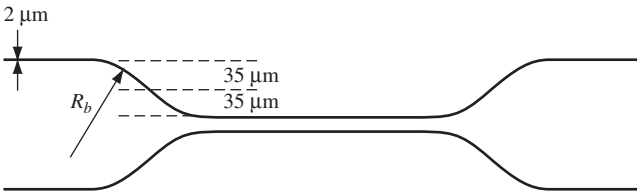


FIGURE 7.34: Directional coupler with input and output S-bends for fan out to fibers.

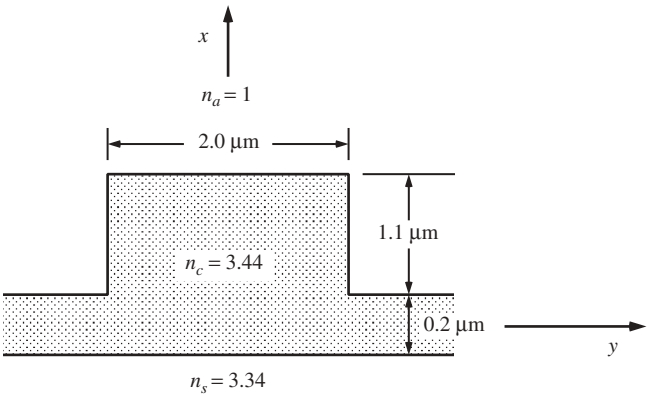


FIGURE 7.35: Rib waveguide cross section.

14. Find the effective index, $\bar{n}$, and the field profile, $U(x, y)$, at 1.55 μm of the fundamental mode of the rib waveguide shown in Fig. 7.35 by numerically solving the scalar wave equation using the finite-difference technique. *Hint*: This is a strongly guiding waveguide, so the computational window boundaries on which $U = 0$ can be placed fairly close to the core of the waveguide. In the initial trial, a 4.8-μm-wide (y-direction) by 2.3-μm-high (x-direction) window together with a grid size given by $\Delta x = 0.1$ μm and $\Delta y = 0.4$ μm should be sufficient. Place the window to leave 0.25 μm above, 0.65 μm below and 1.4 μm on either side of the high-index rib. For more details see M. J. Robertson et al., Semiconductor waveguides: analysis of optical propagation in single rib structures and directional couplers, *IEE Proceedings, Part J*, **132**(b), (1985).

15. Find the effective index, $\bar{n}$, and the 1-D field profile, $U(x)$, at 1.55 μm of the fundamental mode of a rib waveguide curve with whose radius is $R = 400$ μm. The index of the cladding is 3.19, and the index of the core is 3.25. Solve numerically, from the scalar wave equation using the finite-difference technique.

16. Consider the slab guide geometry shown below. The free-space wavelength is 1.3 μm.

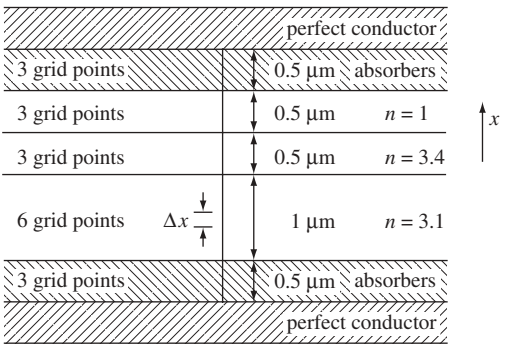
Choose the computational window as shown in the figure, i.e., 1 μm above the core in the air and 1.5 μm below the core in the substrate. Introduce absorbing regions 0.5 μm thick on the boundaries of the computational window. In the absorbing regions we need to introduce a negative imaginary part to the actual index distribution. For example, the indices of the absorbers in the air and in the substrate can be chosen as $1 - j0.08$ and $3.1 - j0.08$ respectively. Now set up a grid such that $\Delta x = 1/6$ μm. This should result in 18 grid points as indicated in the figure. Excite this geometry at $z = 0$ with a Gaussian beam whose profile is given as,

$$\psi(x, 0) = \exp\left[\frac{-(x - \frac{10}{6})^2}{(\frac{1}{6})^2}\right],$$

where x is in micrometers.

This is a Gaussian whose center is located on the lower grid point in the core, i.e., on the grid point closer to the substrate interface. Using the FD-BPM propagate this Gaussian 100 μm along the slab. You can choose $n_r = 3.1$ and $\Delta z = 5$ μm. At the end of 100 μm compare the resulting field distribution with the analytical one discretized on the same mesh. In this case you may find it advantageous to calculate the inverse of the matrix **A** once, since the geometry is invariant along the z-direction. You can improve your accuracy by increasing the number of mesh points and the propagation distance, but this will increase the computational effort.

17. Using BPM code (a free MATLAB script is available on Wiley's download site), calculate the optimum length of a 2-D 1×2 MMI splitter, whose background index is 3.1752 and the core index is 3.2625. This device operates at 1.55 μm, and the input and output waveguides are 1 μm wide.



CHAPTER EIGHT

Photonic Integrated Circuits

8.1 INTRODUCTION

In the last two chapters, we introduced techniques to analyze waveguides and complex waveguide junctions. In this chapter we shall apply these techniques to practical examples of photonic integrated circuits (PICs). The purpose is to illustate the importance of these techniques and generate proficiency in their use, rather than to give a complete summary of PIC technology. In keeping with the active device theme of this text, we shall emphasize PICs that are diode lasers of a relatively complex design, as well as PICs that include diode lasers.

To qualify as a PIC, a device must have at least two distinct functional sections, with an optical waveguide junction there between. In Chapter 3 a few examples have already been introduced. For example, Fig. 3.22 shows a three-section DBR laser with separate gain, modal phase, and Bragg wavelength control electrodes. Each contacts a section with somewhat different optical waveguide properties and with a clear optical function.

We shall begin with a brief review of two-and-three section DBR reflective grating-based multisection laser photonic integrated circuits. The main concepts of tuning and operation will then be expanded to encompass various widely tunable laser implementations. In digital communication systems, transmitter signal quality, power, and chirp are of key interest. As we have discussed in Chapters 2 and 5, directly modulated lasers have severe limitations in terms of bandwidth and output signal qualitty. The natural progression in PIC development will then lead us toward a discussion on externally modulated lasers (EMLs), which decouple the functions of light sourcing and light modulation and allow for their independent optimization

Diode Lasers and Photonic Integrated Circuits, Second Edition.
Larry A. Coldren, Scott W. Corzine, and Milan L. Mašanović.
© 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

in a more complex PIC. This will lead us to introducing new important PIC building blocks, modulators, and semiconductor optical amplifiers.

The second part of this chapter will deal with even more functionally advanced PICs, which generally include both transmitter and receiver functions on the same chip. Some of these devices, like triplexers, have the potential to be deployed in vast quantities, in applications like fiber-to-the-home. Optical transceivers, wavelength converters and optical routing PICs are still more in the research arena, and they can potentially enable size, weight, power, and cost reduction in future optical communication systems.

Finally, we will devote a separate section to PICs for optical coherent communication. Optical coherent communications enable even better utilization of available capacity for information transport on the existing fiber infrastructure, while simultaneously enabling new powerful techniques for transmission impairment control and mitigation. This is an arena of active research, development, and deployment for some of the most complex PICs to date.

8.2 TUNABLE, WIDELY TUNABLE, AND EXTERNALLY MODULATED LASERS

Tunable diode lasers allow for the output wavelength of the laser to be controlled postfabrication, which makes them attractive in a number of applications, spanning optical communications, sensing and medical imaging. These laser structures also form a basis for more complex photonic integrated circuits, such are externally modulated lasers.

Tunability in semiconductor lasers can be achieved by changing the spectral response of the mode filter inside the cavity. In this section, we will introduce and analyze a number of different photonic integrated circuits that are based on tunable lasers. We will start with the simplest two-section DBR laser structure, introduced in Chapter 3, and finish with functionally complex widely tunable, externally modulated lasers and transceivers.

8.2.1 Two- and Three-Section In-plane DBR Lasers

The simplest tunable diode laser is an in-plane, two section DBR laser, illustrated in Fig. 8.1. It consists of a gain section, a semiconductor distributed grating reflector, and one waveguide junction between them, forming a simple photonic integrated circuit.

From discussion in Chapter 3, we recall that the grating provides a frequency-selective reflection, and that the other reflection comes from the output facet. Assuming that the transverse waveguide layer is within the i-region of a *pin* heterojunction diode, its refractive index can be changed by either applying a forward current to inject carriers (free-carrier plasma effect) or a reverse bias to increase the electric field (electro-optic effect). Forward biasing is used more commonly, since it introduces lower absorption losses (due to free carrier absorption)

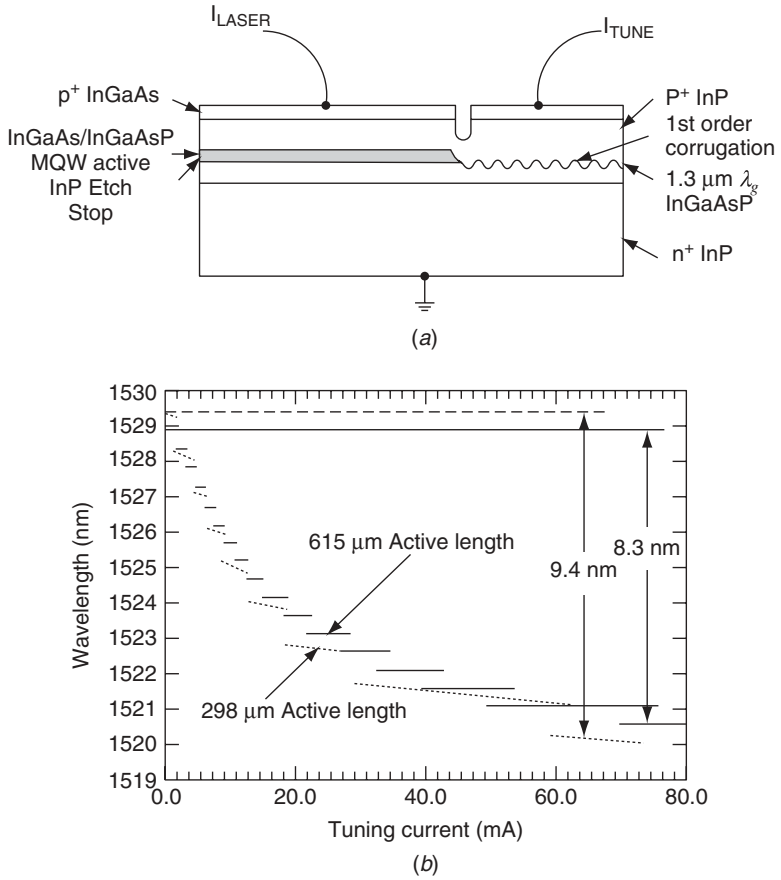


FIGURE 8.1: (a) Schematic of the axial structure of a two-section DBR laser, and (b) output tuning characteristic versus the tuning current to the grating section. After [1]. (*Reprinted, by permission, from Appl. Phys. Lett.*)

as opposed to reverse biasing which triggers the Franz–Keldysh effect, discussed in the upcoming section on electroabsorption modulators. Thus, the center (Bragg) frequency of the grating reflector can be tuned in direct proportion to the index change.

The laser is controlled by biasing each of the sections independently: the gain section for laser gain, and the mirror section for Bragg wavelength control. However, each electrode will also influence at least another aspect of laser operation: for example, shifting the grating index will also change the effective length of the cavity, thus moving the cavity modes, as well as the center frequency of the grating. The mode locations will move only slightly along with the grating tuning according to the fraction of the mode volume that resides in the grating. As the grating current is adjusted, the filter will move across the mode comb, alternately

selecting axial modes. Its center wavelength, λ_B, tunes according to

$$\frac{\Delta\lambda_B}{\lambda_B} = \frac{\Delta\bar{n}_{gr}}{\bar{n}_{gr}}, \tag{8.1}$$

where $\bar{n}_{gr}$ is the effective index in the grating. A given mode will lie at the grating's Bragg frequency only at one point near the middle of its selection during this tuning of the grating filter. The frequency spacing between selected modes at these points is

$$\Delta f = \frac{c}{2[(\bar{n}_a)_g L_a + (\bar{n}_{gr})_g L_{eff}]}, \tag{8.2}$$

where $(\bar{n}_a)_g$ and $(\bar{n}_{gr})_g$ are the group indices in the active and mirror sections. Figure 8.1 illustrates an actual experimental example of such a two-section discretely tunable DBR.

To allow the cavity modes to be shifted independent of the DBR mirror's center frequency, thus achieving continuous tuning, a third, phase section, needs to be incorporated into a DBR laser. Figure 8.2 shows a general three-section DBR laser: one each for gain, mode phase shifting, and shifting of the frequency-selective

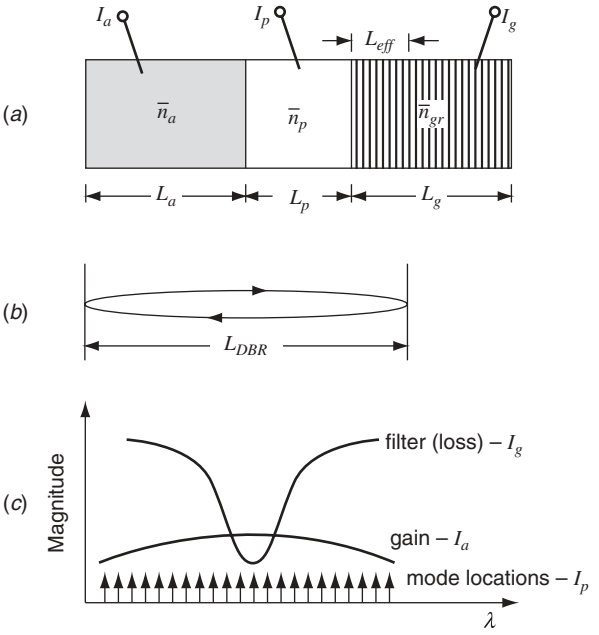


FIGURE 8.2: Three-section DBR. (a) Laser schematic; (b) effective cavity that determines mode spacing and tuning characteristics; (c) sketch of relative spectral characteristics of the cavity filter, gain, and modes along with currents that predominately affect the specified element.

mirror. In this structure, we can move the mode comb along with the grating filter by simultaneously applying a current to the phase-shift section to achieve continuous tunability over some wavelength range. Because this section only encompasses a portion of the mode volume, we expect that the index change here will have to be larger than in the grating. Specifically, we desire the wavelength of the mode, λ_m, to be the same as the Bragg wavelength of the DBR mirror, λ_B, for best modal purity (best MSR). For this ideal continuous wavelength tuning, the relative change in wavelength of the mode should equal the relative change in the grating's Bragg wavelength, or

$$\frac{\Delta\lambda_m}{\lambda_m} = \frac{\Delta\lambda_B}{\lambda_B} = \frac{\Delta\bar{n}_{gr}}{\bar{n}_{gr}}. \tag{8.3}$$

Using Eq. (3.76) we can then solve for the required change in index of the phase shift region to accomplish this alignment,

$$\frac{\Delta\bar{n}_p}{\bar{n}_{gp}} = \frac{\Delta\lambda_m}{\lambda_m}\left(1 + \frac{\bar{n}_{ga}L_a}{\bar{n}_{gp}L_p}\right) - \frac{\Delta\bar{n}_a}{\bar{n}_{ga}}\frac{\bar{n}_{ga}L_a}{\bar{n}_{gp}L_p}. \tag{8.4}$$

Thus, to obtain continuous single-mode tuning in the three-section DBR, the primary tuning signal is applied to adjust the index of the grating, and then a secondary signal is derived to adjust the phase shift region to satisfy Eq. (8.4). In practice, it is not necessary to measure all the independent parameters in Eq. (8.4) to derive the desired tracking signal. This is because the output power experiences a local maximum when this ideal alignment is obtained because the cavity loss is minimized there. Thus, a simple feedback loop can be constructed that senses the output power and adjusts the signal to the phase control electrode to locate a local maximum. Figure 8.3 illustrates one simple control loop design that has been demonstrated. In principle, this feedback loop could be internal to the laser package.

To accomplish the ultimate goal of a tunable laser with only two control inputs—one for power level and another for wavelength—the second independent feedback loop in Fig. 8.3 is added to control the current to the gain section so that a given reference current from a power-monitoring detector is maintained. This latter feedback loop is already commonly applied to most lasers in lightwave systems environments to level the output power. In practice, control curcuits tend to be somewhat more complex than suggested by the figure.

Two- and three-section DBR lasers require a cleaved facet to operate properly. In addition the output power of a properly designed two-and-three section DBR laser will be larger from the cleaved facet. This represents a complication in terms of using these DBR laser embodiments as integrated sources in more complex PICs. One solution to this problem is to apply a high reflection coating on the cleaved facet [1]. Another solution is to utilize a four section DBR laser, where another DBR mirror is added to replace the reflection from the facet. In a four-section DBR laser, the Bragg wavelength tuning is accomplished by tuning both

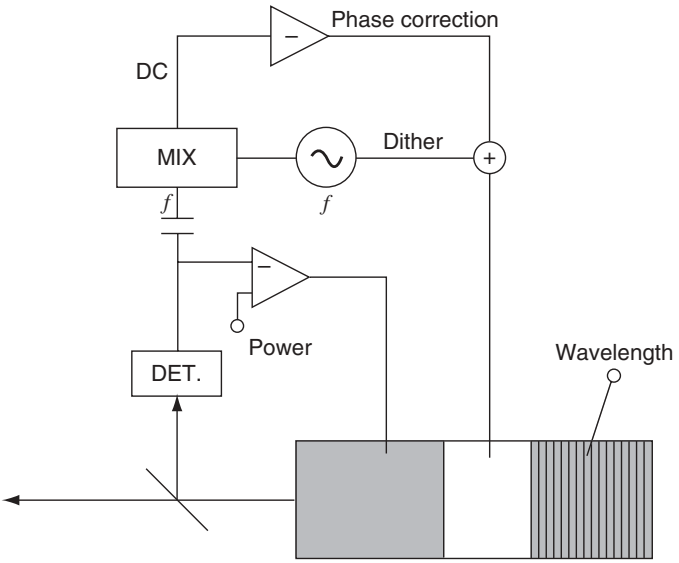


FIGURE 8.3: Circuit that uses first-order feedback to set the power level and a primitive AFC to lock the mode wavelength to the Bragg wavelength of the grating, an alignment where a local maximum in output power exists.

mirrors. Although not very commonly used in standard DBR lasers, this approach represents a basis for widely tunable laser designs to be discussed in the next section.

The maximum possible tuning range of a DBR laser is limited by the maximum amount of index change, Eq. (8.1), to around 8 nm for 1550 nm operation. Therefore, techniques to extend the tuning range by modifying the DBR laser design have been developed, and a number of widely tunable diode lasers have been commercialized. These devices will be discussed in more depth later in this section.

Example 8.1 We are designing a tunable two-section 1.55 μm InGaAs/InP DBR (a gain region with a cleave at one end and passive grating at the other), Fig. 8.1, whose axial mode spacing is 50 GHz. A four-quantum-well offset gain region with characteristics as given by Fig. 4.31 is used. The waveguide of bandgap wavelength 1.25 μm is clad by InP, as illustrated in Fig. 8.2. The product of transverse and lateral confinement factors is found to be 6%. The grating region is formed by removing the quantum wells and etching a fundamental-order triangular sawtooth grating with a peak-to-peak depth of 50 nm on the top side of the remaining 0.3 μm quaternary waveguide. It is desired to tune the output over 12 of the axial modes. We are using a BH waveguide whose width is 3 μm. The active region injection efficiency is 0.8, and the internal loss along the entire device length is 10 cm^{-1}. The grating length needs to be selected to give a power reflection of 70% in the absence of loss.

Problem: (a) Determine the grating and gain region lengths. (b) Determine the threshold current and the differencial efficiency.

Solution: (a) We first need to determine the necessary grating length. The net power reflection gives us the value for κL_g, and κ can be calculated from the waveguide geometry. For the passive grating region we will first calculate the effective index and the effective waveguide width using Appendix 3 material and an interpolated index of $n_1 = 3.37$ from Table 1.1 for the waveguide slab: $V = k_0 d\sqrt{3.37^2 - 3.17^2} = 1.39$, $b = 0.300$, and $\bar{n} = 3.231$. Also, the transverse k-vector and decay constants are $k_x = 3.879\ \mu\text{m}^{-1}$, and $\gamma_x = 2.539\ \mu\text{m}^{-1}$, respectively. Thus, from Eq. (6.13), $d_{eff} = 1.088\ \mu$m.

The grating coupling constant can be found from Eq. (6.55), using $G = 8/\pi^2$ for a sawtooth pattern and a = 25 nm,

$$\kappa = Gk_0 a \frac{n_1^2 - \bar{n}^2}{2\bar{n} d_{eff}} = 8/\pi^2 25\ \text{nm} \frac{3.37^2 - 3.231^2}{2 \cdot 3.231 \cdot 1.088\ \mu\text{m}} = 0.0107\ \mu\text{m}^{-1}.$$

From $r_g^2 = 0.70 \approx \tanh^2 \kappa L_g$, we determine the grating length as $L_g = 113\ \mu$m. Also, we can calculate the grating mirror effective length, $L_{eff} = 0.5(\tanh \kappa L_g)/\kappa = 39.1\ \mu$m.

For 50 GHz mode spacing, we use Eq. (8.2) to find L_{DBR},

$$L_{DBR} = L_a + L_{eff} = c/(2\bar{n}_g \Delta f) = 789\ \mu\text{m}$$

Subtracting the grating penetration depth, L_{eff}, the active length becomes $L_a = L_{DBR} - L_{eff} = 789\ \mu\text{m} - 39.1\ \mu\text{m} = 749.9\ \mu$m.

(b) The threshold modal gain Γg_{th} is

$$\Gamma g_{th} = \alpha_i + \alpha_m = 10\ \text{cm}^{-1} + \frac{1}{0.0789\ \text{cm}} \ln \frac{1}{\sqrt{0.32 \times 0.70}} = 19.48\ \text{cm}^{-1}.$$

Therefore, $g_{th} = (750/710)(1/0.06)19.48\ \text{cm}^{-1} = 342.85\ \text{cm}^{-1}$. From Fig. 4.31, we have that $J_{th} = 150\ \text{A/cm}^2$ per well. For four wells, $J_{th} = 600\ \text{A/cm}^2$. Then, $I_{th} = wL_a J_{th}/\eta_i = 16$ mA. The differential efficiency for the cleaved end # 1 is

$$\eta_{d1} = F_1 \eta_i \frac{\alpha_m}{\Gamma g_{th}} = 0.4F_1.$$

The fraction coupled out end # 1 is,

$$F_1 = \frac{t_1^2}{(1 - r^2) + \frac{r_1}{r_g}(1 - r_g^2)} = 0.77,$$

giving the total differential efficiency of $\eta_{d1} = 30.8\%$.

8.2.2 Widely Tunable Diode Lasers

The fundamental limit in the tuning range of DBR lasers, set by the maximum amount of index shift achievable, needs to be overcome by a change in the laser mirror design. A number of different techniques for implementation of widely tunable diode lasers have been proposed and commercialized. A majority of techniques rely on the ability to generate mirror reflectivity spectra consisting of a number of coarsely spaced reflectivity peaks, a reflectivity comb. From basic Fourier analysis, we know that a periodic function in the time domain will produce a comb function in the frequency domain. This principle forms the basis for sampled grating DBR mirrors, one of the first, and most widely commercially deployed implementations. Thus, in the next section, we will first analyze widely tunable sampled grating DBR lasers, and then we will discuss other device variations based on the same principle of operation. Then we will provide an overview of other key PIC laser implementations.

Sampled Grating and Other Extended Tuning Range DBR Lasers Sampled grating DBR lasers build on the principles of basic four-section DBR designs. The key difference is that the amplitude or the phase of the gratings in the front and rear mirrors are periodically modulated. The simplest form of modulation is blanking, which leaves the mirrors consisting of a number of short grating patches, or bursts, followed by blank areas of waveguide. This periodic spatial modulation creates a corresponding reflection spectrum with periodic maxima in the frequency domain. A lateral cross section of an SGDBR laser, with relevant sampled mirror dimensions is shown in Fig. 8.4. To gain a better understanding of how the spectra are produced in frequency domain, let us examine a special case of weak reflections in a very long grating with very short grating bursts, and a grating period Λ. This is illustrated in Fig. 8.6. For a continuous grating with period Λ, the Fourier space image is a delta function, centered around the grating Bragg frequency $1/\Lambda$. To obtain a sampled grating in real space, we need to multiply the continuous grating with a sampling function, consisting of a number of periodic rectangular sampling windows of width Z_1. The Fourier space image of this sampling function is a sampling comb, with the spacing between the peaks

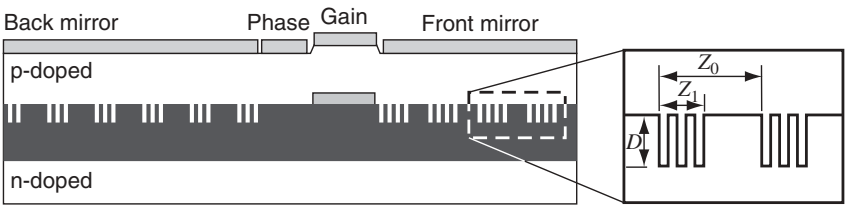


FIGURE 8.4: Sampled grating DBR laser axial cross-section illustration, with relevant dimensions for the sampled mirror defined. Note that the dimensions are not to scale.

defined by the inverse of the distance between adjacent sampling windows Z_0. In real space, this multiplication provides a sampled grating, whose burst length Z_1 is determined by the sampling window length, and whose burst spacing is determined by the distance Z_0. To obtain the frequency response of the sampled grating in Fourier space, we need to convolve the grating and sampling function Fourier responses, a delta funtion and a sampling comb, since this convolution corresponds to the operation of multiplication in real space. Therefore, the resulting spectrum will be a sampling comb whose components are still spaced by $1/Z_0$ in frequency space, but which is centered around the position of the original grating delta function, $1/\Lambda$. The envelope of the reflectivity spectrum is determined by the grating burst length and shape—introducing a phase modulation inside the grating burst allows us to modify the envelope function from a parabola to a hat-top type.

For a finite-length grating, the reflection coefficient for each peak in the comb of reflection maxima has the same form as the unsampled grating (i.e., Eq. (6.48)); however, in this case the total reflection spectrum is composed of a superposition of these reflection components, one for each peak (reflection order) of the comb. Also, the coupling constant, κ_p, and the normalized propagation constant, σ_p, are functions of the duty factor of the sampling function, Z_1/Z_0, and the peak (or order) number, p, in the comb of reflection peaks. For simple periodic blanking of a grating, the coupling constant is given by

$$\kappa_p = \kappa_g \frac{Z_1}{Z_0} \frac{\sin(\pi p Z_1/Z_0)}{(\pi p Z_1/Z_0)} e^{-j\pi p Z_1/Z_0}, \tag{8.5}$$

where κ_g is the coupling constant for the continuous, unsampled grating, Z_1 is the length of the grating burst, and Z_0 is the sampling period. From Eq. (6.48) the reflection coefficient for one of the reflection orders is

$$r_p = \frac{-j\kappa_p^* \tanh[\sigma_p L_g]}{\sigma_p + j\delta_p \tanh[\sigma_p L_g]}, \tag{8.6}$$

where

$$\sigma_p = \sqrt{|\kappa_p|^2 - \delta_p^2},$$
$$\delta_p = \frac{2\pi n}{\lambda} - j\frac{\alpha}{2} - \frac{\pi}{\Lambda} - \frac{\pi p}{Z_0},$$

and α is the net propagation power loss. The net total sampled-grating reflection coefficient is

$$r_g = \sum_p r_p. \tag{8.7}$$

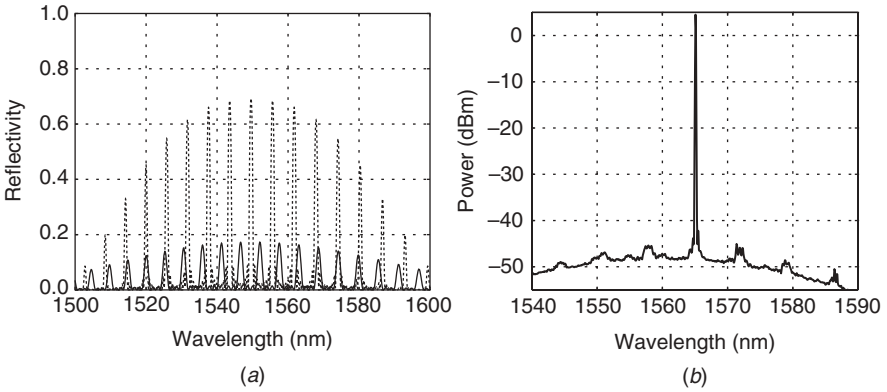


FIGURE 8.5: (a) Example front (solid) and back (dotted) sampled grating DBR mirror spectra calculated using S and T matrix formalism. The grating parameters are: for the front mirror, 5 bursts, 9 cm^{-1} loss, $Z_0 = 68.5$ μm and $Z_1 = 4$ μm; for the back mirror, 10 bursts, 9 cm^{-1} loss, $Z_0 = 61.5$ μm and $Z_1 = 6$ μm. (b) Measured output spectrum from an SGDBR laser. Note that all the widely tunable Vernier effect based lasers discussed in this chapter will have similar spectra with similar performance.

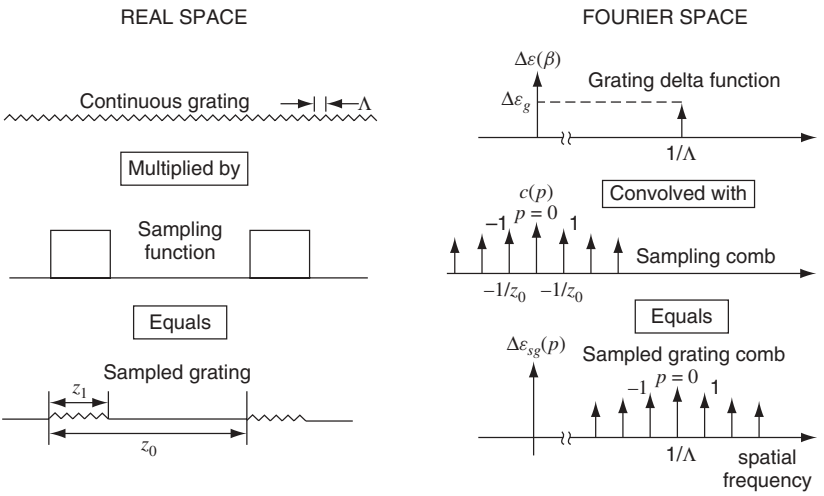


FIGURE 8.6: Sampling of a continuous grating is accomplished by multiplying it by a sampling function. In the Fourier domain, the delta function spectrum of the continuous grating is replicated into a comb.

In Eq. (8.7), we should only use the largest r_p at any wavelength in the summation to be consistent with the assumptions of the coupled-mode formalism, which only considers one Fourier order at a time.

More accurate reflectivity spectra can be obtained using S and T matrix models for sampled grating mirrors. An example of calculated reflection spectra for two sampled grating mirrors that could be used as a low reflectivity front mirror and

a high reflectivity back mirror is shown in Fig. 8.5(a). The mirror parameters are given in the caption of the figure.

Example 8.2 A 1550 nm sampled-grating DBR mirror has been implemented using a passive waveguide consisting of a 1.3Q μm bandgap InGaAsP material, 350 nm thick. The passive waveguide modal loss is 5 cm^{-1}. The injection efficiency is 0.7 for both active and tuning sections. The grating burst length in the mirror is $Z_1 = 5$ μm, coupling coefficient $\kappa = 300$ cm^{-1}, and the mirror has $N_B = 10$ bursts with a sampling period of $Z_0 = 50$ μm.

Problem: (a) Determine the spatially averaged coupling constant for this mirror. (b) Determine the effective length of this mirror. (c) Determine the maximum reflectance of this mirror. (d) Determine the full-width at half maximum (FWHM) of the reflectance peaks for this mirror. (e) Determine the wavelength spacing between peaks in this mirror. (f) What is the overall FWHM envelope bandwidth of this mirror?

Solution: (a) Since the mirror is sampled, the reflections along the mirror are distributed, yielding a burst reflection every $Z_0 - Z_1$. Therefore, we can calculate a spatially averaged coupling constant $\bar{\kappa}$,

$$\bar{\kappa} = \frac{N_{burst} \cdot L_{burst} \cdot \kappa}{L_{mirror}},$$

where $L_{mirror} = (N_{burst} - 1) \cdot Z_0 + L_{burst} = 455$ μm.

Applying this to our mirror design, we have that

$$\bar{\kappa}_B = \frac{N_B \cdot Z_1 \cdot \kappa}{(N_B - 1) \cdot Z_0 + Z_1} = \frac{10 \cdot 5 \cdot 300}{9 \cdot 50 + 5} \text{ cm}^{-1} = 32.97 \text{ cm}^{-1}.$$

(b) The effective back mirror length can be obtained from Eq. (3.63),

$$L_{eff} = \frac{1}{2\bar{\kappa}_B} \tanh \bar{\kappa}_B L_{mirror} = \frac{1}{2\bar{\kappa}_B} \tanh \bar{\kappa}_B \cdot ((N_B - 1) \cdot Z_0 + Z_1)$$
$$= \frac{1}{2 \cdot 32.97 \text{ cm}^{-1}} \tanh(32.97 \text{ cm}^{-1} \cdot 455 \text{ μm}) = 137 \text{ μm}.$$

(c) First, we can determine the lossless value of the mirror reflectance (power reflectivity), using Eq. (3.55), and realizing that the total reflection from a lossless sampled mirror will be equal to the reflection of a continuous grating whose length is equal to the sum of the burst lengths. Thus, we have that

$$R'_B = |r'_B|^2 = \tanh^2 \kappa N_B Z_1 = \tanh^2(300 \cdot 10 \cdot 5 \cdot 10^{-4}) = 0.819.$$

To incorporate the losses into the mirror reflectance, using the material from Chapter 3, we use the effective length of the sampled mirror, and Eq. (3.61),

to get

$$R_B = R'_B \cdot e^{-2\alpha_i L_{eff}} = 0.819 \cdot e^{-2\cdot 5\cdot 0.0137} = 0.7144.$$

(d) Using the Eq. (3.57) for full width at half maximum of a DBR, we have that

$$\Delta\lambda_{FWHM} = \frac{\lambda^2}{2 \cdot \bar{n}_g L_{eff}} \frac{3.76}{\pi} = \frac{1.55^2\ \mu\text{m}^2}{2 \cdot 3.8 \cdot 2 \cdot 137\ \mu\text{m}} \frac{3.76}{\pi} = 2.761\ \text{nm}$$

(e) The wavelength spacing between the peaks will be determined by the sampling period Z_0, as the resonant cavity is formed between each set of adjacent bursts. Using the expression for mode spacing, Eq. (2.32), we have that

$$\Delta\lambda_{peak} = \frac{\lambda^2}{2 \cdot \bar{n}_g Z_0} = \frac{1.55^2\ \mu\text{m}^2}{2 \cdot 3.8 \cdot 50\ \mu\text{m}} = 6.32\ \text{nm}$$

(f) The bandwidth of the envelope is determined by the length of the burst, and we calculate it using the expression for mode spacing again, Eq. (2.32),

$$\Delta\lambda_{env} = \frac{\lambda^2}{2 \cdot \bar{n}_g Z_1} = \frac{1.55^2\ \mu\text{m}^2}{2 \cdot 3.8 \cdot 5\ \mu\text{m}} = 63.22\ \text{nm}$$

Sampled-grating DBR lasers are four-section devices that utilize two DBR mirrors generally sampled at different periods Z_0 and with different burst lengths Z_1. By sampling the gratings at different periods, reflection maxima with different wavelength periods are created in each mirror. Thus, as shown in Fig. 8.5, if a particular reflection maximum from one mirror is aligned with one in the second mirror, the others will be misaligned, and the product of the two reflectivities, which determines the cavity loss, will only have one large maximum. Based on our side-mode supression ratio analysis from Chapter 3, this ensures that the laser will still be a good single-frequency laser with very good MSR. Once the two mirror peaks are aligned, if both mirrors are tuned together and the mirror peaks are kept in alignment at all times, we will be changing the Bragg frequencies continuously, and we will get ~8 nm of continuous single-mode tunability. Tuning the phase electrode will ensure optimum alignment between the cavity mode and the mirror peaks, optimizing the side mode suppresion ratio and output power.

If the index of one sampled mirror is tuned differently from the other, we will be changing the Bragg wavelegth of that mirror and moving the whole spectral comb, and adjacent maxima of the two mirrors will successively line up. We refer to this differential adjustment to obtain alignment of different reflection maxima as *channel changing*, whereas the above joint adjustment of the two mirrors is *fine tuning*. If we wish to have full wavelength coverage, the channel spacing

between mirror reflection maxima should be less than the fine-tuning range of ~8 nm. Since the periodicities of the maxima may be only slightly different, only a relatively small differential index change is necessary to switch channels. Thus, with a differential effective index change of less than ~0.1%, it is possible to switch the alignment point across many different reflection maxima channels. As can be seen, this sliding-scale action is very similar to the function of a vernier scale. Figure 8.5b gives a typical experimental output spectrum for an SGDBR laser design. Large SMSR is obtained, with spectral features attributed to other mirror burst visible in the scan. Large tuning ranges with very good side mode suppression are a characteristic of these lasers. The output power depends on the tuning range selected and can be independently controlled through integration with a booster semiconductor optical amplifier, discussed later in this Chapter [2]. This device is currently being marketed and deployed [3].

Several different variations of sampled grating DBR laser design are shown in Fig. 8.7. All these designs yield devices with excellent side mode suppresion ratio, and operating spectra similar to the example shown in Fig. 8.5. The devices in part (a) and part (b) of the figure are a three section DBR laser and a sampled grating DBR laser based on amplitude modulated DBR mirrors, respectively.

Figure 8.7 (c) shows another two mirror DBR device, a digital supermode DBR laser [4]. The back mirror on this device can be either an amplitude or a phase-modulated DBR mirror. A phase modulated mirror can be engineered to have a flat mirror reflectivity envelope, compared to a sampled grating DBR mirror, thus keeping the threshold condition more constant over the wide tuning range. The front mirror consists of a continuously chirped DBR grating, over which is placed a series of independent contact pads. The front grating, when not tuned, does not provide enough reflectivity for lasing to occur. With proper biasing of different front mirror electrodes, several short grating sections can be aligned in the frequency domain, producing a total reflectivity which is sufficient to allow lasing and good side mode supression for the laser. Differential tuning between mirrors is accomplished by biasing different sections of the front mirror and enables a large tuning range. Tuning in between mirror peaks is accomplished by tuning both mirrors simultaneously. The phase section again is used to align the cavity mode to the mirrors and thus maximize the SMSR. This device is currently being marketed and deployed [5].

Part (d) of the figure illustrates the design and operation of a superstructure grating DBR laser [6]. This type of device utilizes two phase-modulated DBR mirrors, but its tuning mechanism is identical to that of an SGDBR laser.

8.2.3 Other Extended Tuning Range Diode Laser Implementations

In the remainder of this section, we will examine other, non purely DBR tuning based widely tunable laser implementations. A number of these devices also utilize Vernier effect to increase the tuning range.

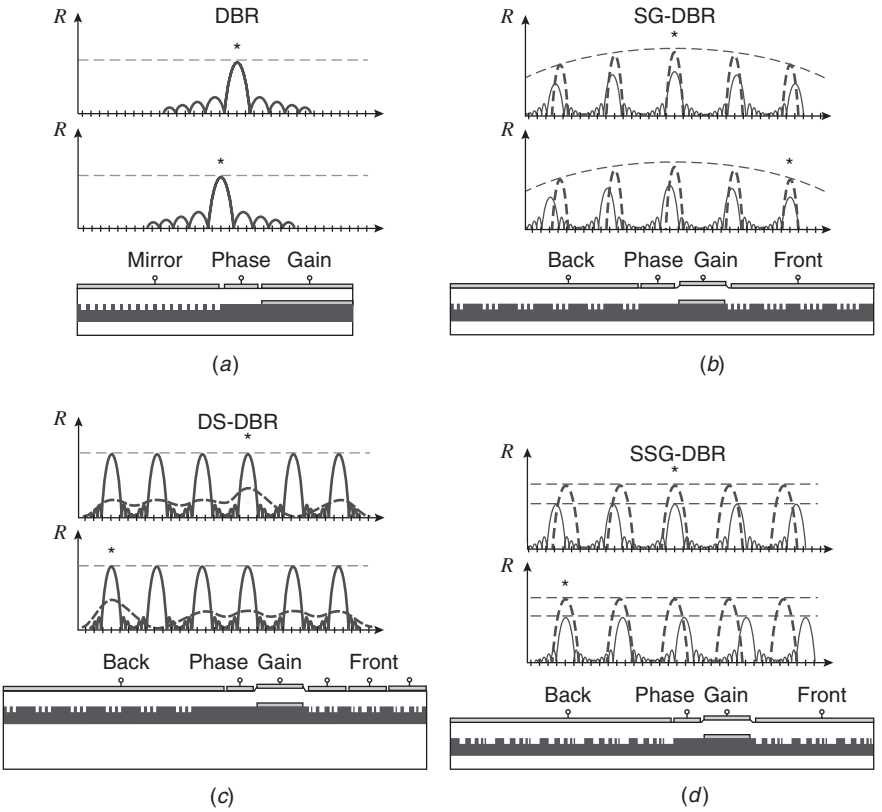


FIGURE 8.7: (a) Three-section DBR laser. Tuning is accomplished by biasing of the DBR mirror to set the wavelength, and the phase section to align the cavity mode to the mirror and thus maximize the SMSR. (b) Sampled grating DBR laser, using two amplitude modulated DBR mirrors. Tuning between supermode peaks in each mirror is accomplished by differentially tuning one mirror versus the other. Tuning in between mirror peaks is accomplished by tuning both mirrors simultaneously. The phase section is used to align the cavity mode to the mirrors and thus maximize the SMSR. (c) Digital supermode DBR laser, using one sampled DBR mirror and one multisection phase-modulated grating mirror. Differential tuning between mirrors is accomplished by biasing different sections of the front mirror. Tuning in between mirror peaks is accomplished by tuning both mirrors simultaneously. The phase section is used to align the cavity mode to the mirrors and thus maximize the SMSR. (d) Superstructure grating DBR laser, using two phase-modulated DBR mirrors. The tuning mechanism is identical to that of an SGDBR laser. The phase section is used to align the cavity mode to the mirrors and thus maximize the SMSR. Note that the dimensions on these drawings have been distorted to enable depiction of necessary details.

Example 8.3 A 1550 nm sampled-grating DBR laser has been implemented using an offset quantum-well gain region, grown on top of a common passive waveguide. The passive waveguide consists of a 1.3 μm bandgap InGaAsP material, 350 nm thick, as described in Example 8.2. The gain region incorporates 7-7 nm thick InGaAs quantum wells, separated by 7 nm thick 1.3 μm barrier material. The active region transverse confinement factor is 4.3%, and the waveguide transverse confinement factor is 60.33%. The material gain within each well can be approximated by $g = 650 \ln(J/60)$ cm^{-1}, where J is the current density per well given in A/cm^2. The injection efficiency is 0.7 for both active and tuning sections. The active section average modal loss is 15 cm^{-1}, and the passive section modal loss is 5 cm^{-1}. We assume that the tuning sections are tuned by current injection, that the recombination is radiative with $B = 0.5 \cdot 10^{-10}$ cm^3/s in the Q-material, and that the mode effective index and material loss change by $\Delta\bar{n}/\Gamma = -0.01 \cdot \Delta N$ and $\Delta\alpha = 5$ cm$^{-1}\Delta N$, respectively, for the increase in carrier density given in units of 10^{18} cm^{-3}. The mode effective index is 3.25. The SGDBR laser has the gain section of $L_g = 500$ μm, the phase tuning section length is $L_p = 75$ μm. The grating burst length in both mirrors is $L_b = 5$ μm, and the coupling coefficient is $\kappa = 300$ cm^{-1}. The ridge width is 3 μm. The back mirror has 10 bursts with a sampling period of $Z_{SB} = 50$ μm (same as that in Example 8.2), while the front mirror has 3 bursts with a sampling period of $Z_{SF} = 40$ μm.

Problem: (a) What is the laser threshold current and differential effiency for no tuning current? (b) What is the axial mode spacing? (c) How much current is required on the back mirror to move from one mirror peak alignment to the next (assuming the center peaks are aligned at some wavelength)?

Solution: The waveguide and the grating coupling constant for this example are the same as those used in the previous example, which will allow us to use some of the results and expressions while solving this problem.

(a) To solve for the threshold modal gain, we use the effective mirror model. For the back mirror, from Example 8.2, we already have that $r'_B = \sqrt{R'_B} = 0.905$ and $L_{effB} = 137$ μm. For the front mirror, since this mirror is short and weak, we can compute its effective length as $L_{effF} = \frac{1}{2}(L_B + 2 \cdot Z_{SF}) = 42.5$ μm. Additionally, the lossless reflectivity for the front mirror is

$$r'_f = \tanh \kappa N_B \cdot L_B = \tanh(300 \cdot 3 \cdot 5 \cdot 10^{-4}) = 0.422.$$

The total cavity length is given by $L_c = L_g + L_p + L_{effF} + L_{effB} = (500 + 75 + 42.5 + 137)$ μm $= 754.5$ μm. The average internal loss is given as

$$\langle \alpha_i \rangle = \frac{\alpha_{ip} \cdot (L_p + L_{effF} + L_{effB}) + \alpha_{ia} \cdot L_g}{L_c}$$
$$= \frac{5 \cdot 254.5 + 15 \cdot 500}{754.5} \text{ cm}^{-1} = 11.627 \text{ cm}^{-1},$$

and the mirror loss is

$$\alpha_m = \frac{1}{L_c} \ln \frac{1}{r'_f \cdot r'_B} = \frac{1}{754.5\ \mu\text{m}} \ln \frac{1}{0.422 \cdot 0.904} = 12.77\ \text{cm}^{-1}.$$

From here, we can compute the differential efficiency,

$$\eta_d = \eta_i \frac{\alpha_m}{\alpha_m + \alpha_i} = 0.7 \frac{12.77}{24.40} = 0.37.$$

This differential efficiency is from both ends, but most of the light will be coming out of the front mirror, since its reflectance is much lower.

The threshold current can be calculated from the modal gain of the laser. First, we need to take into account that the total confinement factor is $\Gamma = \frac{L_A}{L_C} \cdot \Gamma_{xy} = \frac{500}{754.5} \cdot 0.043 = 0.0285$. Now, we can calculate the threshold current,

$$\begin{aligned} I_{th} &= N_w \cdot w \cdot L_g \frac{1}{\eta_i} 60\ \text{A/cm}^2\, e^{\frac{\alpha_i + \alpha_m}{\Gamma \cdot 650\ \text{cm}^{-1}}} \\ &= 7 \cdot 3 \cdot 500 \cdot 10^{-8}\ \mu\text{m}^2 \frac{1}{0.7} \cdot 60\ \text{A/cm}^2 \cdot e^{\frac{24.40}{0.0285 \cdot 650}} = 33.60\ \text{mA}. \end{aligned}$$

(b) The axial mode spacing, per effective mirror model, is

$$\Delta\lambda_{mode} = \frac{\lambda^2}{2 \cdot \bar{n}_g \cdot L_c} = \frac{1.55^2\ \mu\text{m}}{2 \cdot 3.8 \cdot 754.5} = 0.4\ \text{nm}.$$

(c) In Example 8.2, we have calculated the wavelength spacing between peaks of the back mirror, $\Delta\lambda_{peakB} = 6.32$ nm. By analogy, for the front mirror, the peak spacing will be given by $\Delta\lambda_{peakF} = \frac{\lambda^2}{2 \cdot \bar{n}_g Z_{SF}} = \frac{1.55^2\ \mu\text{m}^2}{2 \cdot 3.8 \cdot 40\ \mu\text{m}} = 7.90$ nm. Starting from the center peaks that are aligned, in order to align the next set of peaks, we need to tune the back mirror by $\Delta\lambda = \Delta\lambda_{peakF} - \Delta\lambda_{peakB} = 1.58$ nm. From Eq. (3.75),

$$\Delta\bar{n}_{bgd} = \bar{n}_{gd} \frac{\Delta\lambda}{\lambda} = 3.25 \cdot \frac{1.58}{1550} = 0.00331.$$

This corresponds to a carrier concentration change of $\Delta N = \frac{\Delta\bar{n}_{gd}}{\Gamma_{wgd} \cdot 0.01} = 5.52 \cdot 10^{17}\ \text{cm}^{-3}$. Finally, we can calculate the tuning current required to produce this carrier concentration,

$$\begin{aligned} I_{bm} &= \frac{q \cdot V \cdot B \cdot (\Delta N)^2}{\eta_i} \\ &= \frac{1.6 \cdot 10^{-19} \cdot 0.007 \cdot 500 \cdot 7 \cdot 3 \cdot 10^{-12} \cdot 0.5 \cdot 10^{-10} \cdot (6.42 \cdot 10^{17})^2}{0.7}\ \text{mA} \\ &= 2.16\ \text{mA}. \end{aligned}$$

Y-branch SGDBR Laser A Y-branch SGDBR laser utilizes a Mach-Zehnder interferometer based cavity, enhanced with two sampled grating DBR mirrors to yield Vernier response required for wide tunability. A schematic of a Y-branch SGDBR laser [8] is shown in Fig. 8.8. A gain and a phase section are followed by a 1×2 MMI splitter. Each output waveguide of the splitter contains a sampled grating DBR mirror, with a unique sampling period. The reflections from the two gratings are added together (as opposed to multiplied in an SGDBR laser). Still, the tuning principle remains the same: large wavelength hops can be achieved by differentially tuning the two SGDBR mirrors, and the continuous tuning in between can be accomplished by concurrent tuning of both mirrors. This is illustrated in Fig. 8.8. The two phase controls are required since there is a need to control both the relative phases of the two reflections, and the absolute phase of combined reflections to satisfy the phase condition at a given wavelength. The additive effect of the reflections leads to better supression of the first side modes, compared to the multiplicative effect used in other modulated DBR lasers. However, none of these lasers are limited in performance by the low SMSR. This device is currently being marketed and deployed [9].

Double Ring Resonator (DRR) Widely Tunable Lasers A single ring resonator structure was introduced as an example in Chapter 6. Its periodic output transfer function is suitable for use in the filter sections in tunable lasers. By cascading two ring resonators of different dimensions, we will obtain two different free spectral ranges in the responses, which can be exploited using the Vernier effect again, to achieve wide tunability in a laser. One such double ring resonator (DRR) laser is shown in Fig. 8.9. It consists of a gain and a phase region coupled to a set of two ring resonators, which are terminated by the back facet [10]. The rings are coupled to the bus waveguides through deeply etched MMI couplers. Only when two filters are aligned, light is reflected back from the back facet, thus the laser can be tuned by controlling the resonance frequencies of the ring resonator filters. In a stand-alone DRR tunable laser, the output is coupled from the facet following the gain region. To enable integration with other components, while optimizing the thermal frequency drift, a version of the DRR tunable laser with a front etched gap mirror has been implemented, which is used to define the laser cavity between the cleaved back facet and the front etched mirror [10].

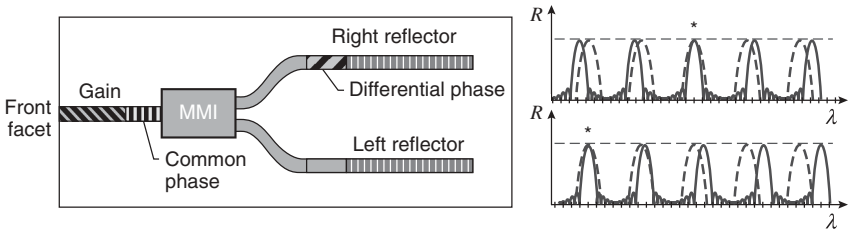


FIGURE 8.8: (left) Schematic layout of the modulated grating Y-branch laser [8] (right) Tuning principle of the Y-branch laser.

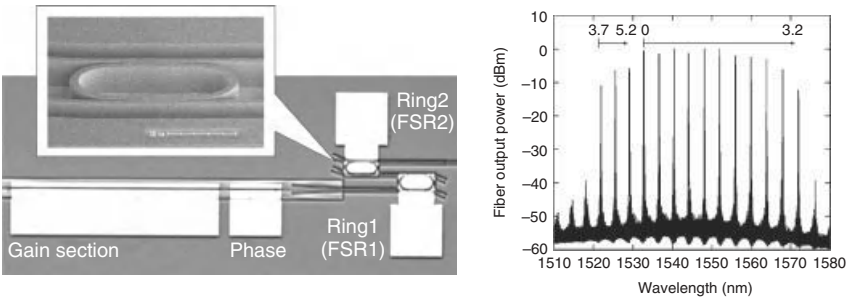


FIGURE 8.9: (left) Photograph of a fabricated DRR tunable laser, consisting of a gain section, phase section, and two microring resonators with different free spectral range (right) Superimposed tuning spectra. (© IEEE 2009, [10].)

To understand how the Vernier effect works in this case, we will expand the results from the ring resonator Example 6.6 in Chapter 6, to calculate the S_{21} parameter of the DRR.

Example 8.4

Problem: For a ring resonator-based mirror structure consisting of two rings whose lengths are L_{r1} and L_{r2}, calculate the overall reflection coefficient S_{11}, and the effective mirror length L_{eff}. The length of the input waveguide is L_1, the length of the second interconnecting waveguide is L_2 and the length of the waveguide connecting the second output to the facet is L_3. The facet reflection coefficient is r_b. The schematic of this double ring resonator configuration is shown in Fig. 8.10.

Solution: In this double ring resonator configuration, at resonance, the normalized field couples into the first ring resonator and then couples out of it, coupling in and out of the second ring resonator, and reflecting from the facet, back through the whole structure. The field at wavelengths that do not couple into both of the resonators are lost. Using the material from Chapter 3 and Chapter 6, we can obtain the total response of this structure through a combination of Mason's rule and scattering matrix theory. First, we note that the structure at hand consists of two ring resonators in series, terminated by a reflection. The total S_{21} parameter for the two ring structures in series, calculated between the input into the waveguide leading into the ring 1 and the output from the ring 2, will be given as a product of the individual S_{21} parameters. We also note that the rings are reciprocal structures.

Furthermore, we can easily calculate the S_{11} parameter for the whole dual ring and reflector structure by noting that from the definition of the scattering parameter S_{11}, $S_{11} = \frac{b_1}{a_1}$, and that $a_1 = b_2 \cdot r_b$ at the output of ring 2. Therefore,

$$S_{11} = S_{21-1} \cdot S_{21-2} \cdot r_b \cdot S_{21-2} \cdot S_{21-1} = r_b \cdot (S_{21-1} \cdot S_{21-2})^2,$$

where S_{21-2} includes the propagation of the field from the ring output to the reflection point.

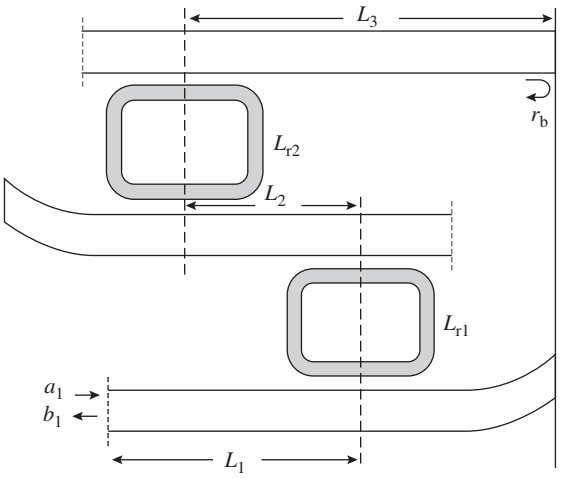


FIGURE 8.10: A double ring resonator structure used in the Example 8.4.

Converting the ring structure into a signal flow diagram with feedback (which we again leave as an excercise to the reader), and using Mason's rule, we have that for the ring 1,

$$S_{21-1} = \frac{(jc)^2 \cdot e^{-j\tilde{\beta}L_1} \cdot e^{-j\tilde{\beta}L_{r1}/2}}{1-(1-c^2)\cdot e^{-j\tilde{\beta}L_{r1}}}.$$

Therefore, based on our analysis, the S_{11} parameter for the total system is

$$S_{11} = r_b \cdot \left(\frac{(jc)^2 \cdot e^{-j\tilde{\beta}L_1} \cdot e^{-j\tilde{\beta}L_{r1}/2}}{1-(1-c^2)\cdot e^{-j\tilde{\beta}L_{r1}}}\right)^2 \cdot \left(\frac{(jc)^2 \cdot e^{-j\tilde{\beta}(L_2+L_3)} \cdot e^{-j\tilde{\beta}L_{r2}/2}}{1-(1-c^2)\cdot e^{-j\tilde{\beta}L_{r2}}}\right)^2.$$

To determine the effective mirror length for this structure, L_{eff}, we start with the definition of effective length, $L_{eff} = -\frac{1}{2}\frac{\partial\phi}{\partial\tilde{\beta}}$. We need to determine the total phase of the S_{11} parameter. We are going to assume that the waveguides are lossless, which will make the propagation constant $\tilde{\beta} = \beta$ real.

At the resonance, which is where the two rings need to operate when producing a reflection required for the laser, $\beta L_{r1} = \beta L_{r2} = 2\pi$. For small deviations in the wavelength (β), the expression for the phase of S_{11} can be linearized.

The tangent of the angle of the first factor within the paranthesis, in the denominator of the S_{11} is given as

$$\tan\phi_1 = \frac{(1-c^2)\sin(\beta L_{r1})}{1-(1-c^2)\cos(\beta L_{r1})}.$$

In the small angle approximation, we have that $\tan\phi_1 \approx \phi_1$, and that $\sin(\beta L_{r1}) \approx \beta L_{r1}$ and $\cos(\beta L_{r1}) \approx 1$. Therefore,

$$\phi_1 = \frac{1-c^2}{c^2}\beta L_{r1}.$$

Similarly,

$$\phi_2 = \frac{1-c^2}{c^2}\beta L_{r2}.$$

Squaring the exponential phase term, and bringing it into the numerator will double these two angles, and change the sign on them. Therefore, the total phase of S_{11} is given by

$$\phi(S_{11}) \approx -2\beta\left(L_1 + L_2 + L_3 + \frac{L_{r1}}{2} + \frac{L_{r2}}{2} + L_{r1}\cdot\frac{1-c^2}{c^2} + L_{r2}\cdot\frac{1-c^2}{c^2}\right).$$

Finally, the expression for the effective length of this mirror element is

$$L_{eff} = -\frac{1}{2}\frac{\partial\phi}{\partial\beta} = L_1 + L_2 + L_3 + (L_{r1} + L_{r2})\left(\frac{2-c^2}{2c^2}\right).$$

Grating-Coupled Sampled Reflector (GCSR) Laser As discussed in Chapter 6, directional couplers can become wavelength filters if the two coupled waveguides are dissimilar. Also, the percentage of tuning can be larger than the fractional index change, so a laser with a tuning range, $\Delta\lambda/\lambda > \Delta\bar{n}/\bar{n}_g$, can be constructed. One difficulty with these devices is that the bandwidth of the filter also tends to become large as the tuning range is increased. Thus, it can be difficult to construct lasers with both large tuning range and adequate side mode suppression ratio.

Grating-Coupled sampled reflector (GCSR) based laser is a historically important four-section widely tunable laser device, shown in Fig. 8.11. While one of the laser cavity filters is again implemented using a sampled grating mirror, the other filtering element consists of a grating assisted coupler. The structure contains two waveguide layers, and the role of the grating coupler is to, in a wavelength selective fashion, couple the mode between the bottom layer, where the gain region is

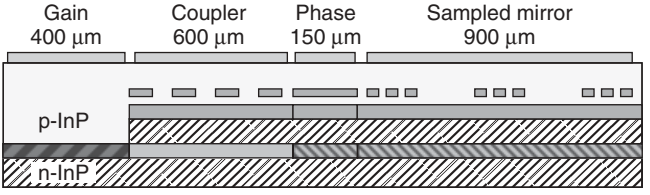


FIGURE 8.11: A grating-coupled sampled reflector laser schematic.

located, to the top layer, where the phase electrode and the sampled grating mirror are implemented. The lasing wavelength is controlled by current injection into the coupler region and the SGDBR mirror. The phase electrode is used to achieve continuous fine tuning, in the same fashion as in other three- and four-section DBR tunable lasers. Wide tunability with these devices has been achieved, as well as performance required for commercial deployment. Using Eq. (6.101), we can relate the mode spacing of such laser, $\Delta\lambda_m$, to the FWHM of the grating filter, $\Delta\lambda_{1/2}$,

$$\frac{\Delta\lambda_m}{\Delta\lambda_{1/2}} = \frac{1}{1.6F}\frac{L_c}{L_T}, \tag{8.8}$$

where L_T is the total effective cavity length experienced by the laser mode. That is, the right side of Eq. (8.8) is the reciprocal of the number of modes under the filter's passband. As the tuning enhancement factor F is increased, so are the number of included axial modes. Fortunately, only a fraction of these modes near the passband center are important in the spectrum of a laser. Comparing the central mode to the next adjacent mode, an approximate analytic expression can be derived [12] for the mode suppression ratio (MSR),

$$\text{MSR} = \frac{2P_{out}}{\hbar\omega\upsilon_g n_{sp}\Gamma g_{th}\ln(1/R)}\left(\frac{L_c}{L_T}\frac{1}{F}\right)^2\frac{L_T}{L_T - L_c}. \tag{8.9}$$

Here, we observe a quadratic dependence on the coupler length and inverse tuning enhancement.

The main drawback of this laser design was in their fabrication process complexity, where multiple growth and regrowth steps were required [11]. The following example illustrates the key concepts related to GCSR laser design and operation.

Example 8.5 A grating coupled sampled reflector laser consists of a 500-μm-long gain section, followed by a a 500-μm-long grating section, a 75-μm-long phase section, and a 500-μm-long sampled grating mirror. Since the scope of this example is a GCSR laser, we are given the penetration length for the SGDBR mirror, $L_{eff} = 151$ μm, as well as the FWHM for the central peak, $\Delta\lambda_{1/2SGDBR} = 1.05$ nm. The front mirror is an uncoated cleaved facet. The GACC bottom waveguide has the same phase and group index values as for the gain section (3.4/4.0), while the top has the same index values as the SGDBR passive sections (3.3/3.8). Assume the GACC is designed with a coupling coefficient appropriate for full coupling over its 500 μm length.

Problem: (a) What is the GACC coupling coefficient for full coupling over its length? (b) What is the GACC tuning enhancement facetor, F? (c) What is the GACC filter FWHM wavelength bandwidth? (d) What is the axial cavity mode spacing $\Delta\lambda_m$, and how many cavity modes are within $\Delta\lambda_{1/2}$ of the back mirror?

(e) What is the maximum index tuning in the coupler region required to achieve a full 50 nm tuning range?

Solution: (a) The coupling constant can be computed from Eq. (6.83),

$$\kappa_c = \left(\frac{2}{\pi} 500\ \mu\text{m}\right)^{-1} = 31.4\ \text{cm}^{-1}.$$

(b) From the definition of the tuning enhancement factor, Eq. (6.44),

$$F = \frac{\bar{n}_{g_1}}{\bar{n}_{g_1} - \bar{n}_{g_2}} = \frac{3.8}{3.8 - 4.0} = -19$$

(c) The FWHM filter bandwidth has been defined in Eq. (6.101),

$$\Delta\lambda_{FWHM} = 0.8 \cdot \frac{F}{\bar{n}_{g1} \cdot L_c} \lambda^2 = 0.8 \cdot \frac{19}{3.8 \cdot 500\ \mu\text{m}} (1.55\ \mu\text{m})^2 = 19.2\ \text{nm}$$

(d) The cavity consists of the full lengths of the gain region, coupler, phase section, and the penetration length of the SGDBR mirror. Since the top and the bottom waveguides have different group indices, we need to take into account that the lasing mode propages through the phase section, SGDBR mirror and one half of the coupler length in the top waveguide, and the gain section and one half of the coupler length in the bottom waveguide. Thus, we have that $L_T = 500\ \mu\text{m} + 500\ \mu\text{m} + 75\ \mu\text{m} + 151\ \mu\text{m} = 1226\ \mu\text{m}$,

$$\begin{aligned}\Delta\lambda_m &= \frac{\lambda^2}{2(\bar{n}_{g1}(L_g + 1/2 \cdot L_c) + (\bar{n}_{g2}(L_{eff} + 1/2 \cdot L_c + L_p)} \\ &= \frac{1.55^2}{2 \cdot ((4.0 \cdot 750) + (3.8 \cdot 476)} \cdot 10^3\ \text{nm} = 0.250\ \text{nm}\end{aligned}$$

Therefore, a total of $\frac{\Delta\lambda_{1/2SGDBR}}{\Delta\lambda_m} = 4.20$ cavity modes will be within the back mirror peak.

(e) Due to the tuning enhancement produced the grating coupler filter, the required index tuning range will be reduced, and wide tunability will be possible. Given that $\Delta\lambda/\lambda = F \cdot \frac{\Delta\bar{n}}{\bar{n}_{g1}}$ we have that

$$\Delta\bar{n} = \bar{n}_{g1} \cdot \frac{\Delta\lambda}{\lambda} \cdot \frac{1}{F} = 3.8 \cdot \frac{50}{1550} \cdot \frac{1}{19} = 0.00645$$

Arrayed Waveguide Grating-Based Multiwavelength Lasers In wavelength division multiplexed optical communication systems, only discrete wavelengths on a determined grid are used for data transmission. Therefore, a laser that can emit on a number of these discrete wavelengths (simultaneously or one at a time) would be useful. This multifrequency laser can be implemented by using an arrayed

waveguide or curved Echelle grating as the cavity filter element. One way to achieve single mode operation is to introduce a gain block into each of the grating's output waveguides. Two different laser implementations, using this principle and different diffraction gratings are illustrated in Figures 8.12 and 8.13. [13] By biasing each of the individual gain blocks, their emission will be filtered according to the grating's transfer function, and with proper feedback and amount of gain, lasing at channels defined by the demultiplexer design will occur. Each channel is controlled by its own gain block, therefore, simultaneous lasing on multiple wavelengths is possible. In addition limited tuning of the lasing wavelength for an individual channel is possible, through thermal or index tuning of the grating elements. Additional gain blocks can be added to boost the laser output power, as shown in the case of the widely tunable grating cavity laser in Fig. 8.13, and the laser output can be modified to support monolithic integration with other elements [24, 23].

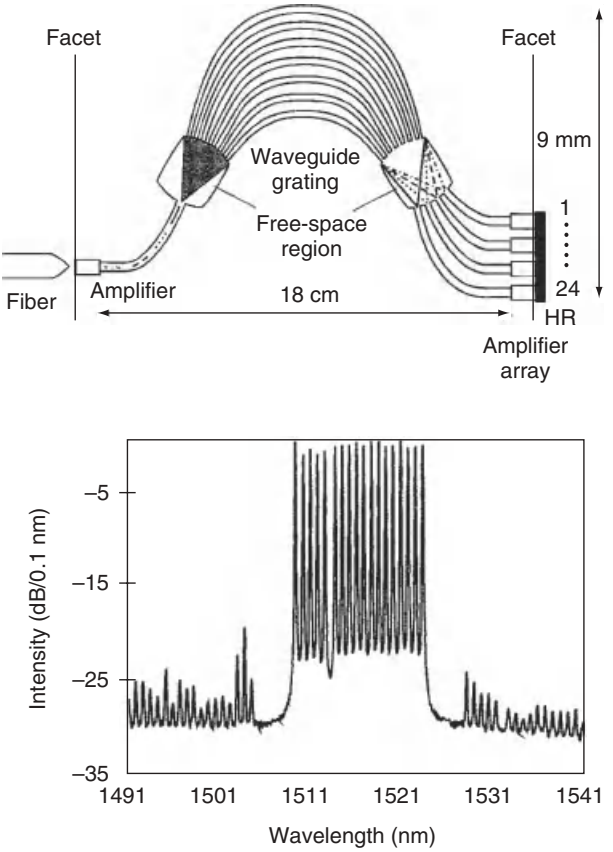


FIGURE 8.12: (left) A schematic of an arrayed waveguide grating-based laser, where the lasing modes are selected by providing gain into a particular channel of the demultiplexer. (right) Output spectrum with all laser channels lasing simultaneously. (© IEEE 1996, [13].)

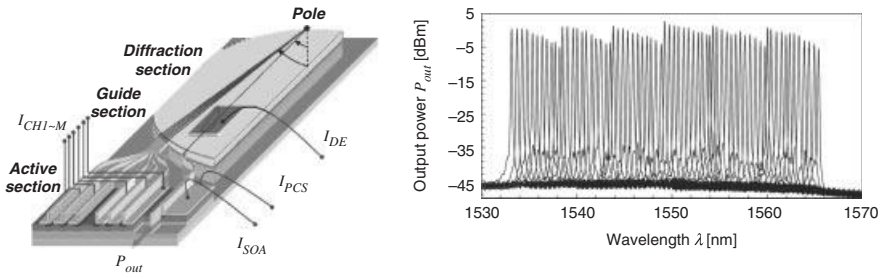


FIGURE 8.13: (left) A schematic of a widely tunable Echelle grating cavity based laser. The lasing modes are selected by providing gain into a particular channel of the demultiplexer. (right) Output spectrum with all laser channels superimposed. (© IEEE 2006, [14].)

8.2.4 Externally Modulated Lasers

In Chapter 5, we discussed frequency response of directly modulated semiconductor lasers. To overcome the distortion, extinction, chirp, and bandwidth limitations of direct modulation, it is desirable to apply the modulation external to the laser cavity, utilizing a separate on-chip modulator component. Figure 8.14 illustrates the maximum transmission distance based on the chirp α parameter of the modulated signal source at 10 Gbps. For directly modulated lasers, we expect the α parameter to be between 3 and 9, allowing for transmission over up to 15 kilometers with 2 dB of power penalty. On the other hand, in an externally modulated laser, with an intensity modulator, the value of the α parameter can be reduced to close to 0, allowing for transmission over around 95 kilometers.

Externally modulated lasers (EMLs) represent one of the more important PICs that have been developed and commercially deployed. External modulator placement allows for separate optimization of the modulator, higher modulation bandwidth, and better modulated signal quality and integrity. Throughout the operation, the laser is operated in the CW regime and can be optimized for the highest output power and SMSR, without regard for its modulation bandwidth. The laser in an EML can be a narowly tunable laser, like a DFB or multimirror DBR laser, or a widely tunable laser, such as the SGDBR or a DS-DBR laser. Any laser structure can be used, as long as the front mirror is lithographically defined. As shown in Fig. 8.15, in an EML we again have at least three functional waveguide sections butted together. The modulator section can be designed to modulate the intensity, the phase of the laser signal, or both, depending on the type of optical link. Phase change induced intensity modulation, using interferometric modulator structures, allow for tailoring of the modulated signal chirp and achieving low negative α parameter values, therefore maximizing the transmission distance (120 kilometers for α between -0.5 and -0.7 to precompensate for the dispersion of standard optical fiber. Initially, we will focus on the these intensity modulated EMLs, whereas phase modulation will be discussed later in a separate section on coherent systems.

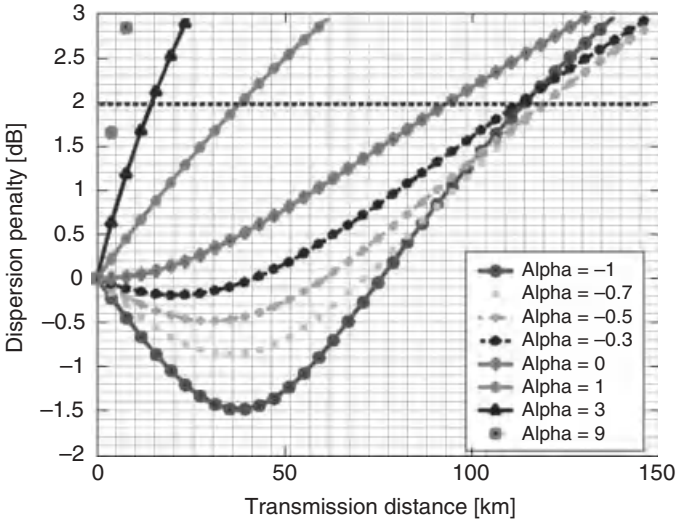


FIGURE 8.14: Maximum transmission distance for 10 Gbps nonreturn to zero modulation as a function of the chirp (α) parameter.

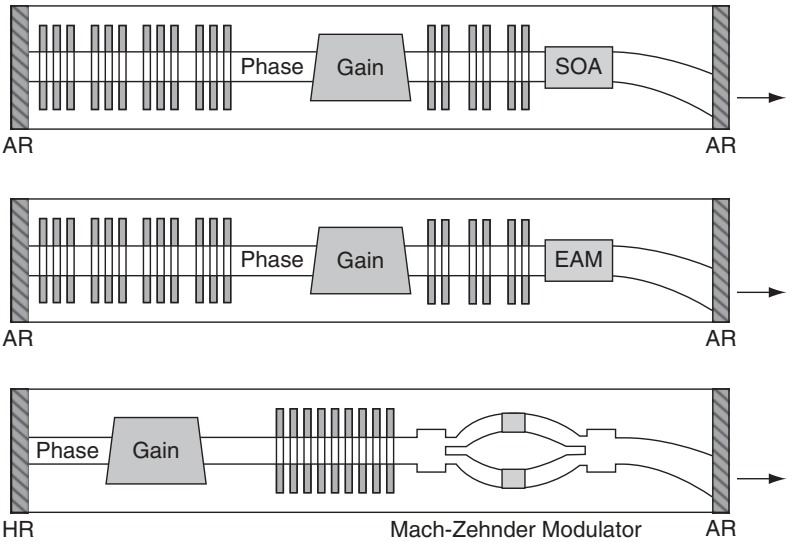


FIGURE 8.15: Schematics of more complex PICs incorporating a tunable DBR laser. The top illustrates integration with an optical amplifier, the center illustrates integration with an electroabsorption modulator, and the bottom illustrates integration with a Mach-Zehnder phase modulator.

Electroabsorption Modulators Electroabsorption modulators are technologically the simplest semiconductor external modulator structure to implement. It consists of a straight section of waveguide, with an active region of slightly larger bandgap energy than the photon energy of the lightwave to be modulated. The active region can consist either of bulk or quantum well material and in its very basic implementation, the waveguide itself can be used as the active region. The application of a reverse bias field lowers the absoption edge and leads to increased absorption of the incoming light. Detailed experimental investigation of the absorption edges and coefficients as function of wavelength and reverse bias are shown in Fig. 8.16, for both bulk and quantum-well-based modulator active regions. The effect of band shift with reverse bias is more pronounced for a quantum well modulator active region.

In a bulk modulator, the application of a reverse bias field lowers the absorption edge via the Franz–Keldysh effect, which, as indicated in Fig. 8.17, describes the electron–hole excitation with below-bandgap photons due to the possibility of lateral carrier tunneling with an applied electric field. It can equivalently be understood in terms of the extension of carrier wavefunctions into the forbidden gap when the field is applied.

The reduction in required photon energy for ionization, $\Delta E/q$, is given by the product of the applied electric field, $\mathscr{E}$, and the effective tunneling distance, Δx. Tunneling is important at distances up to several nanometers, and electric fields of $\sim 2-3\times 10^5$ V/cm are easily applied in PIN structures. Thus, a lowering of the absorption edge by $\mathscr{E}\Delta x \sim 10$ meV (or increasing the absorption wavelength $\sim$10 nm near 1.3 μm) is obtainable. As a result, the attenuation constant for a given wavelength within several nanometers of the absorption edge can be increased from a few cm^{-1} to well over 100 cm^{-1} for a few volts applied in practical PIN DH waveguides. For a desired modulation ratio, P_0/P_{in}, and available change in modal loss, $\langle \Delta\alpha_{FK}\rangle$, the required modulator length is given by

$$L_m = \frac{1}{\langle \Delta\alpha_{FK}\rangle} \ln \frac{P_0}{P_{in}}. \tag{8.10}$$

Modulator lengths $\sim$300 μm are typical for practical lumped Franz–Keldysh modulators. Another important design consideration is that the zero field loss, $\langle\alpha_0\rangle$, must be sufficiently small so that the available output power is not reduced too much. This restricts just how close the operating wavelength can be to the unperturbed absorption edge.

In the quantum-well case, the absorption is primarily due to excitonic effects. Excitonic effects, where holes and electrons due to their spatial overlap bind, were discussed in Chapter 4. This effect of absorption with reverse bias inside quantum wells has been dubbed the quantum-confined Stark effect (QCSE), to emphasize the different physics involved. As can be seen, the primary difference with the quantum-well absorption edge is that it is much sharper. As a result, larger field effects are possible over a narrow wavelength range. However, because the effect is more confined to the proximity of the absorption edge, devices using it require

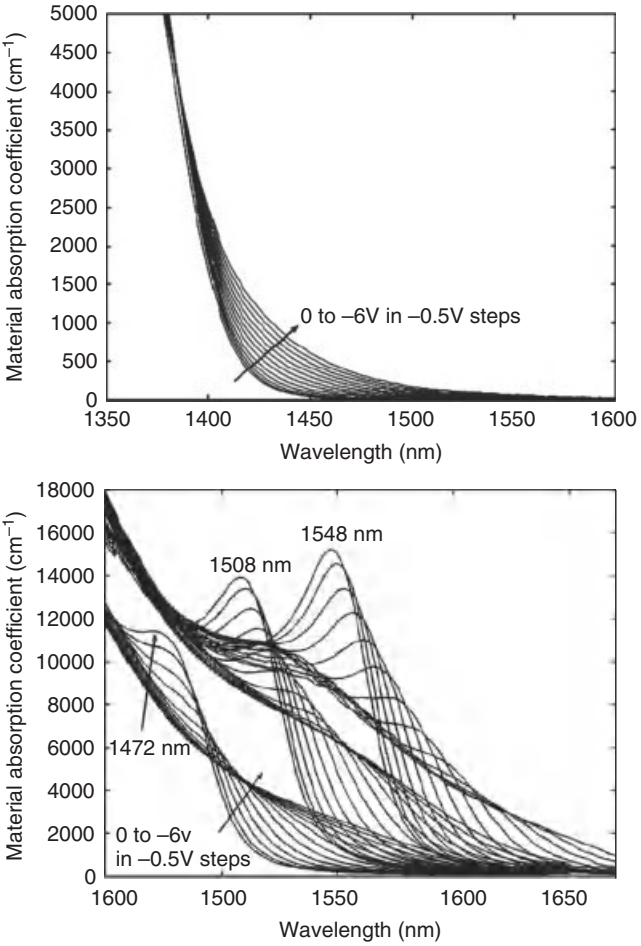


FIGURE 8.16: Absorption edges of (top) bulk and (bottom) quantum-well material as reverse voltages are applied. Applied voltages are given in the figures. For (b) $d_{QW} =$ 8.0 nm [15]. (*Reprinted, by permission, from Elsevier Solid State Electronics © 2007.*)

well-defined wavelengths and are more temperature sensitive as compared to bulk Franz–Keldysh effect devices.

Although both bulk and quantum-well-based externally modulated lasers have been implemented and introduced as commercial products, the QCSE devices tend to be a more popular choice. One such device, designed for uncooled operation, is shown in Fig. 8.18, consisting of a DFB laser integrated with a multiquantum well electroabsorption modulator.

For higher speed of operation, greater than 40 GB/s, traveling wave modulator designs need to be used—where the group velocity of the electrical modulation signal is matched to that of the optical signal.

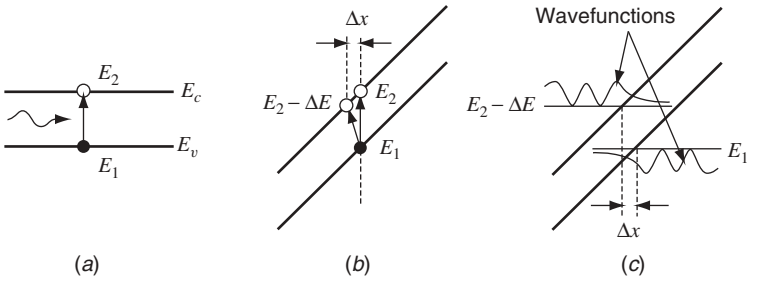


FIGURE 8.17: Schematic of photon absorption to generate an electron–hole pair; (a) with no electric field, and (b) and (c) with a strong applied field so that photons lower in energy by ΔE can be absorbed; (b) illustrates the tunneling model and (c) the equivalent wavefunction overlap model.

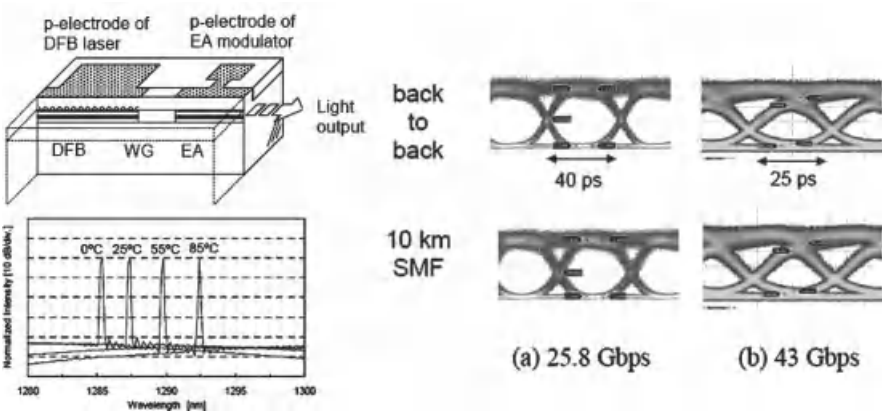


FIGURE 8.18: Schematic of a monolithically integrated modulated laser and electroabsorption modulator. The laser is a 400 μm long DFB, and the 100 μm long multiple quantum well modulator is quantum confined Stark effect based. This EML is designed for uncooled operation, and the spectra at different temperatures, and eye diagrams at 85° are shown. (© IEEE 2009, [17].)

Example 8.6

Problem: An electroabsorption modulator section in an EML needs to be designed so that the device has a peak power of 5 mW and an ON/OFF ratio of 10 : 1, with a 2 V peak-to-peak applied modulation. The characteristics of the modulator region are given by reference [16], its width is 0.3 μm, the confinement factor is $\Gamma = 0.66$, and the residual loss $\alpha_i = 10\ \text{cm}^{-1}$. Determine the length of the modulator.

Solution: To determine the characteristics of the modulator, we need to use Eq. (8.15) with absorption numbers from [16]. First, we need to assume a zero-bias built-in voltage of 0.8 V for a relatively heavily doped PIN structure.

With the modulation swing of 2 V, the internal modulator field varies between 0.8 V/0.3 μm and 2.8 V/0.3 μm. From Ref. [16], for 1.45 μm bandgap material at 1.55 μm, we obtain $\Delta\alpha_{FK} = 140\ \text{cm}^{-1}$. Using Eq. (8.10), we have

$$L_m = \frac{1}{0.66 \cdot 140\ \text{cm}^{-1}} \ln(10) \approx 250\ \mu\text{m}.$$

With a 10 cm^{-1} residual loss, the net ON-loss in the modulator is 1.1 dB. Thus, we must get 6.4 mW out of the laser mirror to get 5 mW to the output facet.

Mach-Zehnder Interferometer Modulators Although an electroabsorption-based modulator represents the simplest design choice for an externally modulated laser, its performance in terms of signal extinction ratio, output chirp, and required bias voltage is not suitable for all optical transmission applications. A better alternative in this case is a Mach-Zehnder interferometer-based modulator structure, which exploits the phase difference between the two branches of the modulator to interfere the signal at the output, and thus produce an amplitude modulated signal at the output. A Mach-Zehnder modulator (MZM) structure is illustrated in Fig. 8.15. The phase in the modulator branch can be controlled through a bias voltage or current, and the output transfer function is a sinusoidal curve in the ideal case.

In semiconductors, there is generally also some absorption change as well with the bias. The device can be biased to destructive interference at low reverse voltage bias, which allows it to overcome the insertion loss limitations of the EAMs while achieving high extinction ratio through interference. Finally, the chirp of this type of device can be tailored by design to be in the low negative values, and this allows for maximizing the transmission distance, as illustrated in Fig. 8.14. To maximize the bandwidth, traveling-wave electrode implementations can be used in MZMs as well. The efficiency and further bandwidth improvements are possible through different drive and biasing schemes, such as: push-pull, in which case one branch of the interferometer is driven with the data signal, and the other simultaneously with a negative of the data signal, and series push-pull, in which case we utilize a single drive voltage, but the electrodes of the two branches are connected in series, effectively reducing the capacitance in half, and thus doubling the maximum possible bandwidth.

Integration of MZMs with a laser source on the same chip requires implementation of light splitters, combiners, and curved waveguides. Great care needs to be taken to minimize the back reflections into the laser cavity. The first widely tunable EML implementation, consisting of a sampled grating DBR laser and a Mach-Zehnder modulator, is shown in Fig. 8.19. This device utilized a series push-pull modulator implementation, and it has been successfully commercialized.

Example 8.7 A Mach-Zehder modulator has been fabricated in InP, in a surface ridge platform, with an undoped 1.42Q InGaAsP waveguide. The branch waveguide cross section is shown in Fig. 8.20: the ridge width is 3 μm, the ridge height is

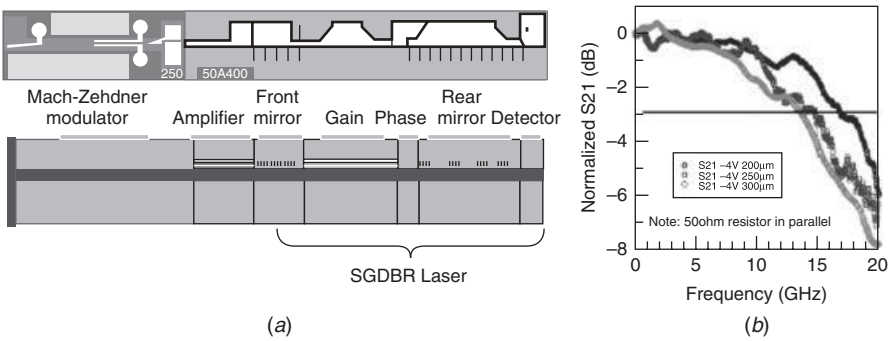


FIGURE 8.19: The first implementation of a widely tunable externally modulated laser utilizing a Mach-Zehnder interferometer. (a) Photonic chip photograph, top view (3.4 × 0.5 mm) and a cross-sectional schematic (b) RF bandwidth as function of MZI electrode length. (© IEEE 2003, [20].)

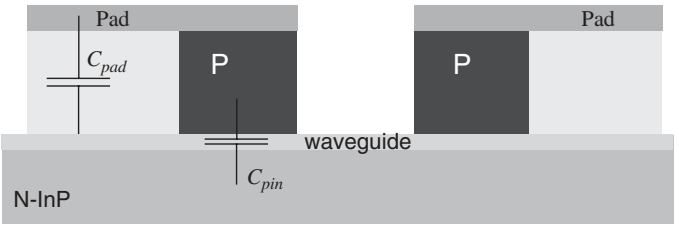


FIGURE 8.20: Cross-section of the Mach-Zehnder modulator from Example 8.7.

2 μm, and the modulator length is 250 μm. The undoped (I) waveguide thickness is $h = 350$ nm, its refractive index 3.45, and it is located on top of an N-doped buffer layer and an N-doped substrate. The top P metal pad is located on top of a 2-μm-high, low dielectric constant $\varepsilon_r = 2.65$ material, whose role is to reduce the pad capacitance. The pad area for bonding is $S_{pad} = 50 \cdot 50\ \mu m^2$. The modulator is terminated with a 50 Ω resistor, which is bonded in parallel with the modulator onto a ceramic carrier.

Problem: What is the estimated modulator bandwidth?

Solution: The maximum bandwidth will be determined by the overall capacitance and resistance of the modulator,

$$f_{3db} = \frac{1}{2\pi RC}.$$

In our case, we will assume that the resistance will be set by the termination resistor. As a very simple model for the capacitance, we will only include two major contibuting components, as illustrated in Fig. 8.20: the reverse-biased PIN junction capacitance and the bonding pad capacitance. Both of these capacitances

can be modeled using the parallel plate capacitor model. In an ideal PIN diode, the depletion region exists almost completely within the intrinsic region. This depletion region is much larger than in a PN diode, and almost constant-size, independent of the reverse bias applied to the diode. For the pad capacitance, the bottom electrode of the capacitor will be formed by the N-doped semiconductor. For a more precise model, we would also need to include the fringing capacitance effects dependent on the electrode/modulator geometric ratios, and the reverse bias dependence in the PIN juction capacitance [18].

The PIN junction capacitance is given by

$$C_{PIN} = \varepsilon_0 \varepsilon_{wgd} \frac{L \cdot w}{h} = 8.85 \cdot 10^{-14} \text{ F/cm} \cdot 11.90 \frac{250 \cdot 3 \ \mu\text{m}^2}{0.35 \ \mu\text{m}} = 0.225 \text{ pF}.$$

The pad capacitance is given by

$$C_{pad} = \varepsilon_0 \varepsilon_r \frac{S_{pad}}{h} = 8.85 \cdot 10^{-14} \cdot \text{F/cm} \cdot 2.65 \frac{2500 \ \mu\text{m}^2}{2.35 \ \mu\text{m}}$$
$$= 0.025 \text{ pF}$$

Finally, the maximum bandwidth of this modulator is

$$f_{3db} = \frac{1}{2\pi RC} = \frac{1}{2\pi \cdot 50\Omega \cdot 0.25 \text{ pF}} = 12.70 \text{ GHz}.$$

This result overestimates the bandwidth somewhat, since we have not taken into account the fringing capacitance from the P electrode.

8.2.5 Semiconductor Optical Amplifiers

Semiconductor optical amplifiers (SOAs) are optical active regions in a semiconductor that are used without any optical feedback, unlike the gain regions in semiconductor lasers. An optical signal launched into the input of an SOA will experience gain through stimulated emission, as explained in Chapter 2. In addition optical noise, caused by the spontaneous emission, will be added to the signal and amplified through stimulated amplification while propagating through the SOA. The lack of optical feedback and carrier clamping will allow for a larger fraction of carriers to recombine through spontaneous emission.

Stand-alone SOA devices exist as commercial products. From the photonic integration perspective, these amplifier blocks have become indispensable in complex PICs, since they enable increased output power, and signal preamplification and the ability to control the output power levels from individual transmitters by altering it external to the cavity.

Oftentimes, for fabrication ease, an SOA block is realized in the same gain material as that in the laser. Thus, the optical gain bandwidth of an SOA should be as large as the laser's gain material. In fact, depending on how the amplifier

section is biased, it may be possible to have an even larger optical bandwidth than the laser. This fact can be appreciated by reviewing the example gain spectra given in Chapter 4.

The gain of a semiconductor amplifier of length L is given by

$$G_0 = e^{\Gamma g(N)L}, \tag{8.11}$$

where Γ is the confinement factor and g is the material gain for the given pumping level.

As the input (or output) power level of the amplifier becomes relatively large, amplifier saturation sets in. When the stimulated emission term becomes comparable to the other carrier recombination terms in the carrier rate equation, the carrier density and, hence, the gain will decrease with increasing optical power. Substituting for the photon density using Eq. (2.38), the carrier rate equation becomes

$$\frac{dN}{dt} = \frac{\eta_i I}{qV} - \frac{N}{\tau} - g\frac{P\Gamma_{xy}}{wdh\nu}. \tag{8.12}$$

At low optical powers in the steady state, $N = N_0 = \eta_i I\tau/(qV)$, and over some bias range we can approximate the gain by $g = a(N - N_{tr})$. Therefore, the steady-state gain can be written as $g_0 = a[\eta_i I_\tau/(qV) - N_{tr}]$. Then, solving for the steady-state carrier density and gain for higher optical powers where the stimulated emission term must be included, we obtain

$$N = \frac{\eta_i I\tau}{qV} - g\frac{P\Gamma_{xy}\tau}{wdh\nu}, \tag{8.13}$$

and

$$g = \frac{g_0}{1 + P/P_s}, \tag{8.14}$$

where

$$P_s = \frac{wdh\nu}{a\Gamma_{xy}\tau}. \tag{8.15}$$

For quantum-well separate-confinement waveguide amplifiers typical values lead to $P_s \sim 1 - 20$ mW.

Now, to get the net amplifier response, we must integrate over the length of the gain region using

$$\frac{dP}{dz} = gP. \tag{8.16}$$

Inserting Eq. (8.14) into Eq. (8.16) and performing the integration over a length L, we obtain an implicit relation for the large-signal gain, $G = P(L)/P(0) = P_o/P_{in}$,

$$G = G_0 \exp\left[-\frac{G-1}{G}\frac{P_o}{P_s}\right], \tag{8.17}$$

where $G_0 = \exp(gL)$ is the unsaturated gain for $P_o \ll P_s$. From this we can derive the output saturation power, P_{o_s}, where the gain has fallen to half of G_0,

$$P_{o_s} = \frac{G_0 \ln 2}{G_0 - 2} P_s. \tag{8.18}$$

Thus, we note that P_{o_s} is slightly smaller than P_s. The saturation power can be modified by changing the optical confinement of the active region—lower confinement leads to higher output saturation power.

Another issue with the semiconductor optical amplifiers is the broadband spontaneous emission level. Although it is easy to keep the spectral density level low, the integrated spontaneous energy can be comparable to the energy at the desired output wavelength if the input signal level is not large. Careful spatial filtering can be used to eliminate out-of-band spontaneous energy, but there is still the inherent spontaneous emission into the mode of interest. It can be shown [25] that the noise figure, $F_A = (\text{SNR})_{in}/(\text{SNR})_{out}$ must be $\geq 2n_{sp}$, the population inversion factor introduced in Chapter 4. In fact, a quantitative measure of this spontaneous emission relative to the gain can be obtained from the details in Chapter 4. The minimum noise figure is increased further by internal loss, α_i, and any facet reflections. Neglecting the facet feedback, we can write the noise figure as [25],

$$F_A = 2n_{sp}\left(\frac{g}{g-\alpha_i}\right) = 2\,\frac{f_2(1-f_1)}{f_2 - f_1}\left(\frac{g}{g-\alpha_i}\right). \tag{8.19}$$

Typically, $F_A \sim 5$ dB in semiconductor amplifiers. In direct detection systems, the amplified spontaneous emission will cause additional noise terms to appear in the receiver, due to signal-spontaneous and spontaneous-spontaneous beating. To minimize those factors, reflection of the SOA's interfaces has to be reduced, cavity length and lossed minimized, and injection level maximized [19]. It is interesting to note that in the case where the SOAs are used to boost the laser output power, the stimulated emission in the SOA overtakes the spontaneous emission and removes the carriers before they can recombine spontaneously, thus significantly reducing the added noise to the laser signal.

For PICs containing optical receivers, polarization independent SOAs are of interest. In that case, the SOA structures need to be tailored carefully, so that the combination of the difference in optical confinement between a TE and a TM mode, and the gain between two polarizations yield low polarization dependence. Straining the active region, as discussed in Chapter 4, is often used as an approach in this situation.

8.2.6 Transmitter Arrays

One of the key performance metrics for EMLs is their maximum data modulation rate. One way to increase the amount of data being transmitted by a single EML chip is to realize a WDM array of EMLs on a single photonic chip. The advantages of this approach, enabled by photonic integration, are in the reduced module cost and overall footprint per transmitted bandwidth.

An example of a 40-channel transmitter PIC is shown in Fig. 8.21. This chip consists of 40 individual EML transmitters, which are all coupled to a single output using an AWG multiplexer. An EML branch consists of a DFB laser with and EAM modulator, as well as a power flattening SOA, and two power monitor photodiodes—one behind the DFB laser and one after the modulator. To achieve multiplexing, each DFB laser is tuned to a separate wavelength. Also shown in the figure are the L-I curves for all 40 lasers, as well as the eye diagrams for all channels, which were used to characterize the device performance. A photonic IC of this level of integration represents the current state of the art; however, it is to be expected that more advanced and densely integrated devices will come in the future.

8.3 ADVANCED PICS

In the last section, we focused our attention on the photonic ICs that generate light. The transmitter array PIC shown in Fig. 8.21 is used to send information on multiple optical channels in the same optical fiber. At the receiving end, an optical receiver of similar complexity is required. This PIC is shown in Fig. 8.22.

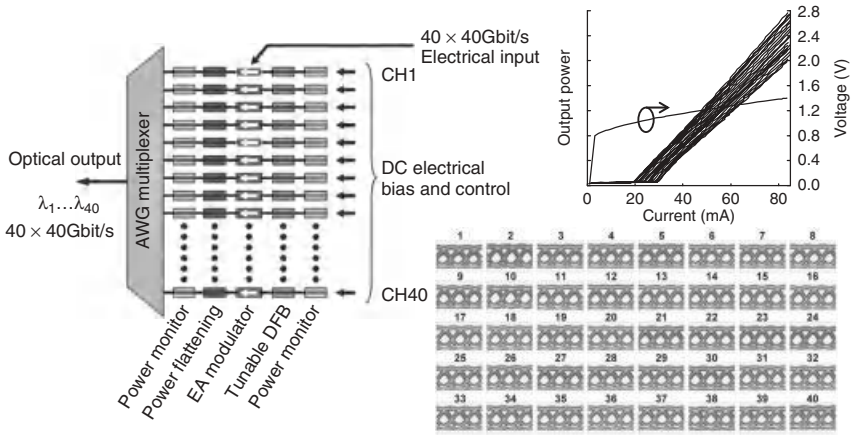


FIGURE 8.21: A 40-channel transmitter PIC, consisting of an array of 40 Gbps DFB-EAM EMLs, which are combined using an AWG multiplexer. The LIV curves and eye diagrams at 40 Gbps for each channel are shown. (© IEE 2006, [21].)

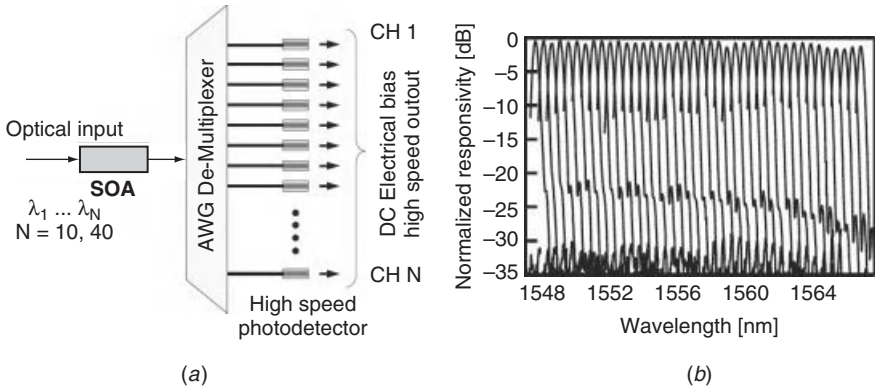


FIGURE 8.22: (a) A 40-channel receiver PIC, consisting of a preamplifier SOA, an arrayed waveguide grating demultiplexer and high-speed photodiodes operating at 40 Gbps. (b) Normalized spectral response for all 40 channels. (*Reprinted by permission from OSA, [22].*)

It consists of an SOA preamplifier followed by an AWG demultiplexer, an array of 40 photodiodes, which are followed by transimpedance amplifiers, and additional electronic circuitry needed to recover the information from the optical domain.

In this section, we wish to move beyond just tunable laser and externally modulated laser PICs, or simple and arrayed optical receivers, to even more functionally complex active photonic ICs that also may incorporate all these and other functions in a single integrated device. We start this section by introducing the basics of photodetectors, one of the key new ingredients, and then we examine examples of transceivers, triplexers and active optical switches.

8.3.1 Waveguide Photodetectors

The role of an optical receiver is conversion of an optical signal back into the electrical domain, where transmitted data recovery is performed. The key photonic component of every optical receiver is a photodiode, which utilized the photoelectric effect to convert incoming photons back into the electrons. Photodiodes have become important building blocks for contemporary photonic integrated circuits, for realization of preamplified receivers and complex receiver arrays, as illustrated in Fig. 8.22, as well as transceivers and triplexers, which will be discussed in the next section.

The photocurrent of photodetector is directly proportional to the incident optical power,

$$I_p = R \cdot P_{in} \tag{8.20}$$

where R is the responsivity of the photodetector. Another key figure of merit for a photodetector is quantum efficiency η, which represent the ratio in number of

electrons generated per number of incident photons, per unit time,

$$\eta = \frac{I_p/q}{P_{in}/h\nu} = R\frac{h\nu}{q}. \tag{8.21}$$

Responsivity increases with the wavelength increase, since more photons are present for the same optical power. In a good photodiode, the goal is to absorb all the photons, such that the quantum efficiency can approach 100%. Depending on the absorption coefficient α of the photodiode material, its thickness or length L will be determined by the minimum acceptable quantum efficiency, since

$$\eta = \frac{P_{in} - P_t}{P_{in}} = \frac{P_{in} - P_{in} \cdot e^{-\alpha L}}{P_{in}} = 1 - e^{-\alpha L}, \tag{8.22}$$

where P_{in} is the input optical power, and P_t is the transmitted optical power at the output of the photodiode's active region. Although stand-alone avalanche photodiodes allow for maximum receiver sensitivities, the vast majority of photodiodes used today in photonic ICs are based on a simpler, reverse biased PIN structure.

The speed of response of a photodiode will be determined by its resistance, capacitance, as well as the carrier transit time, which depends on the thickness of the photodiode active region. An illustration of the evolution of photodiodes from a simple vertical design to several waveguide-based structures is shown in Fig. 8.23. Historically, the photodetectors were designed as top illuminated structures, as shown in Fig. 8.23(a). For higher bit-rate operation, 10 Gbps and beyond, waveguide photodiode structures, (c)–(d) have become necessary, since they allow for overcoming of bandwidth limitations imposed by the responsivity maximization requirements and finite drift times in vertical structures. In addition waveguide photodiodes are a natural match for planar integration, as they utilize a planar waveguide, which is most often adjacent to the absorbing layers.

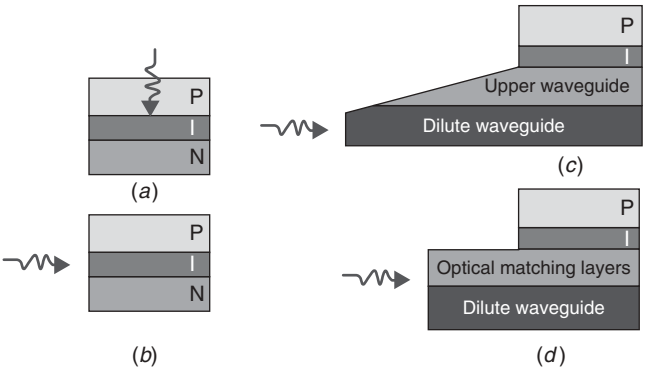


FIGURE 8.23: Illustration of (a) normal incidence, (b) edge coupled waveguide, (c) tapered evanescently coupled waveguide, and (d) multimode input waveguide evanescently coupled photodiodes. (*Reprinted by permission from Elsevier, [28].*)

Simple edge coupled diodes suffer from large coupling losses, which reduce their external quantum efficiency, and thus limit their power handling capabilities to the maximum optical intensity, which does not degrade the input facet. Therefore, evanescently coupled waveguide photodiodes were developed, and they allow for about four times higher diode saturation currents. This coupling technique enables simple integration of optical spot size converters, which significantly improves the external coupling efficiency. The maximum bandwidth of all these structures can be further extended via implementation of traveling wave electrodes, which match the velocity of propagation of the electric and optical wave.

Another fundamental limitation of optical receiver performance in general is the noise introduced into the signal in the process of its conversion from optical into the electrical domain. The amount of noise will directly determine the minimum input power into the receiver that can be detected with a sufficiently low bit error rate.

Thermal noise, result of random motion of electrons at any finite temperature, dominates as a noise source and generally limits the performance in practical direct detection systems based on PIN photodiodes. Shot noise is another noise mechanism important in photodiodes. It is a consequence of photon–electron conversions occuring at random, discrete times. Shot noise becomes important for large photodiode input powers, as in coherent systems, which are the last topic of this chapter, where it dominates. In that case, an order of magnitude lower optical powers is required to achieve the same bit error rate, compared to a thermal noise dominated receiver [26]. This higher receiver sensitivity is one of the key features of coherent receivers and systems in general.

Photodetectors with high bandwidths and high saturation power are of particular interest for shot-noise limited operation and in optical transceiver and receiver applications, as their photocurrent swing is used to electrically drive optical modulators. With high optical powers, in a PIN structure, slowing of the response is observed due to a large space charge created inside the active region, which effectively reduces the reverse bias field, and consequently increases the carrier transit time. A very popular way to realize high power, high bandwidth photodiodes involves uni-traveling carrier (UTC) photodiode epitaxial structure designs [27]. Band diagrams of a UTC-PD and a conventional PIN-PD are schematically shown in Fig. 8.24. The absorption region in a UTC-PD is a narrow-gap, p-type layer, surrounded by a diffusion blocking layer and by a wide bandgap collector. The holes, which move with much lower velocity, are quickly captured by the p-contact, whereas electons are swept accross the collector in a strong drift current. This unique design enables the use of UTC photodiodes in a variety of high-speed applications, at bit rates of several hundred gigahertz.

Example 8.8 Parameters of a waveguide photodetector need to be determined for operation at 1550 nm, 1 Gbps, and with quantum efficiency of $\eta = 90\%$. The overlap between the waveguide and the absorption layer is $\Gamma_{xy} = 5\%$, the absoption coefficient of the active region is $\alpha = 5000\ \text{cm}^{-1}$, and the capacitance of the waveguide under reverse bias can be taken to be approximately $1\ f\text{F}/\mu\text{m}$. The

photodiode is followed by a FET amplifier stage, connected to the photodiode through a load resistor R_L.

Problem: (a) Determine the length of this waveguide photodiode. (b) Determine the value of the load resistor R_L. (c) Determine the voltage at the FET gate for an input power of 1 μW.

Solution: The waveguide detector design involves a trade-off of its quantum efficiency and capacitance, since both increase with length. Receiver sensitivity, which depends on the load resistance, is inversely proportional to the capacitance, since the load resistor must be chosen to provide enough bandwidth to receive the highest frequency components of the data. The required detector length for a desired quantum efficiency, η_D, is given by

$$L_D = \frac{1}{\Gamma_{xy}\alpha} \ln \frac{1}{1-\eta} = \frac{1}{0.05 \cdot 5000} \ln \frac{1}{1-0.9} \text{ cm} = 92.12 \text{ μm},$$

For an RC-limited receiver the bandwidth is equal to the bit rate, B (i.e., a receiver bandwidth twice the highest fundamental frequency component in the data). The required load resistor, R_L, is

$$R_L = \frac{1}{2\pi B C_D}.$$

The capacitance for this length of the waveguide under reverse bias is 92 fF. Thus, for a data rate of 1 Gbps, the load resistor value is,

$$R_L = \frac{1}{2\pi B C_D} = \frac{1}{2\pi (10^9/\text{s})(92 \cdot 10^{-15} \text{ F})} = 1.73 \text{ kΩ}.$$

From the expression for responsivity, the value of the photocurrent is given by

$$I_p = \frac{e\eta}{h\nu} P_{in} = \frac{0.9 \cdot 1.55}{1.240 \text{ μm}} 1 \text{ μA} = 1.125 \text{ μA}.$$

For a FET amplifier front end, the received signal is proportional to the signal voltage on the gate. This voltage is equal to the detector photocurrent, I_p, times this load resistance. For an input optical power of P_i, we have

$$V_G = I_p R_L = 1.95 \text{ mV}.$$

8.3.2 Transceivers/Wavelength Converters and Triplexers

Integration of transmitter and receiver functions on a single chip is the next logical progression in PIC development. It follows the development of related fiber optic subsystems, where transceivers and transponders, consisting of transmitter and receiver photonics and neccessary electronic circuitry, are a key building blocks for

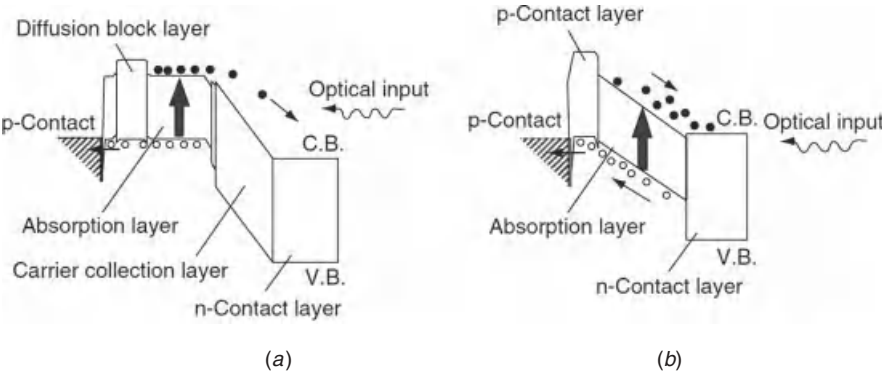


FIGURE 8.24: Band structure of (a) uni-traveling carrier and (b) conventional PIN photodiodes. (*Reprinted by permission from Elsevier, [28].*)

optical networks of all scales. In the fiber-to-the-home market, more cost-effective triplexer devices are used for similar purpose.

Tunability in such devices remains key, due to the need to address the wavelength space offered through wavelength division multiplexing. Increasingly, wavelength routing is being used in optical networks to manage traffic and connections. Although most of the current networks are optical link based, where a particular laser in the network remains tuned to the same wavelength for extended periods of time (minutes to hours), it is envisioned that in the future more of the switching and routing functions in the optical networks will be pushed into the physical layer. This type of solution could reduce the overall switching power/per bit required for traffic processing (due to higher communication line rate that optic provides relative to electronics), reduce the footprint of switching equipment through high-density photonic integration, as well as allow for better network utilization, by more efficiently combining optical packets from different sources in wavelength and time domain. To satisfy these new network functionalities, fast wavelength switching, tuning, and ability to copy optical data packets from one wavelength to another quickly (wavelength conversion) and efficiently become critical. In this context, a term of tunable wavelength converter is used in literature, for a device that allows data to be transferred from an input wavelength to a tunable output wavelength without passing the signal through electronics. Functionally, a tunable wavelength converter is identical to an optical transceiver and represents an optically controlled optical gate.

As an example of a tunable transceiver/wavelength converter PIC, we are showing a photocurrent-driven device, Fig. 8.25. This device consists of a widely tunable SGDBR laser, followed by a booster SOA, monolithically integrated with a traveling wave electroabsorption modulator (TW-EAM). The EAM is driven by on on-chip preamplified optical receiver, consisting of a couple of input SOAs, followed by a high-speed photodiode. The electrode is shared between the photodiode in the receiver and the TW-EAM, and the electrical signal generated by the detected photocurrent is used to modulate the output of the tunable laser and thus generate

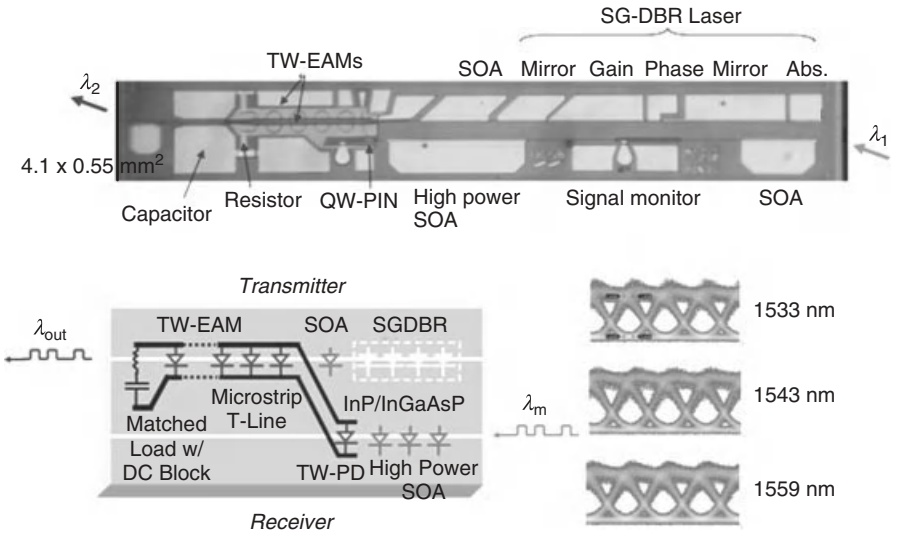


FIGURE 8.25: A widely tunable monolithic transceiver/wavelength converter, consisting of an SGDBR laser based traveling wave EML, monolithically integrated with a preamplified receiver in InP [40].

an output signal on a different wavelength. Electrical termination for the traveling wave modulators is also integrated on chip. This particular device design was demonstrated to operate up to 40 Gbps data rate, and its appeal is in the ability to work over a range of input and output wavelengths and bandwidths, with small footprint, and without a filter to reject the original input signal [40]. One issue is the receiver's polarization sensitivity, which can be addressed through use of some of the approaches discussed earlier in this chapter.

Tunable transceivers and wavelength converters can be combined with optical routing elements, such is an AWGR discussed in Chapter 6, to produce even more complex optical functions on a single chip. An example of high density integration of optical switching and routing functions on a single chip in indium phosphide is shown in Fig. 8.26. This device consists of an array of eight widely tunable wavelength converters, which are connected to an 8×8 arrayed waveguide grating router. Therefore, an optical signal from any of the eight inputs can be optically routed to any of the eight outputs, by properly selecting the operating wavelength of the input laser. The device was named MOTOR, as in Monolithic Tunable Optical Router [41].

The wavelength converter design used in this device is based on a sampled grating DBR laser monolithically integrated with a differential Mach-Zehnder interferometer SOA design [42]. The input signal is split on chip into two separate waveguides, and one of the waveguides is delayed relative to the other. The interferometer is normally off, with no light coming out. The first input pulse disturbs the balance of the interferometer, and opens the output. The amplitude of the second

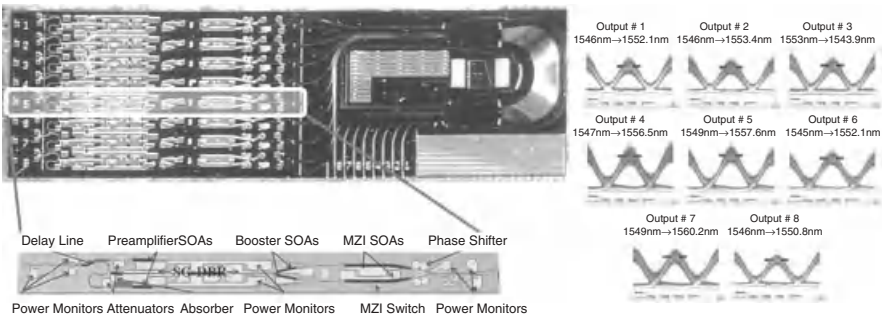


FIGURE 8.26: (a) A photograph of the Monolithic Tunable Optical Router chip (MOTOR), consisting of eight widely tunable, differential Mach-Zehnder interferometer SOA based wavelength converters, integrated with an 8×8 arrayed waveguide grating router. (b) Details of different functional blocks of tunable wavelength converters. (c) Open 40 Gbps eye diagrams for a single device input, shown at all eight output ports [41].

pulse is adjusted in a way that it closes the interferometer. This differential arrangement allows for overcoming of the limit to the carrier lifetime processes [43], and operation at 40 Gbps for this particular device.

Due to complexity of functions required in the MOTOR chip, several different waveguide implementations were used: highly confining waveguides in the delay lines, surface ridge waveguides in the wavelength converter sections, and low-loss buried rib waveguides in the AWGR sections. The experimental results, eye diagrams of the routed signal at eight different outputs, are also shown in the figure.

A related device to an optical transceiver is a triplexer, which is widely used in passive optical networks, including the highest volume fiber-to-the-home markets. A triplexer operates by sending the information on one wavelength (generally, 1310 nm) and receiving the information on a different wavelength (1490 nm). An example of a monolithically integrated triplexer in InP is shown in Fig. 8.27. This device contains a spot size converter, which simplifies and improves the coupling into a fiber, a wavelength sensitive light splitter, which separates the 1310 nm from the 1490 nm, and, on the receiver side, a wavelength selective absorber followed by a PIN receiver, and on the transmitter side, a DFB laser with a back side power monitor. In this particular implementation, the full device is realized using a single epitaxial growth, with a number of sucessive etch steps.

8.4 PICS FOR COHERENT OPTICAL COMMUNICATIONS

Meeting the demand for optical bandwidth through simple wavelength division multiplexing is going to be a continuing challenge in the future. Introduction of new and expanded services, such as HD-TV, IP-TV, on-line video streaming applications, social networking, computerized stock trading and 3D-TV, will continue

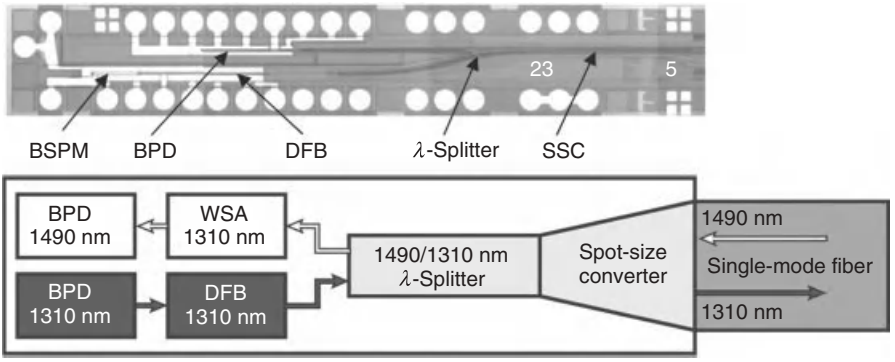


FIGURE 8.27: A monolithic triplexer chip in InP, consisting of a DFB laser transmitter at 1300 nm with a back-side power monitor (BSPM), a broadband photodiode (BPD), preceeded by a wavelength selective absorber (WSA), and a wavelength splitter/combiner, followed by a spot-size converter. (*Part reprinted, by permission from [44].*)

to maintain the extreme 100x bandwidth increase/decade slope [45]. To make better use of available bandwidth in optical fibers, we are now looking to improve the spectral efficiency (SE) of transmitted signals—the net bits/s of data rate per Hz of communication channel bandwidth, or more simply the channel rate/channel spacing ratio. This is being done with advanced (vector) modulation formats and coherent detection, making digital optical communication systems resemble much more the radio frequency systems. In the next sections, we will discuss the basics of optical coherent communication systems and some of the latest developments and anticipated future PICs in this arena.

8.4.1 Coherent Optical Communications Primer

Coherent optical communications rely on digital modulation, a term used in radio, satellite, and terrestrial communications to refer to modulation in which digital states are represented by modification of carrier amplitude, frequency, and phase simultaneously or separately. A common name for this arbitrary carrier phase and magnitude modulation is vector modulation. Different modulation states are represented with components of the electric field vector on the complex plane, using in-phase and quadrature (I-Q) constellation diagrams, illustrated in Fig. 8.28 for three different modulation formats. In optical communication systems, the frequency of the carrier (laser wavelength) is usually fixed; thus we only need to consider the phase and magnitude changes. The unmodulated carrier is then the phase and frequency reference, aligned along the I axis, and the modulated signal is interpreted relative to the carrier. Q represents the quadrature (90° out of phase) component. A discrete point, modulation state on the I-Q diagram, can be represented by vector addition of a specific magnitude of the in-phase carrier with a specific magnitude of the quadrature carrier.

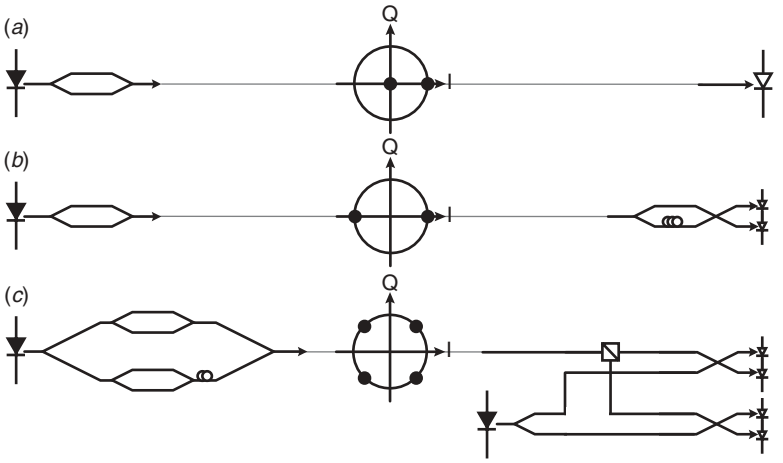


FIGURE 8.28: (a) Amplitude modulation (on-off keying) based noncoherent system with direct detection (b) Differential phase shift keying based coherent system with self-homodyne detection, without the need for a local oscillator (c) Quadrature phase shift keying coherent system with intradyne coherent detection, using a local oscillator matched in frequency and phase to the input signal.

Figure 8.28 illustrates some of the unique, commonly used vector modulation based links. The first link, part (a), utilizes simple binary amplitude modulation, the most widely exploited, noncoherent modulation format in optical communications to date. On the I-Q diagram, the field vector changes its amplitude from 0 to the maximum amplitude, along the I axis, as binary digital signals are translated into a stream of light pulses. The transmitter in this case is a simple amplitude modulator, and the detection of the signal is achieved through direct detection, as shown in the link schematic.

Link (b) represents the next level of sophistication—a simple coherent system, in which the amplitude of the signal remains constant, but the phase of the carrier is differentially changed by π in between bits, to reflect the change in adjacent bit value. To detect this type of signal, one approach is differential detection, where the signal is interfered with a delayed version of itself to produce an amplitude response at the receiver. No local oscillator is required in the receiver in this case. This system is limited to a particular bit rate, as it relies on exactly one bit delay for signal detection. Note that the receiver consists of two photodiodes that are connected in series, forming a balanced receiver, examined later in this section.

Figure 8.28c illustrates a more complex and flexible system, where the carrier phase is modulated to one of four possible values—thus the name of this type of modulation is quadrature phase shift keying (QPSK). The advantage of this approach is in the fact that with the same bit rate as on-off keying, we can transmit twice the amount of information, since with each detected symbol (1 out of 4 possible phase values), we can recover two bits of information, a major benefit of

vector modulation and coherent systems. The transmitter for this modulation format is relatively simple, consisting of two nested Mach-Zehnder phase modulators, which are delayed by 90° with respect to one another, allowing independent I and Q component modulation. The main complexity results from the receiver, where the incoming signal needs to be phase matched, locked and mixed with a local oscillator laser. In addition the signals in the receiver need to be mixed and delayed properly, so that both the I and the Q signal components can be extracted independently, in the two sets of balanced receivers shown in the schematic. Any changes in phase of the incoming signal, caused by the laser phase noise, need to be tracked and neutralized in the receiver. A number of different techniques can be used to accomplish this. Optical phase-locked loops use optical feedback to control the phase of the local oscillator laser [46], showing promise for low power, simple implementation of true coherent receivers. Digital signal processing can be used to perform real time phase tracking and control as well, at the expense of electronic chip sophistication and power consumption [47].

Although these three examples show the progression from simple, noncoherent amplitude modulated system, through a differentially modulated coherent system to a true I-Q coherent system, it is important to emphasize that many other different vector modulation formats and links are possible and used: differential QPSK, where the phase changes to one of four states are recorded only when the adjacent bit changes; or quadrature amplitude modulation (QAM), where both the amplitude and the phase of the sum I and Q components are changed, resulting in a multitude of points on the constellation diagram, and further improvement of spectral efficiency, up to the fundamental limit imposed by the fiber dynamic range to around 10 bits/s per Hz of bandwidth. Finally, additional doubling in spectral efficiency for each of the coherent links can be accomplished by multiplexing two signals onto two degenerate orthogonal polarizations of an optical fiber, creating a polarization-multiplexed link.

It is interesting to note that coherent technology was being widely explored in the 1980s because of its promise of increased transmission distance due to improved receiver sensitivity. Er-doped fiber amplifiers (EDFAs) had not been developed at the time, and wavelength division multiplexing (WDM) was expensive due to repeater cost and complexity (de-multiplexing, optical-electrical conversion, amplification, electrical demultiplexing to a lower data rate, regeneration, multiplexing back up, electrical-optical conversion, and multiplexing into optical fiber).

Coherent approaches have promised to double the repeater separation, and allow placing of the WDM channels closer together because the channel filtering could be done by a fixed IF filter in the RF-domain after heterodyne down-conversion by tuning the optical local oscillator (LO), much as in a radio. Bulk optical heterodyne receivers were quickly found to be very difficult to make due to stablity issues. Thus, some efforts were initiated to explore the possibility of monolithic integration of PICs for coherent communications. However, the invention of erbium doped fiber amplifiers and inexpensive, AWG-based multiplexers and demultiplexers channeled the development toward modern WDM systems for much of the 1995–2010 timeframe. The need for improving the spectral efficiency of transmission in the future

has led to renewed interest and intense research on optical coherent systems, as well as to the recent first deployment of this technology.

8.4.2 Coherent Detection

The entire concept of optical coherent systems is inspired by radio and microwave communications, where phase-modulated, coherent links are the mainstream technology. The key idea behind coherent detection is to combine the input signal coherently with a locally generated continuous optical field (local oscillator) at the receiver and before the signal is detected, as shown in Fig. 8.29. This action achieves two effects: It amplifies the detected signal through mixing, allowing improved receiver sensitivity, and it enables the demodulation of phase and amplitude/phase modulated signals, which is not possible through direct detection. This is a key enabler for achieving improved spectral efficiency in a coherent link. Taking the time harmonic field at the detector generated by a local oscillator $\mathscr{E}_{LO}(x,y,z,t) = \hat{\mathbf{e}}_i E_{LO} U(x,y) e^{j(\omega_{LO} t - \Phi_{LO})}$, and mixing it with a time harmonic field of the input signal, $\mathscr{E}_{IN}(x,y,z,t) = \hat{\mathbf{e}}_i E_{IN} U(x,y) e^{j(\omega_{IN} t - \Phi_{IN})}$, we can obtain the response of the photodetector keeping in mind that its current is proportional to the number of photons, and thus intensity of the electric field,

$$\begin{aligned} I(t) &= RK(\mathscr{E}_{LO} + \mathscr{E}_{IN})^* \cdot (\mathscr{E}_{LO} + \mathscr{E}_{IN}) = \\ &= P_{LO} + P_{IN} + 2\sqrt{P_{LO} P_{IN}} \cos((\omega_{IN} - \omega_{LO})t \pm \Phi_{IN} \mp \Phi_{LO}). \end{aligned} \tag{8.23}$$

Here, we have used K as the constant of proportionality between the optical power and the electric field intensity. From Eq. (8.23), we can draw two important conclusions. First, the time variation of the photodetector current will be caused by the time variation of the relative phase terms of the input and local waves, as well as on the frequency difference between the two waves, the third term in the equation. For high bit rate optical coherent systems, we are primarily interested in the case of homodyne detection, where $\omega_{IN} = \omega_{LO}$. Thus, phase modulation can be easily detected since it is being converted into amplitude modulation by this scheme. Second, the photocurrent detected is proportional to $2\sqrt{P_{LO} P_{IN}}$, which enables us to amplify the input signal by using a high-power local oscillator. This allows us to operate a coherent receiver in the shot noise limited regime, thus improving the optical receiver sensitivity.

8.4.3 Coherent Receiver Implementations

A couple of different coherent receiver architectures, for differential and regular detection were described in Fig. 8.28. The most basic coherent receiver implementation, as illustrated in Fig. 8.29, consists of simple building blocks that can be realized as a PIC—a laser, light coupler, a photodiode and passive waveguides. A higher performance implementation, also illustrated in a historic example shown in Fig. 8.30 uses a balanced receiver pair on the detection side. The local oscillator signal and the input signal are mixed inside a 2×2 coupler element, and detected

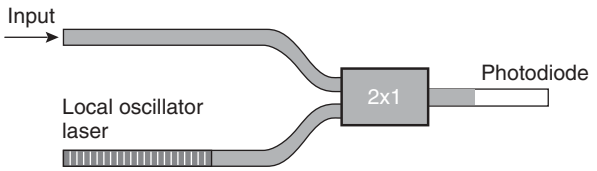


FIGURE 8.29: An illustration of the principle of coherent detection—the input signal is mixed with a local oscillator signal into a photodiode.

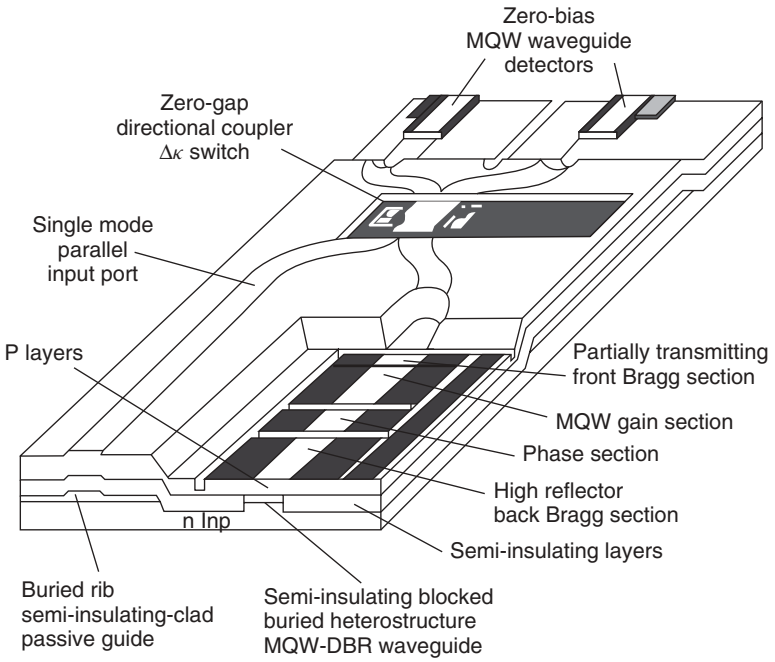


FIGURE 8.30: Schematic diagram of a MQW balanced heterodyne receiver photonic integrated circuit, containing a continuously tunable LO, a low-loss buried-rib parallel input port, an adjustable 3 dB coupler, and two zero-bias MQW waveguide detectors. (*Reprinted, by permission, from [29].*)

by two individual photodiodes, connected in series. With this photodiode configuration and the phase differences introduced by the optical coupler, it is possible to easily obtain an output signal which will be given by the difference of the two photocurrents-thereby canceling out the current contributions and the intensity of noise from the local oscillator, and adding the photocurrents resulting from the signal modulation. This type of architecture therefore allows for complete rejection of the CW signal, and conversion of phase modulation into amplitude modulation.

Ideally, the local oscillator would be a widely tunable laser, allowing for realization of a tunable coherent receiver.

The first photonic IC with a coherent balanced receiver implementation, shown Fig. 8.30, includes all the optical components of a coherent receiver: an integrated local DFB oscillator laser, a 3 dB coupler, and a balanced detector pair. The preamplifier shown in this example can be used to compensate for the coupling and waveguide losses.

An example of a high speed, integrated I-Q receiver in InP, is shown in Fig. 8.31. This device consists of two input waveguides, a 90° hybrid implementation using a 4×4 MMI coupler, and two balanced photodiode pairs. The optical hybrid allows for mixing of the local oscillator L and the input signal S, and for balanced detection of both the in phase I and the quadrature Q components of the input signal. This is accomplished through precise phase control in signal splitting, which results in the following signal combination at each of the photodiodes from top down, assuming that the signal is coupled to the top input waveguide: $L+S$, $L-S$, $L+jS$ and $L-jS$. The outputs from two balanced receiver pairs will be $2S$ and $2jS$, the in-phase and quadrature components of the input signal.

A more elaborate concept PIC implementation of a polarization multiplexed receiver, which includes two full I-Q receivers integrated with a local oscillator, and which could be used in polarization multiplexed optical links, is shown in Fig. 8.32. This concept pol-muxed receiver has two inputs, assuming that the polarization splitting is occuring outside of the chip. The polarization multiplexed signal results in two constellation diagrams, one on each polarization, as illustrated in the figure. Integration of polarization demultiplexing elements on this PIC is possible as well, in which case the chip would have a single input. In addition as

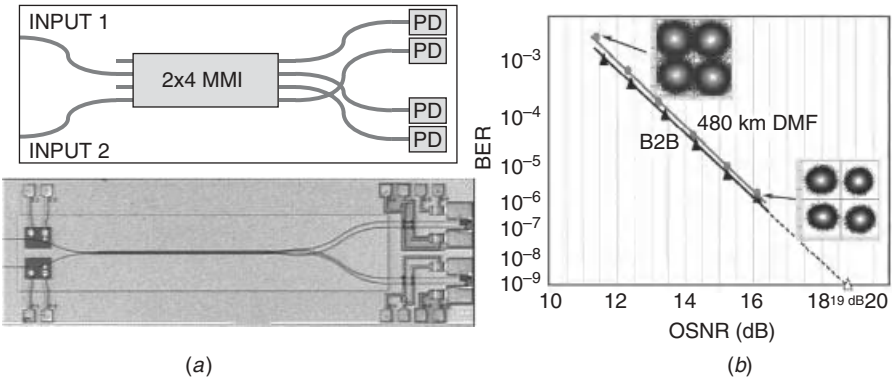


FIGURE 8.31: An integrated I-Q receiver in InP (a) Receiver architecture schematic, showing two inputs, a 90° hybrid implementation using a 4×4 MMI coupler, connected to two balanced photodiode pairs; a device photograph on the bottom. (b) Results of receiver operation at 50 Gbps, showing the bit error rate and constellation diagrams as function of optical signal to noise ratio. (© IEEE 2010, [48].)

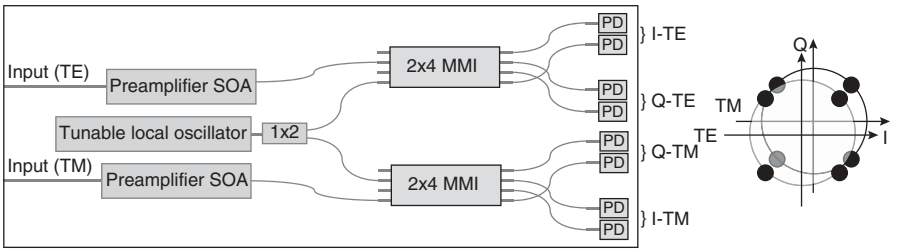


FIGURE 8.32: Schematic of a polarization multiplexed, balanced coherent I-Q receiver, with an integrated widely tunable local oscillator laser and two complete I-Q receivers, one for each polarization. Although this schematic shows two separate inputs for two separate polarizations of light, with the assumption that polarization demultiplexing is done outside of the PIC, integration of polarization demultiplexing elements on the chip is possible as well, in which case the chip would have a single input. A QPSK pol-muxed constellation diagram is shown on the right.

with other PIC-based devices further parallel integration into device arrays is possible. The key issue with tunable laser integration for coherent receiver purposes is that of sufficiently narrow linewidth and low phase noise, one of the topics of ongoing research.

8.4.4 Vector Transmitters

As discussed in one of the previous sections, arbitrary vector modulation can be generated using the combination of both amplitude and phase modulation. One popular way to accomplish this task is to use the nested Mach-Zehnder modulator structure shown earlier in Fig. 8.28c. Because this structure assigns the I axis to one MZM and the Q axis to a second MZM, it can modulate the resultant vector to any (I,Q) point in the plane of the I-Q diagram. For QPSK modulation, four equal amplitude (I-Q) points are accessed. A 10-wavelength transmitter PIC utilizing this type of nested I-Q Mach-Zehnder structure for QPSK modulation is shown in Figure 8.33 [49, 50]. Polarization multiplexing is implemented in this example to double the transmission rate, requiring a pair of identical nested I-Q MZ structures for each of the 10 DFB lasers on the chip. The constellation diagrams in the figure illustrate the four constellation points accessed by each of the 20 IQ modulators. Each individual I and Q MZ modulator is running at 14.25 Gbps, but as discussed in Section 8.4.1, with QPSK modulation, we double the amount of information transmitted (by effectively combining the I and Q signals in phase quadrature) resulting in 28.5 Gbps per IQ modulator. This chip utilizes DFB lasers for the light source, but one could imagine replacing them with the widely tunable variety. It is easy to envision that in the future, widely tunable EML vector transmitters will be developed. As with coherent receivers, the key issue with tunable laser integration for coherent transmitter purposes is that of achieving sufficiently narrow linewidth and low phase noise.

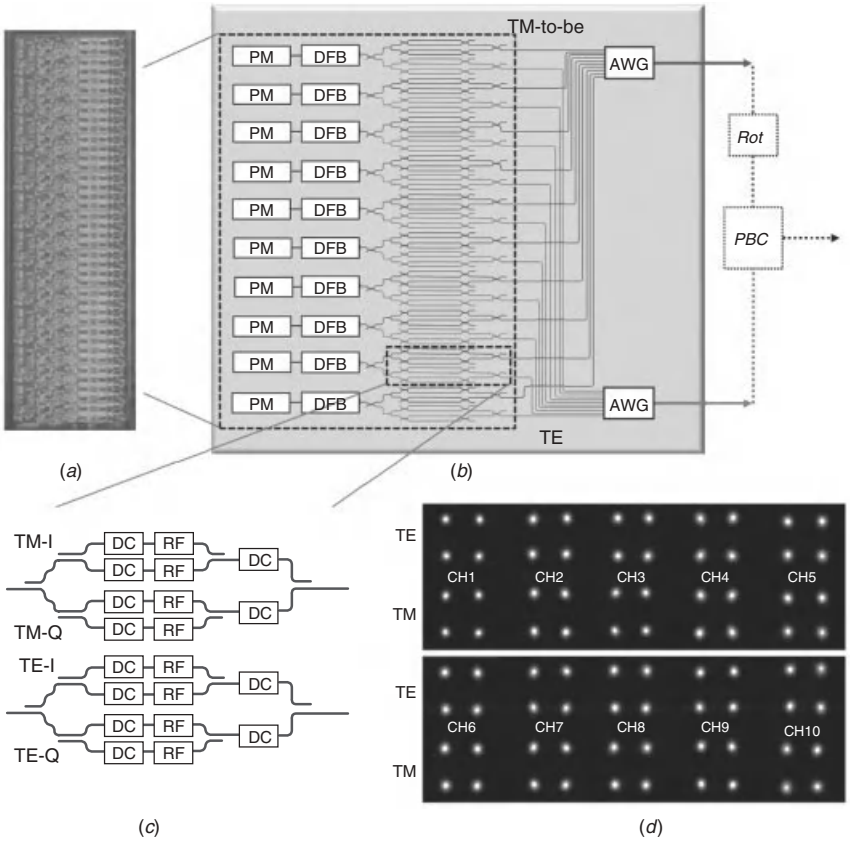


FIGURE 8.33: (a) Photograph of the active block of a 10-wavelength PM-QPSK transmitter PIC, utilizing nested IQ Mach-Zehnder modulator EML devices, (b) schematic layout of PIC illustrating the TE and TM-to-be duplicate sets of modulators, AWGS, and output waveguides, to support polarization-multiplexed operation through off-chip polarization beam combining, (c) schematic of TE/TM nested IQ MZ modulator section of one wavelength showing RF and DC controls, and (d) IQ constellation diagrams for all 20 QPSK data streams, each IQ stream running at 28.5 Gbps for an aggregate 570 Gbps transmission capability across the 10 wavelengths. Reprinted by permission from [49, 50].

REFERENCES

1. Johnson JE, Ketelsen LJ-P, Ackerman DA, Liming Zhang Hybertsen MS, Glogovsky KG, Lentz CW, Asous WA, Reynolds CL, Geary JM, Kamath KK, Ebert CW, Park M, Przybylek GJ, Leibenguth RE, Broutin SL, Stayt JW Jr, Dreyer KF, Peticolas LH, Hartman RL, Koch TL. Fully sxtabilized electroabsorption-modulated tunable DBR laser transmitter for long-haul optical communications. *IEEE Journal of Selected Topics in Quantum Electronics* 2001;**7**(2): 168–177.
2. Mason B, Barton J, Fish GA, Coldren LA, Denbaars SP. Design of sampled grating

DBR lasers with integrated semiconductor optical amplifiers. *IEEE Photonics Technology Letters* 2000;**12**(7): 762–764.

3. JDS Uniphase Corporation. Available at http://www.jdsu.com.
4. Ward AJ, Robbins DJ, Busico G, Barton E, Ponnampalam L, Duck JP, Whitbread ND, Williams PJ, Reid DCJ, Carter AC, Wale MJ. Widely tunable DS-DBR laser with monolithically integrated SOA: design and performance. *IEEE Journal of Selected Topics in Quantum Electronics* 2005;**11**(1): 149–156.
5. Oclaro. Available at http://www.oclaro.com.
6. Ishii H, Tanobe H, Kano F, Tohmore Y, Kondo Y, Yoshikuni Y. Quasicontinuous wavelength tuning in super-structure-grating (SSG) DBR lasers. *IEEE Journal of Quantum Electronics* 1996;**32**(3): 433–441.
7. Matsuo S, Segawa T. Microring-Resonator-Based Widely Tunable Lasers. *IEEE Journal of Selected Topics in Quantum Electronics* 2009;**15**(3): 545–554.
8. Wesstrom J-O, Sarlet G, Hammerfeldt S, Lundqvist L, Szabo P, Rigole P-J. Optical Fiber Communication Conference 2004 Feb 23–27;1:2.
9. Ignis ASA. Available at http://ignis.com/products/.
10. Matsuo S, Segawa T. Microring-resonator-based widely tunable lasers. *IEEE Journal of Selected Topics in Quantum Electronics* 2009;**15**(3): 545–554.
11. Rigole P-J, Nilsson S, Backbom L, Klinga T, Wallin I, Stalnacke B, Berglind E, Stoltz B. 114 nm wavelength tuning range of a vertical grating assisted codirectional coupler with super structure grating distributed Bragg reflector. *IEEE Photonics Technol Lett* 1995;**7**(7): 697–699.
12. Chuang Z-M, Coldren LA. Design of widely tunable semiconductor lasers using grating-assisted codirectional-coupler filters. *IEEE J Quantum Electron* 1993;**29**: 1071.
13. Zirngibl M, Joyner CH, Doerr CR, Stulz LW, Presby HM. An 18-channel multifrequency laser. *IEEE Photonics Technology Letters* 1996;**8**(7): 870–872.
14. Oh Kee Kwon, Johg Hoi Kim, Kang Ho Kim, Eun Deok Sim, Kwang Tyong Oh. Widely tunable multichannel grating cavity laset. *IEEE Photonics Technology Letters.* 2006;**18**(16): 1699–1701.
15. Morrison GM, Raring JW, Wang CS, Skogen EJ, Chang Y-C, Sysak M, Coldren LA. Electroabsorption modulator performance predicted from band-edge absorption spectra of bulk, quantum-well, and quantum-well-intermixed InGaAsP structures. *Solid-State Electronics* 2007;**51**(1): 38–47.
16. Tanaka H, Suzuki M, Matsushima Y. Optical short pulse generation by a DFB laser/EA modulator integrated light source. *IEEE J Quantum Electron* 1993;**29**: 1708.
17. Makino S, Shinoda K, Kitatani T, Hayashi H, Shiota T, Tanaka S, Aoki M, Sasada N, Naoe K. High-speed electroabsorption modulator integrated DFB laser for 40 Gbps and 100 Gbps application. Indium Phosphide & Related Materials. IEEE International Conference on IPRM 2009 May 10–14. p 362–366.
18. Sysak MN. Monolithically integrated wavelength converters using a dual quantum well integration platform [dissertation]. Santa Barbara (CA): University of California Santa Barbara; 2005.
19. Numai T. *Laser diodes and their applications to communications and information processing*. Hoboken, NJ: Wiley; 2010.
20. Barton JS, Skogen EJ, Masanovic ML, Denbaars SP, Coldren LA. A widely tunable high-speed transmitter using an integrated SGDBR laser-semiconductor optical amplifier

and Mach-Zehnder modulator. *IEEE Journal of Selected Topics in Quantum Electronics* 2003;**9**(5): 1113–1117.

21. Nagarajan R, Kato M, Pleumeekers J, Evans P, Lambert D, Chen A, Dominic V, Mathur A, Chavarkar P, Missey M, Dentai A, Hurtt S, Back J, Muthiah R, Murthy S, Salvatore R, Grubb S, Joyner C, Rossi J, Schneider R, Ziari M, Kish F, Welch D. Single-chip 40-channel InP transmitter photonic integrated circuit capable of aggregate date rate of 1.6 Tbit/s. *Electronics Letters* 2006;**42**(13): 771–773.
22. Nagarajan R, Kato M, Hurtt S, Dentai A, Pleumeekers J, Evans P, Missey M, Muthiah R, Chen A, Lambert D, Chavarkar P, Mathur A, Back J, Murthy S, Salvatore R, Joyner C, Rossi J, Schneider R, Ziari M, Kish F, Welch D. *Monolithic, 10 and 40 channel InP receiver photonic integrated circuits with on-chip amplification*. National Fiber Optic Engineers Conference, OSA Technical Digest Series. Washington, DC: Optical Society of America; 2007.
23. Bernasconi P, Yang W, Zhang L, Sauer N, Buhl L, Kang I, Chandrasekhar S, Neilson DT. Monolithically integrated 40 Gb/s wavelength converter with multi-frequency laser. Optical Fiber Communication Conference 2005 Mar 6–11;5: 3.
24. den Besten J. Integration of multiwavelength lasers with fast electro-optical modulators [dissertation]. Eindhoven, Netherlands: Technical University Eindhoven; 2004.
25. Saitoh T, Mukai T. Traveling-wave semiconductor laser amplifiers. In: Yamamoto Y, editor. *Coherence, amplification and quantum effects in semiconductor lasers*. New York: Wiley; 1991. Chapter 7.
26. Agrawal G. *Fiber optic communication systems*. New York: Wiley; 2002.
27. Kodama S, Ito H. UTC-PD-based optielectronic components for high-frequency and high-speed applications. *IEICE Transactions on Electronics* 2007;**E90-C(2)**: 429–435.
28. Kaminow I, Li T, Willner AE. *Optical fiber telecommunications V A: components and subsystems*. 5th ed. Amsterdan: Elsevier; 2008.
29. Koch TL, Koren U, Miller BI. High performance tunable 1.5 μm InGaAs/InGaAsP multiple quantum well distributed Bragg reflector lasers. *Appl Phys Lett* 1988;**53**: 1036.
30. Coldren LA, Corzine SW. Continuously-tunable single-frequency semiconductor lasers. *EEE J Quantum Electron* 1987;**23**: 903.
31. Idler W, Schilling M, Baums D, Laube B, Wunstel K, Hildebrand O. Y laser with 38 nm tuning range. *Electron Lett* 1991;**27**: 2268.
32. Alferness RC. Koren U, Buhl LL, Miller BI, Young MG, Koch TL, Raybon G, Burrus CA. Broadly tunable InGaAsP/InP laser based on a vertical coupler filter with 57-nm tuning range. *Appl Phys Lett* 1992;**60**: 3209.
33. Jayaraman VJ, Chaung Z-M, Coldren LA. Theory, design, and performance of extended tuning range semiconductor lasers with sampled gratings. *IEEE J Quantum Electron* 1993;**29**: 1824.
34. Knüpfer B, Kiesel P, Kneissl M, Dankowski S, Linder N, Weimann G, Döhler GH. Polarization-insensitive high-contrast GaAs/AlGaAs waveguide modulator based on the Franz-Keldysh effect. *IEEE Photonics Tech Lett* 1993;**5**: 1386.
35. Boyd GD, Livescu G. Electro-absorption and refraction in Fabry-Perot quantum well modulators: a general discussion. *Optical and Quantum Electron* 1992;**24**: S147.
36. Huang TC, Chung Y, Coldren LA, Dagli N. Field-induced waveguides and their application to modulators. *IEEE J Quantum Electron* 1993;**29**: 1131.

37. Koch TL, Koren U, Gnall RP, Chao FS, Hernandez-Gil F, Burrus CA, Young MG, Oron M, Miller BI. GaInAs/GaInAsP multiple-quantum-well integrated heterodyne receiver. *Electron Lett* 1989;**25**: 1621.

38. Feit MD, Fleck JA Jr. Computation of mode properties in optical fiber waveguides by a propagating beam method. *Appl Opt* 1980;**19**: 1154.

39. Feit MD, Fleck JA Jr. Computation of mode eigen functions in graded-index optical fibers by the propagating beam method. *Appl Opt* 1980;**19**: 2240.

40. Dummer MM, Klamkin J, Tauke-Pedretti A, Coldren LA. A bit-rate-transparent monolighically integrated wave-length converter. 34th European Conference on Optical Communication, 2008 Sept 21–24; 2008. p 1–4.

41. Nicholes SC, Masanovic ML, Jevremovic B, Lively E, Coldren LA, Blumenthal DJ. The world's first InP 88 monolithic tunable optical router (MOTOR) operating at 40 Gbps line rate per port. Conference on Optical Fiber Communication; 2009 Mar 22–26. p 1–3.

42. Vikrant L, Misanovic ML, Summers JA, Fish G, Blumenthal DJ. Monolithic wavelength converters for high-speed packet-switched optical networks. *IEEE Journal of Selected Topics in Quantum Electronics* 2007;**13**(1): 49–57.

43. Masanovic ML, Lal V, Summers JA, Barton JS, Skogen EJ, Rau LG, Coldren LA, Blumenthal DJ. Widely tunable monolithically integrated all-optical wavelength converters in InP. *Journal of Lightwave Technology* 2005;**23**(3): 1350–1362.

44. Tolstikhin VI, Moore R, Pimenov K, Logvin Y, Wu F, Watson CD. One-step growth optical tranceiver PIC in InP. 35th European Conference on Optical Communication; 2009 Sept 20–24. p 1–2.

45. Coldren L, Nicholes S, Johansson L, Ristic S, Guzzon R, Norberg E, Krishnamachari U. High performance InP-based photoic Ics—a tutorial. *IEEE Journal of Lightwave Technology*. Forthcoming.

46. Johansson LA, Krishnamachari U, Ramaswamy A, Klamkin J, Ristic S, Chou H-F, Coldren LA, Rodwell M, Bowers JE. Linear coherent optical receivers. 2010 Conference on OFC/NFOEC; 2010 Mar 21–25. p 1–3.

47. O'Sullivan M. Expanding network application with coherent detection. National Fiber Optic Engineers Conference; 2008.

48. Bottacchi S, Beling A, Matiss A, Nielsen ML, Steffan AG, Unterborsch G, Umback A. Advanced photoreceivers for high-speed optical fiber transmission systems. *IEEE Journal of Selected Topics in Quantum Electronics* 2010;**16**(5): 1099–1112.

49. Corzine SW, Evans P, Fisher M, Gheorma J, Kato M, Dominic V,Samra P, Nilsson A, Rahn J, Lyubomirsky I, Dentai A, Studenkov P, Missey M, Lambert D, Spannagel A, Muthiah R, Salvatore R, Murthy S, Strzelecka E, Pleumeekers JL, Chen A, Schneider R, Nagarajan R, Ziari M, Stewart J, Joyner CH, Kish F, Welch DF. Large-scale InP transmitter PICs for PM-DQPSK fiber transmission systems. *IEEE Photonics Technology Letters* 2010;**22**(14): 1015–1017.

50. Evans P, Fisher M, Malendevich R, James A, Studenkov P, Goldfarb G, Vallaitis T, Kato M, Samra P, Corzine S, Strzelecka E, Salvatore R, Sedgwick F, Kuntz M, Lal V, Lambert D, Dentai A, Pavinski D, Behnia B, Bostak J, Dominic V, Nilsson A, Taylor B, Rahn J, Sanders S, Sun H, Wu K-T, Pleumeekers J, Muthiah R, Missey M, Schneider R, Stewart J, Reffle M, Butrie T, Nagarajan R, Joyner C, Ziari M, Kish F, Welch D. Multi-channel coherent PM-QPSK InP transmitter photonic integrated circuit

(PIC) operating at 112 Gb/s per wavelength. Optical Fiver Communication Conference, Post Deadline Paper PDPC7; 2011 March.

READING LIST

Coldren LA. Lasers and modulators for OEICs. In: Dagenais M, Crow J, Leheny R, editors. *Integrated optoelectronics*. New York: Academic Press; 1994.

Ebeling KJ. *Integrated opto-electronics*. Berlin: Springer-Verlag; 1993. Chapters 12 and 13.

Koch TL, Koren U. Photonic integrated circuits. In: Dagenais M, Crow J, Leheny R, editors. *Integrated opto-electronics*. New York: Academic Press; 1994.

Kaminow I, Li T, Willner AE. *Optical fiber telecommunications V A: components and subsystems*. 5th ed. Amsterdam: Elsevier; 2008.

PROBLEMS

These problems may draw from material in previous chapters and appendices.

1. Design a tunable two-section 1.55-μm InGaAsP/InP DBR (a gain region with a cleave at one end and a passive grating at the other) which has output frequencies equally spaced by 100 GHz when the index of the grating is tuned. Assume a 4-quantum-well gain region with characteristics as given by Fig. 4.31 with quaternary barriers of bandgap wavelength 1.25 μm. This active region is placed on top of a 0.25-μm waveguide also of 1.25 μm bandgap material and all clad with InP. (You must determine the confinement factor, effective index, etc.) The grating region is formed by removing the quantum wells and etching a fundamental-order triangular sawtooth grating with a peak-to-peak depth of 30 nm on the top side of the remaining 0.25 μm quaternary waveguide. This again is coated with InP. It is desired to tune the output over eight of the axial modes (spaced by 100 GHz). Assume a BH waveguide width of 1.5 μm, an internal loss of 10 cm^{-1} along the entire device length, an internal carrier injection efficiency of 70%, and provide a grating length to give a power reflection of 50% in the absence of loss.

(a) Determine the grating and gain region lengths.

(b) Plot the power out of the cleaved end versus current into the gain section.

(c) Plot the output frequency deviation versus the current injected into the grating assuming both radiative and Auger recombination in the 1.25 μm bandgap Q-material.

2. Design a 1.55 μm DFB laser with an integrated amplifier used as a modulator. The DFB laser is a quarter-wave-shifted design with each grating half having $\kappa L = 0.7$. Its total length is 400 μm. The active region consists of four quantum wells in a separate-confinement waveguide as described by the gain curve of Fig. 4.31, and the transverse confinement factor is found to be 6%. This active

region is the same throughout, and there is no significant gap between the laser section and the amplifier section. Assume ideal AR coatings on the output facets, an internal efficiency of 70%, and an internal loss of 15 cm^{-1} throughout. The laser is biased to output 3 mW cw into the amplifier section. We desire the amplifier-modulator to operate as fast as possible, to have a 10 dB optical on/off ratio, and to have an on-level output of 10 mW.

(a) Determine the length and width of the amplifier (same width for laser—-cannot exceed 5 μm for single lateral mode).

(b) Determine the total laser current and the on/off level currents to the amplifier.

(c) What is the maximum modulation data rate?

(d) Can the modulation rate be improved for different laser and modulator biases?

3. It is desired to design a quarter-wave-shifted 1.55 μm DFB with a maximum overall power efficiency (power out/power in) at an output power of 10 mW. (Assume a series resistance that scales inversely with active area, which equals 10Ω at an area of 100 μm^2. Estimate the junction voltage by the approximate quasi-Fermi level separation.) At 300 K, we assume a gain as determined by Fig. 4.31, a transverse confinement factor of 6%, an internal efficiency of 70%, an internal loss of 15 cm^{-1}, and a $\kappa = 50$ cm^{-1}. However, we also empirically decrease the gain curve and the internal efficiency by 1% for each 1°C temperature rise. We assume that the thermal impedance can be estimated by Eq. (2.69) with a substrate thickness, $h = 100$ μm and effective thermal conductivity $\xi = 0.68$ W/cm-°C. Assume lateral cladding is InP.

(a) What length and width is optimum?

(b) Plot the resulting L–I curve.

4. For the sampled grating DBR laser from Example 8.3,

(a) How much current is required on the phase section to tune from one axial mode to the next (assuming a pair of mirror peaks are aligned at some wavelength)?

(b) Assuming that the central mirror peaks and an axial cavity mode are aligned at zero current, what currents are required on the back mirror, the front mirror and the phase section to continuously tune the laser output by exactly 1 nm.

(c) What are the laser threshold current and differential efficiency for tuning currents of both mirrors and the phase section sufficient to change all of the modal indices by 0.005?

5. It is desired to design a 1.55-μm grating assisted codirectionally coupled laser with a tuning enhancement factor of $F = 15$ and a maximum MSR at an output of 5mW. We assume a gain characteristic as in Fig. 4.31, a transverse gain region confinement factor of 6%, an internal efficiency of 70%, $n_{sp} = 1.2, n_{g1} = 4$, an internal loss of 15 cm^{-1} in the active and passive sections of the top guide, an

internal loss of 5 cm^{-1} in the lower guide, a design for 100% guide-to-guide coupling, and an active region length of 300 μm. Also, the operating current density in the active region is limited to 8 kA/cm^2.

(a) What is the corrugation period in the coupler?

(b) What is the required κ for 100% coupling?

(c) What is the optimum coupler region length for maximum MSR?

(d) Plot the L–I characteristic for a 3-μm active region width from the left facet when the coupler is tuned for alignment of a mode with its filter peak.

6. For the GCSR laser from Example 8.5, the internal loss in the gain section is 15 cm^{-1}, and the passive tuning sections have a loss of $(5 + 5000\Delta\bar{n}/\bar{n}_g\ cm^{-1})$, where $\Delta\bar{n}/\bar{n}_g$ is the relative modal index change due to carrier injection tuning. What are the threshold current and the differential efficiency for:

(a) no tuning?

(b) for max tuning—required for full wavelength coverage up to 50 nm?

(c) Plot the P-I curves for no tuning and maximum tuning cases.

7. For the Mach-Zehnder modulator from Example 8.7, determine the maximum length of the electrode to achieve 25 GHz modulation bandwidth?

8. We wish to optimize the design of a sampled-grating DBR laser for a tuning range of 50 nm to around 1550 nm operating wavelength range. The back mirror, phase section, and gain section have already been designed. We need specify the design of the front mirror for maximum power out at 5 kA/cm^2 into the gain electrode, and verify the characteristics of the design for the desired 50-nm tuning range. The waveguides are all 2 μm wide, and lateral current spreading can be neglected. The common passive waveguide thickness is 400 nm in all sections. The internal loss in the gain section is 15 cm^{-1} above threshold, and the passive tuning sections (both mirrors and phase) have a loss of $(5 + 5000\Delta\bar{n}/n)\ cm^{-1}$, where $\Delta\bar{n}/n$ is the relative modal index change due to carrier injection tuning. $\Delta n/n$ is also limited to 0.5% in all tuning sections. The phase and group modal indices are 3.3 and 3.8 in the passive sections, and 3.4 and 4.0 in the gain section, respectively.

The back mirror consists of: 10 grating bursts (samples), 50 μm burst (sampling) period, 5 nm burst length, 300 cm^{-1} grating coupling constant (within burst) and internal (injection efficiency) of 70%.

The phase section is 75 μm long.

The gain section is 500 μm long, contains 5 QWs characterized by Fig. 4.31, the net transverse/lateral confinement factor for the 5 QWs 5%, and the internal efficiency 70%.

The front mirror has the same grating coupling constant and burst length as back mirror.

(a) What is the back mirror peak power reflection magnitude, R_1, at Bragg with no tuning, and with maximum tuning?

(b) What is the FWHM, $\Delta\lambda_{1/2}$, of a back mirror reflection peak?

(c) What is the wavelength spacing $\Delta\lambda_p$ between the back mirror reflection peaks?

(d) At what wavelength deviation from the Bragg does the peak reflection envelope fall to 0.5 R_1?

(e) What is the grating sampling period in the front mirror for the desired 50-nm tuning range with the given maximum in $\Delta n/n = 0.005$?

(f) What is the magnitude of the front mirror reflection, R_1, for maximum power out at 5 kA/cm^2 into the gain section terminal, assuming no tuning currents? How many sampling periods are desired in the front mirror?

(g) What is the axial cavity mode spacing? How many modes are within the FWHM width of the back mirror?

(h) What is the threshold current and differential efficiency for (1) no tuning? (2) maximum tuning required for full wavelength coverage up to 50 nm?

(i) Plot the P–I curves for no tuning and max tuning cases.

(j) Sketch superimposed front and rear mirror reflection spectra indicating widths and spacings as well as mode locations. Is the design good? (Good MSR, reasonable performance at all wavelengths, etc.) How could it be improved?

9. The ring-resonator mirror laser, shown in Fig. 8.34 has the following parameters: circumference of ring 1 (including couplers and phase shifters) is 300 μm; circumference of ring 2 (including couplers and phase shihers) is 400 μm; length of each gain section is 300 μm; length of each phase section is 75 μm; total length of straight guides between couplers is 450 μm (includes space between elements); length of each coupler is $l_c = 50$ μm; cross coupling of each coupler is $-j0.5 \cdot e^{-j\tilde{\beta}\cdot l_c}$. Assume symmetry where necessary, and neglect reflections at the ends of the straight waveguides.

(a) Assuming zero current to the gain and phase sections, what is the transfer function from the straight guides at plane A to B through the couplers and ring 2 (i.e., S_{BA})? Give an analytical expression, and then evaluate it with the numbers given both on and off of a ring resonance.

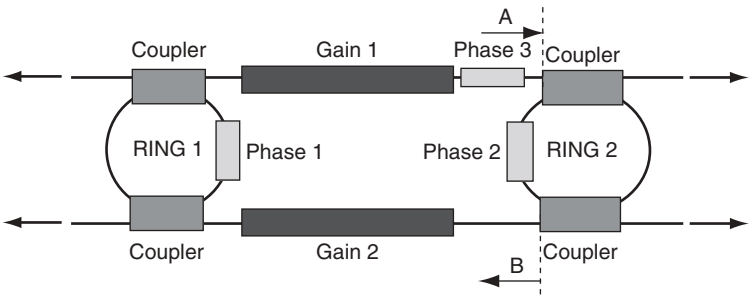


FIGURE 8.34: Ring resonator mirror laser.

(b) What is the resonance spacing in each of the two rings, again assuming zero current to the gain and phase sections.

(c) Assuming current to the gain sections for lasing, what is the axial laser mode spacing?

(d) Plot and label superimposed spectra of the ring resonances with the laser mode locations.

(e) Assuming a pair of ring resonance peaks are aligned at zero phase currents, how much cunent is required on the phase section in ring 1 (phase 1) to move its resonances so the next resonance peaks align (i.e., vernier tuning between the resonances of the two rings)?

(f) Assuming a pair of resonance peaks and a laser cavity mode are aligned with zero phase currents, what currents are required on phase 1, phase 2, and phase 3 to continuously tune the laser output by exactly 1 nm?

(g) What is the laser threshold current and differential efficiency out of one of the four output guides for no tuning currents?

(h) What is the laser threshold current and differential efficiency (from one output) for tuning currents on all phase sections sufficient to change all of the modal indices by 0.005?

(i) Assuming aligmnent of a resonance of the two rings and a laser cavity mode, estimate the MSR for (1) the adjacent laser cavity mode, and (2) the laser cavity mode nearest the partial overlap of the adjacent resonances of the rings.

APPENDIX ONE

Review of Elementary Solid-State Physics

A1.1 A QUANTUM MECHANICS PRIMER

A1.1.1 Introduction

In quantum mechanics, the properties and motion of particles are defined in terms of a wave (or state) function, Ψ; its magnitude squared gives the probability density of finding a particle at some point in time in a volume element dV. Or, put another way, the density of particles at some point in space is proportional to $\Psi^*\Psi$. Note that $\int \Psi^*\Psi \, dV = 1$, for properly normalized state functions, since the probability of the particle being somewhere is unity. In our case we are interested in both electrons and photons as particles. For photons this description is roughly equivalent to standard electromagnetic theory where the wavefunction is analogous to a normalized electric field. Maxwell's equations give the description of photon fields. In this appendix we shall focus more specifically on the properties of electrons.

In quantum mechanics, measurements are limited in accuracy by the uncertainty principle, $\Delta x \, \Delta p_x \geq \hbar/2$, where p_x is the momentum in the x-direction. (In electromagnetic theory the equivalent statement is that $\Delta x \, \Delta k_x \geq 1/2$.) The expected (or mean) value of some observation is calculated by operating on the wavefunction with the operator, A, corresponding to the observable, a. The operation to obtain

Diode Lasers and Photonic Integrated Circuits, Second Edition.
Larry A. Coldren, Scott W. Corzine, and Milan L. Mašanović.
© 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

the mean value is analogous to a standard weighted average,

$$\langle a \rangle = \int \Psi^* A \Psi \, dV, \tag{A1.1}$$

where a is a possible observation of the operator A. In many cases the operator simply multiplies the observable variable, in others it is more complex, such as momentum, p, where it is $-i\hbar\nabla$.

The motion of particles is governed by Schrödinger's equation,

$$\frac{-\hbar^2}{2m}\nabla^2\Psi + V\Psi = i\hbar\frac{\partial\Psi}{\partial t}, \tag{A1.2}$$

where m is the particle's (e.g., electron) mass, V is the potential energy operator (same as observable), and $(-\hbar^2/2m)\nabla^2$ is the kinetic energy operator $(= p^2/2m)$. Together, these two form the overall energy operator, the so-called Hamiltonian. The state function can be expressed as the product of space-dependent and time-dependent factors, $\Psi(r,t) = \psi(r)w(t)$. If we substitute into Eq. (A1.2) and divide by ψw, we obtain a function on the left that only depends on r and a function on the right that only depends on t. Thus, to be valid for all r and t, each side must equal a constant, E:

$$\frac{-\hbar^2}{2m}\frac{\nabla^2\psi}{\psi} + V = \frac{i\hbar}{w}\frac{\partial w}{\partial t} = E. \tag{A1.3}$$

From this we immediately have

$$w(t) = Ce^{-i(E/\hbar)t}, \tag{A1.4}$$

from which we can identify $E = \hbar\omega$, where ω is the radian frequency of oscillation. For the time-independent part,

$$\frac{-\hbar^2}{2m}\nabla^2\psi + V\psi = E\psi. \tag{A1.5}$$

The general solution for a uniform potential can be written as the sum of two counterpropagating plane waves,

$$\psi(r) = Ae^{ikz} + Be^{-ikz}, \tag{A1.6}$$

where

$$k^2 = \frac{2m}{\hbar^2}(E - V), \tag{A1.7}$$

is found by substituting back into Eq. (A1.5).

A1.1.2 Potential Wells and Bound Electrons

Electrons are confined by some potential depression in most situations. The most fundamental example is the atom, where electrons are bound by the confining potential of the positively charged nucleus. For the simple case of the hydrogen atom, $V(r) = -q^2/[4\pi\varepsilon_0 r]$, and analytic solutions to Schrödinger's time-independent Eq. (A1.5) can be found. However, for atoms with higher atomic numbers and many electrons, only numerical solutions are possible. Nevertheless, the electron always experiences some sort of confining potential.

When solids are formed from these atoms, the more weakly bound electrons near the exterior of the atom are significantly influenced by the attractive potential of neighboring atoms. In fact, in covalently bonded solids, the outer valence electrons are shared by many atoms, and they develop wavefunctions that extend throughout the crystal. In such cases, the details of the original atomic confining potential are lost. Thus, we shall not dwell on that problem unduly. Rather, we shall investigate the properties of an electron in a simple rectangular potential well to develop the concepts of confined wavefunctions and discrete energy levels common to atoms.

It will later be shown that by coupling together a series of such wells, a periodic potential is formed that leads to electronic properties very similar to those in real crystals. Thus, we can learn much about the properties of electrons in solids by taking this course. As is well known, one of the key results is that electrons in solids can behave much like free electrons with plane wave solutions and a parabolic E–k relationship as illustrated by Eq. (A1.7). However, they appear to have an *effective mass*, m^*, that is different from the free electron mass. Also, this effective mass approximation is usually limited to relatively low kinetic energies. Finally, we shall consider quantum-confined structures that include heterostructures to form much larger potential wells than for a single atom. Nevertheless, the mathematics is very similar, and we will be able to apply much of what we develop in this section.

First, consider the one-dimensional potential well of width l shown in Fig. A1.1. The simplest method of solution is to recognize that there are three separate regions of uniform potential, where the solution to Schrödinger's equation will have the form of Eq. (A1.6). Then, if we assume that the effective electron mass is the same in all regions, we can develop a complete wavefunction by requiring that the value and slope of the constituent solutions in each of the three regions match at the boundaries. That is, we would not expect any discontinuity in the probability density function. Looking for bound solutions, for which $E < V_0$, we can rewrite the general solution Eq. (A1.6) in each region. In the central region II,

$$\psi_{\mathrm{II}} = \begin{cases} A\cos kx & \text{(symmetric solutions)} \\ A\sin kx & \text{(antisymmetric solutions),} \end{cases} \tag{A1.8}$$

where $k^2 = 2mE/\hbar^2$. In region III,

$$\psi_{\mathrm{III}} = Be^{-\gamma x}, \tag{A1.9}$$

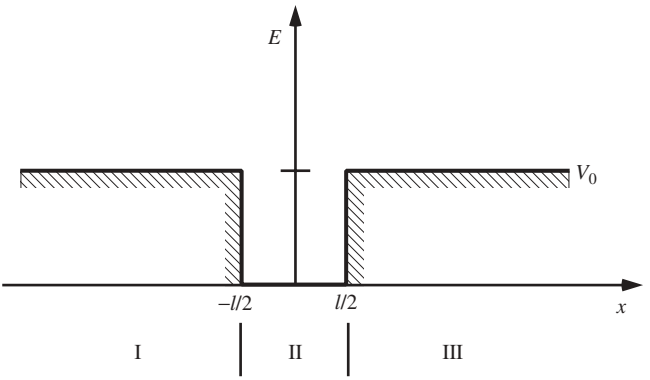


FIGURE A1.1: One-dimensional potential well for electron confinement.

where $\gamma^2 = 2m(V_0 - E)/\hbar^2$. In region I, $\psi_{\mathrm{I}} = Be^{\gamma x}$, but by symmetry, we only need to use the single boundary condition at $x = l/2$ between regions II and III. At $x = l/2$, we have that $\psi_{\mathrm{II}} = \psi_{\mathrm{III}}$ and $\psi'_{\mathrm{II}} = \psi'_{\mathrm{III}}$. For the symmetric solutions, this gives

$$A \cos\left(\frac{kl}{2}\right) = Be^{-\gamma l/2}, \tag{A1.10a}$$

and

$$Ak \sin\left(\frac{kl}{2}\right) = B\gamma e^{-\gamma l/2}. \tag{A1.10b}$$

Dividing Eq. (A1.10a) by Eq. (A1.10b), we obtain the characteristic equation,

$$k \tan\left(\frac{kl}{2}\right) = \gamma. \tag{A1.11}$$

Similarly, for the antisymmetric solutions, we obtain

$$k \tan\left(\frac{kl}{2} - \frac{\pi}{2}\right) = \gamma, \tag{A1.12}$$

where $\cot x = -\tan(x - \pi/2)$ has been used to illuminate the similarity between the symmetric and antisymmetric characteristic equations.

The electron energy, E, appears on both sides of these characteristic equations via k and γ, implying that only discrete values of E will satisfy the requirement that the wavefunction and its derivative be continuous across the boundaries. Because the tangent function is periodic, multiple solutions can be found for E, leading to a discrete set of wavefunctions that satisfy the boundary conditions.

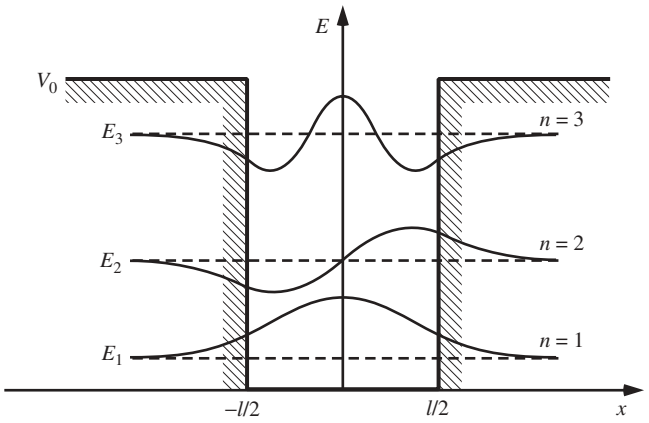


FIGURE A1.2: Energy levels and wavefunctions of one-dimensional potential well. Three bound solutions illustrated.

Figure A1.2 shows the first few wavefunctions drawn schematically on their respective energy levels over the potential well for reference. These clearly represent bound solutions. There are solutions with $E > V_0$ but they are not bound, and their wavefunctions extend to $\pm\infty$. An interesting property of all the solutions is that they must be *orthogonal*. That is, if we multiply one wavefunction by the complex conjugate of another and integrate over all space, the integral must be zero. If the wavefunctions are normalized so that the integral of the product of a wavefunction times its own complex conjugate is unity, then the wavefunctions would be *orthonormal*.

To determine the bound solutions, we need to solve the characteristic equations. For infinitely large V_0, such that the wavefunction goes to zero at the boundaries (i.e., $\psi(l/2) = 0$), the characteristic equation for both symmetric and antisymmetric cases becomes simply

$$\frac{kl}{2} = \frac{n\pi}{2}, \quad n = 1, 2, 3, \ldots \tag{A1.13}$$

where odd (even) quantum numbers correspond to symmetric (antisymmetric) states. The corresponding discrete energy levels in terms of the quantum numbers are

$$E_n = n^2 E_1^{\infty}, \tag{A1.14}$$

where

$$E_1^{\infty} = \frac{\hbar^2 k_1^2}{2m} = \frac{\hbar^2 \pi^2}{2ml^2} = 3.76(m_0/m)(100\ \text{Å}/l)^2 \text{ in meV}.$$

When V_0 is reduced from infinity, the discrete energies can still be found using (A1.14), however, the quantum numbers in this case are no longer simple integers, but are real numbers that we will refer to as n_{QW}. For example, if $V_0 = 25E_1^\infty$, the infinite-barrier integer quantum numbers $n = 1, 2, 3, 4, 5$ become $n_{QW} = 0.886$, 1.77, 2.65, 3.51, 4.33.

To calculate n_{QW} for an arbitrary V_0, we need to solve the characteristic equations given in Eqs. (A1.11) and (A1.12). Using Eq. (A1.14), combined with the definitions for k and γ given below Eqs. (A1.8) and (A1.9), the characteristic equations can be conveniently normalized:

$$\tan\left[\frac{\pi}{2}n_{QW}\right] = \frac{1}{n_{QW}}[n_{\max}^2 - n_{QW}^2]^{1/2} \quad \text{(symmetric)} \tag{A1.15}$$

$$\tan\left[\frac{\pi}{2}(n_{QW} - 1)\right] = \frac{1}{n_{QW}}[n_{\max}^2 - n_{QW}^2]^{1/2} \quad \text{(antisymmetric)} \tag{A1.16}$$

where

$$n_{QW} \equiv \sqrt{\frac{E_n}{E_1^\infty}} \quad \text{and} \quad n_{\max} \equiv \sqrt{\frac{V_0}{E_1^\infty}}. \tag{A1.17}$$

These equations can be solved graphically by plotting both the left-hand side (LHS) and the right-hand side (RHS) as a function of n_{QW}. Figure A1.3 illustrates this procedure for four different values of V_0.

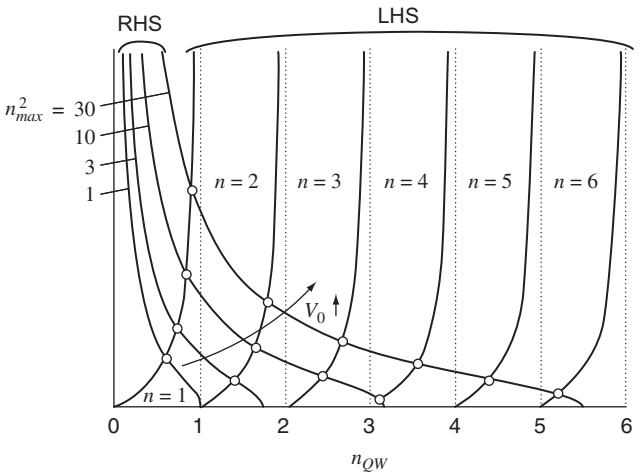


FIGURE A1.3: Graphical solution to Eqs. (A1.15) and (A1.16). The intersections between the LHS and RHS of the equations yield the possible values of n_{QW} for a given $n_{\max}$ (or equivalently V_0). The odd (even) quantum numbers displayed next to each tangent curve correspond to the LHS of the symmetric (antisymmetric) characteristic equation.

Note that only a finite set of quantum numbers exist for a given potential barrier, V_0. The normalized variable, $n_{\max}$ when rounded up to the nearest integer, yields the largest number of bound states possible for a given V_0. For example, with $V_0 = 3E_1^{\infty}$, from Eq. (A1.17), we find that $n_{\max} = \sqrt{3} \approx 1.73$. Thus, only two bound states are possible under these circumstances. This is perhaps demonstrated more clearly by plotting the possible n_{QW} as a continuous function of $n_{\max}$. Figure A1.4 gives all possible solutions for $n_{\max} \leq 6$ (which covers nearly all practical ranges of interest). Note that all quantum numbers approach their integer limit as $n_{\max}$ increases toward infinity. In addition, the quantum numbers cease to satisfy the equations (indicated by the open circles) when a given quantum number approaches the integer value of the next lowest state. The lowest quantum number can be

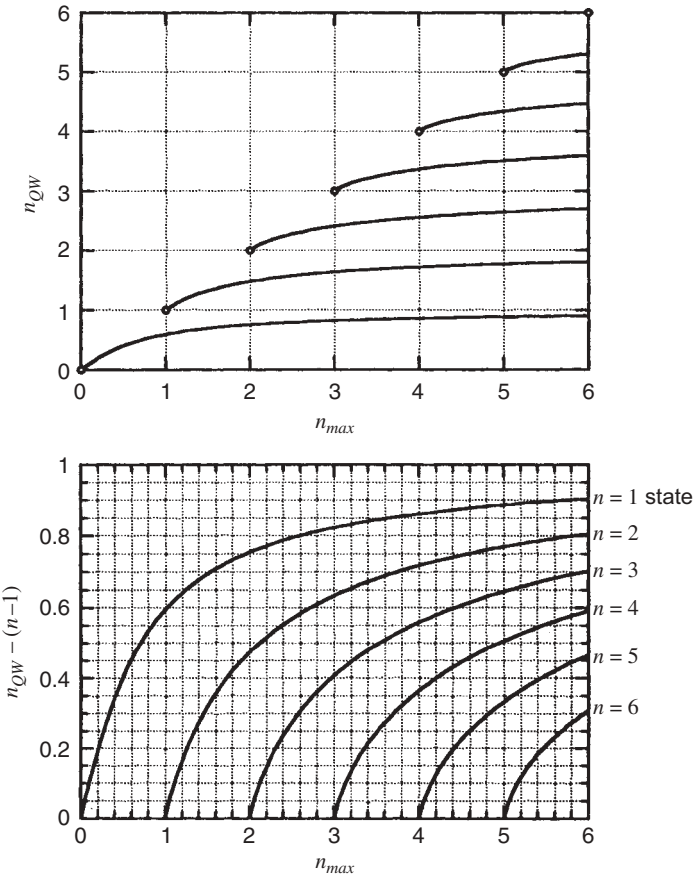


FIGURE A1.4: Plot of quantum numbers as a function of the maximum allowed quantum number that is determined by the potential height, V_0. The quantum numbers are related to V_0 and E through Eq. (A1.17). The lower plot gives a close-up view of the curves (which have been shifted vertically to fit on the same scale).

approximated to within ±1% using the following formula:

$$n_{QW} \approx \frac{2}{\pi} \tan^{-1}\left[n_{\max}\left(1 + 0.6^{n_{\max}+1}\right)\right]. \tag{A1.18}$$

A1.2 ELEMENTS OF SOLID-STATE PHYSICS

A1.2.1 Electrons in Crystals and Energy Bands

Electrons in crystals experience a periodic potential originating from the regularly spaced wells at the lattice ions. Figure A1.5 gives a schematic picture along one dimension of such a lattice. As predicted in Chapter 1, when N_A atoms are coupled in such a manner, each atomic energy level of the constituent atoms splits into a *band* of N_A discrete levels. However, this splitting is only significant for the uppermost energy levels where the two atoms interact.

There are several approaches that have been applied to solve this problem. The Kronig–Penney model approximates the actual periodic potential of Fig. A1.5 by a square wave potential, then uses the single rectangular well solution above as a starting point. However, the result is a complex transcendental equation that must be solved numerically. A second approach, which provides better closed-form analytic solutions, is the coupled-mode approach of Feynman et al. [1]. For accuracy some fairly complex functions need to be evaluated, but by leaving them in general form, we can still get a good picture of the nature of the solutions.

The first step is to go back to Schrödinger's equation and consider a possible general solution for a perturbed system, such as the atom that has been placed into a crystal. The isolated atom had a set of orthonormal wavefunction solutions just as we obtained for the rectangular potential well. When we perturb the original

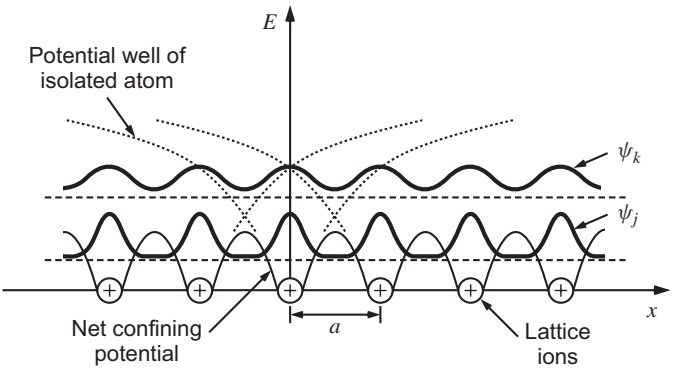


FIGURE A1.5: Schematic of net potential variation along a one-dimensional crystal lattice. Wavefunctions of nominally bound (ψ_j) and free (ψ_k) electron states are illustrated.

potential, a new set of orthonormal wavefunctions will exist. But now it may be impossible to solve Schrödinger's equation. It is common to use a superposition of the original set of orthonormal functions to express the new solutions. We shall use this kind of "normal mode expansion" later when we discuss optical solutions.

In the present case we let

$$\Psi = \sum_j w_j(t)\psi_j(r), \tag{A1.19}$$

plug into Eq. (A1.2), multiply by ψ_k^*, and integrate. Then, we have

$$\sum_j w_j(t) \int \psi_k^* H \psi_j \, dV = i\hbar \sum_j \frac{dw_j}{dt} \int \psi_k^* \psi_j \, dV, \tag{A1.20}$$

where the Hamiltonian, $H = [(-\hbar^2/2m)\nabla^2 + V]$, in which V includes the perturbation. Since the original basis functions are orthonormal, the last integral is zero[1] unless $j = k$. Using the shorthand notation,

$$H_{kj} \equiv \int \psi_k^* H \psi_j \, dV, \tag{A1.21}$$

we finally have

$$i\hbar \frac{dw_k}{dt} = \sum_j H_{kj} w_j. \tag{A1.22}$$

This is the desired coupled-mode equation, which is independent of the spatial variables. It illustrates that the probability density will flow back and forth among the various original states as a function of time to form the new states. Note that with $k = j$ in Eq. (A1.21) we have the equation to determine the expected value of energy for that original wavefunction—the eigenvalue E that we have been evaluating previously. Thus, the diagonal terms in the H_{kj} matrix are these energy eigenvalues for the respective unperturbed states. The off-diagonal terms represent the coupling strength between the various states. They determine the magnitude of the energy splitting experienced by some original state. It is also important to realize that for most of what we are doing here we do not have to know the actual form of the wavefunctions or even the magnitudes of the matrix elements, H_{kj}. Experimental measurements are usually used to determine the actual values.

Our first example is that of coupling just two identical atoms together. For sufficiently weak coupling, we can approximate the effect on a particular state by using only the basis function for that state from each atom in the summation. (Clearly, for vanishingly small coupling, these give the exact solution.) Then, Eq. (A1.22)

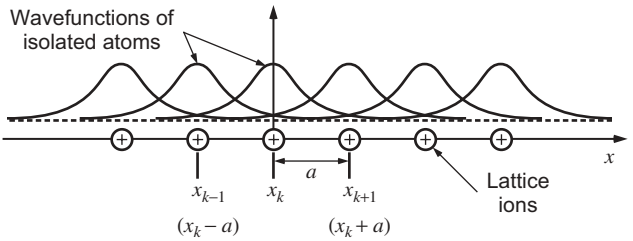


FIGURE A1.6: One-dimensional lattice of coupled atoms to derive energy bands using the coupled-mode approach.

can be expanded into two coupled-mode equations:[1]

$$\begin{aligned} i\hbar\frac{dw_1}{dt} &= H_{11}w_1 + H_{12}w_2, \\ i\hbar\frac{dw_2}{dt} &= H_{21}w_1 + H_{22}w_2. \end{aligned} \tag{A1.23}$$

Letting the energy of the state in question $H_{11} = H_{22} = E_0$, the coupling energy $H_{12} = H_{21} = \Delta E$, and then, assuming solutions $w_j(t) = C_j \exp(-iEt/\hbar)$ and plugging into Eq. (A1.23), we obtain a characteristic equation from which we must have

$$E = E_0 \pm \Delta E. \tag{A1.24}$$

Thus, the original energy level at E_0 for the isolated atom has split into two levels spaced equally on either side by the magnitude of the off-diagonal matrix element, ΔE. This same process for N_A atoms leads to N_A levels spaced symmetrically about the original level.

Now we are ready to illustrate how *energy bands* are formed when a large number of atoms are coupled together in a crystal. First we consider a simple one-dimensional crystal. Figure A1.6 illustrates a row of atoms spaced by a distance a, similar to the situation of Fig. A1.5.

Our first approximation will be to neglect the perturbation from all atoms except nearest neighbors. Then we can consider a general atom, the kth atom, which can represent every atom in this long chain. From Eq. (A1.22), taking $H_{11} = H_{kk} = E_1$ and $H_{12} = H_{kk\pm1} = \Delta E$,

$$i\hbar\,\frac{dw_k}{dt} = \Delta E\, w_{k-1} + E_1 w_k + \Delta E\, w_{k+1}. \tag{A1.25}$$

Again letting $w_j(t) = C_j \exp(-iEt/\hbar)$ and plugging into Eq. (A1.25), we obtain a characteristic expression,

$$EC_k = E_1 C_k + \Delta E[C_{k-1} - C_{k+1}]. \tag{A1.26}$$

[1]Actually, in some important cases of interest, ψ_k and ψ_j may include basis functions of laterally displaced atoms to better approximate the perturbed solution. Thus, for some terms in the summation the integral is only small rather than identically zero.

However, since the subscript k corresponds to the general lattice location, x_k, and the neighbors are at $x_k \pm a$, we can rewrite Eq. (A1.26) letting $C_k \Rightarrow C(x_k)$ and $C_{k\pm1} \Rightarrow C(x_k \pm a)$. Then, we have a difference equation in terms of the spatial variable x. This is solved by letting $C(x_k) = K \exp(ikx_k)$. Finally plugging in this assumed solution,

$$Ee^{ikx_k} = E_1 e^{ikx_k} + \Delta E[e^{ik(x_k - a)} + e^{ik(x_k + a)}],$$

or,

$$E = E_1 + 2\Delta E \cos ka. \tag{A1.27}$$

Equation (A1.27) indicates that in this infinite one-dimensional crystal a continuum of energy values between $E = E_1 \pm 2\Delta E$ is allowed. This is the familiar energy band that solid-state and semiconductor engineers are always referring to. (A later section of this appendix will remind us that for finite crystals, the discrete levels in any real situation really are very closely spaced.)

This same development of bands happens for all of the higher-lying energy levels when atoms are bonded together to form crystals. Thus, the next higher-lying band at energy E_2 also splits into a band due to nearest-neighbor coupling energy $\Delta E'$. Therefore, it provides a new band with $E' = E_2 + 2\Delta E' \cos ka$, where in direct bandgap semiconductors, the sign of $\Delta E'$ is reversed. Also, the overlap of wavefunctions is larger for the higher lying energy levels. Thus, according to Eq. (A1.21), the coupling energy is larger, and the bands become wider. Figure A1.7 illustrates these two bands. As indicated, one period of the plot is sometimes referred to as a Brillouin zone. Since the curves repeat themselves for larger k-values, we usually need concern ourselves only with the first Brillouin zone.

In semiconductors all states of all bands up to the valence band are full, and in the next higher-lying band, called the conduction band, they are empty at $T = 0\,\mathrm{K}$. We could imagine that Fig. A1.7 represents the conduction and valence bands of a direct bandgap semiconductor such as GaAs or InP. The potentials affecting electrons in such semiconductors are a little more complicated than described by this simple example, in which only nearest-neighbor interactions are considered. So, the E–k plots are not perfect sine waves. Also, in these materials the valence band actually divides into two bands called the *light-hole* and *heavy-hole* bands. These originate because of the asymmetric wavefunctions involved, and the difference in overlap that can occur for different relative orientations when Eq. (A1.21) is evaluated.

For a real three-dimensional crystal with lattice constants a, b, and c, the same procedures can be carried out using a three-dimensional version of Eq. (A1.22) with coupling coefficients $\Delta E_x, \Delta E_y$, and ΔE_z, and a three-dimensional envelope wavefunction,

$$w(x, y, z, t) = e^{-iEt/\hbar} e^{i(k_x x + k_y y + k_z z)}, \tag{A1.28}$$

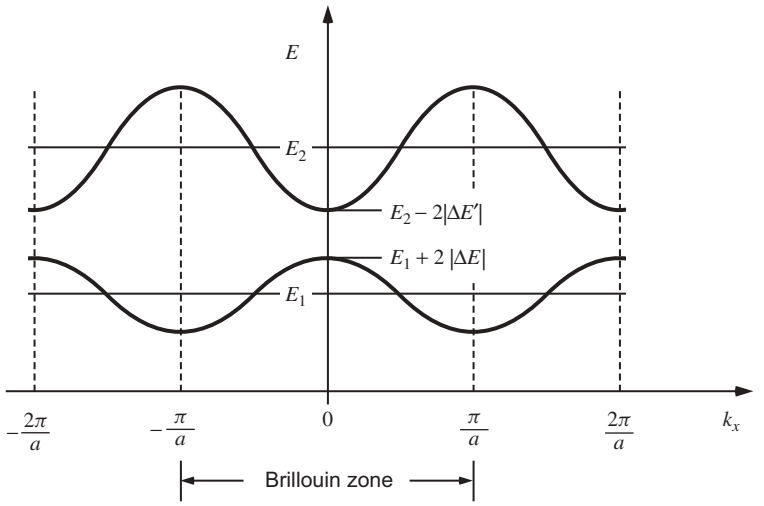


FIGURE A1.7: Energy bands created by a one-dimensional chain of coupled atoms. Two bands resulting from two original discrete states are shown.

to obtain

$$E = E_1 + 2\Delta E_x \cos k_x a + 2\Delta E_y \cos k_y b + 2\Delta E_z \cos k_z c. \tag{A1.29}$$

The real situation is still more complex than the first-order calculation resulting in Eq. (A1.29). Figure A1.8 illustrates the actual band structure for both GaAs and InP along the ⟨100⟩ and ⟨111⟩ directions.

A1.2.2 Effective Mass

Near the top of the valence band and near the bottom of the conduction band it is sometimes possible to approximate the shape of these E–k extrema by parabolas. In these cases the concept of an effective mass is useful, and simple expressions for the density of states are possible. However, the concept of an effective mass has also been extended to limited regions within nonparabolic bands where the parabolic approximation is still valid.

To determine an expression for the effective mass and show that the parabolic band is desired, we follow a semiclassical approach in which we calculate the acceleration of an electron in a solid under the force of an applied electric field. The force $q\mathscr{E}$ on a particle may be classically expressed as the time rate of change of its momentum, p. Quantum mechanically $p = \hbar k$. Thus, the force is

$$F = q\mathscr{E} = \hbar \frac{dk}{dt}. \tag{A1.30}$$

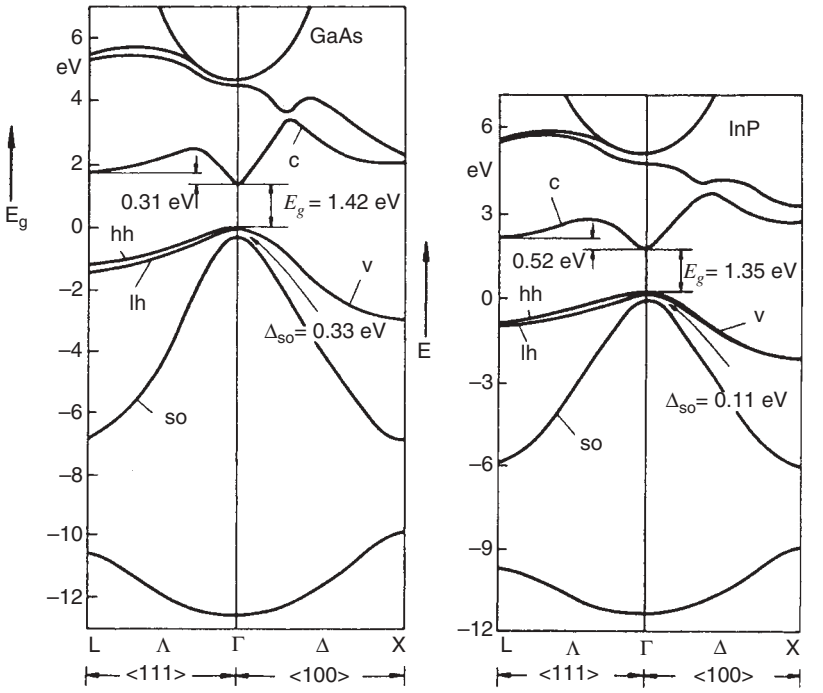


FIGURE A1.8: Band structure of GaAs and InP. The conduction band as well as the heavy-hole, light-hole, and split-off valence bands are labeled by, *c, hh, lh*, and *so*, respectively. (*Reprinted, by permission, from K. J. Ebeling, Integrated Optoelectronics, Springer-Verlag, 1993.*)

The velocity of particles is defined by their group velocity, $v_g = d\omega/dk = (1/\hbar)\, dE/dk$, which shows the proportionality of velocity to the slope of the E–k characteristic. Since the acceleration, *acc.*, is the time derivative of the velocity, we can write

$$acc. = \frac{dv_g}{dt} = \frac{dv_g}{dk}\frac{dk}{dt} = \frac{1}{\hbar}\frac{d^2E}{dk^2}\frac{dk}{dt}. \tag{A1.31}$$

Dividing Eq. (A1.30) by (A1.31), and defining an effective mass, $m^* = F/acc.$, we obtain

$$m^* = \frac{\hbar^2}{d^2E/dk^2}. \tag{A1.32}$$

Thus, for parabolic bands, as observed for uniform potentials (e.g., Eq. (A1.7)), the electron will move much like a free particle, but with an effective mass, m^*, related to the curvature of the band. For nonparabolic bands, m^* is not constant and the local slope and curvature of the E–k relationship must be used to obtain the velocity and acceleration of a particle with energy E.

A1.2.3 Density of States Using a Free-Electron (Effective Mass) Theory

We just learned above that an electron in a crystal can behave much like a free electron moving in a region of uniform potential if it is at a point on the E–k diagram that is parabolic. This is a remarkable result, since we know that the potential within a crystal is very nonuniform. Nevertheless, this revelation allows us to treat some very complex problems. For example, if we consider a crystal of finite dimensions, d_x, d_y, d_z, we can more or less ignore the crystal lattice potential which is periodic on the scale of the lattice constant a, provided that $d_j \gg a$ (Fig. A1.9). But, we must use a different effective mass as determined by the curvature of the E–k diagram.

By considering electron energies near band extrema, where the E–k curve tends to be parabolic, we can now consider reusing some of the same physics that we developed in Section A1.1.2 for electrons in very simple potential wells that had uniform potential regions. That is, we can now find the states in finite pieces of crystal or pieces with potential wells created by double heterostructures as described in Chapter 1. The simplest case is when the potential barriers are large so that we can assume that the bound wavefunctions go to zero at the boundaries. Then, from Eqs. (A1.13) and (A1.14), we have

$$E = \frac{\hbar^2 k^2}{2m^*} = \frac{\hbar^2}{2m^*}[k_x^2 + k_y^2 + k_z^2], \tag{A1.33}$$

where we have included all three dimensions for completeness and assumed that the effective mass is the same in all directions. Applying the boundary conditions for a large barrier, $k_j d_j = n_j \pi$,

$$E = \frac{\hbar^2 \pi^2}{2m^*}\left[\left(\frac{n_x}{d_x}\right)^2 + \left(\frac{n_y}{d_y}\right)^2 + \left(\frac{n_z}{d_z}\right)^2\right]. \tag{A1.34}$$

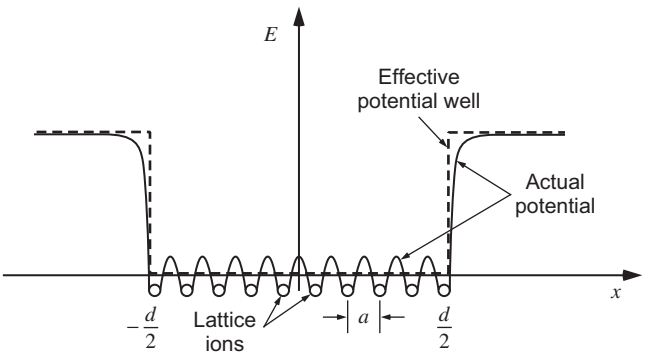


FIGURE A1.9: Potential plot for crystal (or quantum-well) of thickness d. The dashed well is an approximation to the actual potential.

From Eq. (A1.34) we note that we have an energy state for each (n_x, n_y, n_z) set of quantum numbers. To determine the size of the energy spacing between states, we can evaluate the coefficient, $\hbar^2\pi^2/2m_0 = 376$ meV-nm^2, where we have used the free electron mass rather than the effective mass. In GaAs the electron mass, $m^* = 0.067m_0$, so we should use $15(376) = 5640$ meV-nm^2 for the coefficient in the conduction band. From this we can see that the energy separation between states is quite small. For example, for a cube with $d_j = 1$ μm, the difference in energy between the first two states, $E(211) - E(111) \approx 17 \times 10^{-3}$ meV. Since $kT \approx 26$ meV at room temperature, we see that this energy difference is less than one thousandth of a kT. On the other hand, for a cube with dimensions $d_j \sim 10$ nm, this energy difference is ~ 170 meV, or more than $6kT$ at room temperature.

From the above, we conclude that for dimensions $d_j \gtrsim 1$ μm, quantum effects are not going to be very noticeable at room temperature, and the E–k diagram can be treated as describing a continuum of states. We shall refer to this as the *bulk* regime. On the other hand for $d_j < 100$ nm, the discreteness of the energy levels indicated in Eq. (A1.34) must be considered. We shall refer to this as the *quantum-confined* regime.

Even though the states may be very closely spaced in the so-called bulk regime, we still need to be able to count them to determine the carrier density and the energy to which they would have to be filled for a given carrier density. In the smaller structures, we again need an effective method of counting states. The method commonly used is to define a *density of states*, ρ, which when integrated over some range gives the number of states in that range. The density of states can be expressed in terms of a number of variables (e.g., E, p, or k) in a number of different coordinate systems. If N_s is the number of states up to some point, we can generally state that

$$N_s(u) = V \int_0^u \rho(u)\, du, \tag{A1.35}$$

where u is the desired variable and V is the volume. Once we have this definition, we can then state that

$$\rho(u)\, du = \frac{1}{V}\, dN_s(u). \tag{A1.36}$$

It should be realized that $\rho(u)\, du$ can be defined and used regardless of the size regime in which we find ourselves. For the smaller structures, we find that it contains discontinuities and even impulse functions, but it is still a good function.

To determine $\rho(u)\, du$ for the various cases of interest, we follow a standard procedure: (1) determine the number of states by calculating the volume in state or n-space, $N_s(n)$; (2) substitute for the desired variable, $n = f(u)$, which gives $N_s(u)$; and (3) apply Eq. (A1.36) to get the desired $\rho(u)\, du$. A few examples are useful for future reference.

For the first example, we consider *bulk* dimensions, a *spherical* coordinate system, and *energy* as the variable. Spherical coordinates imply that we are considering

a spherical state space. Equation (A1.34) is written in rectangular coordinates, but as stated after it, each set of quantum numbers, or each volume element in n-space, represents a state that can be occupied by an electron. Figure A1.10 illustrates this n-space. The first step is to calculate the volume $N_s(n)$

$$N_s(n) = \tfrac{4}{3}\pi n^3 \cdot 2 \cdot \tfrac{1}{8}. \tag{A1.37}$$

The first factor just gives the standard expression for volume. However, we must multiply by 2, since two states actually exist at each allowed energy because of spin degeneracy. And only positive quantum numbers are allowed, so we have the factor of $\frac{1}{8}$. Now, for the second step, we use Eq. (A1.34) in spherical coordinates, (identical to Eq. (A1.14)), solve for n in terms of E, and plug back into Eq. (A1.37):

$$N_s(E) = \frac{\pi}{3}\left(\frac{2m^* E d^2}{\hbar^2 \pi^2}\right)^{3/2}. \tag{A1.38}$$

For the third step, we now apply Eq. (A1.36), and use $V = d^3$, without loss of generality

$$\rho(E)\,dE = \frac{1}{2\pi^2}\left[\frac{2m^*}{\hbar^2}\right]^{3/2} E^{1/2}\,dE. \tag{A1.39}$$

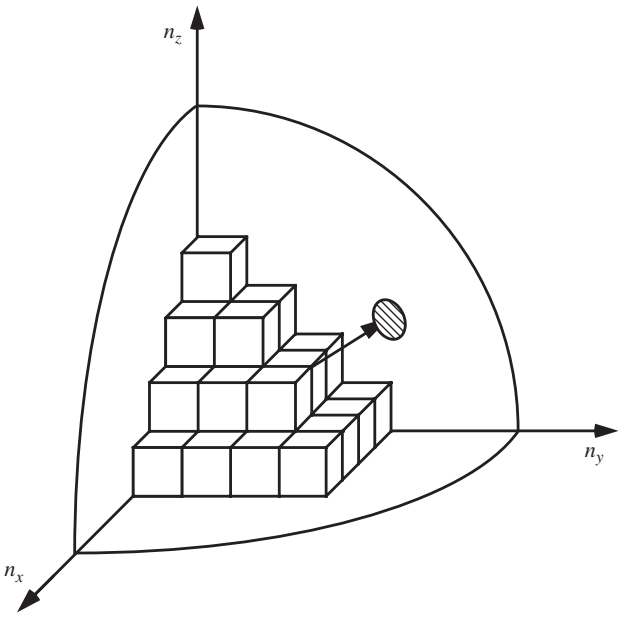


FIGURE A1.10: State space or n-space in spherical coordinates. Each block corresponds to a particular state and has unit dimensions.

Equation (A1.39) is our final result. It will be of much use in calculating carrier densities, gain, and other quantities associated with bulk active regions in lasers.

For the second example, we consider *bulk* dimensions, *rectangular* coordinates, and *momentum* as the variable. Then, the volume in n-space is, $N_s(n) = 2n_x n_y n_z$. The momentum in each direction, j, is

$$p_j = \hbar k_j = \frac{\hbar \pi n_j}{d_j} = \frac{h n_j}{2 d_j}. \tag{A1.40}$$

Solving for n_x, n_y, and n_z, and plugging into $N_s(n)$, we get

$$N_s(p_x, p_y, p_z) = 2\left(\frac{2}{h}\right)^3 p_x p_y p_z (d_x\, d_y\, d_z). \tag{A1.41}$$

Applying Eq. (A1.36) to (A1.41) gives the desired density of states,

$$\rho(p_x, p_y, p_z) dp_x\, dp_y\, dp_z = 2\left(\frac{2}{h}\right)^3 dp_x\, dp_y\, dp_z. \tag{A1.42}$$

The density of all states with a given momentum, $\rho(p)$, can be obtained from this result by setting $dp_x dp_y dp_z = 4\pi p^2 dp/8$. The factor of 8 is required because (A1.42) defines the density of *standing wave* states, $\rho_{SW}(p_x, p_y, p_z)$, which do not distinguish between positive and negative values of momentum. Hence, the density is limited to the first quadrant. We can also define a density of *plane wave* states, $\rho_{PW}(p_x, p_y, p_z)$, which can travel in any direction. Using periodic boundary conditions, we have $k_j = 2\pi n_j / d_j$ instead of $k_j = \pi n_j / d_j$, but now we consider both positive and negative values of n_j as unique states. In three dimensions then, $\rho_{PW}(p_x, p_y, p_z) = \rho_{SW}(p_x, p_y, p_z)/2^3$, and is distributed over all quadrants of momentum space. So for $\rho_{PW}(p_x, p_y, p_z)$, we can set $dp_x\, dp_y\, dp_z = 4\pi p^2\, dp$. In either case, we obtain $\rho_{PW}(p) = \rho_{SW}(p) = 8\pi p^2/h^3$, from which we also obtain $\rho(k) = (k/\pi)^2$ (since $\rho(k)\, dk = \rho(p)\, dp$).

For the third example, we consider a *quantum well* (small dimension in one direction), *cylindrical* coordinates, and *energy* as the variable. We shall let d_x be the small dimension. As always the energies are given by Eq. (A1.34), but we need to develop a density of states for the y–z plane, which will be summed for each n_x. Figure A1.11a gives a plot of the energy, E, relative to the $k_x - k_z$ plane for a quantum-well region using Eq. (A1.34). The lowest-lying states for $n_x = 1$ and 2 are labeled by the quantum numbers (n_x, n_y, n_z). Since d_x is small, there are no states near $k = 0$. Figure A1.11b replots the energy versus k_z (or k_y) in a two-dimensional graph for clarity. In part (a) only positive k_x and k_z are shown.

Now to determine the density of states for this quantum well, we start with the disk of Fig. A1.12 and determine $N_s(n)$. The volume of the unit height disk in the first quadrant, multiplied by 2 for spin, is

$$N_s(n_{yz}) = \frac{\pi}{2}(n_y^2 + n_z^2). \tag{A1.43}$$

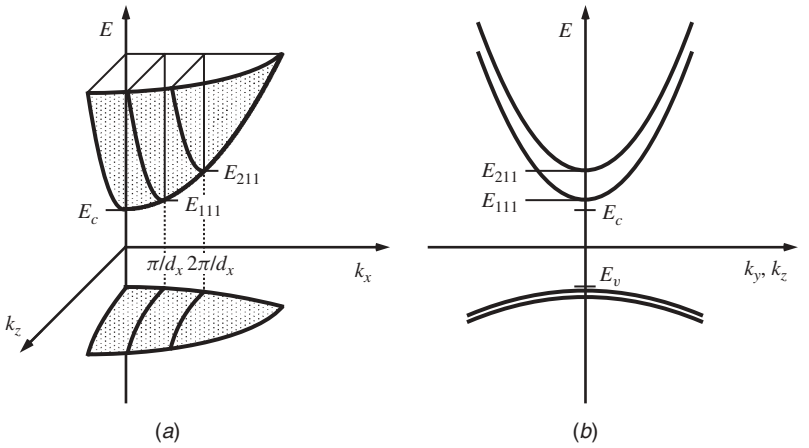


FIGURE A1.11: (a) Three-dimensional E–k plot showing discrete jumps in k_x due to small d_x. (b) Projection perpendicular to the k_z axis.

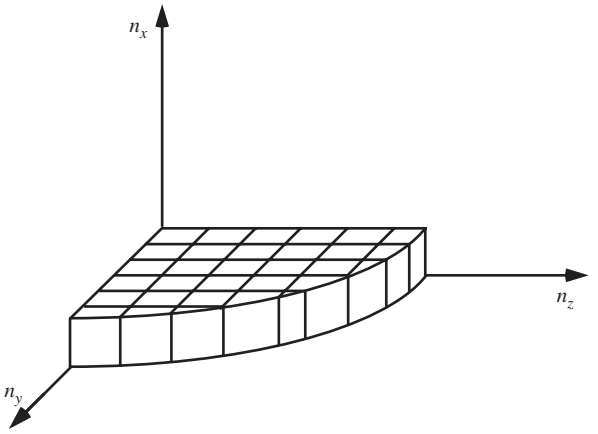


FIGURE A1.12: Two-dimensional state space that occurs for each n_x in a quantum well.

Using the y and z terms from Eq. (A1.34) for E_{yz} and $(n_y^2 + n_z^2)$, letting $d^2 = (d_y^2 + d_z^2)$, and assuming that the effective masses are the same in y and z,

$$N_s(E_{yz}) = \frac{m^* d^2}{\pi \hbar^2} E_{yz}. \tag{A1.44}$$

Again, we apply Eq. (A1.36) and recognize that this density of states is for $n_x = 1$. Thus, generally

$$\rho(E) = \frac{1}{d_x} \sum_{n_x} \frac{m^*}{\pi \hbar^2} \mathscr{H}(E - E_{n_x}), \quad \text{(for QW)} \tag{A1.45}$$

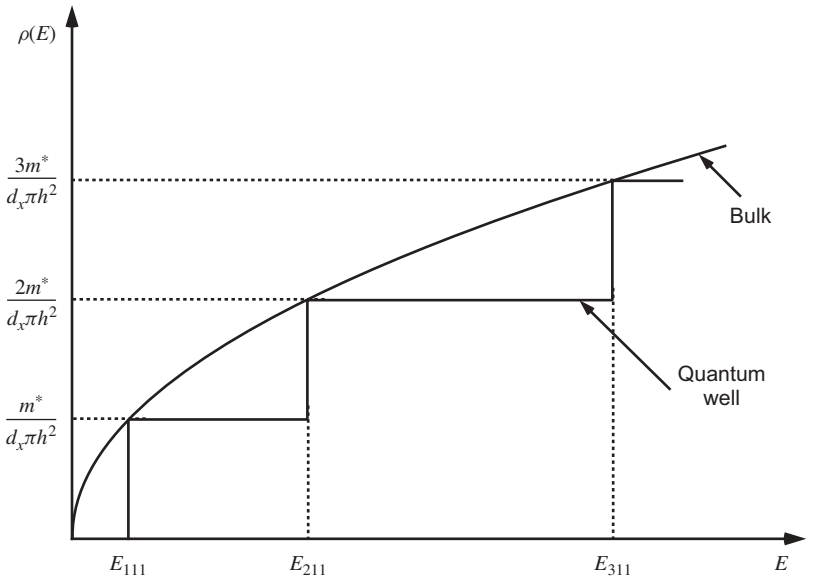


FIGURE A1.13: Density of states for an infinite-barrier quantum well and bulk material. If the barrier is not infinite, the quantum-well energies decrease slightly. If desired, the density of state plateaus can be decreased by using an effective $d_x = d_x^*$ (a different one for each state) so that the extrema continue to intersect the bulk characteristic.

where n_x and E are related by Eq. (A1.14) with $l = d_x$, and $\mathscr{H}(E - E_{n_x})$ is the Heaviside unit step function.

Figure A1.13 compares the densities of states for the bulk and quantum-well active regions, Eqs. (A1.39) and (A1.45), respectively. As can be seen the bulk curve forms an envelope for the steps of the quantum-well case, which correspond to the energies where the quantum numbers are $(n_x, 1, 1)$.

REFERENCES

1. Feynman RP, Leighton RB, Sands ML. *The Feynman lectures on physics.* New York: Addison-Wesley; 1964.

READING LIST

Solymar L, Walsh D. *Lectures on the electrical properties of materials*. New York: Oxford University Press; 1984. Chapters 3 to 8.

Kittel C. *Introduction to solid state physics*. New York: Wiley; 1986. Chapters 7 and 8.

Kroemer H. *Quantum mechanics*. Englewood Cliffs, NJ: Prentice Hall; 1994. Chapter 7.

APPENDIX TWO

Relationships between Fermi Energy and Carrier Density and Leakage

A2.1 GENERAL RELATIONSHIPS

In Appendix 1 the groundwork of energy bands and densities of states within these bands was outlined. Here we attempt to understand how carriers fill these states in semiconductors. This will allow us to relate the carrier density to the ranges of energies that must be occupied. They key missing element is the state occupation probability, *f(E)*, at energy E. For a large density of states, this function is equivalent to the fraction of states occupied at energy E. For particles in solids the relevant function is the Fermi–Dirac distribution,

$$f(E) = \frac{1}{e^{(E-E_F)/kT} + 1}, \tag{A2.1}$$

where the energy E_F is called the Fermi level. This probability distribution is one-half at $E = E_F$ and closely approaches unity for E a few kT below E_F, and zero for E a few kT above E_F. In fact, for E a few kT on either side of E_F, the asymptotic approach to these limits is exponential. These regions are sometimes referred to as the Boltzmann tails, since they imitate the classical Maxwell–Boltzmann distribution. At $T = 0$ K, *f(E)* is essentially a step function, stepping from unity to zero at $E = E_F$.

Diode Lasers and Photonic Integrated Circuits, Second Edition.
Larry A. Coldren, Scott W. Corzine, and Milan L. Mašanović.
© 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

Now, we can write the carrier density, N, as the integral over energy of the density of filled states in the conduction band,

$$N = \int \rho_c(E) f(E)\, dE, \tag{A2.2}$$

and similarly, we can write the hole density, P, as the integral of the density of unfilled states in the valence band,

$$P = \int \rho_v(E)[1 - f(E)]\, dE, \tag{A2.3}$$

where $\rho_c(E)$ and $\rho_v(E)$ are the density of states (per unit volume per unit energy) in the conduction and valence bands, respectively. Figure A2.1 gives schematics of the density of filled and unfilled states versus electron energy in the conduction and valence bands, respectively, for a bulk semiconductor under nonequilibrium conditions at $T = 0$ K and 300 K. Figure A2.2 gives analogous plots for a quantum well.

In both cases we have only shown the heavy-hole band in the valence band. In reality, the light-hole band also is a significant part of $\rho_v(E)$, and some of the holes occupy states there. The light-hole band is a little less important in quantum wells because of the separation of the lighthole band from the heavy-hole band at the finite k_x where the lowest energy is found.

Equations (A2.2) and (A2.3) give a relationship between the carrier densities and the Fermi level, E_F, for a known density of states. As a very simple example consider a quantum well at $T = 0$ K with an electron density of N. In this case the

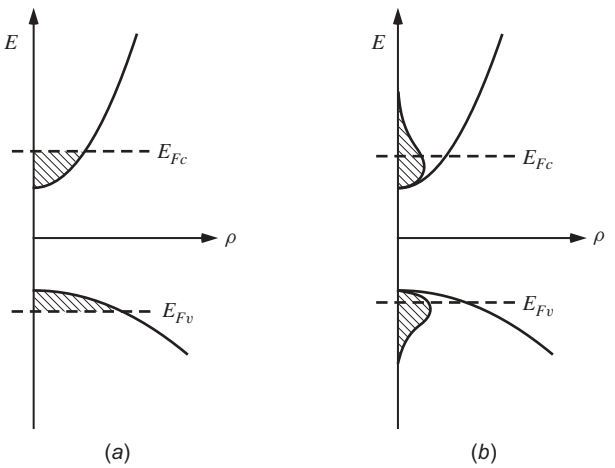


FIGURE A2.1: Density of filled states in the conduction and valence bands of a bulk semiconductor under nonequilibrium conditions at (a) $T = 0$ K and (b) $T = 300$ K. Under nonequilibrium conditions, such as when a current is flowing, separate Fermi functions and quasi-Fermi levels, E_{Fc} and E_{Fv}, are used for the electrons and holes, respectively.

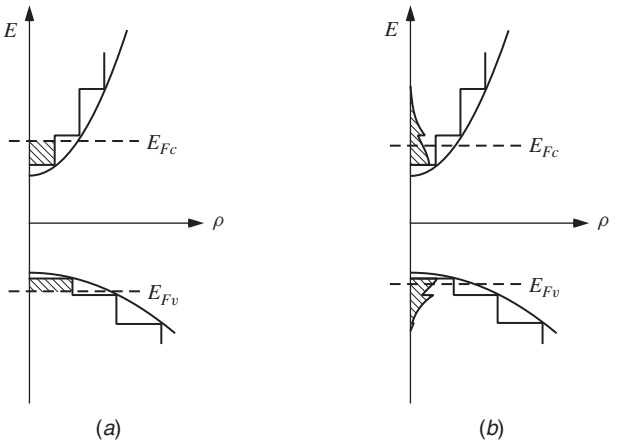


FIGURE A2.2: Density of filled states in the conduction and valence bands of a quantum well under nonequilibrium conditions at (a) $T = 0$ K and (b) $T = 300$ K.

step-like Fermi function merely sets the limits of integration from the first allowed state, E_{111} to E_F as shown in Fig. A2.2(a). From Eqs. (A1.45) and (A2.2), and assuming we do not approach $n_x = 2$, we have

$$N = \frac{m^*}{\pi\hbar^2 d_x}(E_F - E_{111}). \tag{A2.4}$$

Thus, we can solve for E_F as

$$E_F = E_{111} + \frac{\pi\hbar^2 d_x}{m^*}N. \quad (\text{QW}; n_x = 1;\ T = 0\ \text{K}) \tag{A2.5}$$

At finite temperatures, the Fermi function is no longer a simple step function, and we must use the general form Equation (A2.1) inside the integral. This makes the integration more complex, but using a quantum-well density of states function, the integral still has a closed-form solution. Repeating the calculation at finite temperature, we have

$$N = \frac{1}{d_x}\sum_{n_x}\frac{m^*}{\pi\hbar^2}\int_{E_{n_x 11}}^{\infty}\frac{dE}{e^{(E-E_F)/kT}+1}, \tag{A2.6}$$

where $\mathscr{H}(E - E_{n_x})$ in Eq. (A1.45) has shifted the lower integration limit. Making the substitution $u = \exp[-(E - E_F)/kT]$, allows the integration to be readily performed. Then, for the situation of Fig. A2.2b,

$$N = \frac{kTm^*}{\pi\hbar^2 d_x}\sum_{n_x}\ln\left[1 + e^{(E_F - E_{n_x 11})/kT}\right]. \quad (\text{QW}) \tag{A2.7}$$

For $n_x = 1$ and $(E_F - E_{111}) \gg kT$, note that Eq. (A2.7) reduces to Eq. (A2.4).

A2.2 APPROXIMATIONS FOR BULK MATERIALS

Unfortunately, the finite-temperature calculation for carrier density is not as simple using a bulk-like density of states function, as in Fig. A2.1b. The general form for the carrier density in bulk material is given by

$$N = \frac{1}{2\pi^2}\left(\frac{2m^*}{\hbar^2}\right)^{3/2}\int_{E_c}^{\infty}\frac{\sqrt{E-E_c}}{1+e^{(E-E_F)/kT}}\,dE, \tag{A2.8}$$

where E_c is the conduction band edge. Defining $v = (E_F - E_c)/kT$ and $y = (E - E_c)/kT$, we can write the integral in a more normalized form:

$$N = N_c\frac{2}{\sqrt{\pi}}\int_0^{\infty}\frac{\sqrt{y}}{1+e^{y-v}}\,dy \equiv N_c\frac{2}{\sqrt{\pi}}F_{1/2}(v), \tag{A2.9}$$

$$N_c \equiv 2\left(\frac{m^*kT}{2\pi\hbar^2}\right)^{3/2} = 2.51\times10^{19}\times\left[\frac{m^*}{m_0}\right]^{3/2}\times\left[\frac{T}{300\,K}\right]^{3/2}\ \text{cm}^{-3}, \tag{A2.10}$$

where $F_{1/2}(v)$ is known as the Fermi–Dirac integral of order 1/2 (referring to the $y^{1/2}$ in the numerator), and N_c is the "effective" density of states in the conduction band. For $v \ll 0$ (in other words, for $E_F \ll E_c$), the 1 in the denominator can be neglected reducing the Fermi function to the *Boltzmann* exponential limit, allowing the integral to be evaluated in closed-form. The result reduces to

$$N \approx N_c e^{(E_F-E_c)/kT}. \quad \text{(bulk, Boltzmann limit)} \tag{A2.11}$$

For larger v, other approximation techniques must be used. Tables A2.1 and A2.2 summarize various approximations to the Fermi–Dirac integral (and their range of validity) which have been formulated over the years (see Blakemore [1] for an overview). The first table finds N in terms of v, whereas the second table finds v in terms of N.

The error associated with the various approximations listed in Table A2.1 are plotted in Fig. A2.3. The Boltzmann limit discussed earlier is valid for large negative v, but quickly begins to overestimate the carrier density as v approaches zero (that is, as E_F approaches the band edge). At the other extreme, the Sommerfeld expansion (which essentially determines the zero-temperature carrier density with correction terms added to account for the finite-temperature smoothing of the Fermi function) can be used to determine the Fermi–Dirac integral for large positive v. The first-order terms of the Boltzmann and Sommerfeld limits are plotted in (a). As can be observed, the two limits fail to cover the range $-2 < v < 5$. Unger [2] introduced a function z, which led him to a useful expansion of the Fermi–Dirac integral for small v. The fractional error of the Unger expansion is plotted in (b). The multiple curves are plotted using the second-order expansion coefficient as a curve-fitting parameter. Increasing a_2 from its true value of ~0.14645 to 0.15 to 0.1537 reveals the trend toward reducing the initial hump in the curve

TABLE A2.1: Summary of Approximations for Carrier Density $N = N_c \frac{2}{\sqrt{\pi}} F_{1/2}(v)$

Formal Name	Mathematical Expression		Range Over Which Fractional Error* < 1%
Fermi–Dirac Integral	$F_{1/2}(v) = \int_0^\infty \frac{\sqrt{y}}{1+e^{y-v}}\,dy$ $v \equiv (E_F - E_c)/kT$ $y \equiv (E - E_c)/kT$		exact definition
Boltzmann approximation	$F_{1/2}(v) \approx \frac{\sqrt{\pi}}{2} e^v$		$v < -3.5$
Unger approximation	$F_{1/2}(v) \approx \frac{\sqrt{\pi}}{2} z[a_1 + a_2 z + \cdots]$ $a_1 = 1$ $a_2 = \frac{1}{2}\left(1 - \frac{1}{\sqrt{2}}\right) \approx 0.146\ 45$ $z \equiv \ln(1 + e^v)$	1st order ($a_2 = 0$) 2nd order ($a_2 = 0.146\ 45$) 2nd order ($a_2 = 0.15$)	$v < -2.7$ $v < 1.7$ $v < 7.6$
Sommerfeld approximation	$F_{1/2}(v) \approx \frac{2}{3} v^{3/2}[a_1 + a_2 v^{-2} + \cdots]$ $a_1 = 1$ $a_2 = \frac{\pi^2}{8} \approx 1.2337$	1st order (0) 2nd order (1.233 7) 2nd order (1.3)	$v > 11$ $v > 2.9$ $v > 1.7$
Modified Sommerfeld approximation	$F_{1/2}(v) \approx \frac{2}{3} z^{3/2}\left[1 + \frac{\pi^2}{8} z^{-2} + \cdots\right]$ $z \equiv \ln(1 + e^v)$	1st order (0) 2nd order (1.233 7)	$v > 11$ $v > 3$
Global approximation (one of many)	$F_{1/2}(v) \approx [(1.3 + 0.3x)F_L^p + (0.76 + 0.24x)F_H^p]^{1/p}$ $F_L = \frac{\sqrt{\pi}}{2} z$ $F_H = \frac{2}{3} z^{3/2}$ $p = 3 + x - 0.19 e^{-(v-3.25)^2/8}$ $x \equiv \tanh(0.8(v - 2.7))$ $z \equiv \ln(1 + e^v)$	note that $z \to e^v$ $x \to -1$ $(v \ll 0)$ $z \to v$ $x \to 1$ $(v \gg 0)$	error <0.06% $-\infty < v < \infty$

*error $\equiv |F_{1/2} - F_{\text{calc}}|/F_{1/2}$

TABLE A2.2: Summary of Approximations for Fermi Level $v = f(N/N_c)$

Formal Name	Mathematical Expression		Range Over Which Abs Error* < 0.04†
Inverse Fermi–Dirac integral	$v = F_{1/2}^{-1}[F_{1/2}(v)]$		exact definition
	$= F_{1/2}^{-1}\left[\frac{\sqrt{\pi}}{2} N/N_c\right] \equiv f(r)$		
	$v \equiv (E_F - E_c)/kT, r \equiv N/N_c$		
Boltzmann approximation	$v \approx \ln r$		$v < -2.1$
Joyce–Dixon approximation	$v \approx \ln r + A_1 r + A_2 r^2 + \cdots$	1st order	$v < 2.1$
	$A_1 = 1/\sqrt{8}$	2nd order	$v < 4.2$
	$A_2 \approx -4.950\ 09 \times 10^{-3}$	3rd order	$v < 5.6$
	$A_3 \approx 1.483\ 86 \times 10^{-4}$	4th order	$v < 6.8$
	$A_4 \approx -4.425\ 63 \times 10^{-6}$		
Padé approximation	$v \approx \ln r + A_1 r$	$K_1 = 4.7$	
	$+[K_1 \ln(1 + K_2 r) - K_1 K_2 r]$	error <0.0004	$v < 9.8$
	$K_1 \approx 4.7 \leftrightarrow 5.2$	$K_1 = 4.9$	
	$K_2 = \sqrt{2\lvert A_2\rvert/K_1}$	error <0.02	$v < 18.8$
	A_1, A_2 as defined above	$K_1 = 5.2$	
		error <0.1	$v < 28$
Inverse second-order Unger approximation	$v \approx \ln\left\{\exp\left[\frac{1}{2a_2}(\sqrt{1 + 4a_2 r} - 1)\right] - 1\right\}$	$(a_2 = 0.146\ 45)$	$v < 2.8$
		$(a_2 = 0.15)$	$v < 7.4$
Inverse first-order Sommerfield approximation	$v \approx \left(\frac{3\sqrt{\pi}}{4} r\right)^{2/3}$		$v > 20$
Nilsson's global approximation	$v \approx \frac{\ln r}{1 - r^2} + \frac{f_H}{1 + (0.24 + 1.08 f_H)^{-2}}$		error < 0.006
			$-\infty < v < \infty$
	$f_H = \left(\frac{3\sqrt{\pi}}{4} r\right)^{2/3}$		

*error $\equiv |v - v_{\text{calc}}|$, † $\Delta E_F \approx \pm 1$ meV @ RT.

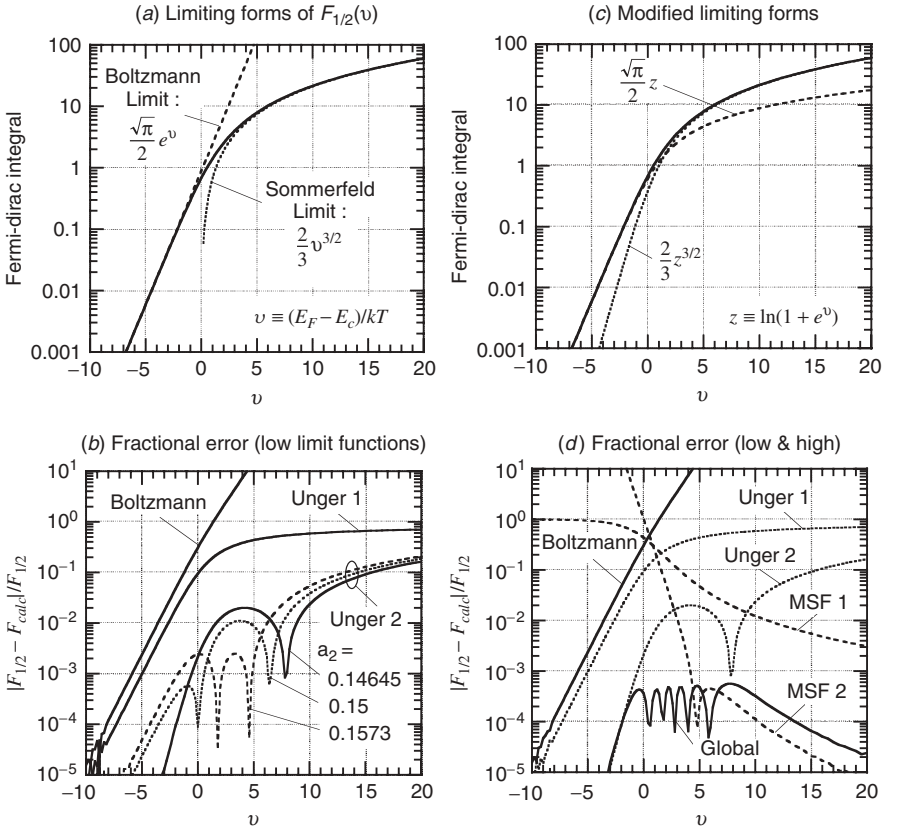


FIGURE A2.3: (a) Plot of Fermi–Dirac integral along with first-order limits for low and high υ. (b) Fractional error of Boltzmann and first- and second-order Unger approximations. (c) Modified first-order limits using the z function substitution introduced by Unger. (d) Fractional error of Boltzmann, first- and second-order Unger, first- and second-order modified Sommerfeld (MSF), and global approximations. (There is little difference between the error in the Sommerfeld and modified Sommerfeld expansions.)

at the expense of increasing the error elsewhere. Note that with $a_2 = 0.15$, the Unger expansion provides an error less than $\sim 1\%$ for υ as large as 7 (actually the initial jump dips below 1% when $a_2 = 0.151$, but there is something to be said for round numbers). What this means is that E_F can penetrate $7kT$ into the conduction band and the Unger approximation will still provide an accurate estimate of the carrier density. Higher-order terms in the Unger expansion can also be used but they do not significantly increase or extend the accuracy of the approximation.

For some applications it is desirable to know the Fermi function for values of υ even larger than the Unger approximation is capable of handling. For these situations, one could switch to the second-order Sommerfeld expansion, but it

would be advantageous to have a single expression to cover the entire range of v. One could think of constructing a global approximation by bridging the two limiting approximations. However, the Sommerfeld expansion is unfortunately invalid for $v \leq 0$. For the purpose of constructing a global approximation, we can substitute z for the two limiting forms given in (a). The resulting expressions are plotted in (c). Note that the "modified" limiting forms are better behaved than the limiting forms in (a). The global approximation given in Table A2.1 bridges the gap between these "modified" limits [3]. The overall expression for the global approximation is a bit complicated, but it does get the job done—it approximates the Fermi–Dirac integral to within 0.06% over the entire range of v, as shown in (d), with excellent convergence on either side of the "trouble area."

Table A2.1 found N in terms of v. However, oftentimes we know the carrier density, and would like to determine the Fermi level. This requires calculating the

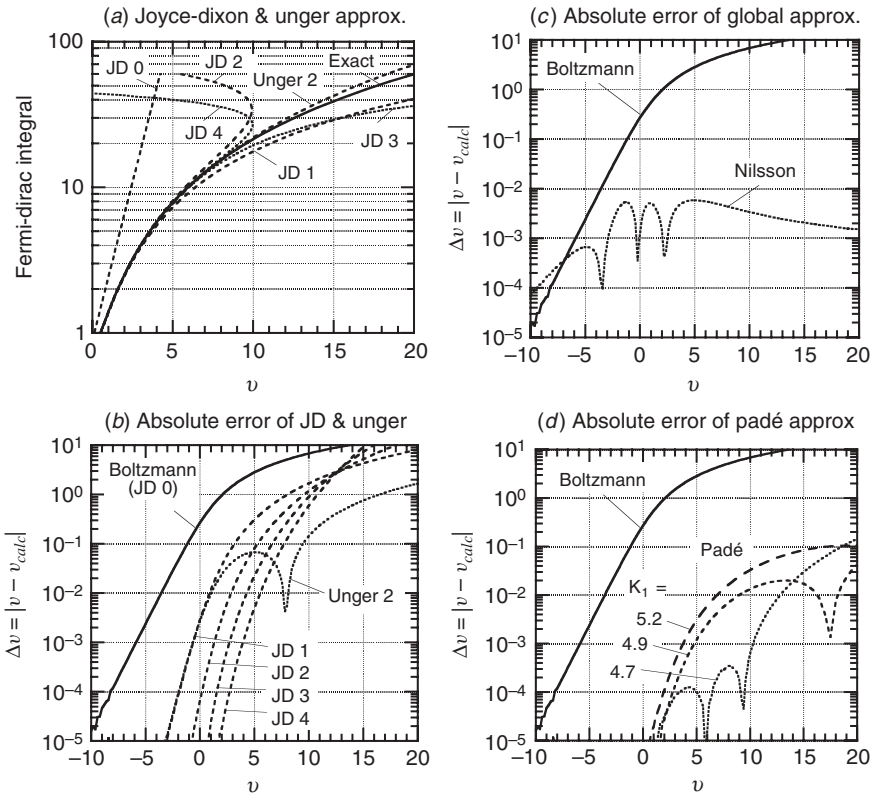


FIGURE A2.4: (a) Plot of Fermi–Dirac integral along with various orders of Joyce–Dixon (JD) approximations and the second-order Unger approximation. (b) Comparison of the absolute error of these approximations. (c) Absolute error of Nilsson's global approximation. (d) Absolute error of the Padé approximation for different values of the adjustable parameter, K_1.

inverse of the Fermi–Dirac integral. Various approximations used to achieve this are summarized in Table A2.2. Of these, the Joyce–Dixon approximation [4] is perhaps the most well known. Its popularity stems from the fact that it simply adds correction terms to the Boltzmann approximation, which is useful for the analytical description of semiconductor devices. The Unger approximation appears in this list as well. In fact, one very attractive feature of the second-order Unger approximation is that it is invertible [5]! Figure A2.4 plots the various approximations for comparison. In (a) we see that the even-order Joyce–Dixon series expansions diverge rapidly for $v > 10$, while the odd-order terms remain well-behaved. The Unger expansion actually tracks the exact solution better than any of the Joyce–Dixon expansions. However, (b) reveals that the absolute error for the second-order Unger expansion is larger than the higher-order Joyce–Dixon expansions for small v. To extend the range, Nilsson [6] constructed a global inverse approximation by attempting to bridge the inverted Boltzmann and Sommerfeld expansions. The error in his global approximation is shown in (c). While the absolute error remains low, the convergence on either side of the "trouble area" is not particularly strong.

The Padé approximation [7] shown in (d) is a variant of the Joyce–Dixon approximation, which avoids the highly divergent properties associated with the even-order Joyce–Dixon series. The Padé approximation is by far the most accurate approximation for $v < 10$ with K_1 set to 4.7. One nice feature of this approximation is that it is adjustable. For example, to extend the range one can sacrifice accuracy slightly by increasing K_1 up to a maximum of 5.2 (beyond which the error becomes very large).

At this point, we can see that the number of approximations available for the Fermi–Dirac integral allow us to determine the relationship between carrier density and Fermi level in bulk material to a very good degree of accuracy. In fact, it could be argued that the quantum-well Fermi level is more difficult to calculate than the bulk Fermi level because the summation in the expression for quantum-well carrier density Eq. (A2.7) prevents one from inverting the equation.

A2.3 CARRIER LEAKAGE OVER HETEROBARRIERS

In double heterostructures the band diagrams shown in Figs. A2.1 and A2.2 for the bulk and quantum-well active layers are not valid for energies greater than the band offsets between the active and cladding materials. That is, once carriers fill to the top of the barrier they are free to diffuse into the cladding regions. At finite temperatures shown in parts (b) of these figures, the carriers in the high-energy tails would ideally extend to energies above the barriers. Thus, in practice there is a "carrier leakage" out of the active region, which results in a leakage current as the carriers diffuse away. With increased temperatures the fraction of carriers in the high-energy tail as well as the carrier leakage current will increase.

Figure A2.5 is a plot of the carrier distribution for an electron density of 2×10^{18} cm^{-3} in a GaAs active region clad by AlGaAs for which the net effective barrier height is 0.3 eV. The horizontal axis is the electron spectral density

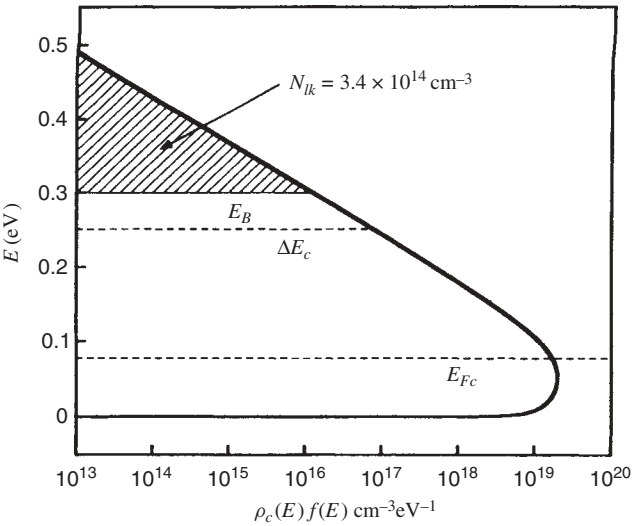


FIGURE A2.5: Room temperature electron energy versus the electron spectral density in the conduction band of GaAs with a total integrated density of $N = 2 \times 10^{18}$ cm^{-3}. The leakage density, N_{lk}, above the 0.3 eV AlGaAs barrier is shaded [8]. The total effective barrier height ($E_B - E_c$) includes residual band bending under forward bias as well as the conduction band discontinuity, ΔE_c.

$\rho_c(E)f(E)$. To calculate the density of electrons above the barrier, N_{lk}, indicated by the shaded region, we integrate the electron spectral density from the top of the effective barrier (which also includes any residual band-bending effects). Thus, as in Eq. (A2.2), but beginning the integration at E_B rather than E_c, we obtain

$$N_{lk} = \int_{E_B}^{\infty} \rho_c(E) f(E)\ dE, \tag{A2.12}$$

or for a bulklike density of states,

$$N_{lk} = \frac{1}{2\pi^2}\left(\frac{2m^*}{\hbar^2}\right)^{3/2} \int_{E_B}^{\infty} \frac{(E - E_c)^{1/2}}{1 + \exp[(E - E_{Fc})/(kT)]}\, dE. \tag{A2.13}$$

For the assumed barrier height of $E_B = 0.3$ eV, and for a Fermi level to yield a total density $N = 2 \times 10^{18}$ cm^{-3} carriers from Eq. (A2.2), one finds that $N_{lk} = 3.4 \times 10^{14}$ cm^{-3} at room temperature as shown in Fig. A2.5.

Some of these high-energy electrons will escape the active region and diffuse into the p-cladding material as illustrated in Fig. A2.6. The fraction that actually do, however, involves a number of factors. For example, some electrons approaching the cladding barrier will be reflected even though they have sufficient energy to pass over the barrier (it is much like photons being reflected off of a dielectric interface). For a simple potential step the reflection coefficient is not significant,

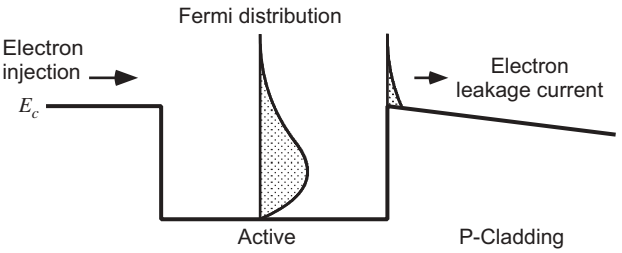


FIGURE A2.6: Schematic band diagram of a double heterostructure showing electron energy distribution in the active layer and leakage current.

dropping below 50% for electrons with energies just 3% greater than the barrier. However, specially designed cladding layers can reflect a much higher portion of the high-energy electrons.

Another factor to consider is that in a thermal distribution, the electron velocities are randomly distributed in all directions (with just a slight preference toward the direction of the current flow). Therefore only a fraction of the high energy electrons are even moving toward the p-cladding barrier. Finally, some electrons that make it into the p-cladding layer will diffuse right back into the active region. Thus, we must consider the flow of electrons in both directions. Thermionic emission theory can be used to estimate the maximum supply rate of electrons to the p-cladding layer by calculating the average thermal velocity of high-energy electrons directed toward the barrier. However, the tricky part is determining how many might be flowing back.

An alternative approach to studying the detailed flow of electrons across the barrier is to simply assume that the electron populations on both sides of the barrier are in thermal equilibrium such that we can match the quasi-Fermi levels across the barrier. If we assume this to be the case, then in the Boltzmann limit, the electron population at the edge of the p-cladding layer is given by

$$N_{p0} = N_c e^{(E_{Fc}-E_B)/kT}. \quad (\text{assuming } E_{Fc}|_{act} = E_{Fc}|_{clad}) \tag{A2.14}$$

This estimate of the actual leakage carrier density that establishes itself in the p-cladding layer is typically 20% to 30% of the value suggested by Eq. (A2.13). It also reveals the strong exponential dependence of the leakage carrier density, and eventually, the carrier leakage current on temperature.

Once we know the electron density at the edge of the p-cladding layer, we can estimate the electron leakage current by assuming that the electrons diffuse away transversely into the p-cladding material as minority carriers with an assumed initial density N_{p0} and a diffusion length L_n. Taking the $x = 0$ origin to be at the barrier on the p-side (opposite side from which they were injected), the distribution of electrons in the p-type cladding layer can then be written as

$$N(x) = N_{p0} e^{-x/L_n}, \tag{A2.15}$$

where we can express $L_n = \sqrt{D_n \tau_n}$, with τ_n being the minority carrier lifetime and D_n being the diffusion constant in the p-cladding material. From the Einstein relationship, $kT/q = D/\mu$, we can use the measured mobility to estimate the diffusion constant.

Assuming that any residual electric field would have a negligible effect, and also assuming a sufficiently thick p-layer, the associated diffusion current of the leakage electrons (neglecting the sign) is given by

$$J_n|_{x=0} = qD_n \frac{dN}{dx} = qD_n \frac{N_{p0}}{L_n} = qL_n \frac{N_{p0}}{\tau_n}. \tag{A2.16}$$

The last form shows that the diffusion current is equivalent to the recombination of a uniform carrier density spread over one diffusion length. For the example of Fig. A2.5 with $N = 2 \times 10^{18}$ cm^{-3} and $E_{Fc} - E_c = 79$ meV, we find that a 0.3 eV barrier yields $N_{p0} = 8.5 \times 10^{13}$ cm^{-3} from Eq. (A2.14) using the N_c of GaAs for simplicity (for AlGaAs, N_c is slightly larger and might need to include other conduction band minima depending on the x value). For $L_n = 5$ μm and $\tau_n = 5$ ns, we obtain $J_n \sim 1.4$ A/cm^2 at room temperature. For holes a similar leakage current will exist. For $P = 2 \times 10^{18}$, we have $E_v - E_{Fv} = -28$ meV. Assuming a net barrier of 0.25 eV and using $N_{v,hh} + N_{v,lh}$ of GaAs for simplicity, we calculate $P_{n0} = 1.4 \times 10^{14}$ cm^{-3}. For $L_p = 1$ μm and $\tau_p = 5$ ns, we find that the hole leakage current at the other barrier is $J_p \sim 0.5$ A/cm^2.

Since typical laser threshold current densities are at least a few hundred A/cm^2, we conclude that for this high barrier GaAs/AlGaAs example, carrier leakage is not significant at room temperature. However, in material systems that do not have the luxury of large heterobarriers (for example, 630–680 nm emission AlInGaP), carrier leakage can be much more of a problem in view of the exponential dependence on barrier height through N_{p0} in Eq. (A2.14). Carrier leakage also increases rapidly with temperature, such that even lasers with relatively high barriers can be affected at high temperatures. Finally, if the cladding material has a high defect density, the reduced minority carrier lifetime can lead to high carrier leakage currents even for a small N_{p0}.

If the resistance in the p-cladding is substantial, an electric field will assist electron diffusion away from the active region, enhancing the leakage rate. If we include this as well as the existence of a contact (which is assumed to be a region of zero lifetime) a distance x_p away, it can be shown that a more general expression for the electron leakage current results [9]:

$$J_n = qD_n N_{p0} \left[\sqrt{\frac{1}{L_n^2} + \frac{1}{L_{nf}^2}} \coth \sqrt{\frac{1}{L_n^2} + \frac{1}{L_{nf}^2}} x_p + \frac{1}{L_{nf}} \right], \tag{A2.17}$$

where

$$L_{nf} \equiv \frac{2kT}{q} \cdot \frac{\sigma_p}{J_{tot}}.$$

σ_p is the conductivity of the p-cladding region, and J_{tot} is the total diode current density. In analogy with the diffusion length, we can think of L_{nf} as the *drift* length, which decreases with increasing total current (as the electric field in the cladding increases). For low currents and high p-type conductivity, $L_{nf} \gg L_n$, and the carrier leakage current is dominated by the diffusion component considered above. However, for high currents and/or low p-type conductivity in the cladding, it is possible for $L_{nf} \ll L_n$, in which case the carrier leakage current becomes dominated by the drift component, which increases with the total current in addition to increasing with N_{p0}. More specifically, if $L_{nf} \ll L_n, x_p$, then $J_n \rightarrow q\mu_n N_{p0} J_{tot}/\sigma_p$. The double dependence on injection level leads to a higher sensitivity to temperature, which can be severe in lasers with low heterobarriers [9]. If the contact is placed close to the active region such that $x_p \ll L_n, L_{nf}$, then the current reduces to $J_n \rightarrow qD_n N_{p0}/x_p$, becoming independent of both drift and diffusion lengths. Finally, it should be realized that for either $x_p \rightarrow 0$ or $L_{nf} \rightarrow 0$, the carrier leakage current does not really go to infinity but becomes limited by the rate at which carriers can be supplied to the cladding region (it becomes thermionic emission-limited rather than drift-diffusion-limited).

So far we have been considering a simple double heterostructure laser for which the active region fills the entire waveguide. In quantum-well lasers, the active region is restricted to the width of the well(s). Thus, carrier leakage has two meanings in this case. Carriers can leak into the separate confinement waveguiding layers as well as leaking out of the entire SCH waveguide region into the doped cladding layers. Carrier populations in the SCH regions lead to recombination that can be approximated using Eq. (A2.16) with the diffusion length replaced by the width of the SCH region, or

$$J_{SCH} = qL_{SCH}\frac{N_{SCH}}{\tau_n}. \tag{A2.18}$$

For $L_{SCH} = 1500$ Å, and $\tau_n = 5$ ns, we find $J_{SCH} \sim 50$ A/cm^2 per 10^{17} cm^{-3} of carrier density. This highlights the importance of maintaining low carrier populations in the waveguide regions of the laser. Of course, depending on the material quality of the SCH regions, the carrier lifetime may be longer or shorter affecting the carrier leakage current accordingly.

In Chapter 2, we introduced an effective recombination rate per unit active volume for carrier leakage, R_l. In terms of the total carrier leakage current, $J_l = J_n + J_p + J_{SCH}$, we can define

$$R_l = \frac{J_l}{qd}, \tag{A2.19}$$

where d is the active region thickness. If lateral carrier leakage is important then it should also be added to R_l with the lateral active width replacing d in Eq. (A2.19). For example, in ridge lasers carriers are free to diffuse laterally since no heterobarrier exists. This lateral diffusion component of R_l is discussed in Chapter 4.

A2.4 INTERNAL QUANTUM EFFICIENCY

We would finally like to consider how carrier leakage affects the internal quantum efficiency, η_i, introduced in Chapter 2. In practice, η_i is defined as the fraction of current above threshold which results in stimulated emission, or $\eta_i(I - I_{th}) = I_{st}$. The output power is then $P_0 = (h\nu/q) \cdot \eta_0 I_{st}$, where η_0 is the *optical* efficiency of the laser cavity introduced in Chapter 5. This gives $P_0 = (h\nu/q) \cdot \eta_i \eta_0 (I - I_{th})$. The slope of the PI curve is related to $\eta_i \eta_0$, from which η_i can be extracted experimentally. This *above-threshold* definition of η_i is slightly different than the one given in Chapter 2; however, the two definitions are the same if η_i is not a function of current beyond threshold.

If we specify the total current as the stimulated current plus the sum of n various other components (i.e. spontaneous, Auger, leakage, etc.), such that $I = I_{st} + \sum I_n$ and $I_{th} = \sum I_{n,th}$, then we can expand $1 - \eta_i$ and rearrange to obtain

$$\eta_i = \frac{I_{st}}{I - I_{th}} = 1 - \frac{\sum (I_n - I_{n,th})}{I - I_{th}}. \tag{A2.20}$$

In this form we clearly see that any currents that clamp at threshold along with the carrier density do not contribute to a reduction of η_i. This point is often misunderstood in the laser community. The common mistake made is to set $\eta_i = I_{rad}/I$, where I_{rad} is the radiative current in the active region. This definition originates from the concept of radiative efficiency in an LED and has nothing to do with the laser performance above threshold. To demonstrate this point, consider a laser in which 90% of the current at threshold is Auger current, while the rest is spontaneous emission current in the active region (there is no leakage current). Because both Auger and spontaneous currents clamp at threshold (i.e. $I_n = I_{n,th}$), Eq. (A2.20) reveals that we can still have $\eta_i = 100\%$ even though the radiative efficiency near threshold is only 10%. The true radiative efficiency in this case does not approach 100% until the stimulated emission current becomes much larger than threshold.

The main question is which currents *do* contribute to a reduction of η_i? The simple answer is those that continue to increase above threshold. Now because the carrier density, N, in the active region clamps at threshold, all currents that depend monotonically on N should have no effect on η_i, including R_{sp}, R_{nr}, as well as carrier leakage out of the active region, R_l (since N_{p0} defined in the last section should in principle clamp along with N). However, the clamping of the modal gain (which is what really clamps) does not always result in a complete clamping of the carrier density in all of the various regions of the laser.

For example, there is always some small compression of the gain with increased photon density (see Chapter 5), which requires additional carriers to restore the threshold modal gain. Also, spatial hole burning of the carrier density profile by the optical mode can result in even larger changes in the local carrier density. These

second-order effects can allow R_{sp}, R_{nr}, and R_l to increase above their threshold values, reducing η_i according to Eq. (A2.20). However, these effects are typically not very significant.

The more important factor to consider is the incomplete clamping of carrier densities in the surrounding regions, N_{SCH} and N_{p0}, defined in Section A2.3. For example, one model of the SCH-quantum well *pn*-junction suggests that carriers in the well couple to a second carrier pool in the SCH region which is not clamped at threshold, but related to the well carriers through a set of capture lifetimes (see Chapter 5). The carrier density in the SCH region and corresponding current defined in Eq. (2.19) can therefore increase with injected current above threshold, while the gain remains fixed in the quantum well [10].

From another point of view, there is no guarantee that the quasi-Fermi levels in the SCH and cladding regions clamp along with the quasi-Fermi levels in the active region at threshold. As a result, recombination rates away from the active region can potentially reduce η_i, if significant carrier populations exist there and if the quasi-Fermi levels in those regions remain unclamped or only partially clamped above threshold. Unfortunately, realistic modeling of the extent of clamping throughout the lasing junction is a very complex undertaking. Thus, the extent to which such effects contribute to a reduction of η_i is hard to predict. Experimental values for η_i are commonly in the range of 70% to 80% but can often be as low as 40% or as high as 95%.

In Chapter 2, we specifically assume that the leakage rate of carriers out of the active region, R_l, depends monotonically on the carrier density in the active region. This is assumed mainly to highlight the fact that *in principle*, carrier leakage does not affect the internal quantum efficiency. In view of the preceding discussion, this may not always be the case since some carrier leakage currents may clamp or partially clamp and others may not. In any case, only the amount of leakage *beyond* threshold, $R_l - R_{l,th}$, reduces η_i. For leakage currents that partially clamp at threshold, this value will typically be much smaller than that suggested by the magnitude of R_l itself.

We can modify the derivation in Chapter 2 to include these effects by first expanding $\eta_i(I - I_{th})$ in Eq. (2.36) into $(I - I_L) - (I_{th} - I_{L,th}) = (I - I_{th}) - (I_L - I_{L,th})$, where I_L is the current leakage which does not generate carriers in the active region (see Fig. 2.2). Then by adding recombination terms $R_l - R_{l,th}$ for carrier leakage and similarly for all other recombination rates, we can use Eq. (A2.20) to reduce the carrier density rate equation back to Eq. (2.36), with the new η_i encompassing all currents that do not clamp at threshold.

Finally, it is interesting to note that even if the carrier density clamps everywhere, the *drift* carrier leakage defined in Eq. (A2.17) will continue to increase above threshold due to its dependence on the total current density. Therefore, if drift carrier leakage is significant, it can lead to a noticeable reduction of η_i, even for a fixed N_{p0} above threshold.

REFERENCES

1. Blakemore JS. Approximations for Fermi-Dirac integrals, especially the function used to describe electron density in a semiconductor. *Solid State Electron* 1982;**25**: 1067.
2. Unger K. Spontane und induzierte emission in laserdioden II. Berücksichtigung der zustandsdichteschwänze. *Z Physik* 1967;**207**: 322.
3. Corzine SW. Unpublished.
4. Joyce WB, Dixon RW. Analytic approximations for the Fermi energy of an ideal Fermi gas. *Appl Phys Lett* 1977;**31**: 354.
5. Unger K. Reversible formulae to approximate Fermi integrals. *Phys Stat Solidi (b)* 1988;**149**: K141
6. Nilsson NG. An accurate approximation of the generalized Einstein relation for degenerate semiconductors. *Phys Stat Solidi (a)* 1973;**19**: K75.
7. Aguilera-Navarro VC, Estevez GA, Kostecki A. A note on the Fermi–Dirac integral function. *J Appl Phys* 1988;**63**: 2848.
8. Ebeling KJ. *Integrated opto-electronics*. Berlin: Springer-Verlag; 1993. Chapter 9.
9. Bour DP, Treat DW, Thornton RL, Geels RS, Welsh DF. Drift leakage current in AlGaInP quantum-well lasers. *IEEE J Quantum Electron* 1993;**29**: 1337.
10. Hirayama J, Yoshida J, Miyake Y, Asada M. Carrier capture time and its effect on the efficiency of quantum-well lasers. *IEEE J Quantum Electron* 1994;**30**: 54.

READING LIST

Agrawal GP, Dutta NK. *Semiconductor lasers*. 2nd ed. New York: Van Nostrand Reinhold; 1993. Chapter 3.

APPENDIX THREE

Introduction to Optical Waveguiding in Simple Double-Heterostructures

A3.1 INTRODUCTION

Starting from Maxwell's equations a wave equation (sometimes referred to as the Helmholtz equation), which is very analogous to Schrödinger's equation, can be derived:

$$\nabla^2 \mathscr{E} = \mu\varepsilon \frac{\partial^2 \mathscr{E}}{\partial t^2}, \tag{A3.1}$$

where ε is the dielectric constant and μ is the magnetic permeability. In this derivation it is assumed that ε is uniform in space, and both μ and ε can be complex, although for the semiconductor materials of interest, $\mu \approx \mu_0$. The imaginary part of ε includes the gain or loss that can occur in these materials. We are searching for time-harmonic fields propagating in the z-direction, so we try

$$\mathscr{E}(x, y, z, t) = \hat{\mathbf{e}}_i E_0 U(x, y) e^{j(\omega t - \tilde{\beta} z)}, \tag{A3.2}$$

as a solution to Eq. (A3.1). The unit vector $\hat{\mathbf{e}}_i$ gives the polarization. E_0 has units of volts, *U(x,y)* has units of per unit length, and it is usually assumed to be normalized such that $\int |U(x, y)|^2 \, dA = 1$. Plugging Eq. (A3.2) into (A3.1) and factoring out

Diode Lasers and Photonic Integrated Circuits, Second Edition.
Larry A. Coldren, Scott W. Corzine, and Milan L. Mašanović.
© 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

common terms, we find that the transverse amplitude function, *U(x, y)*, must satisfy

$$\nabla^2 U(x,y) + \left[\tilde{n}^2 k_0^2 - \tilde{\beta}^2\right] U(x,y) = 0, \tag{A3.3}$$

where the square of the index of refraction $\tilde{n}^2 = \varepsilon/\varepsilon_0$, the free-space propagation constant, $k_0^2 = \omega^2 \mu_0 \varepsilon_0 = (2\pi/\lambda)^2$, and both $\tilde{n}$ and the propagation constant in the z-direction, $\tilde{\beta}$, are complex. Expanding $\tilde{\beta} = \beta + j\beta_i$, we see from the form of Eq. (A3.2) that $\beta = \omega/v_p$, where v_p is the velocity of a phase front, or the phase velocity. We can also introduce an effective index of refraction, $\bar{n}$, so that $\beta = k_0 \bar{n} = \omega \bar{n}/c = 2\pi \bar{n}/\lambda$. In Chapter 2 we modeled how β_i contains the gain and internal loss terms in a laser.

A3.2 THREE-LAYER SLAB DIELECTRIC WAVEGUIDE

As shown in Fig. A3.1, the double heterostructure used in diode lasers provides a three-layer slab configuration in which each layer has a different index. More generally, in lasers we may be interested in regions where the index varies both transversely and laterally. Thus, we might modify Eq. (A3.3) to give

$$\nabla^2 U(x,y) + [\tilde{n}^2(x,y) k_0^2 - \tilde{\beta}^2] U(x,y) \approx 0. \tag{A3.4}$$

However, as indicated by the $\approx$ symbol, this now only approximately satisfies Maxwell's equations, since our derivation assumed that ε (and therefore n) was uniform in space.

In the current one-dimensional slab case, we can obtain an exact solution by solving Eq. (A3.3) for uniform n in each of the three regions and then matching the boundary conditions at the interfaces. By noting that the ratio of the real to the imaginary part of the index is very large in all practical cases of interest, we can replace $\tilde{\beta}$ by β, with some assurance that the mode shape, *U(x, y)*, will not be

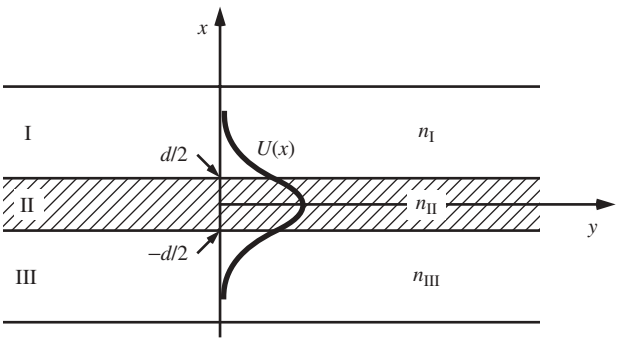


FIGURE A3.1: Schematic of a three-layer slab waveguide. Indices are assumed to be uniform in the z-direction.

significantly in error. We shall still use Eq. (A3.2) to include the gain or loss in the propagating mode.

The solution procedure is very similar to how we solved for the confined states of the electron in a one-dimensional rectangular well in Appendix 1. Indeed, there is no difference in the form of the wave equation for the transverse electric field, Eq. (A3.3), and the time-independent Schrödinger's equation, Eq. (A3.5). Thus, the solutions will have the same form if the boundary conditions are analogous. As we shall see, they are analogous for the TE modes, which are polarized in the y-direction, but a little different for the TM modes, which are polarized in the x-direction.

A3.2.1 Symmetric Slab Case

For a symmetric three-layer slab guide ($n_{\mathrm{I}} = n_{\mathrm{III}}$), we follow the solution given in Appendix 1 for a one-dimensional potential well very closely. In the central region, we assume solutions of the form,

$$U_{\mathrm{II}}(x) = \begin{cases} A\cos k_x x \,(\text{symmetric solutions}) \\ A\sin k_x x \,(\text{antisymmetric solutions}) \end{cases}. \tag{A3.5}$$

In region I,

$$U_{\mathrm{I}}(x) = Be^{-\gamma x}. \tag{A3.6}$$

After substituting Eqs. (A3.5) and (A3.6) into (A3.3) with $n = n_{\mathrm{II}}$ and n_{I}, respectively, we find that

$$k_x^2 = k_0^2 n_{\mathrm{II}}^2 - \beta^2, \tag{A3.7}$$

and

$$\gamma^2 = \beta^2 - k_0^2 n_{\mathrm{I}}^2.$$

In region III, $U_{\mathrm{III}} = Be^{\gamma x}$, but by symmetry in this case ($n_{\mathrm{I}} = n_{\mathrm{III}}$), we only need to use the single-boundary condition at $x = d/2$ between regions II and I.

For the TE modes at $x = d/2$, we have that $U_{\mathrm{II}} = U_{\mathrm{I}}$ and $U'_{\mathrm{II}} = U'_{\mathrm{I}}$. These conditions derive from the requirements that both the tangential electric and magnetic fields, respectively, are equal at the boundary. For the symmetric solutions, this gives

$$A\cos\frac{k_x d}{2} = Be^{-\gamma d/2}, \tag{A3.8}$$

and

$$Ak_x \sin\frac{k_x d}{2} = B\gamma e^{-\gamma d/2}. \tag{A3.8}$$

Dividing Eq. (A3.8) by Eq. (A3.8a), we obtain the characteristic equation,

$$k_x \tan \frac{k_x d}{2} = \gamma. \tag{A3.9}$$

Similarly, for the antisymmetric solutions, we obtain $k_x \cot k_x d/2 = -\gamma$. Both can be included in a characteristic equation by recognizing the $\pi/2$ shift between the tan and cot functions. That is, after substituting for k_x and γ from Eqs. (A3.7) and using $\beta = k_0 \bar{n}$,

$$\frac{k_0 d}{2}[n_{\mathrm{II}}^2 - \bar{n}^2]^{1/2} = \tan^{-1}\left(\frac{\bar{n}^2 - n_{\mathrm{I}}^2}{n_{\mathrm{II}}^2 - \bar{n}^2}\right)^{1/2} + (m-1)\frac{\pi}{2}, \tag{A3.10}$$

where $m = 1, 2, 3, \ldots$ for the fundamental and higher-order modes. For the TM modes, the continuity of the tangential electric and magntic fields at the boundaries results in an additional factor of $(n_{\mathrm{II}}^2/n_{\mathrm{I}}^2)$ on the right side of Eq. (A3.9) and inside the brackets of the $\tan^{-1}$ function in Eq. (A3.10).

The solution to this transcendental equation is done graphically as was done for confined electrons in Appendix 1 and illustrated in Fig. A1.3. As might be expected, the results are analogous for this symmetrical slab case.

A3.2.2 General Asymmetric Slab Case

To add additional generality to the present case, we can repeat the preceding procedures for an antisymmetric three-layer slab waveguide. That is, $(n_1 \neq n_{\mathrm{III}})$. For the TE modes, we find that the characteristic equation, Eq. (A3.9), becomes

$$\tan k_x d = \frac{(\gamma_{\mathrm{I}}/k_x) + (\gamma_{\mathrm{III}}/k_x)}{1 - \gamma_{\mathrm{I}}\gamma_{\mathrm{III}}/k_x^2}, \tag{A3.11}$$

where γ_{I} and γ_{III} are the decay constants in the upper and lower cladding regions, respectively, defined as in Eq. (A3.7). Equation (A3.11) also can be solved graphically, but it is convenient to define a normalized frequency, V, propagation parameter, b, and asymmetry parameter, a, in order to display the results. These normalized parameters are defined as follows:

$$\begin{aligned} V &\equiv k_0 d (n_{\mathrm{II}}^2 - n_{\mathrm{III}}^2)^{1/2}, \\ b &\equiv \frac{\bar{n}^2 - n_{\mathrm{III}}^2}{n_{\mathrm{II}}^2 - n_{\mathrm{III}}^2}, \end{aligned} \tag{A3.12}$$

and

$$a \equiv \frac{n_{\mathrm{III}}^2 - n_{\mathrm{I}}^2}{n_{\mathrm{II}}^2 - n_{\mathrm{III}}^2}.$$

Figure A3.2 gives plots of the normalized propagation parameter as a function of the normalized frequency for a range of normalized asymmetry parameters.

For TM modes we can use Figure A3.2 with some small error due to the neglected dielectric constant ratio that should multiply γ_{I} and γ_{III} in the dispersion relationship. The error becomes vanishingly small in weak dielectric guides.

An approximative expression for the fundamental mode of a symmetric slab waveguide that links the normalized frequency V with the propagation parameter b is given by

$$b = 1 - \frac{\ln\left(1 + \frac{V^2}{2}\right)}{\frac{V^2}{2}}. \tag{A3.13}$$

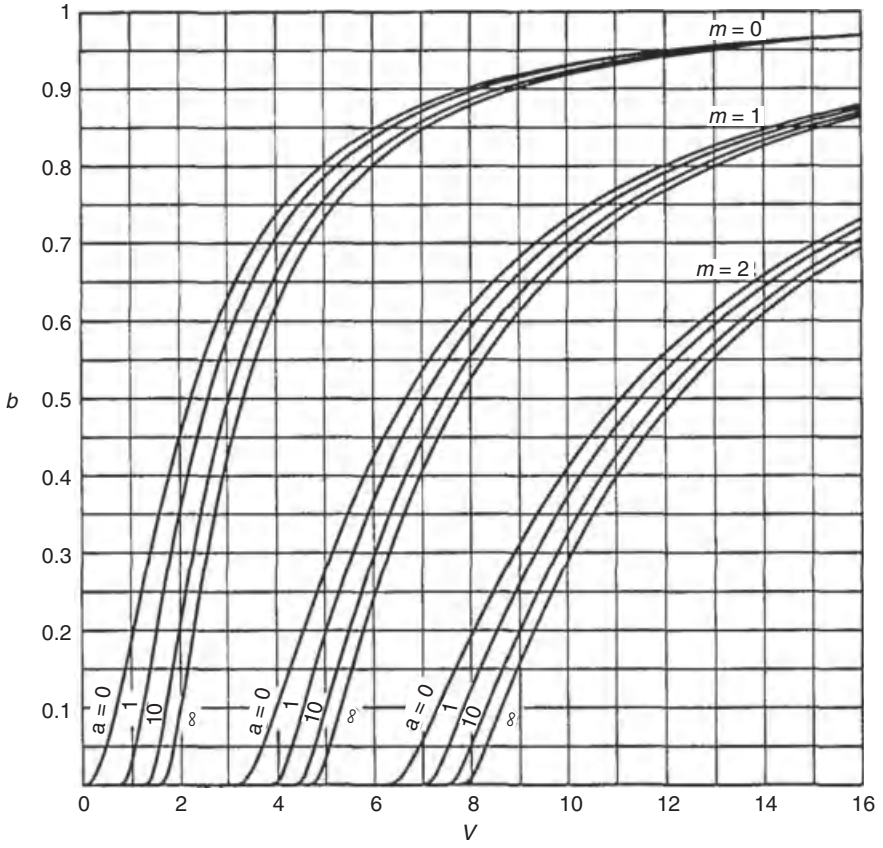


FIGURE A3.2: Normalized propagation parameter vdrsus normalized frequency for a range of asymmetries for the first three TE modes. After Kogelink and Ramaswamy [1]. (*Reprinted, by permission, from Applied Optics*.)

A3.2.3 Transverse Confinement Factor, Γ_x

The transverse confinement factor for the three-layer slab waveguide is defined as the fraction of the optical energy that is contained in the active slab region. As derived in Appendix 5, Eq. (A5.13), this fraction can be approximated as (neglecting $n/\bar{n}$)

$$\Gamma_x = \frac{\int_{-d/2}^{d/2} |U(x,y)|^2 \, dx}{\int_{-\infty}^{\infty} |U(x,y)|^2 \, dx}. \tag{A3.14}$$

For the symmetric case ($a = 0$), we can use Eqs. (A3.7) and (A3.5) in Eq. (A3.14) to get

$$\Gamma_x = \frac{1 + 2\gamma d/V^2}{1 + 2/\gamma d}. \tag{A3.15}$$

For the fundamental mode with relatively small index differences, Eq. (A3.15) can be approximated by

$$\Gamma_x \approx \frac{V^2}{2 + V^2}. \tag{A3.16}$$

Figure A3.3 compares the approximate formula to the exact confinement factor (neglecting $n/\bar{n}$) for several values of n_{III}.

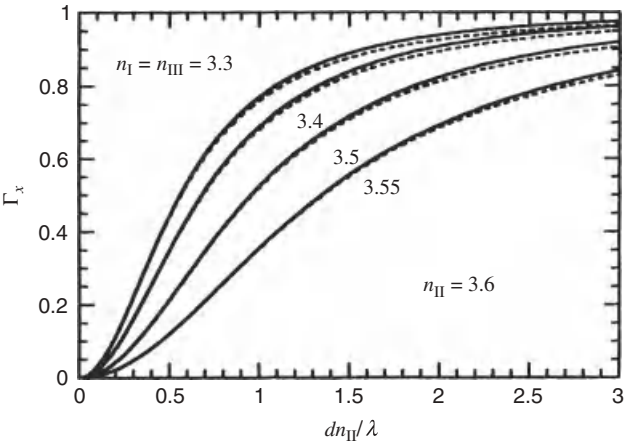


FIGURE A3.3: Comparison between the exact confinement factor (solid curve) and the approximate formula (Eq. (A3.25); dashed curve) for several values of the cladding index as a function of the guide thickness for the fundamental slab mode.

A3.3 EFFECTIVE INDEX TECHNIQUE FOR TWO-DIMENSIONAL WAVEGUIDES

As indicated in Chapters 1 and 2, practical diode lasers usually employ waveguiding in the lateral y-direction as well as the transverse x-direction. These are referred to as either two-dimensional or channel waveguides. After Eq. (A3.4) it was suggested that exact analytic solutions to this two-dimensional problem depicted in Fig. A3.4 are impossible. The problems arise in matching the lateral boundary conditions for all values of x.

In the limiting case of very strong index discontinuities at the active–cladding interfaces, the field within the active region will fall to zero at the boundaries, and the boundary conditions can be met around the perimeter. That is, for symmetric modes in region II-2 (the active region),

$$U(x, y) = U_0 \cos k_x x \cos k_y y, \quad \text{(region II-2)} \tag{A3.17}$$

and for antisymmetric modes, sine functions replace the cosines. Including both in sequence, the boundary conditions give $k_x d = \pi m_x$ and $k_y w = \pi m_y$, where the m_js are the respective mode numbers ($m_j = 1$ is the first symmetric mode in either direction, $m_j = 2$ is the first antisymmetric mode, etc.). Thus, the general solution for the symmetric modes is

$$U_{m_x m_y}(x, y) = U_0 \cos \frac{\pi m_x x}{d} \cos \frac{\pi m_y y}{w}. \quad (\Delta n \to \infty) \tag{A3.18}$$

Unfortunately, we cannot make the large index discontinuity assumption in most cases. (However, a semiconductor–air interface is sufficiently large to neglect fields outside the semiconductor provided it is thicker than a few wavelengths.) Thus, we must employ some other approximation. The most common of these is the *effective*

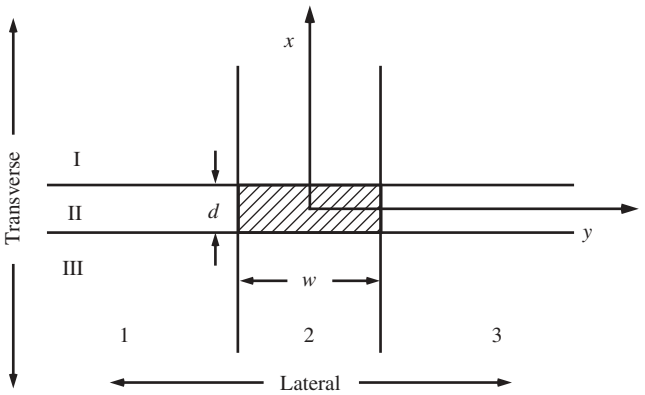


FIGURE A3.4: Waveguide cross section perpendicular to the z-direction. The transverse and lateral directions are divided into regions I, II, and III and 1, 2, and 3, respectively.

index technique. This technique involves a sequential solution to the problem of Fig. A3.4. It is most accurate when the transverse slab mode solutions in regions 1, 2, and 3 are nearly the same, and when $w/d \gg 1$.

The effective index technique is the limiting form of a rigorous technique to match the fields in the lateral direction for all values of x along the interfaces between regions 1, 2, and 3 at $y = \pm w/2$ illustrated in Fig. A3.4. This rigorous technique uses a superposition of the simple (uniform-y) slab mode solutions along x derived in Section A3.2 to express the overall channel waveguide mode shape in each of the three lateral regions. To be completely rigorous, a complete set of slab modes along x must be used in each region (including the radiation modes) to accurately synthesize any arbitrary mode shape. Another consideration is that all components of the channel waveguide mode must propagate along z with the same phase velocity, or the same propagation constant, β.

To adjust the β's of the various slab modes in the expansion to the same value, we must add a y-component to the propagation direction. For example, if β_m is the propagation constant solution of the mth slab mode, and β is the propagation constant along z, then for a slab mode propagating along z, $\beta_m = \beta$ (as we assumed earlier). Now, if we tilt the propagation direction slightly toward y, then $\beta_m^2 = \beta^2 + k_{ym}^2$. Thus, k_{ym} can be used to make up the difference between the fixed β and each slab mode's β_m. For each slab mode in region 2, we therefore define $k_{ym} = \sqrt{\beta_{m2}^2 - \beta^2}$, which produces cosine and sine profiles along y (for $\beta_{m2} > \beta$). In regions 1 and 3, we anticipate that $\beta > \beta_{m1}, \beta_{m3}$ so we define $\gamma_{ym1} = \sqrt{\beta^2 - \beta_{m1}^2}$ and $\gamma_{ym3} = \sqrt{\beta^2 - \beta_{m3}^2}$. This yields the familiar evanescent decaying solutions along y in these regions.

Adding the y profiles to each of the slab mode solutions along x, we can now express the overall channel waveguide mode as a weighted sum of the slab modes within each lateral region:

$$\mathscr{E}_1(x,y) = \sum_m B_{m1} U_{m1}(x) e^{\gamma_{ym1} y},$$

$$\mathscr{E}_2(x,y) = \sum_m A_m^e U_{m2}(x) \cos k_{ym} y + \sum_m A_m^o U_{m2}(x) \sin k_{ym} y,$$

$$\mathscr{E}_3(x,y) = \sum_m B_{m3} U_{m3}(x) e^{-\gamma_{ym3} y}.$$

B_{m1}, A_m^e, A_m^o, and B_{m3} are the weighting or *expansion* coefficients, which are adjusted to accurately synthesize the overall channel waveguide mode in each of the lateral regions. $U_{m1}(x), U_{m2}(x)$, and $U_{m3}(x)$ are the transverse mode shapes of the mth slab mode in regions 1, 2, and 3.

The final task is to match the boundary conditions at the planes between regions 1 and 2, and 2 and 3 using $\mathscr{E}_1, \mathscr{E}_2$, and $\mathscr{E}_3$. If the electric field is predominantly polarized along x (TM transverse mode), the matching of the tangential electric and magnetic fields is to a good approximation equivalent to matching $\mathscr{E}_x$ and

$\partial \mathscr{E}_x/\partial y$ at each boundary (we neglect the small $\mathscr{E}_z$ and $\mathscr{E}_y$ components). If the lateral waveguide is symmetric, we can define laterally symmetric and antisymmetric solutions, which require matching at only one boundary. For the symmetric solutions, we can set $B_{m1} = B_{m3} \equiv B_m, A_m^o = 0$, and $A_m^e \equiv A_m$. Matching these fields and their y-derivatives at the $y = w/2$ boundary yields

$$\sum_m A_m U_{m2}(x) \cos \frac{k_{ym} w}{2} = \sum_m B_m U_{m3}(x) e^{-\gamma_{ym} w/2}, \tag{A3.19}$$

and

$$\sum_m A_m U_{m2}(x) k_{ym} \sin \frac{k_{ym} w}{2} = \sum_m B_m U_{m3}(x) \gamma_{ym} e^{-\gamma_{ym} w/2}, \tag{A3.19a}$$

where

$$k_{ym}^2 = \beta_{m2}^2 - \beta^2 \quad \text{and} \quad \gamma_{ym}^2 = \beta^2 - \beta_{m3}^2. \tag{A3.19b}$$

β_{m2} and β_{m3} are the mth slab mode propagation constant solutions in regions 2 and 3, respectively. The A_m and B_m expansion coefficients can only satisfy Eqs. (A3.19) and (A3.18) at all values of x for discrete values of β, which yield the guided mode solutions of the overall channel waveguide.

If the electric field is instead predominantly polarized along y (TE transverse mode), as is more commonly the case, the matching of the tangential electric and magnetic fields across the $y = w/2$ boundary is approximately equivalent to matching $\varepsilon \mathscr{E}_y$ and $\partial \mathscr{E}_y/\partial y$. Thus for TE transverse modes, the dielectric constant ratio $\varepsilon_3(x)/\varepsilon_2(x)$ should appear on the right side of Eq. (A3.19a). This factor however can often be ignored without introducing much error.

To solve Eq. (A3.19), we need to have as many equations as there are expansion coefficients. If M modes are included in the summation in region 2, and M modes in region 3, then there are $2M$ unknown coefficients overall, and two equations at each x location. To get enough equations along the lateral boundary then, we must evaluate the lateral boundary conditions at M values of x. The resulting system of equations can be solved uniquely only when the determinant of the matrix formed by the factors multiplying each expansion coefficient goes to zero. The zero crossing can be found numerically by scanning the value of β across some appropriate range. As already suggested, if the transverse slab mode shapes in the adjacent lateral regions are similar, only a few slab modes are required to give good results.

The effective index technique results when we use only one transverse slab mode of region 2 and region 3 to approximate the channel waveguide field. Because only one unknown coefficient exists in each region (two total), we only need to evaluate the boundary conditions at one point along x (usually at $x = 0$). In this limit,

Eqs. (A3.19) and (A3.18) reduce to Eqs. (A3.8) and (A3.8), used earlier to solve for the one-dimensional slab modes. Thus, the solution to the lateral guide problem becomes identical to the transverse slab guide problem, except that Eq. (A3.19) replace Eq. (A3.7), or β_{m2} replaces k_{II} and β_{m3} replaces k_{III} $(= k_{\mathrm{I}})$. More generally, the effective index of the slab mode in each lateral region is used in place of the index of the medium in the normalized parameters defined in Eq. (A3.12).

In summary, the solution sequence for the effective index technique is as follows:

1. Three-layer slab problems are first solved across the smallest dimension (x-direction in Fig. A3.4) using $n_{\mathrm{I}}, n_{\mathrm{II}}, n_{\mathrm{III}}$, as if the regions were uniform in the other longer (y) dimension. This is repeated in each of the three lateral regions listed in Fig. A3.4 to produce $\bar{n}_1, \bar{n}_2$, and $\bar{n}_3$. The effective index in the central (# 2) region is used for the transverse k_x and γ_x in the final solution.
2. The three effective indices thus obtained ($\bar{n}_1, \bar{n}_2$, and $\bar{n}_3$) are now used in a new three-layer slab problem for the other (y) direction. The result is a final effective index for the two-dimensional problem. This is used for the lateral k_y and γ_y. The net axial propagation constant, β, thus, as always satisfies

$$[k_0 n_{\mathrm{II}}]^2 = k_x^2 + k_y^2 + \beta^2. \tag{A3.20}$$

This effective index technique can be applied relatively easily by successive use of Fig. A3.2. Although this figure is strictly only for TE modes, it can be approximately used for TM modes if the index differences are small. The thinner transverse dimension determines the designation of TE or TM. The second half of the effective index procedure is really dealing with the opposite polarization from the first. However, if the indices are similar, the difference in the solution is negligible.

The overall solutions for the symmetric modes of $U(x, y)$ are identical to Eq. (A3.17) within the active region, region II-2. In region I-2 the y-dependent factor remains the same, but the x-dependent factor becomes a decaying exponential like Eq. (A3.6). That is,

$$U(x, y) = U_0 \cos \frac{k_x d}{2} e^{-\gamma_x (x - d/2)} \cos k_y y, \quad \text{(region I-2)} \tag{A3.21}$$

where γ_x comes from the first step and k_y comes from the second step in the effective index technique. Note that the magnitude has been matched to the solution in region II-2. For region II-3, the solution in the x-direction remains from the first step and the lateral decay is found from the second step,

$$U(x, y) = U_0 \cos \frac{k_y w}{2} (\cos k_x x) e^{-\gamma_y (y - w/2)}. \quad \text{(region II-3)} \tag{A3.22}$$

The functional forms of $U(x, y)$ in other regions should now be obvious.

The lateral confinement factor Γ_y can be approximated by the lateral equivalent to Eq. (A3.16). That is,

$$\Gamma_y \approx \frac{V_l^2}{2 + V_l^2}, \tag{A3.23}$$

where, $V_l = k_0 w(\bar{n}_2^2 - \bar{n}_3^2)^{1/2}$.

A3.4 FAR FIELDS

The waveguide field profiles given above are assumed to be uniform along the length of the laser as long as the cross section of the guide remains uniform. At the output facet this field emerges from the laser waveguide and diffracts freely into the surrounding dielectric (usually air). In most applications it is desirable to capture this output light into some other waveguide or detector. The amount of light actually captured depends on the size and shape of the beam at the cross section of the capturing object among other things. Thus, it is useful to predict the field profile as it extends beyond the output facet.

In diffraction theory we refer to the field emitted from the laser weaveguide as the *near field* and the diffracted field some distance away as the *far field*. The transition occurs at roughly w^2/λ, where w is some characteristic full width of the near-field pattern. In a real index-guided waveguide, the wavefronts are planar as they approach the emitting facet. They remain approximately planar in the near field, but begin to show noticeable curvature in the transition to the far field. After some distance into the far field, the wavefronts approach a spherical shape with a radius of curvature measured from the center of the emitted mode at the facet where the wavefronts are planar. The narrowest, planar wavefront location is termed the *beam waist* in Gaussian mode theory. Figure A3.5 describes the generation of the far field by the emitted near field.

To determine the far-field pattern, U_F, a distance R and angle θ_R from the origin as defined in Fig. A3.5, we consider each amplitude increment of the near-field pattern, $U(x, y)$, to be a radiating point source. Each point source or "spherical wavelet" then propagates a distance r to contribute a far-field element, dU_F. The total far-field amplitude at any given point is found by coherent addition of all wavelet contributions. The field of each spherical wavelet a distance r away from its source and weighted by the near-field amplitude within the increment $dx\,dy$ can be written as

$$dU_F = dx\,dy\,U(x,y)\frac{e^{-jkr}}{r}\left[\frac{j}{\lambda}\cos\theta_r\right], \tag{A3.24}$$

where $k = 2\pi/\lambda$ and θ_r is the angle between $\mathbf{r}$ and the z-axis (if the medium outside the laser is not air then $\lambda \to \lambda/n$). The added factors in square brackets are mathematical refinements to the intuitive wavelet concept originally conceived

by Huygen. These factors introduced later by Fresnel (as well as Rayleigh and Sommerfeld) allow Huygen's principle to be applied to a wide range of scalar diffraction problems with excellent accuracy. The most significant of these terms is the *obliquity factor*, $\cos\theta_r$, which adds a directivity pattern to each spherical wavelet (it accounts for the reduced apparent area of each incremental emitter when viewed off-axis).

If $x, y \ll R$ for all x and y in which $U(x, y)$ has significant amplitude, we can approximate the distance from the increment to the measurement point as

$$r \approx R - x \sin\theta_x - y \sin\theta_y, \tag{A3.25}$$

where θ_x and θ_y are the angles measured from the z-axis toward the x- or y-axis, respectively. Since the phase factor is the most sensitive, we should use Eq. (A3.25) in the exponent of Eq. (A3.24). However in the denominator, we can set $r \approx R$ and also $\theta_r \approx \theta_R$. With these far-field approximations, we can sum over all wavelet contributions by integrating Eq. (A3.24) over the near-field emission plane to obtain

$$U_F(\theta_x, \theta_y) = \frac{j \cos\theta_R}{\lambda R} e^{-jkR} \iint U(x, y) e^{jk \sin\theta_x x} e^{jk \sin\theta_y y}\, dx\, dy, \tag{A3.26}$$

with $\cos\theta_R = \cos\theta_x \cos\theta_y / (1 - \sin^2\theta_x \sin^2\theta_y)^{1/2}$.

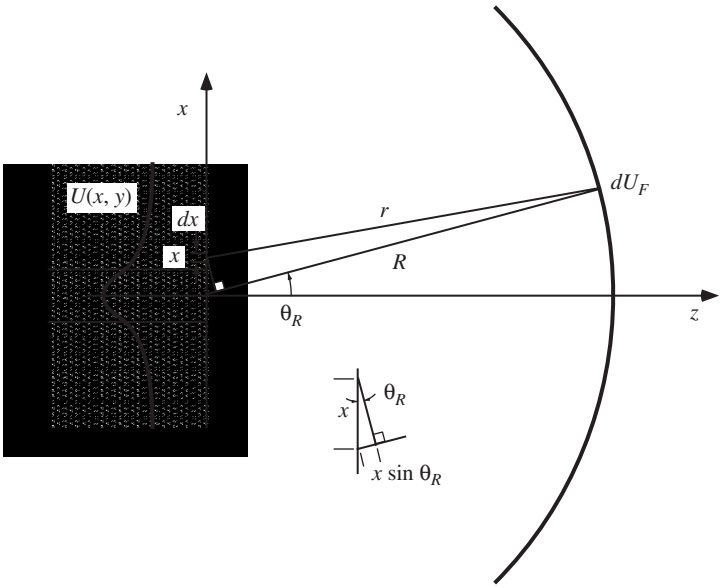


FIGURE A3.5: Illustration of near-field pattern, $U(x, y)$, and far-field element, dU_F, a distance R from the facet at the waveguide axis in the x-z plane.

Concentrating on a single axis by setting $\theta_y = 0$, and taking the magnitude squared, we arrive at the angular power spectrum in the far field,

$$|U_F(\theta_x)|^2 = \frac{\cos^2\theta_x}{\lambda^2 R^2}\left|\int U(x)e^{jk\sin\theta_x x}\,dx\right|^2, \tag{A3.27}$$

where $U(x) \equiv \int U(x,y)\,dy$. First of all we see that the far-field intensity drops as $1/R^2$, like a spherical wave. In addition, for small angles where $\sin\theta_x \approx \theta_x$ and $\cos\theta_x \approx 1$, we observe that $U_F(\theta_x)$ and $U(x)$ are Fourier transform pairs. In this approximation then, the far-field angular spectrum is just the Fourier transform of the near field, suggesting that a *narrow* emitting aperture (i.e., waveguide) leads to a *wide* angular distribution in the far field, and vice versa. For larger angles, the Fourier transform relationship breaks down, but the inverse relationship between the near-field and far-field beam width remains qualitatively valid.

With in-plane lasers, the transverse waveguide width is usually much smaller than the lateral width. This causes the transverse far-field pattern to have a much larger angular spread than the lateral far-field pattern. However, as long as $U(x,y)$ is real (i.e., has planar wave fronts), the far-field wavefronts will be spherical from the constant $\exp[-jkR]$ factor in Eq. (A3.26). In other words, even though the emitted power distribution is asymmetric, the output beam is *not* astigmatic. Thus, the emitted beam can be easily collimated into planar wavefronts with a simple spherical lens. However, the intensity pattern will be elliptical. To correct for this, the collimated beam can be refracted through a wedge with nonparallel planar surfaces to elongate one axis. The result is a circular collimated beam, but the procedure is rather inconvenient. With VCSELs, the transverse and lateral guide widths are usually the same, producing nice circular output beams in the far field.

REFERENCES

1. Kogelnik H, Ramaswamy V. Scaling rules for thin-film optical waveguides. *Appl Opt* 1974;**8**: 1857.

READING LIST

Haus JA. *Waves and fields in optoelectronics*. Englewood Cliffs, NJ: Prentice Hall; 1984. Chapter 6.

Lee DL. *Electromagnetic principles of integrated optics*. New York: Wiley; 1986. Chapter 4.

Yariv A, Yeh P. *Optical waves in crystals*. New York: Wiley; 1984. Chapter 11.

Goodman JW. *Intriduction to Fourier optics*. New York: McGraw-Hill; 1968. Chapter 3.

APPENDIX FOUR

Density of Optical Modes, Blackbody Radiation, and Spontaneous Emission Factor

A4.1 OPTICAL CAVITY MODES

Figure A4.1 shows an optical cavity with dimensions d_x, d_y, and d_z. If we assume that the reflection coefficient at each boundary is real, then for each component of the propagation constant, k_j, the boundary condition for resonance is

$$2k_j d_j = 2m_j \pi, \tag{A4.1}$$

where $j = x, y$, or z, and m_j is the respective mode number. This follows from the fact that the electric field of a resonant mode must replicate its phase after traversing a round-trip in the cavity. Thus, considering all three components, the magnitude of the propagation constant for some resonant mode of the cavity is given by

$$|\mathbf{k}|^2 = \pi^2 \left[\left(\frac{m_x}{d_x} \right)^2 + \left(\frac{m_y}{d_y} \right)^2 + \left(\frac{m_z}{d_z} \right)^2 \right], \tag{A4.2}$$

where $|\mathbf{k}| = \omega n / c$.

Diode Lasers and Photonic Integrated Circuits, Second Edition.
Larry A. Coldren, Scott W. Corzine, and Milan L. Mašanović.
© 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

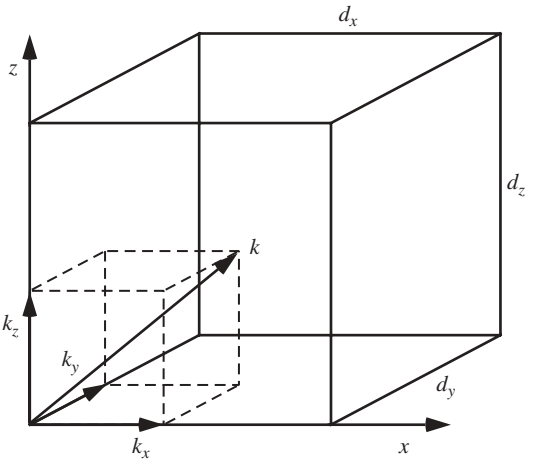


FIGURE A4.1: Optical cavity showing dimensions d_x, d_y, and d_z, and a superimposed k-vector decomposition for an optical mode.

Now just as in the case of electronic states, we wish to calculate the density of optical states or modes. The reasons are much the same. We would like to know the number of optical modes within some energy range, and rather than laboriously counting mode numbers, we would prefer to integrate a density of states. The process and the results are much the same, for there is little difference in concept between the electronic wavefunction and the normalized optical electric field. They both satisfy similar wave equations and boundary conditions as reviewed in Appendices 1 and 3.

However, there is one major difference between photons and electrons. Whereas only one electron can occupy an electronic state (after doubling the possible energy levels because of spin), an unlimited number of photons can occupy the same optical mode (again, after we have doubled the number of allowed k-values because of polarization). That is, there is no Pauli exclusion principle for photons. Thus, after we calculate the number of photon modes, we still have to determine how many photons occupy each mode to get the total number of photons.

Following the same procedure outlined after Eq. (A1.36), we now proceed to calculate the number of optical modes per unit volume per unit frequency. (We could do energy, $E = h\nu$, but let's have a little variety). As in Appendix 1 we shall first consider *bulk* dimensions, a *spherical* coordinate system, but switch to *frequency* as the variable. The first step is to calculate the volume, $N_s(m)$, in mode number space as depicted in Fig. A1.10. Analogous to Eq. (A1.37), we determine the volume of the sphere, multiply by 2 for the two polarization states and divide by 8 to allow only positive mode numbers,

$$N_s(m) = \tfrac{4}{3}\pi m^3 \cdot 2 \cdot \tfrac{1}{8}. \tag{A4.3}$$

Now, in spherical coordinates, $k^2 = (2\pi\nu)^2/(c/n)^2 = \pi^2 m^2/d^2$, where n is the index of refraction, $m^2 = m_x^2 + m_y^2 + m_z^2$, and we have let $d^2 = d_x^2 = d_y^2 = d_z^2$. Solving for $m(\rightarrow 2\ dn\nu/c)$ and plugging into Eq. (A4.3),

$$N_s(\nu) = \frac{\pi}{3}\left(\frac{2\ dn}{c}\right)^3 \nu^3. \tag{A4.4}$$

Finally, we apply Eq. (A1.36) (set $\rho_0(\nu)\ d\nu = (1/V)(dN_s/d\nu)d\nu$), and use $V = d^3$ to obtain

$$\rho_0(\nu)d\nu = \frac{8\pi}{c^3} n^2 n_g \nu^2\ d\nu, \tag{A4.5}$$

where we have defined a *group index*, $n_g = [n + \nu(\partial n/\partial\nu)]$, since the index of refraction can be frequency dependent.

If the dimensions of the optical cavity become small, say $d_j \leq 10\lambda$, the modes will not be so closely spaced, and it is advisable to use Eq. (A4.2) to count the modes within some range of wavelengths or frequencies. This situation is analogous to the electronic "quantum box." Because the states are widely spaced, the density of states just becomes a series of delta functions as each state is encountered. If only one dimension is small, we again have a similar situation to the quantum well for electrons as discussed in Appendix 1.

A4.2 BLACKBODY RADIATION

Equation (A4.5) gives us the desired density of optical states, but as mentioned above, any number of photons can occupy these states, so we need another piece of information to determine the photon density per unit frequency. One interesting boundary condition that can be applied to give a photon density is *thermal equilibrium*. In this case we can use the Maxwell–Boltzmann distribution to give the occupation probability of the optical states as a function of frequency. Maxwell–Boltzmann statistics dictate that a state with energy E will have a probability $\exp(-E/kT)$ of being occupied under thermal equilibrium at temperature T. Thus, we can calculate an average energy per state, $\langle E\rangle$, by taking a weighted average of the possible energies and their probability of occupation. Because light is quantized into photons of energy $h\nu$, the allowed energies are $E_j = jh\nu, j = 0, 1, 2, \ldots$. So we take the sum of allowed energies multiplied by the occupation probabilities and normalize this by dividing by the sum of the probabilities. That is,

$$\langle E\rangle = \frac{0e^{-0} + h\nu e^{-h\nu/kT} + 2h\nu e^{-2h\nu/kT} + \cdots}{1 + e^{-h\nu/kT} + e^{-2h\nu/kT} + \cdots}. \tag{A4.6}$$

Summing the infinite series and dividing, we get the average energy in each mode,

$$\langle E\rangle = \frac{h\nu}{e^{h\nu/kT} - 1}. \tag{A4.7}$$

The *blackbody radiation density*, $W(\nu)d\nu$, is now found by multiplying Eq. (A4.5) by Eq. (A4.7). That is, the equilibrium energy per unit volume per unit frequency is equal to the number of modes per unit volume per unit frequency times the average energy per mode under thermal equilibrium. Thus,

$$W(\nu)\,d\nu = \frac{\rho_0(\nu)h\nu}{e^{h\nu/kT}-1}\,d\nu = \frac{8\pi n^2 n_g h\nu^3/c^3}{e^{h\nu/kT}-1}\,d\nu. \tag{A4.8}$$

A4.3 SPONTANEOUS EMISSION FACTOR, β_{SP}

The spontaneous emission factor can be obtained from the density of optical modes per unit volume per unit frequency, Eq. (A4.5), by integrating over the cavity volume and the bandwidth of the spontaneous emission to find the number of optical modes that must contain the total spontaneous emission. Usually, it is assumed that the coupling to all modes is the same, so the reciprocal of this number of relevant modes is just the *fraction of energy going into each mode*, β_{sp}. Initially, making this assumption, the number of cavity modes in a bandwidth $\Delta\nu_{sp}$ is found by integrating Eq. (A4.5) from ν to $\nu + \Delta\nu_{sp}$ to be

$$N_{sp} = V_c \frac{8\pi}{c^3} n^2 n_g \nu^2 \Delta\nu_{sp}, \tag{A4.9}$$

where V_c is the cavity volume. Taking the reciprocal and using $\Delta\nu_{sp}/\nu = \Delta\lambda_{sp}/\lambda$,

$$\beta_{sp} = \frac{\Gamma_c \lambda^4}{8\pi V n^2 n_g \Delta\lambda_{sp}}, \tag{A4.10}$$

where $\Gamma_c = V/V_c$, and V is the active region volume. For typical values of parameters, $\beta_{sp} \sim 10^{-5}$ to 10^{-4}.

But the assumption that the spontaneous emission is uniformly distributed among the various cavity modes within $\Delta\lambda_{sp}$ is generally not true. In real devices, the emission spectrum follows a more bell-shaped curve as a function of both frequency and wavelength. If the mode of interest falls at the peak of the emission spectrum, then (A4.10) should be modified by setting $1/\Delta\lambda_{sp} \rightarrow$ (peak)/(area), where the area represents the total area under the emission spectrum, or the total emission rate. For example, if the emission spectrum follows a Lorentzian lineshape, we can set (peak)/(area) $= (2/\pi)/\Delta\lambda_{sp,FW}$, where the FW implies the full-width at half-maximum. Thus, an additional factor of $2/\pi$ should appear in Eq. (A4.10). For more complicated lineshapes, the ratio (peak)/(area) will be related to $1/\Delta\lambda_{sp}$, but the exact prefactor will depend on the specific lineshape.

Additional considerations involve the strength of the mode's wavefunction within the active region. The fraction of spontaneous emission represented by (A4.10) refers to the *modal* spontaneous emission rate, which must account for the overlap of the mode of interest with the active region. In Appendix 5 we show how such modal averages are taken. We have intentionally set $1/V_c = \Gamma_c/V$ in

Eq. (A4.10) to emphasize this modal averaging process. One interesting consequence of this substitution is that if the active region is very thin, and the mode for which we wish to calculate β_{sp} has a null at this point, then $\Gamma_c \approx 0$, from which it follows that $\beta_{sp} \approx 0$ as well. This would not be obvious if we had simply used $1/V_c$ in Eq. (A4.10).

Equation (A4.10) represents the semiclassical version of the spontaneous emission factor. In Chapter 4, a more useful version of the spontaneous emission factor derived from quantum mechanical considerations will be introduced. The quantum version is more general and simpler to evaluate.

READING LIST

Verdeyen JT. *Laser electronics*. 2nd ed. Englewood Cliffs, NJ: Prentice Hall; 1989. Chapter 7.

APPENDIX FIVE

Modal Gain, Modal Loss, and Confinement Factors

A5.1 INTRODUCTION

Within a laser cavity the gain and loss are not uniform throughout the volume occupied by the optical modes of interest. In fact, the gain region typically occupies only a few percent of the volume occupied by the optical modes, and the material absorption loss is typically very different in different regions of the cavity. Thus, we must develop some sort of overlap factor, which gives the net gain or loss provided to an optical mode.

To be completely general, let's define the localized *material* gain as a function of all three dimensions of space: $g(x, y, z)$. To provide the net effect of $g(x, y, z)$ on the mode as a whole, a properly weighted average of the gain distribution must be taken throughout the entire cavity volume. Now from Chapter 4, we know that the gain varies according to the stimulated emission rate, which in turn is proportional to the square of the electric field. Thus, it seems reasonable to use the electric field pattern, $\mathscr{E}(x, y, z)$, of the appropriate resonant mode as our weighting function. The standard definition of a weighted average from classical or quantum mechanics leads then to the following definition of *modal* gain:

$$\langle g \rangle = \frac{\int \mathscr{E}^*(x, y, z) g(x, y, z) \mathscr{E}(x, y, z)\, dV}{\int |\mathscr{E}(x, y, z)|^2\, dV}. \tag{A5.1}$$

Diode Lasers and Photonic Integrated Circuits, Second Edition.
Larry A. Coldren, Scott W. Corzine, and Milan L. Mašanović.
© 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

Before applying this result to typical laser structures, it is instructive to examine a little more carefully the appropriateness of this somewhat ad hoc definition.

A5.2 CLASSICAL DEFINITION OF MODAL GAIN

To rigorously determine the appropriate weighting function to be used in defining the modal gain or loss experienced by a waveguide mode, we turn to a classical description of gain and loss in the cavity. If we define $w_E(x,y,z)$ as the energy density throughout the cavity, then gain and loss *per unit time* can be associated with the time rate of change of this local energy density, dw_E/dt. For example, if $dw_E/dt < 0$ at some point in the cavity, energy is being generated locally, which indicates the presence of gain in the material. Integrating dw_E/dt over the entire cavity then allows us to determine the *total* energy generation rate. Defining the modal gain per unit time as the *fractional* generation rate of energy, we have

$$\langle G \rangle \equiv \frac{1}{W_E}\frac{dW_E}{dt} = \frac{\int dw_E/dt\, dV}{\int w_E\, dV}, \tag{A5.2}$$

where W_E is the total energy in the cavity. To check that this classical definition is in agreement with what we know from Chapter 2, we can express the total energy in terms of the photon density using $W_E = h\nu N_p V_p$. Equation (A5.2) then reduces to the familiar rate equation, $dN_p/dt = \langle G \rangle N_p$ where $\langle G \rangle = v_g \langle g \rangle$ and $\langle g \rangle$ is the modal gain (written explicitly as Γg in Chapter 2).

Equation (A5.2) is a rigorous classical definition of the modal gain per unit time. We now wish to express this definition in terms of the electric fields of a given waveguide mode. It can be shown using a variant of Poynting's theorem (which makes use of slowly time-varying phasors) that dw_E/dt is related to the local field strength through the imaginary part of the dielectric constant, ε_i. The exact relation (assuming $\mu = \mu_0$) is given by

$$\frac{dw_E}{dt} = \tfrac{1}{2}\omega\varepsilon_0\varepsilon_i \boldsymbol{\mathscr{E}} \cdot \boldsymbol{\mathscr{E}}^*, \tag{A5.3}$$

where $\boldsymbol{\mathscr{E}}$ is the electric field vector. Equating the imaginary parts of the dispersion relation, $\tilde{k}^2 = \omega^2\mu_0\varepsilon_0\tilde{\varepsilon}$, we find that $\varepsilon_0\varepsilon_i = k(x,y,z)g(x,y,z)/\omega^2\mu_0$ where $g \equiv 2\ \mathrm{Im}\{\tilde{k}\}$. Decomposing the total field into the two counterpropagating fields within the cavity, $\boldsymbol{\mathscr{E}}^+(x,y,z)$ and $\boldsymbol{\mathscr{E}}^-(x,y,z)$, the preceding relation then becomes

$$\frac{dw_E}{dt} = \tfrac{1}{2}(kg/\omega\mu_0)|\mathscr{E}^+ + \mathscr{E}^-|^2. \tag{A5.4}$$

For notational convenience, it is to be understood that if the electric field vector has more than one component, then $|\mathscr{E}|^2$ implies $\boldsymbol{\mathscr{E}} \cdot \boldsymbol{\mathscr{E}}^*$. This expression can be used in the numerator of Eq. (A5.2). Now we need to replace the denominator.

To relate the total energy in the cavity to the electric fields, we start by writing down an integral relationship between the energy density and the Poynting vector taken over any cross-sectional area of the cavity:

$$v_g \int w_E \, dA = \tfrac{1}{2}(\beta/\omega\mu_0) \int [|\mathscr{E}_T^+|^2 + |\mathscr{E}_T^-|^2] \, dA. \tag{A5.5}$$

On the left-hand side, the group velocity of the mode, v_g, converts the integrated energy density to total power flowing through the cross-section. On the right-hand side, the sum of forward and backward time-averaged Poynting vectors defines the local power density, which when integrated also yields the total power flowing through the cross-section. We have written the Poynting vectors explicitly in terms of the *transverse* electric field components and the effective propagation constant of the mode, β.

To determine the total energy, we need to integrate Eq. (A5.5) over the length of the cavity, which we take as the z-direction. To include possible axial variations in the waveguiding structure, we let $v_g \to v_g(z)$ and $\beta \to \beta(z)$. Moving $v_g(z)$ to the right-hand side, integrating over z, and setting $dA\,dz \to dV$, we obtain

$$\int w_E \, dV = \int \frac{\frac{1}{2}\beta(z)/\omega\mu_0}{v_g(z)} [|\mathscr{E}_T^+|^2 + |\mathscr{E}_T^-|^2] \, dV. \tag{A5.6}$$

Using Eqs. (A5.4) and (A5.6) we can write $\langle G \rangle$ in terms of the electric fields. We can then use the group velocity in the active waveguide to convert the modal gain per unit time to modal gain per unit *active* length: $\langle g \rangle = \langle G \rangle / v_{ga}$. Setting $k = nk_0$ and $\beta(z) = \bar{n}(z)k_0$, we finally obtain

$$\langle g \rangle = \frac{\int gn|\mathscr{E}^+ + \mathscr{E}^-|^2 \, dV}{\int \frac{v_{ga}}{v_g(z)} \bar{n}(z)[|\mathscr{E}_T^+|^2 + |\mathscr{E}_T^-|^2] \, dV}. \tag{A5.7}$$

If the material gain, g, is dependent on the polarization of the electric field (which is true in quantum-well and quantum-wire active materials), then the numerator should be written more precisely as $g_x|\mathscr{E}_x|^2 + g_y|\mathscr{E}_y|^2 + g_z|\mathscr{E}_z|^2$ (where each field component represents a sum, $\mathscr{E}_i^+ + \mathscr{E}_i^-$).

It is interesting to note that the energy generation rate in the numerator of Eq. (A5.7) includes coherent effects such as standing waves, while the denominator does not. This is because gain in a medium is obtained specifically through interactions with the electric field. Hence, the energy generation rate in the cavity is concentrated at the peaks of the electric field standing wave pattern. The stored energy, on the other hand, exists in both the electric and magnetic fields continually shifting back and forth between them. As a result, the time-averaged energy distribution is independent of the standing wave pattern.

When there are no axial variations in group velocity or propagation constant, we can simplify Eq. (A5.7) by setting $v_{ga}/v_g(z) = 1$ and pulling $\bar{n}$ out of the

integral:

$$\langle g \rangle = \frac{\int \bar{g}\,|\mathscr{E}^{+} + \mathscr{E}^{-}|^2\,dV}{\int [|\mathscr{E}_T^{+}|^2 + |\mathscr{E}_T^{-}|^2]\,dV}, \qquad \text{where } \bar{g} = g\frac{n}{\bar{n}}. \tag{A5.8}$$

By defining an *effective* material gain, $\bar{g}$, in our weighted average (which is found by replacing the index with the effective index in the expression for gain given in Chapter 4), the rigorous definition of modal gain can be made to resemble our initial guess Eq. (A5.1). In fact, for TE modes that have no longitudinal electric field components, the correspondence is exact (aside from the incoherent sum in the denominator). For TM modes that do have longitudinal electric fields, there is still a slight difference between (A5.8) and (A5.1). For laser applications, the mode of interest is typically a TE mode which is very well confined to the active region, implying that $\bar{n} \approx n$, and hence $\bar{g} \approx g$. As a result, it is quite common to simply use Eq. (A5.1) with the real material gain, g, in defining the modal gain. This procedure typically introduces little error. To be completely rigorous, however, Eq. (A5.8) (or Eq. (A5.7) if there are axial variations in the waveguiding structure) should be used.

A5.3 MODAL GAIN AND CONFINEMENT FACTORS

Figure 2.6 illustrates a laser cavity in which the gain region intersects the optical mode over some portion of its width and length. Figure A5.1 gives a slightly more detailed version of a typical laser cavity. The ith component of the electric field in such a cavity can be written as

$$\mathscr{E}_i(x,y,z) = U_i(x,y)\sqrt{2}\cos\beta z, \tag{A5.9}$$

where $U_i(x,y)$ is the normalized transverse electric field profile, derived in Appendix 3 for a three-layer slab, and $\beta(= 2\pi/\lambda_z \equiv 2\pi\bar{n}/\lambda)$ is the z-component of the propagation constant k. In writing (A5.9), we have neglected the imaginary part of β, and hence are ignoring the growth of the fields in the cavity (see

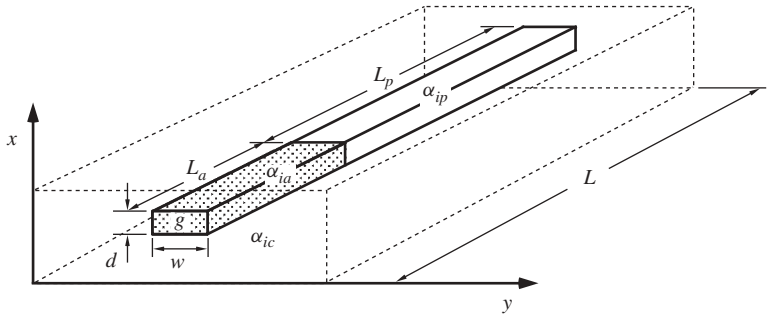


FIGURE A5.1: General laser cavity with active and passive axial sections.

Section A5.5). The simple cosine function also indicates that an infinite standing wave ratio has been assumed in the z-direction. Such a situation will actually exist only over ranges in the cavity where the two counterpropagating waves are equal in magnitude (see Fig. A5.3). For a cavity with large mirror reflectivities ($r > 0.9$), such as in a vertical cavity laser, this range extends over the entire cavity. For a laser with low mirror reflectivities ($r > 0.2$), an infinite standing wave ratio exists only over a small portion of the cavity. In either case, if the axial (z-direction) integration is over many wavelengths, the contribution from this factor averages out.

If we assume that the group velocity and propagation constant are the same (or very similar) in both active and passive sections then we can use Eq. (A5.8) to define the modal gain. Plugging Eq. (A5.9) into Eq. (A5.8) and considering the $e^{+j\beta z}$ and $e^{-j\beta z}$ components of the $\cos \beta z$ separately in the denominator, we obtain

$$\langle g \rangle = \frac{\displaystyle\iiint_{xyz} \bar{g}(x,y,z)|U(x,y)|^2 2\cos^2 \beta z \, dx \, dy \, dz}{L \displaystyle\iint_{xy} |U(x,y)|^2 \, dx \, dy}. \tag{A5.10}$$

We have dropped the i subscript on the transverse field pattern, assuming that only one electric field component exists. If $U(x, y)$ is normalized, the denominator integral will be unity. However, we shall carry it along to include cases where it is not normalized.

In applying Eq. (A5.10) to the laser cavity in Fig. A5.1, a few further assumptions are usually made. For in-plane lasers, the gain can be assumed to be constant within the active region, which has transverse thickness d, width w, and length L_A. Therefore, the gain can be removed from the integral, replacing the limits of integration by these dimensions. Assuming the origin to be in the center of the active region, and also resetting $\bar{g} = gn/\bar{n}$ (where n is the refractive index of the active material), we have

$$\langle g \rangle = g \frac{n}{\bar{n}} \frac{\displaystyle\int_{-L_a/2}^{L_a/2} \int_{-w/2}^{w/2} \int_{-d/2}^{d/2} |U(x,y)|^2 2\cos^2 \beta z \, dx \, dy \, dz}{L \displaystyle\iint_{xy} |U(x,y)|^2 \, dx \, dy}. \tag{A5.11}$$

The integration over x and y yields a transverse confinement factor Γ_{xy} and a transverse modal gain, $\langle g \rangle_{xy} = \Gamma_{xy} g$, which defines the incremental rate of growth of the fields in the active section (aside from any losses present). The integration over z gives an axial confinement factor Γ_z and the overall cavity modal gain:

$$\langle g \rangle = \Gamma_z \langle g \rangle_{xy} = \Gamma_{xy} \Gamma_z g, \tag{A5.12}$$

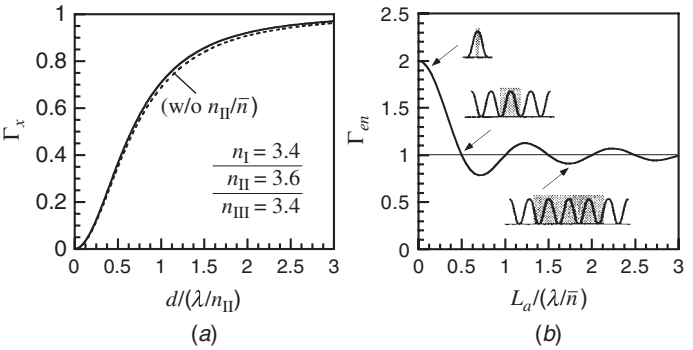


FIGURE A5.2: (a) Transverse confinement factor for a typical three-layer slab waveguide versus normalized waveguide thickness. (b) Enhancement factor vs. normalized active length (where $\bar{n}$ is the effective index of the guide). The insets display the standing wave pattern in the cavity and its overlap with the active material.

where

$$\Gamma_{xy} = \frac{n}{\bar{n}} \frac{\int_{-w/2}^{w/2} \int_{-d/2}^{d/2} |U(x,y)|^2 \, dx \, dy}{\iint\limits_{xy} |U(x,y)|^2 \, dx \, dy}, \tag{A5.13}$$

$$\Gamma_z = \frac{1}{L} \int_{-L_a/2}^{L_a/2} 2\cos^2 \beta z \, dz = \frac{L_a}{L} \left[1 + \frac{\sin \beta L_a}{\beta L_a} \right]. \tag{A5.14}$$

The ratio $n/\bar{n}$ in Eq. (A5.13) is usually close to unity and is typically neglected. The axial confinement factor can be further separated into a fill factor, $\Gamma_{fill} = L_a/L$, and an enhancement factor, $\Gamma_{enh} = 1 + \sin(\beta L_a)/\beta L_a$, such that $\Gamma_z = \Gamma_{fill}\Gamma_{enh}$. Figure A5.2 plots the one-dimensional transverse confinement factor, Γ_x, for a symmetric three-layer slab waveguide geometry (with and without $n/\bar{n}$), and Γ_{enh} in general.

A5.4 INTERNAL MODAL LOSS

Equation (A5.7) with g replaced by α_i gives the correct expression for the internal modal absorption loss, $\langle \alpha_i \rangle$. Unfortunately, it does not simplify quite as much as the modal gain, even if it is taken as constant within certain regions. This is because the passive loss has a value everywhere, unlike the gain. However, because we have already defined the transverse and axial confinement factors, we can still construct an abbreviated form for the modal loss for this particular case, if we assume that the loss is constant within the active and passive channel waveguide regions as well as in all of the surrounding cladding material. We

take the respective loss values as $\alpha_{ia}, \alpha_{ip}, \alpha_{ic}$, as labeled in Fig. A5.1. Then using Eq. (A5.13) we can define *transverse* modal averages within both active and passive sections as

$$\langle \alpha_i \rangle_{xy}^{a} = \Gamma_{xy}\alpha_{ia} + (1 - \Gamma_{xy})\alpha_{ic},$$
$$\langle \alpha_i \rangle_{xy}^{p} = \Gamma_{xy}\alpha_{ip} + (1 - \Gamma_{xy})\alpha_{ic}, \tag{A5.15}$$

The complete modal average over the entire cavity (assuming the group velocity is similar in both sections) is then found using Eq. (A5.14):

$$\langle \alpha_i \rangle = \Gamma_z \langle \alpha_i \rangle_{xy}^{a} + (1 - \Gamma_z)\langle \alpha_i \rangle_{xy}^{p}. \tag{A5.16}$$

Combining these equations, we obtain:

$$\langle \alpha_i \rangle = \Gamma\alpha_{ia} + \Gamma_{xy}(1 - \Gamma_z)\alpha_{ip} + (1 - \Gamma_{xy})\alpha_{ic}, \tag{A5.17}$$

where the cavity confinement factor, $\Gamma = \Gamma_{xy}\Gamma_z$. In some cases, the modal loss expression can become even more complicated than listed here. For example, the P and N cladding regions on either side of the active region may have different losses. Furthermore, in separate confinement lasers (including quantum-well lasers), there are additional waveguide (barrier) layers that have different losses than the active and cladding materials. In these cases the transverse confinement factor must be calculated for every layer that has a unique loss value. In VCSELs, the loss is often a function of the axial direction. In such cases, the axial confinement factor for the internal loss must be calculated by integrating Eq. (A5.14), including $\alpha_i(z)$ as well as additional axial variations of the standing wave pattern such as an exponential decay into a distributed Bragg reflector, as considered in Chapter 3. Because the exact form for modal loss is very cavity-specific, we will continue to refer to cavity modal loss as $\langle \alpha_i \rangle$ in most expressions.

A5.5 MORE EXACT ANALYSIS OF THE ACTIVE/PASSIVE SECTION CAVITY

Sections A5.3 and A5.4 make two simplifying assumptions in defining the various axial averages: (1) the propagation constant β in Eq. (A5.9) is purely real, and (2) the group velocities in the active and passive sections are the same. In this section we will remove these two assumptions. However, to retain relatively simple expressions we must assume that no reflections exist at the active-passive interface. For this case, $\bar{n}|\mathscr{E}_T|^2$ is preserved across the interface, and multiple bounces in the cavity are eliminated. Standing wave effects are also ignored since the following analysis applies primarily to in-plane lasers. However, footnotes are provided to indicate where standing wave effects would alter the result. We will first examine the axial confinement factor and then describe

the threshold condition and differential efficiency more carefully than considered in Chapter 2.

A5.5.1 Axial Confinement Factor

If we include the growth of the fields within the active section (i.e., we allow β to be complex in Eq. (A5.9), the axial confinement factor (neglecting coherent effects) is no longer equal to the simple geometric fill factor, L_a/L. This can understood by examining Fig. A5.3, which illustrates the power flowing back and forth in a laser cavity assuming the active section provides net gain and the passive section provides loss. The area under the power flow curves divided by the group velocity in each section is proportional to the total energy in the cavity. From the figure it is evident that the fractional area contained in the active section is nontrivial when the fields display exponential growth and decay characteristics.

To determine the fractional energy contained in the active section (i.e., the axial confinement factor), we need to integrate the power curves along z. The functionality of each curve is provided in Fig. A5.3. Performing the integrations (neglecting coherent effects) and weighting the area in each section by the group velocity in

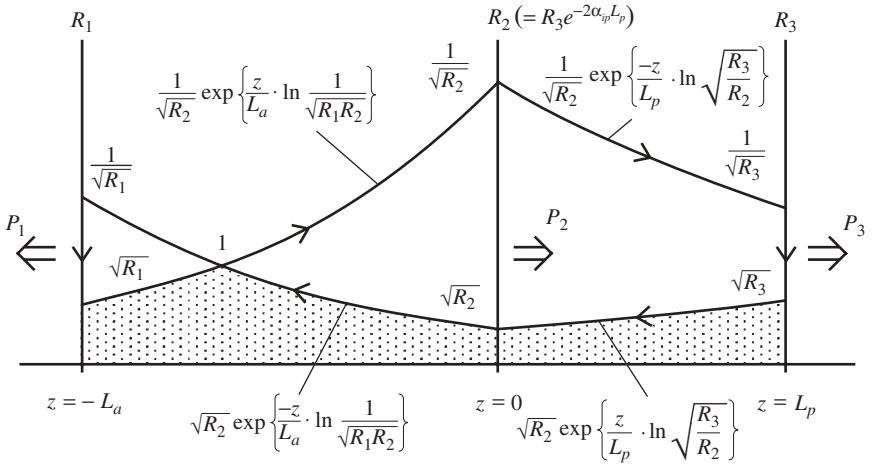


FIGURE A5.3: Illustration of power flow in a laser cavity with active and passive sections, assuming no reflections at the active–passive interface. R_2 defines the *effective* power reflectivity of the passive mirror (R_3) as seen by the active section at $z=0$. The magnitudes of each power curve at the interfaces are indicated relative to the power level at the crossing point inside the active section. The functionality of each curve is also provided. In the active section, the threshold condition is used to express the *net* growth rate as: $\langle g\rangle_{xy} - \langle\alpha_i\rangle_{xy}^{a} \rightarrow \ln(1/\sqrt{R_1R_2})/L_a$. In the passive section, the definition of R_2 is used to express the decay rate as $\alpha_{ip} \equiv \langle\alpha_i\rangle_{xy}^{p} \rightarrow \ln(\sqrt{R_3/R_2})/L_p$.

accordance with Eq. (A5.7), the axial confinement factor becomes[1]

$$\Gamma_z = \frac{\frac{P_1+P_2}{\ln\sqrt{1/R_1R_2}}L_a/v_{ga}}{\frac{P_1+P_2}{\ln\sqrt{1/R_1R_2}}L_a/v_{ga} + \frac{P_2-P_3}{\ln\sqrt{R_3/R_2}}L_p/v_{gp}}, \tag{A5.18}$$

where

$$P_i \equiv \frac{1}{\sqrt{R_i}} - \sqrt{R_i} = \text{relative power out mirror } i. \tag{A5.19}$$

Equation (A5.18) is equivalent to $\Gamma_z = W_E^a/W_E$, where W_E^a is the total energy within the active section. If the passive section has no loss, then the second term in the denominator (i.e., W_E^p normalized to the power at the crossing point) reduces to $(1/\sqrt{R_3} + \sqrt{R_3})L_p/v_{gp}$.

Typically the net transverse modal gain in the active section (i.e., $\langle g\rangle_{xy} - \langle\alpha_i\rangle_{xy}^a$) is larger than the loss in the passive section $\langle\alpha_i\rangle_{xy}^p$, implying a larger bowing of the power flow curves and hence a larger reduction of the area under the curves in the active section. As a result, the fill factor tends to be smaller than the geometric fill factor. However, the more linear the curves are, the more Eq. (A5.18) resembles the geometric fill factor. For example, linear curves imply low gain or high mirror reflectives. If we set $R_i = 1 - T_i$ and examine the limits as $R_1, R_2, R_3 \to 1$, we find that $P_i \to T_i, \ln[1/R_1R_2] \to T_1 + T_2$, and $\ln[R_3/R_2] \to T_2 - T_3$. The axial confinement factor in this limit reduces to:

$$\Gamma_z \to \frac{\bar{n}_{ga}L_a}{\bar{n}_{ga}L_a + \bar{n}_{gp}L_p}. \quad \text{(roughly linear growth/decay)} \tag{A5.20}$$

This limiting form is convenient to use even though it usually overestimates the axial confinement factor slightly. Note that the lengths are weighted by the group index. For monolithic active and passive sections, the group index does not vary much, and the group indices can be cancelled out of the expression. However, if an external cavity is coupled to a laser, then $\bar{n}_{gp} \to 1$ and the inclusion of the group indices becomes important.

To include reflections at the active–passive interface in the derivation of the axial confinement factor, one must determine the power distribution throughout the entire cavity and then calculate the fraction of energy contained in the active section.

A5.5.2 Threshold Condition and Differential Efficiency

With the axial confinement factor defined, we can now determine the threshold condition and differential efficiency of the active/passive section cavity shown in Fig. A5.3. The threshold condition can be found by setting the energy generation

[1]If the active section is very short, Eq. (A5.18) can be multiplied by Γ_{enh} as defined in Eq. (A5.14) to approximately account for standing wave effects.

rate equal to the loss rates created by absorption and power coupled out of the cavity (i.e., $P_1 + P_3$). From Eqs. (A5.2) and (A5.12) we can write the total energy generation rate per unit time as $\Gamma_z v_{ga} \langle g \rangle_{xy} W_E$. Defining a similar term for the absorption losses per unit time and dividing by the total energy W_E, we obtain

$$\Gamma_z v_{ga} \langle g \rangle_{xy} = \Gamma_z v_{ga} \langle \alpha_i \rangle_{xy}^a + (1 - \Gamma_z) v_{gp} \langle \alpha_i \rangle_{xy}^p + \frac{P_1 + P_3}{W_E}. \tag{A5.21}$$

The last term corresponds to the fractional mirror loss per unit time. Alternatively, by following the power curves through one round-trip of the cavity similar to the procedure outlined in Chapter 2 (see Section 2.5 and Eq. (2.23)), we obtain the more standard version of the threshold condition:

$$\langle g \rangle_{xy} L_a = \langle \alpha_i \rangle_{xy}^a L_a + \langle \alpha_i \rangle_{xy}^p L_p + \ln \frac{1}{\sqrt{R_1 R_3}}. \tag{A5.22}$$

With this version, if we divide by L we obtain $\Gamma_z = L_a/L$ regardless of any exponential field growth in the cavity or differences in group velocity between the active and passive sections.[2] In contrast, Γ_z in Eq. (A5.21) is the actual axial energy confinement factor (A5.18). The two versions of the threshold condition are equivalent but appear different because the former equates the total generation and loss rates, whereas the latter simply equates the accumulated exponential growth and decay factors. So while the latter version Eq. (A5.22) is more convenient to use, it lacks a one-to-one correspondence with the generation and loss rates of the active/passive section laser.

For example, the $(1/L) \ln \sqrt{1/R_1 R_3}$ term in Eq. (A5.22) is usually associated with the mirror loss. However, with Eq. (A5.21) we can identify the true mirror loss rate per unit *active* length as: $\alpha_m = (P_1 + P_3)/(v_{ga} W_E)$. Using Eq. (A5.18) to set $W_E = W_E^a / \Gamma_z$, the true mirror loss becomes[3]

$$\alpha_m = \frac{P_1 + P_3}{P_1 + P_2} \frac{\Gamma_z}{L_a} \ln \frac{1}{\sqrt{R_1 R_2}}. \tag{A5.23}$$

With no passive section, $\alpha_m \to (1/L_a) \ln \sqrt{1/R_1 R_2}$. Thus, the mirror loss deviates from the standard definition only with the addition of a passive section. With no loss in the passive section, $P_2 = P_3, R_2 = R_3$, and $\alpha_m \to (\Gamma_z/L_a) \ln \sqrt{1/R_1 R_3}$. With high mirror reflectivities and similar group velocities, $\Gamma_z \to L_a/L$. Combining these restrictions we recover the standard definition: $\alpha_m \to (1/L) \ln \sqrt{1/R_1 R_3}$. In

[2]In very short active sections, a more careful round-trip analysis including reflections at the active-passive interface(s) created by the gain discontinuity would yield $\Gamma_{enh} L_a/L$ instead of L_a/L.

[3]For very short active sections where standing wave effects are important, $\Gamma_z = \Gamma_{enh} W_E^a / W_E$ so that $\Gamma_z \to \Gamma_z / \Gamma_{enh}$ in the expression for the true mirror loss. In other words, standing wave enhancements to the gain do not affect the mirror loss.

a VCSEL which typically has small losses and high mirror reflectivities, $\alpha_m \rightarrow \frac{1}{2}(T_1 + T_3)/(L_a + L_p v_{ga}/v_{gp})$.

Because $\Gamma_z \neq L_a/L$, we also know that the total generation rate is not exactly $\langle g \rangle_{xy} L_a/L$ and the total absorption loss rate is not exactly $\langle \alpha_i \rangle_{xy}^a L_a/L + \langle \alpha_i \rangle_{xy}^p L_p/L$ as one is tempted to conclude from Eq. (A5.22). This observation has no effect on the threshold value of $\langle g \rangle_{xy}$ since both Eqs. (A5.21) and (A5.22) predict the same value. However, it does become important when defining the differential efficiency, which involves the ratio of these terms. Using Eqs. (A5.12) and (A5.23), the differential efficiency of the active/passive section cavity becomes

$$\eta_d = \eta_i \frac{\alpha_m}{\langle g \rangle} = \eta_i \frac{\ln \sqrt{1/R_1 R_2}}{\langle g \rangle_{xy} L_a} \frac{P_1 + P_3}{P_1 + P_2} = \eta_i \eta_{da} \eta_{dp}, \tag{A5.24}$$

where

$$\eta_{da} = \frac{\ln \sqrt{1/R_1 R_2}}{\langle g \rangle_{xy} L_a} = \frac{\ln \sqrt{1/R_1 R_2}}{\langle \alpha_i \rangle_{xy}^a L_a + \ln \sqrt{1/R_1 R_2}}, \tag{A5.25}$$

$$\eta_{dp} = \frac{P_1 + P_3}{P_1 + P_2} = \frac{(1 - R_1)/\sqrt{R_1} + (1 - R_3)/\sqrt{R_3}}{(1 - R_1)/\sqrt{R_1} + (1 - R_3 e^{-2\alpha_{ip} L_p})/\sqrt{R_3} e^{-\alpha_{ip} L_p}}. \tag{A5.26}$$

The active and passive section efficiencies, η_{da} and η_{dp}, were expanded using Eqs. (A5.22) and (A5.19) with $R_2 = R_3 e^{-2\alpha_{ip} L_p}$.

To more clearly see the high mirror reflectivity limit, we can rewrite Eq. (A5.26) using hyperbolic definitions and algebraic manipulations to obtain

$$\eta_{dp} = \frac{\sinh \frac{1}{2}[\ln \sqrt{1/R_1 R_3}]}{\sinh \frac{1}{2}[\langle \alpha_i \rangle_{xy}^p L_p + \ln \sqrt{1/R_1 R_3}]} \frac{\cosh \frac{1}{2}[\ln \sqrt{R_1/R_3}]}{\cosh \frac{1}{2}[\langle \alpha_i \rangle_{xy}^p L_p + \ln \sqrt{R_1/R_3}]}. \tag{A5.27}$$

For $R_1, R_3 \rightarrow 1, R_1 \sim R_3$, and $\alpha_{ip} L_p \ll 1$, we can replace the sinh functions by their arguments and set the cosh functions to one, which gives

$$\eta_{dp} \approx \frac{\ln \sqrt{1/R_1 R_3}}{\langle \alpha_i \rangle_{xy}^p L_p + \ln \sqrt{1/R_1 R_3}}, \tag{A5.28}$$

and

$$\eta_{da} \eta_{dp} \approx \frac{\ln \sqrt{1/R_1 R_3}}{\langle \alpha_i \rangle_{xy}^a L_a + \langle \alpha_i \rangle_{xy}^p L_p + \ln \sqrt{1/R_1 R_3}}. \tag{A5.29}$$

In this limit, the differential efficiency reduces to the ratio one would assume using the standard threshold condition Eq. (A5.22).

Another consideration is the fraction of power out of each facet. With $r_i = \sqrt{R_i}$ and $P_i = (1 - r_i^2)/r_i$ from Eq. (A5.19) we quickly find that

$$F_1 = \frac{P_1}{P_1 + P_3} = \frac{(1 - r_1^2)/r_1}{(1 - r_1^2)/r_1 + (1 - r_3^2)/r_3}, \tag{A5.30}$$

$$F_3 = \frac{P_3}{P_1 + P_3} = \frac{(1 - r_3^2)/r_3}{(1 - r_1^2)/r_1 + (1 - r_3^2)/r_3}. \tag{A5.31}$$

When F_1 is multiplied by $t_1^2/(1 - r_1^2)$ and F_3 is multiplied by $t_3^2/(1 - r_3^2)$ to account for possibly lossy mirrors, we obtain the same expression (3.30) as derived in Chapter 3. Thus, the fraction of power out is not affected by the lossy passive section.

In practice, Γ_z does not usually deviate substantially from the geometric fill factor. As a result, the terms comprising the standard version of the threshold condition (A5.22) usually represent reasonable approximations to the generation and loss rates, implying that the differential efficiency is usually well approximated by Eq. (A5.29). It is for this reason that we prefer to use the simpler more intuitive Eqs. (A5.22) and (A5.29) throughout this book. However, there are certain situations where Eq. (A5.29) fails to predict the differential efficiency accurately.

For example, with $R_1 = 0.1$ and $R_3 = 1$, the approximate differential efficiency (A5.29) is within 5% of the exact Eq. (A5.26) for $\alpha_{ip} L_p \leq 1$. For uncoated facets ($R_1 = R_3 = 0.3$), it overestimates the exact value by ~10% for $\alpha_{ip} L_p = 0.5$, and ~30% for a larger loss of $\alpha_{ip} L_p = 1$. For $R_1 = 1$ and $R_3 = 0.1$, the overestimate is close to 25% for $\alpha_{ip} L_p = 0.5$, and is almost 60% for $\alpha_{ip} L_p = 1$. In general, the approximation gets worse for smaller values of R_3 and larger values of $\alpha_{ip} L_p$, but is not significantly affected by the value of R_1. Hence, for $\alpha_{ip} L_p \leq 0.5$ and $R_3 \geq 0.3$, the approximate differential efficiency (A5.29) is generally fairly accurate.

It should be noted in closing this section that if there are reflections at the active–passive interface, then Γ_z and all dependent expressions will be different than given here.

A5.6 EFFECTS OF DISPERSION ON MODAL GAIN

Finally we would like to consider parenthetically the effects of material and waveguide dispersion on the modal gain. Equation (A5.8) reveals that $\langle g \rangle \propto 1/\bar{n}$. Physically, this dependence reflects the fact that the effective index controls the angle at which plane waves bounce down the guide (see Chapter 7). The smaller the effective index, the more bounces the plane waves make per unit length. As a result, the mode effectively "sees" more active material per unit length, which shows up as an enhancement in the modal gain per unit length. In other words, waveguide dispersion slows down the mode, allowing more stimulated transitions to be acquired into the mode per unit length. Thus, the modal gain per unit length is *enhanced by waveguide dispersion*. As for material dispersion, we find that (A5.8) contains

no dependence on the overall group velocity, implying that $\langle g \rangle$ is *independent of material dispersion*.

The gain per unit time defined by $\langle G \rangle = v_g \langle g \rangle$, is inversely related to the group effective index since $\langle g \rangle$ is independent of $\bar{n}_g$. Now because $\bar{n}_g > n$ in most cases, we conclude that *waveguide and material dispersion reduce* the modal gain per unit time. The waveguide dispersion component of $\bar{n}_g$ is roughly compensated by the $1/\bar{n}$ implicit in $\langle g \rangle$. Neglecting waveguide dispersion, we still find that $\langle G \rangle$ is compromised by material dispersion. Physically this occurs because the local energy density in a dispersive medium is $\propto n_g n |\mathscr{E}|^2$ (where n_g is the group index associated with material dispersion exclusively). Thus, an increase in the group index for a given energy density compromises the electric field strength, which reduces the local stimulated emission rate. The result is a reduced modal gain per unit time in the cavity. The modal gain per unit length does not suffer this consequence because as n_g increases, the wave moves slower, allowing more stimulated transitions to be acquired per unit length. The reduction in field strength and the slowing down of the mode offset each other, leaving the modal gain per unit length independent of material dispersion, as concluded above.

APPENDIX SIX

Einstein's Approach to Gain and Spontaneous Emission

A6.1 INTRODUCTION

Equation (2.14) gives us the relationship between gain and the stimulated recombination rate,

$$R_{st} = v_g g N_p. \tag{A6.1}$$

As shown in Fig. 1.6, the net stimulated rate $R_{st} = R_{21} - R_{12}$, is the stimulated emission less the stimulated absorption of photons. Thus, we wish to calculate R_{st}, from which the gain, g, can be obtained. As also suggested in Chapters 1 and 2, the stimulated emission and absorption rates depend on the number of available electronic states and their probability of occupation in both the conduction and valence bands for the transitions to occur. The unknowns are the multiplicative rate constants. Once these are determined, we can calculate R_{st} and the gain, g.

Einstein gave us a technique to calculate these rates without delving too deeply into the details of the stimulated emission physics. His technique is to determine the desired rate constants under a particular set of boundary conditions. Once obtained, however, these constants are generally applicable to other situations. As shown in Fig. A6.1, the medium of interest is placed in a closed cavity, which has neither inputs nor outputs, and held under thermal equilibrium. Then, a dynamic balance equation can be set up that expresses the desired rates in terms of the equilibrium

Diode Lasers and Photonic Integrated Circuits, Second Edition.
Larry A. Coldren, Scott W. Corzine, and Milan L. Mašanović.
© 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

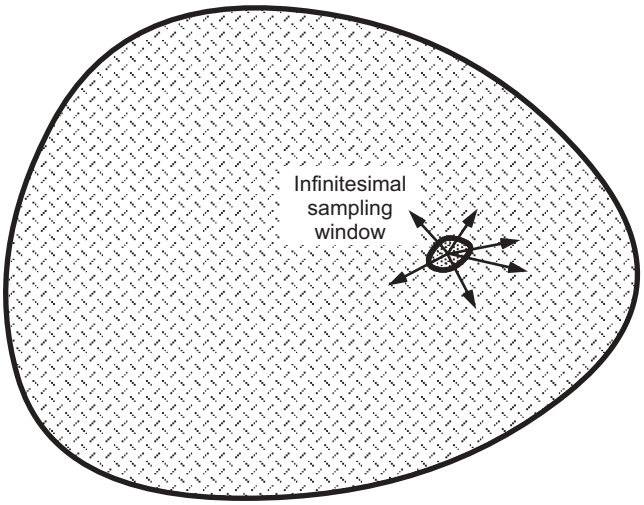


FIGURE A6.1: Schematic of an arbitrary closed cavity with negligible energy in or out. Contents must be in thermal equilibrium.

optical energy density, $W(\nu)$. Since this is known (from Appendix 4), the rate constants can be determined.

In this situation we must include all carrier recombination mechanisms in writing a dynamic balance equation. The nonradiative rates generate heat, which is naturally taken into account in this closed system. By one means or another, they must saturate in equilibrium. Thus, in equilibrium, for a pair of energies, E_2 and E_1, in the conduction and valence bands, respectively, we have

$$\frac{dN_p}{dt} = 0 = R_{21} + R_{sp,21} - R_{12}, \tag{A6.2}$$

where the first two terms represent electrons recombining by either stimulated or spontaneous processes, respectively, and the last term represents electrons being generated by stimulated absorption. The 21 subscript on R_{sp} distinguishes this two-level spontaneous rate from the net recombination between two bands in a semiconductor as we have considered elsewhere. We could summarize Eq. (A6.2) by saying that the downward transition rate (i.e., conduction to valence band) must equal the upward transition rate.

The equilibrium occupation probability at some temperature is given by the Fermi function introduced in Appendix 2

$$f_i = \frac{1}{e^{(E_i - E_F)/kT} + 1}, \tag{A6.3}$$

where $i = 1$ or 2 for the involved transition energies in the valence or conduction bands, respectively. That is, $f_2 \equiv$ fraction of states filled at E_2, and $f_1 \equiv$ fraction of states filled at E_1. Figure A6.2 illustrates the various energy levels for reference.

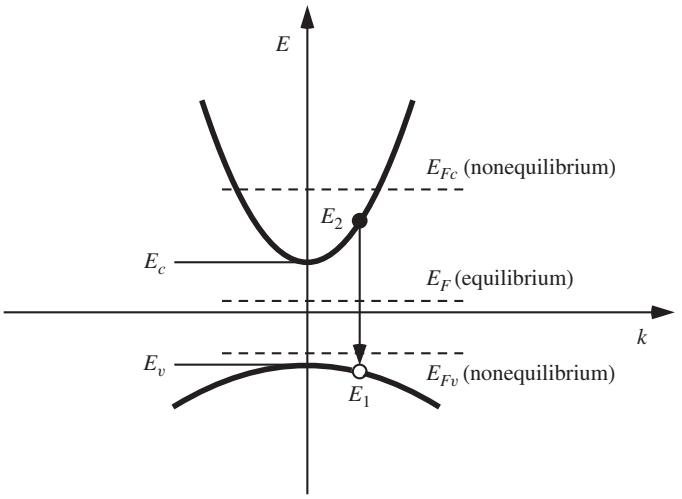


FIGURE A6.2: Energy versus momentum schematic illustrating a transition between two energy states in the conduction and valence bands, respectively. Quasi-Fermi levels for both equilibrium and nonequilibrium carrier densities are also illustrated.

In Appendix 1, we also defined a density of states, $\rho(E)$, to describe the distribution of states in a band. The number of states are equally distributed in k-space, but by integrating $\rho(E)$ over some energy range, the number of states in that range is obtained. As shown in Figure A6.2, the radiative recombination of an electron and hole involves states in the conduction and valence bands with the same k-vector. That is, both energy ($E_{21} = h\nu_{21}$) and momentum ($\hbar\mathbf{k}$-electron $\approx \hbar\mathbf{k}$-hole) conservation must be satisfied. (As discussed in Chapter 4, the photon momentum is negligible.) Thus, we can consider the *density of state pairs, or a reduced density of states*, $\rho_r(E_{21})$, in calculating emission at E_{21} from electron-hole recombination. We shall explicitly derive $\rho_r(E_{21})$ for parabolic bands to obtain an analytic expression, however, the concept is entirely general.

With reference to Fig. A6.2 and using the results of Appendix 1, we can express the transition energy difference, $E_2 - E_1 = E_{21}$ as

$$E_{21} = E_g + \frac{\hbar^2 k^2}{2m_c^*} + \frac{\hbar^2 k^2}{2m_v^*} = E_g + E' \frac{m_v^* + m_c^*}{m_v^*}, \tag{A6.4}$$

where $E' = \hbar^2 k^2 / 2m_c^*$ for parabolic bands. For relatively thick active regions (bulk), we found from Eq. (A1.39) in Appendix 1 that

$$\rho(E')\, dE' = \frac{1}{2\pi^2} \left[\frac{2m_c^*}{\hbar^2} \right]^{3/2} (E')^{1/2}\, dE'. \tag{A6.5}$$

Solving for E' from Eq. (A6.4), and forming dE', we have

$$E' = \frac{m_v^*}{m_c^* + m_v^*}(E_{21} - E_g),$$
$$dE' = \frac{m_v^*}{m_c^* + m_v^*}\, dE_{21}. \tag{A6.6}$$

Finally, plugging Eq. (A6.6) into Eq. (A6.5), we obtain the desired reduced density of states,

$$\rho_r(E_{21}) = \frac{1}{2\pi^2}\left[\frac{2m_r^*}{\hbar^2}\right]^{3/2}(E_{21} - E_g)^{1/2}, \tag{A6.7}$$

where $m_r^* = m_c^* m_v^*/(m_c^* + m_v^*)$.

A6.2 EINSTEIN *A* AND *B* COEFFICIENTS

The general approach of Einstein was to assign rate constants to the three radiative processes appearing in Eq. (A6.2), with the assumption that these rates must be proportional to the carrier density. These were written empirically as $R_{sp,21} = A_{21}N_2, R_{21} = B_{21}\,W(\nu)N_2$, and $R_{12} = B_{12}\,W(\nu)N_1$. Generally speaking, the A rate constant is associated with spontaneous processes, whereas the B rate constants are associated with stimulated processes and hence are weighted by the radiation *spectral* density, $W(\nu)$, which was introduced in Appendix 4. In Einstein's day, most radiative transitions of interest took place between atoms with very isolated, sharp energy levels. The carrier densities therefore referred to the density of *atoms* with electrons in either energy level 1 or 2. In the current context, we must interpret these definitions somewhat differently because in semiconductors, the energy levels are neither isolated nor sharp.

To deal with the continuous nature of energy states in semiconductors, we restrict our attention to a differential population of state pairs existing between E_{21} and $E_{21} + dE_{21}$. Using the reduced density of states function derived earlier, the differential population available for producing downward transitions becomes

$$dN_2 = f_2(1 - f_1) \cdot \rho_r(E_{21})\, dE_{21}, \tag{A6.8}$$

where $dN_2 \equiv$ number of state pairs per unit volume between E_{21} and $E_{21} + dE_{21}$ available to interact with photons near E_{21}, in which the upper state is full and the lower state is empty. Similarly, the differential population available for producing upward transitions becomes

$$dN_1 = f_1(1 - f_2) \cdot \rho_r(E_{21})\, dE_{21}, \tag{A6.9}$$

where $dN_1 \equiv$ number of state pairs per unit volume between E_{21} and $E_{21} + dE_{21}$ available to interact with photons near E_{21}, in which the lower state is full and the upper state is empty.

Another factor we must consider in semiconductors is that the states in $\rho_r(E_{21})$ have a finite lifetime due to collisions with other electrons and phonons. As a result, a given differential population can actually appear over a small range of energies and thus interact with photons spanning some narrow energy range, ΔE_{21}. The probability of finding this population at energies away from E_{21} is characterized by some lineshape function, $\mathscr{L}(E - E_{21})$, which has a full-width half-maximum (FWHM) ΔE_{21} and is centered at E_{21} as shown in Fig. A6.3. The longer the lifetime of a given state, the narrower the spread in energy ΔE_{21}, and hence, the more chance there is of finding the state pair at E_{21}.

A direct consequence of the finite state lifetime is that when we consider interactions between our differential population and photons of a given energy, $h\nu$, we must somehow account for the fact that the population only spends a fraction of its time at that photon energy. In other words, we must weight our differential population by the probability of finding that population between photon energies $h\nu$ and $h\nu + h\,d\nu$, which is given by $\mathscr{L}(h\nu - E_{21})\,h\,d\nu$. The differential population "seen" by photons at energy, $h\nu$, therefore becomes

$$\begin{aligned} dN_2 &\rightarrow dN_2 \cdot \mathscr{L}(h\nu - E_{21})\,h\,d\nu, \\ dN_1 &\rightarrow dN_1 \cdot \mathscr{L}(h\nu - E_{21})\,h\,d\nu. \end{aligned} \tag{A6.10}$$

These are the forms for N_1 and N_2 that we must use when analyzing semiconductors.

With our differential populations defined, we can now make use of Einstein's *A* and *B* coefficients. The *differential* downward transition rates created by our differential dN_2 population can be written as

$$dR_{21} + dR_{sp,21} = \int [dN_2 W(\nu) B_{21} + dN_2 A_{21}] \mathscr{L}(h\nu - E_{21})\,h\,d\nu. \tag{A6.11}$$

The integral over photon frequency is necessary to include contributions that $W(\nu)$ makes over the full range of energies that dN_2 is spread over.

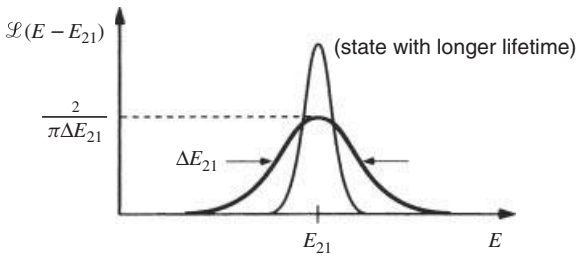


FIGURE A6.3: Plot of lineshape function versus transition energy. We require that it be normalized so that its area equals one (the state must exist somewhere!). (The peak probability shown assumes a Lorentzian distribution.)

The differential transition rates on the left-hand side of (A6.11) have units of [volume^{-1}time^{-1}]. The population density, dN_2, has units of [volume^{-1}]. Therefore, the spontaneous recombination rate constant, A_{21}, has units of [time^{-1}] and is often expressed as an inverse spontaneous lifetime, $1/\tau_{sp}$. In atomic systems, A_{21} does represent the inverse of the spontaneous lifetime of energy level 2. However, in the present context, A_{21} is associated with the spontaneous lifetime (which we will denote as τ_{sp}^{21}) of only the differential population, dN_2, and is *not* equal to the entire band-to-band spontaneous lifetime, τ_{sp}. Later we will determine the relationship between the two-level lifetime τ_{sp}^{21} and τ_{sp}. For the stimulated term, we note that $W(\nu)$ represents the optical energy per unit volume per unit frequency. Therefore, the units of the stimulated recombination rate constant, B_{21}, are [(volume $\times$ frequency)/(energy $\times$ time)].

The differential upward transition rate created by our differential dN_1 population can be written as

$$dR_{12} = \int [dN_1 W(\nu) B_{12}] \mathscr{L}(h\nu - E_{21})\, h\, d\nu, \tag{A6.12}$$

where B_{12} is the stimulated generation rate constant and has the same units as B_{21}.

A6.3 THERMAL EQUILIBRIUM

Einstein's approach allows us to relate the three rate constants appearing in Eqs. (A6.11) and (A6.12) in a straightforward manner. Under thermal equilibrium, we can set (A6.11) equal to (A6.12) according to (A6.2), which we assume holds for differential rates as well as the integrated rates. To remove the integrals over photon frequency, we note that equilibrium blackbody radiation is broadband and varies little over typical linewidths, ΔE_{21}, associated with the lineshape function. This allows us to treat the lineshape as a delta function, or $h\mathscr{L}(h\nu - E_{21}) \rightarrow \delta(\nu - \nu_{21})$. In other words, we simply evaluate all terms under the integral at ν_{21}. The balance equation then reduces to

$$dN_2 W(\nu_{21}) B_{21} + dN_2 A_{21} = dN_1 W(\nu_{21}) B_{12}. \tag{A6.13}$$

Rearranging, we obtain

$$\frac{dN_2}{dN_1} = \frac{W(\nu_{21}) B_{12}}{W(\nu_{21}) B_{21} + A_{21}}. \tag{A6.14}$$

Alternatively, from Eqs. (A6.8) and (A6.9), using (A6.3) and $E_{21} = h\nu_{21}$, we find the simple result

$$\frac{dN_2}{dN_1} = \frac{f_2(1 - f_1)}{f_1(1 - f_2)} = e^{-h\nu_{21}/kT}. \tag{A6.15}$$

Setting Eq. (A6.15) equal to (A6.14) and solving for $W(\nu_{21})$, we obtain

$$W(\nu_{21}) = \frac{A_{21}/B_{21}}{(B_{12}/B_{21})e^{h\nu_{21}/kT} - 1}. \tag{A6.16}$$

From Appendix 4, the blackbody radiation formula that defines the spectral density of photons under thermal equilibrium is given by

$$W(\nu_{21}) = \frac{\rho_0(\nu_{21})h\nu_{21}}{e^{h\nu_{21}/kT} - 1}. \tag{A6.17}$$

Comparing Eq. (A6.16) with Eq. (A6.17), we see that both can be true for all temperatures only if the following two equalities hold:

$$B_{12} = B_{21}, \tag{A6.18}$$

$$A_{21} = \rho_0(\nu_{21})h\nu_{21} \cdot B_{21}. \tag{A6.19}$$

Equation (A6.18) reveals that stimulated emission and stimulated absorption are truly complementary processes associated with the same rate constant. Perhaps more significantly, Einstein's approach establishes a fundamental link between stimulated and spontaneous emission processes through Eq. (A6.19). Thus, by analyzing the system under thermal equilibrium, Einstein's approach allows us to reduce the three differential rate constants to one independent constant, B_{21}. We could have designated A_{21} as the independent constant, however, too often this leads to the *incorrect* conclusion that B_{21} is inversely dependent on the density of optical modes, $\rho_0(\nu_{21})$. More correctly, we should view B_{21} as the rate constant of a single optical mode, and A_{21} as this rate constant multiplied by the equivalent spectral density that induces spontaneous emission into the full density of optical modes near ν_{21}. It is interesting to note that the equivalent spectral density, $\rho_0(\nu_{21})h\nu_{21}$, implies one photon per optical mode. In Chapter 4, more insight into this observation is gained through a quantum mechanical perspective.

A6.4 CALCULATION OF GAIN

Now that we have established the connection between the three rate constants, we can leave the closed-system, thermal-equilibrium environment and proceed to calculate the gain for a *monochromatic* radiation field under *nonequilibrium* conditions. Under forward bias, the Fermi level in the active region splits into two quasi-Fermi levels to reflect the nonequilibrium electron and hole densities. The splitting corresponds roughly to the applied voltage. The nonequilibrium carrier densities are then calculated from

$$N = \int \rho_c(E)f_c(E)\,dE \approx P = \int \rho_v(E)[1 - f_v(E)]\,dE, \tag{A6.20}$$

where we have assumed $N \approx P$ because of negligible doping in the active region. The factors, $\rho_c(E)$ and $\rho_v(E)$, refer to the densities of states in the conduction and valence bands, respectively. Also, $f_c(E)$ and $f_v(E)$ are the Fermi functions, in which E_F is replaced by the quasi-Fermi levels, E_{Fc} and E_{Fv}, for the conduction and valence bands, respectively.

To calculate the gain we turn to Eq. (A6.1) which allows us to relate the gain to the net stimulated emission rate. However, we must keep in mind that contributions to the gain at a particular frequency of radiation, ν_0, will come from many differential populations distributed roughly over energies comparable to the lineshape width, ΔE_{21}. The gain contributed by each population can be written as

$$\begin{aligned} dg(h\nu_0) &= \frac{dR_{st}}{v_g N_p} \\ &= \frac{1}{v_g N_p} \int W(\nu) B_{21} [dN_2 - dN_1] \mathscr{L}(h\nu - E_{21})\, h\, d\nu. \end{aligned} \tag{A6.21}$$

The second equality is obtained by setting $dR_{st} = dR_{21} - dR_{12}$ and using Eqs. (A6.11) and (A6.12) with $B_{12} = B_{21}$. For a monochromatic field, $W(\nu) \to h\nu_0 N_p \delta(\nu - \nu_0)$, where the strength of the delta function is equal to the energy density of the field and ν_0 is the frequency of the wave. With this substitution, the integral reduces to evaluating all photon frequency-dependent factors at ν_0, and we are left with

$$dg(h\nu_0) = \frac{h\nu_0}{v_g} h B_{21} [dN_2 - dN_1] \mathscr{L}(h\nu_0 - E_{21}). \tag{A6.22}$$

The appearance of the lineshape function reminds us that the further away the differential population is in transition energy from the photon energy, the less contribution it makes to the gain at that frequency.

The total gain at $h\nu_0$ is found by integrating dg over all existing populations of state pairs which might possibly interact with the field. Expanding the differential populations in Eq. (A6.22) using Eqs. (A6.8) and (A6.9), simplifying the Fermi factors, and integrating over all possible transition energies, we obtain

$$g(h\nu_0) = \frac{h\nu_0}{v_g} h \int B_{21} \rho_r(E_{21}) (f_2 - f_1) \mathscr{L}(h\nu_0 - E_{21})\, dE_{21}. \tag{A6.23}$$

If the energy-dependent factors, $B_{21}\rho_r(E_{21})(f_2 - f_1)$, are slowly varying compared to the lineshape function, then we can set $\mathscr{L}(h\nu_0 - E_{21}) \to \delta(h\nu_0 - E_{21})$ and the gain expression reduces to

$$g_{21} \approx \frac{h\nu_{21}}{v_g} h B_{21} \rho_r(E_{21}) (f_2 - f_1), \quad (\Delta E_{21} \to 0) \tag{A6.24}$$

where E_{21} is evaluated at the photon energy of interest (i.e., $g_{21} \equiv g(h\nu_0 = E_{21})$).

Equation (A6.23) is the central result of this section. It reveals that the gain is directly proportional to the rate constant B_{21}, the reduced density of states, and the Fermi probability factors. It is immediately apparent from Eq. (A6.24) that to achieve positive gain, we must create enough electrons and holes to allow $f_2 > f_1$. This places a condition on the quasi-Fermi levels, which reduces to the requirement that the quasi-Fermi level separation be larger than the incident photon energy, or $\Delta E_F > h\nu_0$. Chapter 4 considers these issues in more detail.

To fully evaluate the gain, we still need to determine the rate constant, B_{21}. In lasers that use atomic transitions, a measurement of the spontaneous emission linewidth of a given transition can allow us to estimate A_{21} if the broadening of the line is dominated by the spontaneous emission lifetime, $\tau_{sp} = 1/A_{21}$. With this information, B_{21} can readily be determined using Eq. (A6.19). Thus, with Einstein's approach and this one simple measurement, the description of gain and spontaneous emission in atomic systems is completely self-contained.

Unfortunately the situation is not so simple in semiconductors since the spontaneous emission spectrum represents a superposition of transitions from all of our differential populations. The resulting broad emission spectrum prevents us from isolating the linewidth of just one differential population, and therefore prevents us from evaluating A_{21} (and hence, B_{21}) via direct experiment. The approach that must be followed in semiconductors is to estimate the transition rates using other more in-depth theories and then relate the resulting expressions back to B_{21} and A_{21}. Chapter 4 details the theory required to evaluate the transition rates explicitly from a more fundamental quantum mechanical analysis. An explicit expression for B_{21} will be presented there.

We can alternatively express the gain in terms of the spontaneous rate constant, or the two-level lifetime, $\tau_{sp}^{21} = 1/A_{21}$. Using Eq. (A6.19) in Eq. (A6.24), the gain takes the form

$$\begin{aligned} g_{21} &= \frac{A_{21}}{\rho_0(\nu_{21})}\frac{h}{v_g}\rho_r(E_{21})(f_2 - f_1) \\ &= \frac{\lambda_0^2}{8\pi n^2 \tau_{sp}^{21}} h\rho_r(E_{21})(f_2 - f_1), \end{aligned} \tag{A6.25}$$

where we have set $\rho_0(\nu_{21}) = 8\pi n^2/\lambda_0^2 v_g$ according to Eq. (A4.5), with $\lambda_0 \equiv c/\nu_{21}$. Although this expression is equivalent to Eq. (A6.24), it is very deceptive for two reasons.

First of all, a common mistake in the literature is to equate τ_{sp}^{21} with the band-to-band spontaneous lifetime, τ_{sp}, incorrectly linking the gain with the carrier lifetime. In fact, some go even further by linking A_{21} with the overall spontaneous emission bandwidth (analogous to atomic transitions)—a procedure that is completely misguided in semiconductors, but nevertheless encouraged by writing the gain in terms of A_{21}.

The second problem is that when written in this way, one might conclude that the gain varies inversely with the optical mode density. Only on more careful inspection

does one realize that implicit in the two-level lifetime, τ_{sp}^{21}, is an inverse dependence on the mode density (i.e., the higher the mode density, the shorter the lifetime). As a result, the product $\rho_0(\nu_{21})\tau_{sp}^{21}$ (and hence the gain) becomes independent of the optical mode density—a conclusion that is obvious from Eq. (A6.24). Reduced optical mode densities possible in very small VCSEL structures (or "microcavities") have lead some researchers to conclude using Eq. (A6.25) that the gain is increased as a result—again, a conclusion which is misguided. For these reasons, Eq. (A6.24), which more appropriately defines the gain in terms of the single-mode stimulated rate constant, B_{21}, is preferable.

Another issue we need to resolve is whether we should use Eq. (A6.24) or the more complex Eq. (A6.23) to evaluate the gain. The time between collisions for electrons in typical semiconductors is on the order of 0.1 ps, which leads to a FWHM of $\Delta E_{21} \approx 14$ meV (assuming a Lorentzian lineshape). At room temperature, this bandwidth is small enough that we can assume $f_2 - f_1$ to be roughly constant. Furthermore, B_{21} does not have a strong energy dependence. Therefore, our main concern lies with $\rho_r(E_{21})$. In "bulk" active regions, $\rho_r(E_{21})$ varies as $E^{1/2}$ and the rate of change can be neglected in comparison to the bandwidth of the lineshape function. In other words, at room temperature, Eq. (A6.24) can generally be used with bulk active regions.

For quantum wells, the reduced density of states can be found by using the reduced mass in Eq. (A1.45). It is zero up to the first allowed energy states in the conduction and valence bands where $n_z = 1$, i.e., $(E_{c1} - E_{h1})$. There it abruptly increases to

$$\rho_r(h\nu_{21}) = \frac{m_r^*}{\pi \hbar^2 d}, \quad E_{21} > (E_{c1} - E_{h1}), \tag{A6.26}$$

where it remains constant up to the point where $n_z = 2$. There it again increases by the same amount. Thus, it violates the assumptions about being slowly varying over the bandwidth ΔE_{21} made between Eq. (A6.23) and Eq. (A6.24). Therefore, we must use Eq. (A6.23). Unfortunately, the actual lineshape function is not well established. A Lorentzian is often used, but the results are somewhat nonphysical; so other, more complex functions have been developed to better fit the experimental data. The simple Lorentzian with an FWHM of ΔE_{21} takes the form

$$\mathscr{L}(E - E_{21}) = \frac{2/\pi \, \Delta E_{21}}{1 + 4(E - E_{21})^2/\Delta E_{21}^2}. \tag{A6.27}$$

The integration of Eq. (A6.23) with Eq. (A6.27) will smooth the discontinuities in $\rho_r(E_{21})$ that exist in quantum-well (as well as quantum-wire and box) lasers. However, the plateau gain levels obtained by inserting Eq. (A6.26) in Eq. (A6.24) will be correct, as long as we are $> \Delta E_{21}$ away from a step edge. The numerical gain calculations in Chapter 4 illustrate this behavior more quantitatively.

A common feature of all active materials (regardless of the lineshape broadening or reduced dimensionality) is that the gain increases from an initial unpumped absorption level given by $g_{21}(f_2 = 0, f_1 = 1)$, to a transparency gain value given

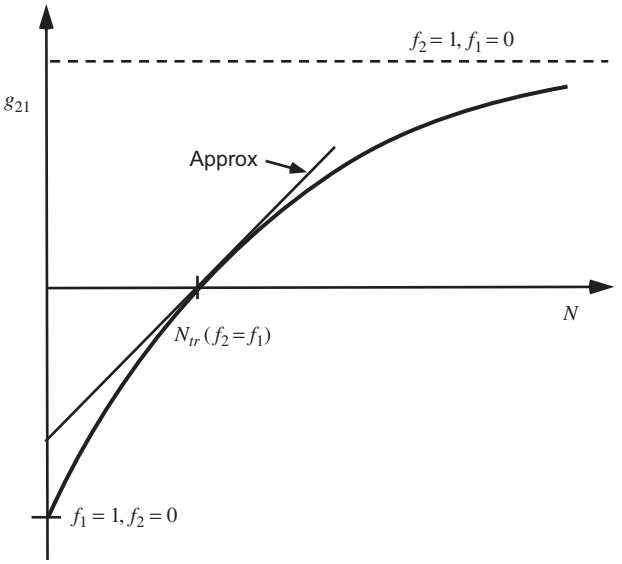


FIGURE A6.4: Illustration of gain versus carrier density. Straight-line approximation valid over limited ranges.

by $g_{21}(f_2 = f_1)$, finally approaching a saturation level of $g_{21}(f_2 = 1,\ f_1 = 0)$ (equal in magnitude to the unpumped absorption level) as more and more carriers are injected into the active region. In Fig. A6.4 we schematically illustrate this characteristic.

In Chapter 4 we shall find that this characteristic can be well approximated by a logarithmic function; however, for many situations only a small portion of the curve near and somewhat above the transparency point is of interest. In these cases, a straight-line approximation is often very useful. That is,

$$g_{21} = a(N - N_{tr}), \tag{A6.28}$$

where a is the differential gain, $\partial g/\partial N$, and N_{tr} is the transparency carrier density.

A6.5 CALCULATION OF SPONTANEOUS EMISSION RATE

In creating a given amount of gain in the semiconductor by increasing dN_2 relative to dN_1, we unfortunately end up creating a large amount of spontaneous emission over a relatively broad range of frequencies. This section deals with developing the relation for the spectrum of spontaneously emitted photons, which when integrated allows us to determine the total number of spontaneous photons being generated per second.

We start by defining the spontaneous emission within a small energy interval to be $R_{sp}^{h\nu}(h\nu)\,h\,d\nu$, where $R_{sp}^{h\nu}$ is the emission rate per unit energy per unit volume occurring at $h\nu$. As with the gain, we need to sum over all differential populations to determine the emission rate at a single frequency. The probability of dN_2 appearing at $h\nu$ is given by $\mathscr{L}(h\nu - E_{21})\,h\,d\nu$. Weighting dN_2 by this factor, multiplying by the spontaneous rate constant, A_{21}, and integrating over all state pairs, we obtain

$$R_{sp}^{h\nu}(h\nu)\,h\,d\nu = \int A_{21}[\mathscr{L}(h\nu - E_{21})\,h\,d\nu]\,dN_2. \tag{A6.29}$$

Canceling out $h\,d\nu$ on both sides, and expanding dN_2, we arrive at the desired result:

$$R_{sp}^{h\nu}(h\nu) = \int A_{21}\rho_r(E_{21})f_2(1-f_1)\mathscr{L}(h\nu - E_{21})\,dE_{21}. \tag{A6.30}$$

Equation (A6.30) reveals that the spontaneous emission spectrum is smoothed in the same manner as the gain spectrum discussed earlier. It is interesting to note that while we must have $E_{21} > E_g$ to have a nonzero reduced density of states, it *is* possible for $h\nu < E_g$ since $\mathscr{L}(h\nu - E_{21})$ can be nonzero for $h\nu - E_{21} < 0$. Thus, spontaneous emission can actually be observed at energies $\approx \Delta E_{21}/2$ below the bandgap. This reflects the uncertainty in the energy of states at the band edge, which results from the finite lifetimes of electrons in those states.

Again if $A_{21}\rho_r(E_{21})f_2(1-f_1)$ is slowly varying compared to the lineshape function, then we can set $\mathscr{L}(h\nu - E_{21}) \to \delta(h\nu - E_{21})$, and Eq. (A6.30) simplifies to

$$R_{sp}^{21} \approx A_{21}\rho_r(E_{21})f_2(1-f_1), \quad (\Delta E_{21} \to 0) \tag{A6.31}$$

where we have defined $R_{sp}^{21} \equiv R_{sp}^{h\nu}(h\nu = E_{21})$ in analogy with g_{21}. Comparing R_{sp}^{21} to g_{21} given in Eq. (A6.24), it is interesting to note that the two are quite similar. In fact, we can express the spontaneous emission at E_{21} in terms of the gain at E_{21} as follows:

$$R_{sp}^{21} = \frac{v_g}{h^2\nu_{21}}\frac{A_{21}}{B_{21}}\frac{f_2(1-f_1)}{(f_2-f_1)}g_{21} = \frac{1}{h}\rho_0(\nu_{21})\cdot v_g n_{sp} g_{21}, \tag{A6.32}$$

where we have made use of the relation between the rate constants expressed in Eq. (A6.19). We have also introduced the *population inversion factor*, which is defined as

$$n_{sp} = \frac{f_2(1-f_1)}{(f_2-f_1)} = \frac{1}{1-e^{(h\nu_{21}-\Delta E_F)/kT}}. \tag{A6.33}$$

The popular usage of the *sp* subscript originates from the fact that n_{sp} was initially referred to as the spontaneous emission factor, but was later changed to eliminate conflict with β_{sp}, which is also defined as the spontaneous emission factor. We now refer to n_{sp} as the population inversion factor because it is the semiconductor

laser equivalent to the ratio $N_2/(N_2 - N_1)$ encountered in atomic laser physics. In atomic systems, when $N_2 - N_1 > 0$, the population is said to be *inverted*, the ratio $N_2/(N_2 - N_1)$ is positive, and optical gain is achieved. Similarly, when $n_{sp} > 0$, a population inversion is established in the semiconductor, indicating a net optical gain.

If the quasi-Fermi level separation in the semiconductor is known, then using Eq. (A6.32), we can determine the spontaneous emission rate if we know the gain, or we can determine the gain if the spontaneous emission rate is known. In Chapter 4 a more thorough investigation of this fundamental relationship between spontaneous emission and gain is provided.

Now we wish to determine the total spontaneous emission rate, R_{sp}, by integrating (A6.30) over all photon energies:

$$\begin{aligned} R_{sp} &= \int R_{sp}^{h\nu}(h\nu)\, h\, d\nu \\ &= \iint A_{21}\rho_r(E_{21}) f_2(1-f_1)\mathscr{L}(h\nu - E_{21}) dE_{21}\, h\, d\nu \\ &\approx \int A_{21}\rho_r(E_{21}) f_2(1-f_1)\left[\int \mathscr{L}(h\nu - E_{21})\, h\, d\nu\right] dE_{21} \\ &= \int A_{21}\rho_r(E_{21}) f_2(1-f_1)\, dE_{21} = \int R_{sp}^{21}\, dE_{21}. \end{aligned} \tag{A6.34}$$

The third equality is found by inverting the order of integration and pulling out all terms not dependent on the photon frequency. The approximate sign is used here because A_{21} is not completely independent of photon frequency ($A_{21} \propto \nu$, from Eq. (A6.19) with $\rho_0(\nu) \propto \nu^2$ and $B_{21} \propto \nu^{-2}$; see Chapter 4). However, in comparison to $\mathscr{L}(h\nu - E_{21})$, this dependence can be neglected. The integral in brackets then reduces to unity, which leads to the fourth equality. In other words, the lineshape broadening has no effect on the total spontaneous emission rate, and we can simply integrate over the simplified R_{sp}^{21} defined in (A6.31).

Setting $A_{21} = 1/\tau_{sp}^{21}$ and $R_{sp} = N/\tau_{sp}$ in (A6.34), we can define the total spontaneous lifetime in terms of the local spontaneous lifetime through the following nontrivial relation:

$$\tau_{sp} \equiv N\left[\int \frac{1}{\tau_{sp}^{21}}\rho_r(E_{21}) f_2(1-f_1)\, dE_{21}\right]^{-1} \tag{A6.35}$$

Generally speaking, the term in brackets will tend to go as N^2 due to the double dependence on the quasi-Fermi levels (f_2 related to N and $1-f_1$ related to P). As a result, the total spontaneous lifetime typically follows a $1/N$ dependence as assumed in Chapter 2.

In performing actual calculations of the total spontaneous emission rate, it is useful to replace A_{21} with the single-mode rate constant, B_{21}. Doing this, Eq. (A6.34)

becomes

$$R_{sp} = \int B_{21} h\nu_{21} \rho_0(\nu_{21}) \rho_r(E_{21}) f_2(1 - f_1)\, dE_{21}. \tag{A6.36}$$

Thus, we see that the spontaneous emission rate includes both optical and electronic density of states functions.

By reducing the cavity size to dimensions on the order of the emission wavelength in the material, it is in principle possible to significantly alter the optical mode density, which allows us to actually alter the spontaneous emission rate. An active field of research that studies these microcavity effects is attempting to reduce the spontaneous emission rate substantially. The motivation for this lies in the following relation:

$$R_{sp} = \eta_i \eta_r \frac{1}{qV}. \tag{A6.37}$$

That is, the total spontaneous emission rate represents the number of carriers lost to spontaneous recombination per second and can therefore be equated with the radiative portion of the injected current. By minimizing R_{sp}, researchers hope to minimize the threshold current of certain types of lasers. In particular, VCSELs represent excellent candidates for such experiments due to their scalable geometry.

In concluding this appendix, it is useful to appreciate that the carrier density, the optical gain, and the radiative current in the active region can all be calculated from the quasi-Fermi levels using Eqs. (A6.20), (A6.23), (A6.36), and (A6.37). Thus, E_{Fc} and E_{Fv} *completely determine* all relevant parameters under nonequilibrium conditions. Furthermore, by invoking charge neutrality, we can find E_{Fc} for a given E_{Fv} using (A6.20), reducing the entire problem to one independent parameter. In other words, we can obtain gain versus current, gain versus carrier density, or current versus carrier density, by scanning E_{Fv} over the appropriate ranges. The linear relationship of gain to carrier density, and the quadratic relationship of current to carrier density discussed in Chapter 2 represent *approximations* to the more rigorous nontrivial relationships derived in this appendix.

READING LIST

Verdeyen JT. *Laser electronics*. 2nd ed. Englewood Cliffs, NJ: Prentice Hall; 1989. Chapters 7 and 11.

APPENDIX SEVEN

Periodic Structures and the Transmission Matrix

A7.1 INTRODUCTION

Distributed Bragg reflectors (DBRs) are important in many laser applications because (1) they can provide extremely high reflectivities, and (2) they can be used as frequency-selective filters. If the dielectric stack is completely periodic, the entire reflection spectrum can be determined exactly. If the dielectric stack is not perfectly periodic, then for practical purposes, only the peak reflectivity can be determined analytically. The main portion of this appendix deals with periodic stacks, first at the Bragg frequency to obtain the peak reflectivity, and then away from the Bragg frequency to determine the entire reflection spectrum. Correspondence is then made between the exact analysis and approximate Fourier and coupled-mode analyses. Finally, nonperiodic dielectric stacks are considered at the Bragg frequency.

A7.2 EIGENVALUES AND EIGENVECTORS

We begin with a homogeneous dielectric slab as shown in Fig. A7.1. Dividing up the slab into half-wave segments, we can assign normalized magnitudes to the forward- and backward-propagating fields at the ith plane:

$$\mathbf{e}_i \equiv \begin{bmatrix} A_i \\ B_i \end{bmatrix}. \tag{A7.1}$$

Diode Lasers and Photonic Integrated Circuits, Second Edition.
Larry A. Coldren, Scott W. Corzine, and Milan L. Mašanović.
© 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

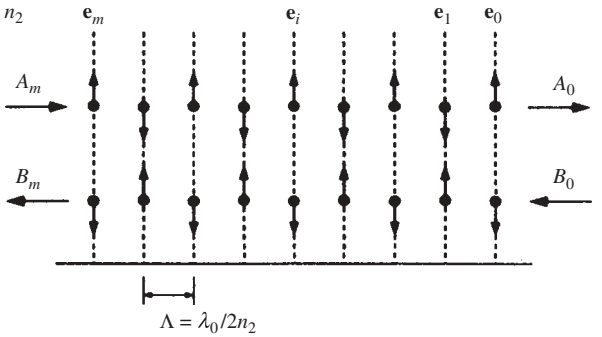


FIGURE A7.1: Propagation of fields through a uniform dielectric slab of index n_2, with reference planes placed at half-wavelength intervals. Arrows indicate the phase of the forward and backward waves.

Here $\mathbf{e}_i$ is defined as a two-component field vector, comprised of a forward wave of normalized amplitude, A_i, and a backward wave of normalized amplitude, B_i.

Now because the planes are spaced by a half-wavelength, the phases of the waves must change by π as we move from one plane to the next, and therefore $\mathbf{e}_{i+1} = -\mathbf{e}_i$. We can also express this relation in terms of the half-wave phase delay transmission matrix introduced in Chapter 3:

$$\mathbf{e}_{i+1} = [\mathbf{T}_{\lambda/2}]\mathbf{e}_i = e^{j\pi}\mathbf{e}_i. \tag{A7.2}$$

Thus, the transmission matrix of a half-wave segment operating on any field vector is reduced to a multiplicative constant. The field at the input can be related to the field at the output as follows:

$$\mathbf{e}_m = [\mathbf{T}_{\lambda/2}]^m\mathbf{e}_0 = e^{jm\pi}\mathbf{e}_0. \tag{A7.3}$$

Thus, the input fields are simply equal to the output fields times the multiplicative constant taken to the mth power, from which the reflection and transmission through the structure can be immediately obtained.

For more complex dielectric structures, the **T**-matrix, which relates the fields at one plane to the fields at the next, is reduced to a multiplicative constant for only two specific field vectors. However, it can be shown that these two field vectors can be weighted and added to construct *any* desired field vector. In mathematical terms, the two fields are referred to as eigenvectors, and the corresponding multiplicative constants are the eigenvalues of the transmission matrix. For a reciprocal matrix with $T_{11}T_{22} - T_{12}T_{21} = 1$, the product of the eigenvalues of the two eigenvectors equals unity, allowing us to express them in the following manner:

$$[\mathbf{T}]\mathbf{e}_+ = e^{+\xi}\mathbf{e}_+ \quad \text{and} \quad [\mathbf{T}]\mathbf{e}_- = e^{-\xi}\mathbf{e}_-. \tag{A7.4}$$

Here $\mathbf{e}_+$ and $\mathbf{e}_-$ represent the two distinct eigenvectors of the **T**-matrix, and $e^{\pm\xi}$ are the corresponding eigenvalues. Written in exponential form, ξ plays the role of a propagation constant. However, the "propagation" we are referring to is from plane to plane, indicating that ξ is more appropriately defined as the *discrete* propagation constant relating the field vector at one plane to the field vector at the next. For a uniform slab with planes spaced by half-wavelengths, Eq. (A7.2) reveals that $\xi = j\pi$. In complex dielectric structures, $\mathrm{Re}\{\xi\} \neq 0$, and $\mathbf{e}_+$ and $\mathbf{e}_-$ grow and decay as we propagate through the stack.

By writing out the equations implicit in Eq. (A7.4), we can solve for the eigenvalues and eigenvectors, or the *eigensystem* in terms of the **T**-matrix components:

$$\begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \begin{bmatrix} A_\pm \\ B_\pm \end{bmatrix} = e^{\pm\xi} \begin{bmatrix} A_\pm \\ B_\pm \end{bmatrix}, \tag{A7.5a}$$

$$e^{\pm\xi} = \tfrac{1}{2}(T_{11} + T_{22}) \pm \sqrt{\tfrac{1}{4}(T_{11} + T_{22})^2 - 1}, \tag{A7.5b}$$

$$\frac{A_\pm}{B_\pm} = \frac{T_{12}}{e^{\pm\xi} - T_{11}} = \frac{e^{\pm\xi} - T_{22}}{T_{21}}. \tag{A7.5c}$$

The eigenvalues are found by moving all terms in Eq. (A7.5a) to the left, setting the resulting determinant to zero, and applying $T_{11}T_{22} - T_{12}T_{21} = 1$ to simplify the square root term. For the eigenvectors, only the ratio of the two components is relevant (the absolute magnitude can be chosen arbitrarily). The two versions of *A/B* in Eq. (A7.5c) are found using the upper and lower equations in Eq. (A7.5a), respectively.

A7.3 APPLICATION TO DIELECTRIC STACKS AT THE BRAGG CONDITION

By filling roughly half of each segment defined in Fig. A7.1 with a different dielectric or index, n_1, we arrive at the quarter-wave distributed Bragg reflector introduced in Chapter 3, as indicated in Fig. A7.2. From Eq. (3.44), the **T**-matrix coefficients of one period of this structure at the Bragg condition are symmetrical such that $T_{11} = T_{22}, T_{21} = T_{12}$, and $T_{11}^2 - T_{21}^2 = 1$. These relations simplify the eigensystem considerably:

$$e^{\pm\xi} = T_{11} \pm T_{21} \quad \text{and} \quad \frac{A_\pm}{B_\pm} = \pm 1. \tag{A7.6}$$

A convenient choice for the two eigenvectors is

$$\mathbf{e}_+ = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{e}_- = \begin{bmatrix} 1 \\ -1 \end{bmatrix}. \tag{A7.7}$$

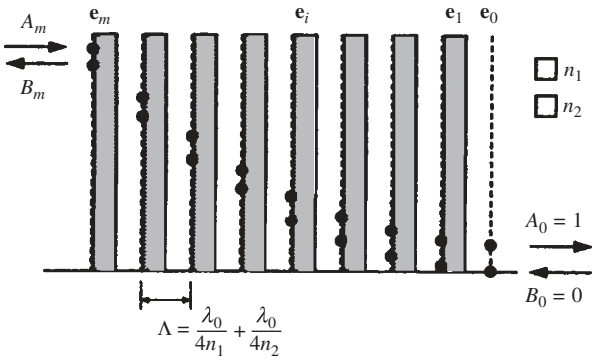


FIGURE A7.2: Forward and backward waves in a grating reflector.

Now getting back to our objective, we want to know the input fields in terms of the output fields of the stack. Assuming no incoming wave from the right ($B_0 = 0$), we can write the field vector at plane 0 as

$$\mathbf{e}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \tfrac{1}{2}(\mathbf{e}_+ + \mathbf{e}_-). \tag{A7.8}$$

To determine the field vector at plane 1, we apply the **T**-matrix to $\mathbf{e}_0$, which upon encountering the two eigenvectors is reduced to the two eigenvalues. Repeated application of the **T**-matrix allows us to determine the field vector at the ith plane:

$$\mathbf{e}_i = [\mathbf{T}]^i \mathbf{e}_0 = \tfrac{1}{2}\left(e^{+i\xi}\mathbf{e}_+ + e^{-i\xi}\mathbf{e}_-\right) = \begin{bmatrix} \cosh i\xi \\ \sinh i\xi \end{bmatrix}. \tag{A7.9}$$

The forward and backward wave components of the field vector are drawn according to Eq. (A7.9) at each plane in Fig. A7.2. Physically, the index discontinuities are feeding power from the forward to the backward wave, causing the forward wave to decay into the stack and the backward wave to grow from zero at the output.

It is apparent in Fig. A7.2 that the more periods there are, the closer the backward and forward wave are in magnitude at the input, and the higher the reflectivity of the stack. Evaluating Eq. (A7.9) at the mth plane and taking the reflected to incident amplitude ratio, we obtain

$$r_g = \frac{B_m}{A_m} = \tanh m\xi, \tag{A7.10}$$

or with $r_1 = \tanh \xi = T_{21}/T_{11}$ (= the reflectivity of one period), we can write

$$r_g = r_1 m_{eff} \quad \text{where} \quad m_{eff} \equiv \frac{\tanh m\xi}{\tanh \xi}. \tag{A7.11}$$

Thus, the total reflectivity is just the reflectivity of one period times some *effective* number of mirror periods seen by the incident field, m_{eff}. As $m \to \infty, m_{eff}$ saturates to a value of $m_{eff} \to 1/\tanh\xi \approx 1/\xi$ (for small ξ). From a physical viewpoint, the forward wave decays approximately exponentially into the mirror, and hence effectively "sees" only $1/\xi$ periods.

To further quantify the reflectivity, we must evaluate the discrete propagation constant. This is accomplished using Eqs. (3.44) to set $T_{11} = -(1 + r^2)/t^2$ and $T_{21} = -2r/t^2$, where r is the reflectivity of the 2–1 interface, and $t^2 = 1 - r^2$. With these substitutions, we have

$$e^{\xi} = T_{11} + T_{21} = -[(1 + r)/t]^2 = n_2/n_1. \tag{A7.12}$$

The first relation reduces to $-(1 + 2r)$ for small interface reflectivities. Solving for the discrete propagation constant, we find $\xi \approx j\pi + 2r$. In addition to the phase shift seen with the homogeneous slab, there is a real part to the discrete propagation constant, which is responsible for the attenuation of the incident wave. The attenuation *per unit length* in this small r limit is just $2r/\Lambda$, or two reflections per half-wavelength—an intuitive result, indeed. Plugging the latter equality of (A7.12) into (A7.11), we find that the effective number of mirror periods at the Bragg frequency becomes

$$m_{eff} = \frac{\tanh\,[m\ln(n_2/n_1)]}{\tanh\,[\ln(n_2/n_1)]}. \tag{A7.13}$$

To determine the reflectivity, we expand (A7.10) into exponential form, and use (A7.12) for e^{ξ}:

$$r_g = \frac{(n_2/n_1)^m - (n_2/n_1)^{-m}}{(n_2/n_1)^m + (n_2/n_1)^{-m}} = \frac{1 - (n_1/n_2)^{2m}}{1 + (n_1/n_2)^{2m}}. \tag{A7.14}$$

As m approaches infinity, this ratio approaches one, with larger differences in the index accelerating the convergence (see Problem A7.1). In the last section of this appendix, we will develop a more generalized version of (A7.14), applicable to more general quarter-wave stacks at the Bragg condition.

A7.4 APPLICATION TO DIELECTRIC STACKS AWAY FROM THE BRAGG CONDITION

The preceding analysis corresponded to the quarter-wave stack at its peak reflectivity, where all interface reflections add in phase. With the thickness of each layer being exactly one quarter-wavelength, the T-matrix reduced considerably allowing us to simplify the description of the eigensystem. In the more general case away from the Bragg condition, $T_{11} \neq T_{22}$ and $T_{21} \neq T_{12}$, and the eigensystem becomes more complex. Rather than repeat the procedure for this more general

case (see Problem A7.2), we will explore an alternative approach, which involves determining the overall grating matrix in terms of the single-period **T**-matrix.

We can relate the grating matrix to the eigenvectors and eigenvalues of the single-period **T**-matrix as follows:

$$[\mathbf{T}_g]\mathbf{e}_\pm = [\mathbf{T}]^m \mathbf{e}_\pm = e^{\pm m\xi}\mathbf{e}_\pm. \tag{A7.15}$$

This equation reveals that the eigenvectors of the grating matrix, $\mathbf{T}_g$, are in fact the same as the eigenvectors of the single period **T**-matrix, with the corresponding eigenvalues taken to the mth power. As a result, from the first equality in Eq. (A7.5c), we can set

$$\frac{T_{g12}}{e^{\pm m\xi} - T_{g11}} = \frac{T_{12}}{e^{\pm\xi} - T_{11}}. \tag{A7.16}$$

Subtracting the minus version of this equation from the plus version, we immediately obtain: $T_{g12}\sinh\xi = T_{12}\sinh m\xi$. If we now add the minus version to the plus version, we find: $(\sinh m\xi)(\cosh\xi - T_{11}) = (\sinh\xi)(\cosh m\xi - T_{g11})$. Solving for the grating matrix coefficients, we obtain

$$\begin{aligned} T_{g11} &= \frac{\sinh m\xi}{\sinh\xi}T_{11} - \frac{\sinh m\xi\cosh\xi - \cosh m\xi\sinh\xi}{\sinh\xi}, \\ T_{g12} &= \frac{\sinh m\xi}{\sinh\xi}T_{12}. \end{aligned} \tag{A7.17}$$

Identical relations exist for T_{g22} and T_{g21} using the second equality in Eq. (A7.5c). The overall reflectivity is then simply given by T_{g21}/T_{g11}, which becomes

$$r_g = \frac{T_{g21}}{T_{g11}} = \frac{T_{21}/T_{11}}{1 - \dfrac{\sinh(m-1)\xi}{T_{11}\sinh m\xi}}, \tag{A7.18}$$

where we have used the identity,

$$\sinh m\xi\cosh\xi - \cosh m\xi\sinh\xi = \sinh(m-1)\xi.$$

While Eq. (A7.17) provides us with the reflectivity of the grating, it is lacking in the sense that it does not provide us with any feel for how the reflectivity changes with frequency. Approximate expressions for r_g characterize the frequency dependence using the detuning parameter, $\delta = \beta - \beta_0$, introduced in Chapter 3. Toward this end, we define a *generalized* detuning parameter, which is characterized by the asymmetry in the diagonal matrix coefficients:

$$\Delta \equiv j\frac{T_{22} - T_{11}}{T_{22} + T_{11}}. \tag{A7.19}$$

In the low interface reflectivity limit near the Bragg frequency, $\Delta \to \delta\Lambda = \pi(\nu - \nu_0)/\nu_0$ (see Eqs. (3.44) with no loss).

In order to use Δ in Eq. (A7.17), we first replace every $\sinh x$ function with $(\cosh x)(\tanh x)$. The $\tanh x$ functions can then be replaced with the effective number of mirror periods, $m_{eff} = \tanh m\xi / \tanh \xi$, and we are left with only $\cosh x$ functions. By adding the plus and minus versions of Eq. (A7.5b), we can set

$$\cosh \xi = \frac{1}{2}(T_{11} + T_{22}) = \frac{T_{11}}{1 + j\Delta} = \frac{T_{22}}{1 - j\Delta}. \tag{A7.20}$$

The latter equalities make use of Eq. (A7.19). Using $\cosh \xi = T_{11}/(1 + j\Delta)$ for T_{g11} and T_{g21}, and $\cosh \xi = T_{22}/(1 - j\Delta)$ for T_{g22} and T_{g12}, Eqs. (A7.17) transform into

$$\begin{aligned} T_{g11} &= (1 + jm_{eff}\Delta)\cosh m\xi, \\ T_{g21} &= \frac{T_{21}}{T_{11}} m_{eff}(1 + j\Delta)\cosh m\xi, \end{aligned} \tag{A7.21}$$

and

$$\begin{aligned} T_{g12} &= \frac{T_{12}}{T_{22}} m_{eff}(1 - j\Delta)\cosh m\xi, \\ T_{g22} &= (1 - jm_{eff}\Delta)\cosh m\xi. \end{aligned} \tag{A7.22}$$

Using Eq. (A7.21) to define the grating reflectivity, $r_g = T_{g21}/T_{g11}$, we obtain

$$r_g = r_1 m_{eff} \frac{1 + j\Delta}{1 + jm_{eff}\Delta}, \tag{A7.23}$$

where $r_1 = T_{21}/T_{11}$, the reflectivity of one period.

In this version of r_g, the frequency dependence is more clearly visible. For example, at the Bragg frequency, $\Delta \to 0$ and Eq. (A7.23) reduces immediately to Eq. (A7.11). For small deviations from the Bragg condition, we can expand the denominator to first order in Δ to obtain

$$r_g \approx r_1 m_{eff}[1 - j(m_{eff} - 1)\Delta] \equiv r_1 m_{eff}[1 - j2\delta L_{eff}]. \tag{A7.24}$$

The latter equality, being the first-order expansion of $e^{-j2\delta L_{eff}}$, suggests that we can approximate the reflection phase deviation by a simple propagation delay associated with some *effective* length, or penetration depth. For small interface reflectivities near the Bragg frequency, $\Delta \approx \delta\Lambda, m_{eff} \gg 1$, and $L_{eff} \approx \frac{1}{2} m_{eff}\Lambda$. A more rigorous analysis which evaluates the exact phase slope at the Bragg frequency reveals that

$$L_{eff} \equiv -\frac{1}{2}\frac{d\phi_m}{d\delta}\bigg|_{\delta=0} = \frac{1}{2} m_{eff}\Lambda \left[\frac{1}{1 + r^2} - \frac{1}{2m_{eff}}\right], \tag{A7.25}$$

where m_{eff} is evaluated at the Bragg condition, and r is the interface reflectivity. For small r, m_{eff} becomes large and the term in brackets reduces to unity.

Problem A7.3 examines the penetration depth concept applied to the energy distribution and absorption loss of a grating at the Bragg frequency. It is shown there that in the small r, large m limit, we can again set $L_{eff} = \frac{1}{2} m_{eff} \Lambda$. L_{eff} is equal to *half* the effective number of mirror periods seen by the field because the energy distribution and power go as the field squared, decaying twice as fast as the field.

To completely specify the grating reflectivity, we need to define explicit relations for r_1, Δ, and m_{eff} in terms of the matrix coefficients developed in Chapter 3. In Eqs. (3.44), two phase factors are used: $\phi_{\pm} = \tilde{\beta}_1 L_1 \pm \tilde{\beta}_2 L_2$. With no loss, we can use Eq. (3.48) to set $\phi_+ = \pi + \delta L$ and $\phi_- = 0$, where $\delta \equiv \beta - \beta_0$, and β is the average propagation constant of the grating defined in Eq. (3.46). For a lossless DBR, Eqs. (3.44) become

$$\begin{aligned} T_{11} &= -\frac{1}{t^2}(e^{j\delta\Lambda} + r^2), \quad T_{22} = -\frac{1}{t^2}(e^{-j\delta\Lambda} + r^2), \\ T_{21} &= -\frac{r}{t^2}(e^{j\delta\Lambda} + 1), \quad T_{12} = -\frac{r}{t^2}(e^{-j\delta\Lambda} + 1). \end{aligned} \tag{A7.26}$$

From these relations we can immediately determine r_1 and Δ:

$$r_1 = \frac{T_{21}}{T_{11}} = 2r \frac{\cos(\delta\Lambda/2) e^{-j\delta\Lambda/2}}{1 + r^2 e^{-j\delta\Lambda}}, \tag{A7.27}$$

$$\Delta = j \frac{T_{22} - T_{11}}{T_{22} + T_{11}} = \frac{\sin \delta\Lambda}{\cos \delta\Lambda + r^2}. \tag{A7.28}$$

To help determine the eigenvalues and m_{eff}, we have

$$\frac{1}{2}(T_{11} + T_{22}) = -\frac{1}{t^2}(\cos \delta\Lambda + r^2), \tag{A7.29}$$

$$\sqrt{\tfrac{1}{4}(T_{11} + T_{22})^2 - 1} = \frac{1}{t^2}[2r^2(1 + \cos \delta\Lambda) - \sin^2 \delta\Lambda]^{1/2}. \tag{A7.30}$$

Using these relations in Eq. (A7.5b), we can calculate both ξ and m_{eff}. The reflection coefficient is now completely defined in terms of fundamental grating parameters: $m, r = (n_2 - n_1)/(n_2 + n_1), t = \sqrt{1 - r^2}$, and $\delta\Lambda = (\beta - \beta_0)\Lambda$, where $1/\beta = \frac{1}{2}[1/\beta_1 + 1/\beta_2]$.

A7.5 CORRESPONDENCE WITH APPROXIMATE TECHNIQUES

In this section, we examine how the exact reflectivity reduces to common approximate expressions. We will consider two limiting cases. The first approximation involves neglecting r^2 and higher terms, while retaining all terms related to $\delta\Lambda$. In this case, the reflectivity reduces to the Fourier limit. The second approximation involves retaining terms up to r^2 and limiting the frequency deviation to $(\delta\Lambda)^2$ terms. In this case, the reflectivity reduces to the coupled-mode limit considered in Chapter 6.

A7.5.1 Fourier Limit

To analyze the grating reflector using Fourier analysis, we send an impulse function into the stack and track the resulting distribution of reflected impulse functions in time. Such an analysis gives us the impulse response of the grating. The Fourier transform of the impulse response should then correspond to the reflection spectrum. While this approach in principle is exact, in practice, the infinite number of impulse functions which appear back at the input after repeated bounces within a multilayer stack makes this approach mathematically intractable in the general case. However, if the interface reflections are small, any impulse functions which make double bounces or more before returning to the input will have negligible magnitude and can be ignored. In this limit, only a small uniform burst of impulse functions will return (assuming the transmission through each interface does not reduce the magnitude of the impulse). This square-shaped envelope of the impulse response when Fourier transformed leads to a sin *x*/*x* type reflection spectrum. It is this functionality we would like to reproduce using the exact expression (A7.23).

To approximate the simplified Fourier analysis, we neglect all but linear terms in r, simulating the single-bounce approximation. However, the frequency variation we retain completely, since this is accurately determined using the Fourier approach. With these approximations, (A7.27) and (A7.28) simplify to

$$\begin{aligned} r_1 &\approx 2r\cos(\delta\Lambda/2)e^{-\delta\Lambda/2}, \\ \Delta &\approx \tan\delta\Lambda. \end{aligned} \tag{A7.31}$$

The discrete propagation constant and the effective number of mirror periods found using (A7.29) and (A7.30) in (A7.5b) and (A7.11) simplify to

$$\begin{aligned} &\tfrac{1}{2}(T_{11}+T_{22}) \approx -\cos\delta\Lambda, \\ &\sqrt{\tfrac{1}{4}(T_{11}+T_{22})^2-1} \approx j\sin\delta\Lambda, \\ &e^{\pm\xi} \approx -\cos\delta\Lambda \pm j\sin\delta\Lambda = e^{j\pi}e^{\mp j\delta\Lambda}, \\ &m_{eff} = \frac{\tanh m\xi}{\tanh\xi} \approx \frac{\tan m\delta\Lambda}{\tan\delta\Lambda}. \end{aligned} \tag{A7.32}$$

In the Fourier limit, the discrete propagation constant becomes simply the phase delay between mirror periods including detuning effects, $j\pi \mp j\delta\Lambda$.

Using these approximations in the expression for reflectivity (A7.23), we obtain

$$r_g \approx 2r\cos(\delta\Lambda/2)e^{-\delta\Lambda/2}\frac{\tan m\delta\Lambda}{\tan\delta\Lambda}.\frac{1+j\tan\delta\Lambda}{1+j\tan m\delta\Lambda}. \tag{A7.33}$$

To reduce this expression to the Fourier limit, we move the cosines in the first fraction to the second fraction and convert the entire second fraction to

exponential form. We then assume that the phase deviation *across one period* can be expanded to second order, such that $\cos(\delta\Lambda/2) \approx 1$ and $\sin\delta\Lambda \approx \delta\Lambda$. The accumulated phase deviation, $m\delta\Lambda$, across the entire grating may be large, so we leave this in general form. With $m\Lambda = L_g$ (the length of the grating), Eq. (A7.33) reduces to

$$r_g|_{Fourier} = 2mr\frac{\sin\delta L_g}{\delta L_g}e^{-j\delta L_g[(m-1/2)/m]}. \tag{A7.34}$$

This result is exactly the result we would obtain from a simplified single-bounce Fourier analysis, revealing the characteristic $\sin x/x$ spectral dependence of weakly reflecting gratings. The $m - 1/2$ term in the phase factor results from our definition of L_g which places the output reference plane one half-period beyond the last reflection of the grating.

A7.5.2 Coupled-Mode Limit

The coupled-mode approach discussed in Chapter 6 analyzes the grating by solving the wave equation assuming the waves are sufficiently slowly varying that second-order derivatives of their magnitude can be neglected. This restriction implies that the interface reflectivities must not be too strong and amounts to neglecting terms higher than r^2. Because the coupled-mode approach includes r^2 terms, it is much more accurate than the Fourier approach for strongly reflecting gratings near the Bragg condition. Another fundamental approximation of coupled-mode theory involves retaining only those Fourier components of the index variation which effectively couple forward- and backward-going waves at the Bragg condition. Away from the Bragg condition, other Fourier components of the index variation begin to contribute to the coupling between forward and backward waves. With these contributions neglected, we would expect that the coupled-mode approach is only good over a limited frequency range, in contrast to the Fourier approach. Therefore, to reproduce the coupled-mode approximation, we will limit the frequency deviation to $(\delta\Lambda)^2$.

Proceeding as before, we retain terms up to r^2 and $(\delta\Lambda)^2$, neglecting all others. We also neglect all products between r and $\delta\Lambda$. Equations (A7.27) and (A7.28) in this case simplify to

$$\begin{aligned} r_1 &\approx 2r \equiv \kappa\Lambda, \\ \Delta &\approx \delta\Lambda. \end{aligned} \tag{A7.35}$$

Here we have introduced a new variable, $\kappa \equiv 2r/\Lambda$, which in coupled-mode theory represents the coupling per unit length between the forward and backward waves. For the square wave profile analyzed here, the *coupling constant*, κ, is just equal to two reflections per grating period. We will use κ in place of r for the rest of this section.

The discrete propagation constant and the effective number of mirror periods found using (A7.29) and (A7.30) in (A7.5b) and (A7.11) simplify to

$$\begin{aligned}
\tfrac{1}{2}(T_{11}+T_{22}) &\approx -(1+(\kappa\Lambda)^2/2-(\delta\Lambda)^2/2),\\
\sqrt{\tfrac{1}{4}(T_{11}+T_{22})^2-1} &\approx [(\kappa\Lambda)^2-(\delta\Lambda)^2]^{1/2}\equiv\sigma\Lambda,\\
e^{\pm\xi} &\approx -(1\mp\sigma\Lambda+(\sigma\Lambda)^2/2)\approx e^{j\pi}e^{\mp\sigma\Lambda},\\
m_{eff} &= \frac{\tanh m\xi}{\tanh\xi}\approx\frac{\tanh m\sigma\Lambda}{\sigma\Lambda}.
\end{aligned}\tag{A7.36}$$

For consistency with coupled-mode notation used in Chapter 6, we have introduced another new variable, $\sigma\equiv\sqrt{\kappa^2-\delta^2}$. The discrete propagation constant is simply $j\pi\mp\sigma\Lambda$, to second order in $\sigma\Lambda$. Within the stopband ($|\delta|<\kappa$), $\sigma=\mathrm{Re}\{\xi/\Lambda\}$ and we can interpret it as the *decay* constant per unit length.

Using the approximations contained in (A7.35) and (A7.36), the general expression for reflectivity (A7.23) reduces to

$$r_g|_{coupled\text{-}mode}=\frac{\kappa}{\sigma}\tanh\sigma L_g\,\frac{1}{1+j\dfrac{\delta}{\sigma}\tanh\sigma L_g}.\tag{A7.37}$$

In deriving this result, we have neglected $j\Delta$ in the numerator in comparison to one, and set $m\Lambda=L_g$. Equation (A7.37) is equivalent to the coupled-mode result, Eq. (6.48), aside from the $-j$ phase factor listed there, which arises from a difference in the choice of input reference plane (see Chapter 6 for details).

A7.6 GENERALIZED REFLECTIVITY AT THE BRAGG CONDITION

The analysis described earlier assumes that the overall grating matrix can be represented by the single-period **T**-matrix taken to the mth power. The disadvantage of this approach is that it assumes the grating is completely periodic with only two alternating index layers. In practice, the input and output layers usually have different refractive indices from those comprising the alternating layers of the grating. More generally, there might be a situation where many of the layers within the grating itself are different. In this section we would like to generalize the results of Section A7.3 to include such nonperiodic gratings. Unfortunately, this type of analysis is not possible away from the Bragg condition, so we focus our attention on determining the reflectivity at the Bragg frequency.

Breaking the grating down into its fundamental components, we find that there are three types of matrices we need to consider: (1) high-to-low interfaces (referenced from left to right), (2) low-to-high interfaces, and (3) quarter-wave phase delays. Denoting these by $\mathbf{T}_{HL}$, $\mathbf{T}_{LH}$, and $\mathbf{T}_{\lambda/4}$, we would like to consider their

effect on the following vectors:

$$\mathbf{e}_+ = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{e}_- = \begin{bmatrix} 1 \\ -1 \end{bmatrix}. \tag{A7.38}$$

Using Table 3.3, it can be shown that

$$\begin{aligned} [\mathbf{T}_{HL}]\mathbf{e}_+ &= e^{+s}\mathbf{e}_+, \quad & [\mathbf{T}_{HL}]\mathbf{e}_- &= e^{-s}\mathbf{e}_- \\ [\mathbf{T}_{LH}]\mathbf{e}_+ &= e^{-s}\mathbf{e}_+, \quad & [\mathbf{T}_{LH}]\mathbf{e}_- &= e^{+s}\mathbf{e}_- \\ [\mathbf{T}_{\lambda/4}]\mathbf{e}_+ &= j\mathbf{e}_-, & [\mathbf{T}_{\lambda/4}]\mathbf{e}_- &= j\mathbf{e}_+ \end{aligned} \tag{A7.39}$$

where for the ith interface

$$s_i = \frac{1}{2}\ln\left[\frac{n_{Hi}}{n_{Li}}\right]. \tag{A7.40}$$

n_{Hi} is the high index of the ith interface, and n_{Li} is the low index. From Eqs. (A7.39), we see that $\mathbf{e}_+$ and $\mathbf{e}_-$ are actually the eigenvectors of both $\mathbf{T}_{HL}$ and $\mathbf{T}_{LH}$, suggesting that s_i is just the discrete *attenuation* constant of the ith interface. For a quarter-wave delay, $\mathbf{e}_+$ and $\mathbf{e}_-$ are not the eigenvectors, however, $\mathbf{T}_{\lambda/4}$ does transform one into the other. In Problem A7.4 the reader is asked to verify these relations.

Equations (A7.39) are all we need to propagate through any structure comprised of index discontinuities and multiples of quarter-wave segments. We are specifically interested in a grating whose index layers follow an HLHL sequence and are each a quarter-wavelength thick. We begin by expanding the output field vector using Eq. (A7.8). Propagating through the first half-wavelength of the structure, assuming interface 0 is high-low, and interface 1 is low-high, we obtain

$$\begin{aligned} [\mathbf{T}_{LH}][\mathbf{T}_{\lambda/4}][\mathbf{T}_{HL}][\mathbf{T}_{\lambda/4}]\mathbf{e}_0 &= [\mathbf{T}_{LH}][\mathbf{T}_{\lambda/4}][\mathbf{T}_{HL}][\mathbf{T}_{\lambda/4}]\tfrac{1}{2}[\mathbf{e}_+ + \mathbf{e}_-] \\ &= [\mathbf{T}_{LH}][\mathbf{T}_{\lambda/4}]\, j\tfrac{1}{2}[e^{-s_0}\mathbf{e}_- + e^{+s_0}\mathbf{e}_+] \\ &= -\tfrac{1}{2}[e^{-(s_1+s_0)}\mathbf{e}_+ + e^{+(s_1+s_0)}\mathbf{e}_-]. \end{aligned} \tag{A7.41}$$

Thus, the s parameters of the two interfaces simply add together. Continuing the above procedure through N interfaces to the input, and taking the ratio of the reflected to incident wave amplitudes, we find

$$|r_g| = \tanh\sum_0^N s_i, \tag{A7.42}$$

where N is the number of quarter-wave layers in the grating. Using Eq. (A7.40) for s_i, we can alternatively express the reflectivity as

$$|r_g| = \frac{1-b}{1+b} \quad \text{where} \quad b = \prod_0^N \left[\frac{n_{Li}}{n_{Hi}}\right]. \tag{A7.43}$$

If n_L and n_H do not change throughout the grating, and there are $N = 2m$ quarter-wave layers with $n_{L0}/n_{H0} = 1$ (see Fig. A7.2), then $b = (n_L/n_H)^{2m}$, and (A7.43) reduces to the more limited (A7.14) derived earlier.

The tools developed here can also be applied to dielectric structures which may have half-wavelength spacings, as well as to structures which do not follow the HLHL alternating sequence (antireflection coatings are an example of this). The interested reader is invited to compare this method to the method discussed by Corzine, Yan, and Coldren (see reading list), which analyzes a wider range of dielectric structures using essentially the same approach.

READING LIST

Corzine SW, Yan R-H, Coldren LA. A tanh substitution technique for the analysis of abrupt and graded interface multilayer dielectric stacks. *IEEE J Quantum Electron* 1991;**27**: 2086.

Babic DI, Corzine SW. Analytic expressions for the reflection delay, penetration depth, and absorptance of quarter-wave dielectric mirrors. *IEEE J Quantum Electron* 1992;**28**: 514.

Yeh P. *Optical waves in layered media*. New York: Wiley; 1988.

PROBLEMS

1. Design the grating mirrors of a VCSEL which meet the following requirements:
   (i) The output coupler mirror must be as close to 99% peak power reflectivity as possible.
   (ii) The back reflector must be greater than 99.9% peak power reflectivity. Specify the mirror designs including number of periods and layer thicknesses for two material systems:
   (a) $Al_{0.2}$ $Ga_{0.8}As$/AlAs alternating layers, which provide a peak reflectivity at 0.87 μm
   (b) InP/InGaAsP(1.3 μm) alternating layers, which provide a peak reflectivity at 1.55 μm

   Table 1.1 may be helpful in establishing the proper design. For the back reflector, assume the grating terminates in the higher index material on both sides. For the output coupler, assume that one side terminates in the higher-index material, and that the other side terminates in air (in your output coupler design, take care to ensure that the index follows the HLHL ordering sequence required for proper phasing of the reflected waves). Include a schematic of the index profile for all finished mirror designs.

   If the region between the two mirrors of the VCSEL is one optical wavelength thick, how long would it take to grow both the shortwavelength and long-wavelength VCSELs assuming a typical growth rate of 1 μm/hour? Can

you comment on the benefits of either the short- or long-wavelength material systems from this perspective?

2. Derive the general expression for the grating reflectivity, Eq. (A7.23), using the techniques developed in Section A7.3. Specifically:
   (i) Determine the eigenvectors using the second version of Eq. (A7.5c).
   (ii) Expand the output vector in terms of the two eigenvectors.
   (iii) Propagate the fields to the input plane.
   (iv) Evaluate the reflected to incident amplitude ratio.

   In defining the eigenvectors you may find it useful to set $e^{\pm\xi} = T_{11}(1 \pm \tanh\xi)/(1 + j\Delta)$ and $T_{22} = T_{11}(1 - j\Delta)/(1 + j\Delta)$ (however, if you do use these relations, verify that they are indeed correct).
3. The concept of a penetration depth not only applies to the linear phase deviation of the grating near the Bragg condition, but also to the energy distribution and absorption loss of the grating at the Bragg frequency.
   (a) The energy penetration depth can be defined as the depth at which the incident energy taken as constant into the mirror is equal to the total energy integrated over the grating length. Setting the incident energy times the energy penetration depth equal to the summation of energy throughout all periods, we have

$$(A_m^* A_m + B_m^* B_m) L_{eff}^e = \sum_{i=1}^{m} (A_i^* A_i + B_i^* B_i)\Lambda. \tag{A7.44}$$

   Here the total energy in each period is approximated by the energy density at the left edge times the thickness of one period. This will be a good approximation if Λ is small compared to the length of the stack (m is large).

   Using Eq. (A7.9) for A_i and B_i at the Bragg frequency, show that the energy penetration depth can be approximated by

$$L_{eff}^e = \frac{\tanh 2m\xi}{2\xi}\Lambda. \tag{A7.45}$$

   Hint: The summation can be approximated with an integral by setting $\Lambda \to dz$ and $i\Lambda \to z$.
   Show that this expression is equivalent to $\frac{1}{2} m_{eff}\Lambda$ in the small r, large m limit.

   (b) The absorption loss penetration depth can be defined as follows. If some small absorption loss, α, is distributed uniformly throughout the grating, the reflectivity of the grating will be reduced somewhat. If we model the reflector as a hard mirror recessed by some penetration depth into the lossy material, then the resulting round-trip attenuation is equal to $e^{-2\alpha L_{eff}^a} \approx 1 - 2\alpha L_{eff}^a$ (for small losses). The definition of L_{eff}^a is found

by equating this effective reduction in *power* reflectivity with the true reduction.

By including a uniform loss in the phase terms, $\phi_{\pm}$, of Eqs. (3.44), show that at the Bragg frequency, $\Delta \approx -j\alpha\Lambda/2$ instead of zero. With this result and Eq. (A7.23), derive an approximate expression for L_{eff}^{a}, assuming the loss terms can be expanded to first order. Show that L_{eff}^{a} is equivalent to $\frac{1}{2}m_{eff}\Lambda$ in the small r limit.

(c) Derive the exact phase penetration depth, L_{eff}^{p}, given in Eq. (A7.25).

4. Verify all relations in Eq. (A7.39). What are the corresponding relations for a half-wave phase delay? Answer this question by:

(a) Directly using the half-wave matrix

(b) Connecting two quarter-wave phase delays in series

Using the half-wave phase delay properties just derived, and other tools developed in Section A7.6, derive the reflectivity of a structure which is composed of:

(a) A low-high interface

(b) A half-wavelength phase delay

(c) A high-low interface (where the high layer of both interfaces is the same)

Express the reflectivity both in terms of the three refractive indices and in terms of the individual interface reflectivities. Repeat the derivation assuming a quarter-wave phase delay instead. How do these expressions compare to the Fabry–Perot expressions derived in Chapter 3?

APPENDIX EIGHT

Electronic States in Semiconductors

A8.1 INTRODUCTION

In Appendix 1, a basic description of electronic states in periodic potentials was given. Important concepts such as energy bands, the electron effective mass, and the density of states were introduced. All these concepts are essential to understanding the manner in which light interacts with semiconductor crystals. However, to provide a quantitative description of optical gain in these materials, a more complete description of the electronic states is required. In this appendix, the wavefunctions of the electrons required to evaluate the transition matrix element (see Appendix 10) will be considered in some detail. In addition, valence band-mixing effects related to the coupling of the heavy-and light-hole bands will be discussed. Finally the effects of strain on the subband structure of strained quantum wells will be treated. Some of the discussions will draw on material introduced in Appendix 1.

A8.2 GENERAL DESCRIPTION OF ELECTRONIC STATES

The electron wavefunctions in the conduction and valence bands of the semiconductor are found by solving the Schrödinger equation, which relates the system Hamiltonian, H_0, of the crystal lattice to the energy, E, of the electron. It can be

Diode Lasers and Photonic Integrated Circuits, Second Edition.
Larry A. Coldren, Scott W. Corzine, and Milan L. Mašanović.
© 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

written as

$$H_0\psi = \left[\frac{\mathbf{p}^2}{2m_0} + V(\mathbf{r})\right]\psi = E\psi, \tag{A8.1}$$

where $\mathbf{p}$ is the momentum operator, $\mathbf{r}$ is the position vector, m_0 is the free electron mass, ψ is the wavefunction of the electron, and $V(\mathbf{r})$ is the potential created by the crystal lattice. Due to the periodicity of $V(\mathbf{r})$, the solutions of the preceding equation are given by Bloch waves of the form

$$\psi = e^{i\mathbf{k}\cdot\mathbf{r}}u(\mathbf{k},\mathbf{r}), \tag{A8.2}$$

where $\mathbf{k}$ is the wavevector of the electron, and u is a Bloch function with the special property that it is periodic with the crystal lattice, and hence repeats itself in each unit cell of the crystal. In most practical cases, we never actually need to know the exact description of the Bloch function. What is important, however, is the *symmetry* properties of the Bloch functions in the conduction and valence bands, which will be considered in Section A8.3.

For analyzing localized electronic states such as those encountered in quantum wells and quantum wires, it is useful to consider linear combinations of the Bloch wave solutions in Eq. (A8.2). Using an arbitrary set of expansion coefficients, $A(k)$, we can express a spatially localized wavefunction as

$$\psi = \int A(k)e^{i\mathbf{k}\cdot\mathbf{r}}u(\mathbf{k},\mathbf{r})\,d^3\mathbf{k} \approx u(0,\mathbf{r})\int A(k)e^{i\mathbf{k}\cdot\mathbf{r}}\,d^3\mathbf{k} \equiv F(\mathbf{r})u(\mathbf{r}). \tag{A8.3}$$

The description of localized states contained in (A8.3) is known as the envelope function approximation. The key assumption here is that within a given energy band, the Bloch function is not a strong function of $\mathbf{k}$ (at least in the proximity of the band edge) and can thus be approximately represented by the band edge ($\mathbf{k} = 0$) Bloch function, $u(0,\mathbf{r}) \equiv u(\mathbf{r})$. This allows us to pull it out of the expansion and define an envelope function, $F(\mathbf{r})$, whose Fourier spectrum is made up of the plane wave components of the solutions in (A8.2). Thus, our generalized *approximate* solutions in a given energy band consist of the band edge Bloch function multiplied by a slowly varying envelope function. We choose the two components to be normalized such that in Dirac notation, we have

$$\langle F|F\rangle \equiv \int_V F^*F\,d^3\mathbf{r} = 1, \quad \langle u|u\rangle \equiv \frac{1}{V_{uc}}\int_{\text{unit cell}} u^*u\,d^3\mathbf{r} = 1. \tag{A8.4}$$

For the envelope functions, V is the volume of the crystal. For the Bloch functions, we only need to consider the volume of a single unit cell of the crystal, V_{uc}. Note that with our chosen definitions, the envelope functions have dimensions of $V^{-1/2}$, whereas the Bloch functions are dimensionless.

From symmetry considerations alone, one can deduce that the conduction band and valence band Bloch functions are orthogonal to each other. In Dirac notation,

this orthogonality can be compactly written as

$$\langle u_c | u_v \rangle = \langle u_v | u_c \rangle = 0. \tag{A8.5}$$

In Section A8.3, a more detailed discussion of the Bloch functions will reveal why this orthogonality exists. The envelope functions on the other hand can be determined explicitly. For example, in bulk material, the normalized envelope functions are simple plane waves:

$$F = \frac{1}{\sqrt{V}} e^{i\mathbf{k}\cdot\mathbf{r}}. \tag{A8.6}$$

The envelope functions in quantum-well material are considered in Appendix 1 and also later in this appendix.

We are particularly interested in the relationship between the energy, E, of the electron or hole given in Eq. (A8.1) and the electron's wavevector, $\mathbf{k}$, given in Eq. (A8.6). It is quite common to assume the bands to be parabolic in both the conduction and valence bands, allowing us to invoke the effective mass concept:

$$E_2(\mathbf{k}) = E_c + \frac{\hbar^2 k_2^2}{2m_c}, \quad E_1(\mathbf{k}) = E_v - \frac{\hbar^2 k_1^2}{2m_v}, \tag{A8.7}$$

where $E_{c,v}$ are the band edge energies, $m_{c,v}$ are the effective masses in the two bands, and $k_{2,1}$ are the magnitudes of the wavevectors of a given electron or hole. However, these expressions are oversimplifications to reality. First of all, $E(\mathbf{k})$ is not the same along all directions of the crystal, and we cannot simply use k in place of $\mathbf{k}$. For example, the heavy-hole band in GaAs is known to be highly anisotropic as a function of the k-vector direction (in comparison, the conduction and light-hole bands are much more isotropic at energies near the band edge). Second, at energies away from the band edge, the band curvature does not remain perfectly parabolic, especially in the light-hole band. In quantum-well and quantum-wire material, the valence band becomes extremely nonparabolic. Thus, in defining the relations for gain, we must keep in mind that (A8.7) is not always a good approximation to reality.

A8.3 BLOCH FUNCTIONS AND THE MOMENTUM MATRIX ELEMENT

To define the electronic states more explicitly, we will first consider the Bloch functions of the various energy bands. The conduction and valence bands are illustrated in Fig. A8.1. The three valence bands are commonly known as the heavy-hole (HH), light-hole (LH), and split-off hole (SO) bands. We can view each energy band as originating from the discrete atomic energy levels of the isolated atoms that compose the crystal as introduced in Appendix 1. In this sense, the conduction band can be thought of as a remnant of an s atomic orbital, while the three valence

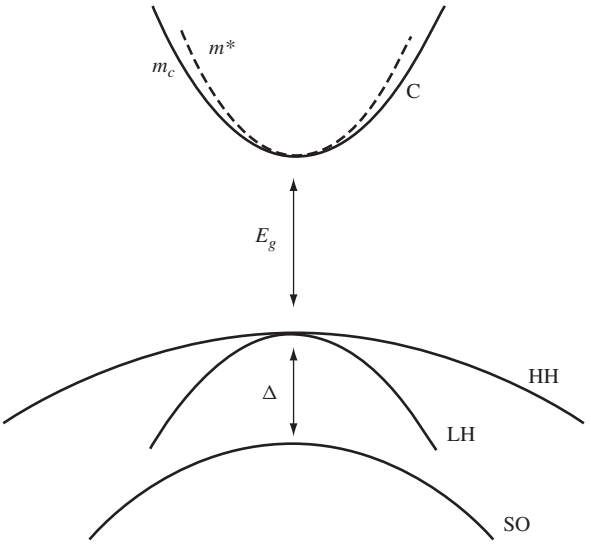


FIGURE A8.1: Typical III–V semiconductor band structure schematic illustrating the conduction (C), heavy-hole (HH), light-hole (LH), and split-off (SO) bands and their relative energy separations. The true (solid) and approximate (dashed) C bands are shown with their corresponding effective masses. (The relative scale of the four bands corresponds roughly to GaAs.)

bands are remnants of the three p atomic orbitals: p_x, p_y, and p_z. The corresponding Bloch functions for these orbitals are denoted here as u_s and u_x, u_y, and u_z. It is very useful to make this correspondence because the Bloch functions retain many of the symmetries that the atomic orbitals possess.

For instance, the conduction band Bloch function, u_s, has even symmetry in all three directions within each unit cell, similar to the spherically symmetric s atomic orbital. In a similar manner, u_z has odd symmetry along z, but even symmetry in the other two directions, within each unit cell, similar to the p_z atomic orbital. From these two facts alone, we can state that the net odd symmetry along z must give $\langle u_s|u_z\rangle = 0$ (where the brackets indicate integration over a unit cell). However, operating on u_z with the momentum operator, p_z, inverts the symmetry along z and the momentum matrix element $\langle u_s|p_z|u_z\rangle$ is therefore, in general nonzero. From these and similar arguments we can immediately write down some useful symmetry relations and definitions [1]

$$\langle u_s|p_i|u_j\rangle = 0, \quad \text{for } i \neq j \tag{A8.8}$$

$$\langle u_s|\mathbf{p}|u_i\rangle = \langle u_s|p_i|u_i\rangle \equiv M, \tag{A8.9}$$

$$\langle u_s|\mathbf{p}|\bar{u}_i\rangle = 0, \tag{A8.10}$$

where $i = x, y, z$, and $u_i, \bar{u}_i$ indicate spin-up and spin-down functions. The third relation comes from the fact that Bloch functions of opposite spin do not interact.

The constant M is defined here as the basis function *momentum matrix element*. Thus, through simple symmetry arguments, the various momentum matrix elements between u_s and the three u_i can now all be related simply to one constant, M.

The valence band Bloch functions u_{hh}, u_{lh}, and u_{so} corresponding to the bands of Fig. A8.1 are usually written as linear combinations of the "basis" functions u_x, u_y, and u_z [2-4] in a manner analogous to the construction of hybrid orbitals in the study of molecular bonds. Spin degeneracy exists in all three bands, so we actually need to define six Bloch functions. Defining the electron's k-vector to be directed along z, the valence band Bloch functions can be written as

$$
\begin{aligned}
u_{hh} &= -\frac{1}{\sqrt{2}}(u_x + iu_y), \qquad \bar{u}_{hh} = \frac{1}{\sqrt{2}}(\bar{u}_x - i\bar{u}_y), \\
u_{lh} &= -\frac{1}{\sqrt{6}}(\bar{u}_x + i\bar{u}_y - 2u_z), \quad \bar{u}_{lh} = \frac{1}{\sqrt{6}}(u_x - iu_y + 2\bar{u}_z), \\
u_{so} &= -\frac{1}{\sqrt{3}}(\bar{u}_x + i\bar{u}_y + u_z), \quad \bar{u}_{so} = \frac{1}{\sqrt{3}}(u_x - iu_y - \bar{u}_z).
\end{aligned}
\tag{A8.11}
$$

The prefactors are normalization constants, which can have arbitrary phase (the phases chosen here are those of Broido and Sham [5]). For **k** directed along another direction we would have to redefine the above relations (however, cyclic permutation of *x, y*, and *z* do yield equivalent relations if we also include the direction of **k** in the permutation).

The above linear combinations of basis functions are known as the angular momentum representation. They are useful when we consider the spin-orbit interaction between the angular momentum of the p orbitals and the spin angular momentum of the electron. The spin-orbit interaction term is "diagonalized" in this representation [2]. The SO band would be degenerate with the HH and LH bands if the spin-orbit interaction did not exist. As it is, the spin-orbit interaction partially removes the degeneracy, suppressing the SO band from the other two by a spin-orbit splitting energy, Δ, which in GaAs is equal to $\sim$0.34 eV.

With the above description of the valence bands, we can obtain a more complete description of the transition matrix element. However, the magnitude of the basis function momentum matrix element is still an unknown. Therefore, we will close this section with an example use of the preceding relations to obtain an approximate expression for the magnitude of $|M|^2$.

Evaluation of $|M|^2$ in bulk material was first obtained by theoretically relating it to the *curvature* of the conduction band [3, 6]. Using a second-order perturbation technique known as the $k \cdot p$ method [1–3], we can express the conduction band effective mass along the electron's k-vector direction (which for the definitions of Eq. (A8.11) is the z-direction) as

$$
\frac{1}{m_{cz}} = \frac{1}{m_0}\left[1 + \sum_{n \neq c} \frac{2}{m_0} \frac{|\langle u_c | p_z | u_n \rangle|^2}{E_c - E_n}\right], \tag{A8.12}
$$

where the summation sums over all n energy bands (not to be mistaken with energy subbands in quantum wells) of the crystal, and the E_n and E_c are the band edge energies of each band. From (A8.12) we see that the deviation of the conduction band effective mass from the free electron mass arises from the interaction between the conduction band and all other energy bands in the crystal.

It is interesting to note that due to the sign of the denominator, contributions from higher energy bands make the effective mass *heavier* and tend to flatten out the conduction band, while contributions from lower-lying energy bands tend to decrease the effective mass, increasing the curvature of the conduction band. In either case we find that the effect of a given band is to repel the conduction band away from it.

Also because of the denominator, only energy bands close in energy to the conduction band will contribute significantly to the summation. By neglecting all but the three valence bands in the summation, we can obtain an approximate closed-form expression for the conduction band effective mass using the relations given in (A8.8) through (A8.10) and (A8.11). Note that the HH Bloch function does not contain u_z, and hence its contribution to the sum is zero. Thus, summing over the LH and SO bands and using the energy separations defined in Fig. A8.1, we obtain

$$\frac{1}{m^*} = \left[1 + \frac{2|M|^2}{m_0}\left(\frac{2}{3}\frac{1}{E_g} + \frac{1}{3}\frac{1}{(E_g + \Delta)}\right)\right]. \tag{A8.13}$$

The approximate conduction band effective mass, m^*, is expected to be lighter than the true effective mass since the effects of any higher energy bands have been neglected in our approximation. The approximate conduction band (with curvature related to m^*) and the true conduction band are both illustrated in Fig. A8.1.

The true conduction band effective mass can be measured experimentally to a good degree of precision using cyclotron resonance techniques [7]. Thus, by assuming that m^* is close to the true effective mass, m_c, we can rearrange (A8.13) to obtain

$$|M|^2 = \left(\frac{m_0}{m^*} - 1\right)\frac{(E_g + \Delta)}{2\left(E_g + \dfrac{2}{3}\Delta\right)} m_0 E_g. \tag{A8.14}$$

So we see that the simple description of the valence bands given in Eq. (A8.8) through Eqs. (A8.10) and (A8.11) has led directly to a formula that can yield a rough estimate of $|M|^2$ [6]. And although the above relation is not exact, it does reveal that $|M|^2$ is roughly proportional to the *ratio* of the energy gap to the conduction band effective mass of the semiconductor [8].

We have derived Eq. (A8.14) as an exercise in using the valence band Bloch functions. It is a useful formula for materials which have not been fully characterized. However, in more common materials, such as GaAs, much more accurate methods of determining $|M|^2$ exist [9–13]. The inaccuracy of Eq. (A8.14)

stems from the fact that the contribution from higher-lying energy bands can be significant, and thus, m^* is not always a very good approximation to the true effective mass, m_c. For example, in GaAs, m^* is approximately equal to $0.0502m_0$ compared to the true effective mass, $0.067m_0$. Thus, using the true effective mass in Eq. (A8.14) leads to an *under*estimation of the matrix element by about 26%. Many previous calculations have failed to recognize this correction (see Yan et al. [14] for a discussion of this) implying that both calculated spontaneous emission (and hence, calculated radiative current density) and optical gain will be underestimated by 26%, a significant factor. The most accurate estimates of $|M|^2$ have actually been obtained using electron spin resonance techniques [9–13]. In Table 4.1, we tabulated the most accurately reported values of $|M|^2$ in several material systems commonly used in semiconductor laser applications.

A8.4 BAND STRUCTURE IN QUANTUM WELLS

With the Bloch functions now defined, we need to concentrate on defining and solving for the envelope functions in the conduction and valence bands.

A8.4.1 Conduction Band

For a nondegenerate energy band (aside from spin), such as the conduction band, it has been shown most notably by Luttinger and Kohn [4] (using a $k \cdot p$ formalism) that an "effective mass equation" or Schrödinger-like equation for the envelope function, F_2, as defined in (A8.3) can be derived. It can be given as

$$(H_c + V)F_2 = E_2(\mathbf{k})F_2, \tag{A8.15}$$

where the Hamiltonian for the conduction band is simply

$$H_c = -\frac{\hbar^2}{2m_c}\nabla^2 \rightarrow \frac{\hbar^2}{2m_c}(k_x^2 + k_y^2 + k_z^2). \tag{A8.16}$$

The arrow indicates the form of the Hamiltonian for plane wave solutions (when V is constant). The potential, V, in this case corresponds to the variation in the material band edge, and the total energy of the electron, $E_2(\mathbf{k})$, is measured relative to the bottom of the conduction band. What is extremely appealing about the effective mass equation as compared to Eq. (A8.1) is that the Bloch functions have been removed from the equation, and the effect of the periodic potential arising from the crystal lattice (and hence, the coupling to other energy bands) is now replaced by a conduction band "effective" mass, m_c. In this approximation, the quantum well created by the interfacing of three materials of different bandgap truly becomes a textbook particle-in-a-box problem with F as the wavefunction, and the material band edges as the potential, V. The solutions to Eqs. (A8.15) and (A8.16) for a quantum well are considered in Appendix 1.

A8.4.2 Valence Band

Degenerate Effective Mass Equation The simplicity of the band structure in the conduction band of a quantum well relies on the assumption that the interaction with other energy bands is weak enough that we can treat it perturbatively by replacing that interaction with a conduction band effective mass. However, for bands degenerate in energy, the assumption of weak interaction is a poor one and cannot be used. Therefore, the effective mass Eq. (A8.15) must be modified to include the strong degenerate band interaction explicitly.

A modified derivation for the degenerate band effective mass equation has also been treated by Luttinger and Kohn [4]. In this case, we still obtain an effective mass equation for each degenerate band similar to Eq. (A8.15), however, as a result of the degeneracy, a coupling term is introduced that couples the equations together. For the degenerate HH and LH bands near the band edge (see Fig. A8.1), this implies that we must work with *four* coupled effective mass equations (we must include the spin degeneracy)! We can actually include as many energy bands in the coupled set of equations as we desire (two equations for each spin-degenerate band). For example, in addition to the HH and LH bands, Eppenga et al. [15, 16] have included the SO band as well as the conduction band in their four-band model, leading to eight coupled equations! However, interaction between the HH and LH bands is by far the strongest, and we do not pay a large penalty by neglecting the other bands. Only when we consider energy levels deep into the valence band (energies comparable to the spin–orbit splitting energy, Δ), do we need to include the SO band explicitly. In GaAs and InGaAs, $\Delta > 300$ meV, implying that for most gain calculations with these materials, we can neglect the SO band entirely. In InP, $\Delta = 100$ meV, however, the well material typically used in this system is closer to lattice-matched InGaAs where again Δ is closer to 300 meV.

The four coupled effective mass equations can be greatly simplified using a method first suggested by Kane [2] and later by Broido and Sham [5]. They pointed out than an appropriate linear combination of the four Bloch functions $(u_{hh}, u_{lh}, \bar{u}_{hh}, \bar{u}_{lh})$, into four new Bloch functions (u_A, u_B, u_C, u_D), decouples the four equations into two identical sets of two coupled equations. Thus, we actually only need to consider two coupled equations in our analysis. However, a price must be paid for this luxury. We must now restrict our attention to analyzing $E_1(k_{xy})$ in a given *plane* in the crystal (the direction of k_{xy} must be specified). Furthermore, the equations remain completely general only for the {100} and {110} planes. However, for the present purposes, these represent only minor restrictions.

The two coupled effective mass equations for the degenerate bands can be expressed as

$$(H_{hh} + V)F_{hh} + WF_{lh} = E_1(\mathbf{k})F_{hh}, \tag{A8.17}$$

$$(H_{lh} + V)F_{lh} + W^{\dagger}F_{hh} = E_1(\mathbf{k})F_{lh}, \tag{A8.18}$$

where $F_{hh,lh}$ are the heavy- and light-hole envelope functions corresponding to two new Bloch functions $u_{A,B}$ to be defined in the next section. These replace the generic

valence band envelope function F_1 in this coupled case. $E_1(\mathbf{k})$ is the total energy of the hole measured from the valence band edge. Also, $W^\dagger$ is the Hermitian conjugate of W. The main difference between the degenerate Eq. (A8.17) and (A8.18) and nondegenerate Eq. (A8.15) effective mass equations lies in the coupling term, W. The energy, $E_1(\mathbf{k})$, is what we would like to solve for, but because of W, we must now solve two equations simultaneously to find it.

The form of the Hamiltonians, $H_{hh,lh}$, are also slightly different from Eq. (A8.16). Let us define k_z and k_t to be two perpendicular k-vector components, with k_z directed along a $\langle 100 \rangle$ direction, and k_t directed either along a $\langle 100 \rangle$ direction or a $\langle 110 \rangle$ direction within the k_x–k_y plane. With these definitions, we can write (assuming plane wave solutions)

$$H_{hh} = (\gamma_1 - 2\gamma_2)k_z^2 + (\gamma_1 + \gamma_2)k_t^2, \tag{A8.19}$$

$$H_{lh} = (\gamma_1 + 2\gamma_2)k_z^2 + (\gamma_1 - \gamma_2)k_t^2, \tag{A8.20}$$

where

$$\gamma_1 - 2\gamma_2 \equiv \frac{\hbar^2}{2m_{hh}}, \quad \gamma_1 + 2\gamma_2 \equiv \frac{\hbar^2}{2m_{lh}}. \tag{A8.21}$$

In writing $H_{hh,lh}$, we are assuming for convenience that $E_1(\mathbf{k})$ measures positive into the valence band. Note that the form of Eqs. (A8.19) and (A8.20) is very similar to Eq. (A8.16). The only surprising feature is the different effective masses used along k_z and k_t. As we shall see in the next section, inclusion of the coupling term compensates for this apparent asymmetry in bulk material. However, in quantum-well material, this asymmetry produces a much lighter HH mass in the plane of the well than along the confinement axis. The material constants $\gamma_{1,2}$ are referred to as the Luttinger parameters [17] and are easily related to the HH and LH effective masses, m_{hh} and m_{lh}, through Eq. (A8.21). A third Luttinger parameter, γ_3, exists in the coupling term, W. The coupling term takes a slightly different form when k_t is directed along $\langle 100 \rangle$ and $\langle 110 \rangle$ directions. We can define the two forms as

$$W = \sqrt{3}k_t(\gamma_2 k_t - i2\gamma_3 k_z) \quad \text{for } \{100\} \text{ planes}, \tag{A8.22}$$

$$W = \sqrt{3}k_t(\gamma_3 k_t - i2\gamma_3 k_z) \quad \text{for } \{110\} \text{ planes}. \tag{A8.23}$$

The Hermitian conjugate is given by

$$W^\dagger = \sqrt{3}k_t(\gamma_{2(3)} k_t + i2\gamma_3 k_z), \tag{A8.24}$$

independent of whether or not k_z is complex, (i.e., we do not take the complex conjugate of the hermitian operator k_z in defining $W^\dagger$). For either of the above forms, it is interesting to note that with $k_t = 0$, the coupling term disappears, and the effective mass equations decouple! Thus, in a quantum well, we can independently solve each of the effective mass Eqs. (A8.17) and (A8.18) for the quantized $E_{1n}(0)$

of the HH and LH valence bands, just as was done for the conduction band in the previous section. However, for finite k_t (i.e., away from the band edge), the equations become coupled and the solutions for $E_{1n}(k_t)$ become more complicated.

Bulk Solutions As in the conduction band, our first step in solving the quantum-well problem is to find the bulk solutions within each material. To find a general relation for $E_1(\mathbf{k})$ in bulk material, it is convenient to cast the effective mass Eqs. (A8.17) and (A8.18), into matrix form:

$$\begin{bmatrix} H_{hh}+V & W \\ W^{\dagger} & H_{lh}+V \end{bmatrix} \begin{bmatrix} F_{hh} \\ F_{lh} \end{bmatrix} = E_1(\mathbf{k}) \begin{bmatrix} F_{hh} \\ F_{lh} \end{bmatrix}. \tag{A8.25}$$

For the bulk solutions we take V to be a constant, V_0. Then in looking at Eqs. (A8.19) and (A8.20), we find that for a given k-vector, the 2×2 Hamiltonian matrix consists of simple constants. The eigenvalue problem in this case is easily solved for the eigenenergies. In general form, the bulk $E_1(\mathbf{k})$ relations for the HH and LH bands are given by

$$E_1(\mathbf{k}) - V_0 = \tfrac{1}{2}(H_{hh}+H_{lh}) \pm \tfrac{1}{2}[(H_{hh}-H_{lh})^2 + 4W^{\dagger}W]^{1/2}. \tag{A8.26}$$

Using Eqs. (A8.19) through (A8.24), $E_1(\mathbf{k})$ can be given explicitly in terms of $\mathbf{k}$ within any {100} plane, by

$$E_1(\mathbf{k}) = \gamma_1(k_z^2+k_t^2) \pm [4\gamma_2^2(k_z^2+k_t^2)^2 + 12(\gamma_3^2-\gamma_2^2)k_z^2k_t^2]^{1/2}, \tag{A8.27}$$

where we have set $V_0 = 0$. A similar equation can be found for $E_1(\mathbf{k})$ in any {110} plane. Figure A8.2 illustrates the constant energy contour curves of (A8.27) for both

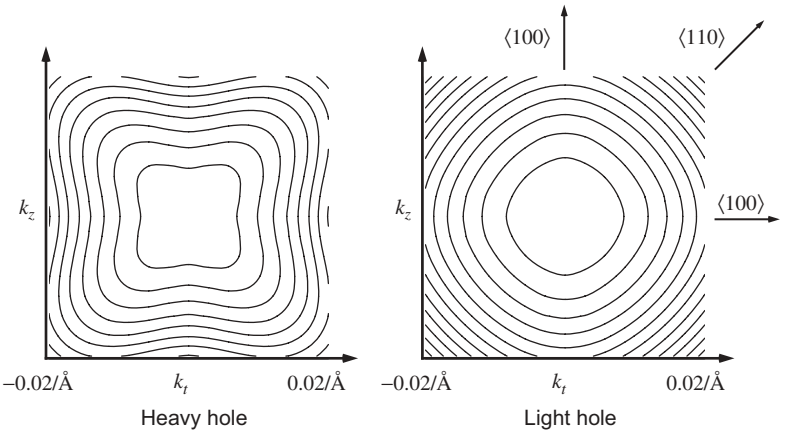


FIGURE A8.2: Contours of constant energy within any {100} plane of k-space for the HH and LH bands in bulk GaAs. The energy spacing between each contour level is 0.5 meVfor the HH band and 3 meVfor the LH band.

the HH band solution (negative root) and the LH band solution (positive root). From these curves, we see that the third Luttinger parameter, γ_3, can be related to the effective mass anisotropy along the $\langle 100 \rangle$ and $\langle 110 \rangle$ directions (because with $\gamma_3 = \gamma_2$, the contour curves would become circles, from Eq. (A8.27)). Looking along either k_z or k_t, (along a $\langle 100 \rangle$ direction), the cross-term in Eq. (A8.27) disappears and we simply have

$$E_1(\mathbf{k}) = (\gamma_1 \pm 2\gamma_2)k^2, \quad \mathbf{k} \text{ directed along any } \langle 100 \rangle \text{ direction.} \tag{A8.28}$$

From the definition of the Luttinger parameters in Eq. (A8.21), we see that the standard HH and LH $E_1(\mathbf{k})$ relations are obtained from Eq. (A8.28). Thus, contrary to what Eqs. (A8.19) and (A8.20) may suggest, the effective masses along k_z and k_t are identical in bulk material (within {100} planes).

To completely specify the bulk solutions, we also need to find the eigenvectors of Eq. (A8.25). With $E_{hh}(\mathbf{k})$ and $E_{lh}(\mathbf{k})$ given by the two roots of Eq. (A8.26), the eigenvectors, apart from a normalization constant, are found to be

$$\psi_1(\mathbf{k}, \mathbf{r}) = \begin{bmatrix} F_{hh} \\ F_{lh} \end{bmatrix} = e^{i\mathbf{k}\cdot\mathbf{r}} \begin{bmatrix} H_{lh} + V_0 - E_{hh} \\ -W^\dagger \end{bmatrix} \equiv e^{i\mathbf{k}\cdot\mathbf{r}} \begin{bmatrix} \Delta_{1h}(\mathbf{k}) \\ \Delta_{1l}(\mathbf{k}) \end{bmatrix}, \tag{A8.29}$$

$$\psi_2(\mathbf{k}, \mathbf{r}) = \begin{bmatrix} F_{hh} \\ F_{lh} \end{bmatrix} = e^{i\mathbf{k}\cdot\mathbf{r}} \begin{bmatrix} H_{lh} + V_0 - E_{lh} \\ -W^\dagger \end{bmatrix} \equiv e^{i\mathbf{k}\cdot\mathbf{r}} \begin{bmatrix} \Delta_{2h}(\mathbf{k}) \\ \Delta_{2l}(\mathbf{k}) \end{bmatrix}, \tag{A8.30}$$

where for either solution, the matrix notation implies

$$\psi = F_{hh}u_A + F_{lh}u_B. \tag{A8.31}$$

The validity of Eqs. (A8.20) and (A8.30) can be checked by substituting them into Eq. (A8.18). In addition, the Hamiltonians in Eqs. (A8.19) through (A8.24) implicitly assumed plane wave solutions, so their inclusion in (A8.29) and (A8.30) is mandatory. Eq. (A8.31) gives the wavefunctions in vector notation. The Bloch functions, $u_{A,B}$, are orthogonal to each other (analogous to two orthogonal unit position vectors in real space) and are given by linear combinations of the valence band Bloch functions defined in Section A8.3. We can write them as [18]

$$u_A = \frac{1}{\sqrt{2}}(\alpha u_{hh} - \alpha^* \bar{u}_{hh}), \tag{A8.32}$$

$$u_B = \frac{1}{\sqrt{2}}(\beta \bar{u}_{lh} - \beta^* u_{lh}). \tag{A8.33}$$

For {100} planes, $\alpha = \beta = 1$. For {110} planes, $\alpha = \exp[i3\pi/8]$, and $\beta = \exp[-i\pi/8]$.

Quantum-Well Solutions To solve the quantum-well problem, we again choose the quantum-well direction to be along z (this is not mandatory, but the effective mass equations would not decouple at the band edge otherwise (see Eq. (A8.22)

through). In this case, k_z is directed along the confinement axis and k_t is the transverse k-vector component which lies in the plane of the well, as shown in Fig. A8.3a. We now construct a general solution out of all bulk solutions that exist at a given energy within each material. From Fig. A8.3b, we see that in general, four plane wave solutions exist at a given energy [19]. The general solution in each region of Fig. A8.3c is then given by a linear combination of these four waves, or

$$\Psi = \sum A_{\pm}\psi_1(\pm k_{hh}, k_t, \mathbf{r}) + \sum B_{\pm}\psi_2(\pm k_{lh}, k_t, \mathbf{r}). \tag{A8.34}$$

The sums are over the plus- and minus-going waves, and the $k_{hh,lh}$ are defined in Fig. A8.3(b). The four coefficients, $A_{\pm}, B_{\pm}$, are unknown constants. Equation (A8.34) is the valence band analog of Eq. (A1.6) in the conduction band. In this sense, Eq. (A8.27) is the more complicated valence band analog of Eq. (A1.7). Both ψ_1 and ψ_2 are two-component vectors, from their definitions in Eqs. (A8.29) and (A8.30). The general solution, Ψ, is also a two-component vector. Each component can be written as

$$F_{hh} = e^{ik_t r_t}\left[\sum A_{\pm}\Delta_{1h}(\pm k_{hh}, k_t)e^{\pm ik_{hh}z} + \sum B_{\pm}\Delta_{2h}(\pm k_{lh}, k_t)e^{\pm ik_{lh}z}\right], \tag{A8.35}$$

$$F_{lh} = e^{ik_t r_t}\left[\sum A_{\pm}\Delta_{1l}(\pm k_{hh}, k_t)e^{\pm ik_{hh}z} + \sum B_{\pm}\Delta_{2l}(\pm k_{lh}, k_t)e^{\pm ik_{lh}z}\right], \tag{A8.36}$$

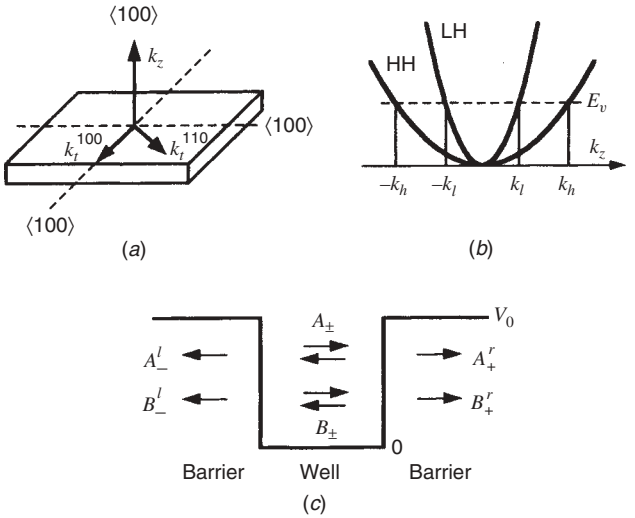


FIGURE A8.3: (a) Coordinate system to be used in the valence band model. The transverse or in-plane k-vector, k_t, can be directed along either a ⟨100⟩ or ⟨110⟩ direction, whereas the confinement axis must be along a ⟨100⟩ direction. (b) Illustration of the four plane wave states that exist at a given energy in the bulk valence band structure with $k_t = 0$. (c) Quantum-well potential and the eight coefficients which are to be used in solving the degenerate effective mass equation (see Eq. (A8.34)). The l and r superscripts indicate left and right coefficient.

where we have pulled out the common transverse plane wave component. Thus, in the case of the valence band, there are four unknown coefficients, $A_{\pm}, B_{\pm}$ in each region, as shown in Fig. A8.3c, as opposed to just two, as was the case in the conduction band.

To find the coefficients, we match the general solutions within each region at both well-barrier interfaces. The boundary conditions for the degenerate effective mass equation involve matching the four following quantities [19] across the interface

$$F_{hh} \quad \text{and} \quad (\gamma_1 - 2\gamma_2)\frac{dF_{hh}}{dz} + \sqrt{3}\gamma_3 k_t F_{lh}, \tag{A8.37}$$

$$F_{lh} \quad \text{and} \quad (\gamma_1 + 2\gamma_2)\frac{dF_{lh}}{dz} - \sqrt{3}\gamma_3 k_t F_{hh}. \tag{A8.38}$$

With $k_t = 0$, the above boundary conditions are identical to those given in Appendix 1 for the nondegenerate case. In addition, for any k_t, if the Luttinger parameters are the same on both sides of the interface, the second boundary conditions reduce to the simple slope continuity conditions. The generalized slope continuity conditions Eqs. (A8.37) and (A8.38) can be derived by integrating Eq. (A8.25) over an infinitesimal thickness that straddles the interface. However, we must first symmetrize the Hamiltonian, which involves setting $(\gamma_1 \pm 2\gamma_2)k_z^2 \rightarrow k_z(\gamma_1 \pm 2\gamma_2)k_z$ in Eqs. (A8.19) and (A8.20), *and* setting $\gamma_3 k_z \rightarrow (\gamma_3 k_z + k_z \gamma_3)/2$ in Eqs. (A8.22) through (A8.24), before setting $k_z \rightarrow -i\,\partial/\partial z$, since $\gamma_1, \gamma_2, \gamma_3$ all depend on z. These symmetrizing substitutions guarantee that the Hamiltonian in Eq. (A8.25) remains hermitian for any arbitrary z-dependence of $\gamma_1, \gamma_2, \gamma_3$. It should be noted that the preceding boundary conditions hold for both {100} and {110} planes. Caution should be issued here that the preceding boundary conditions apply only when the Bloch functions of the two materials are similar (this similarity should be mirrored in the values of the Luttinger parameters).

Applying the four boundary conditions at each interface gives us a total of eight equations. There are four unknown coefficients in each of the three regions, or twelve in total. However, requiring the envelope functions to go to zero at infinity leaves us with a total of eight unknown coefficients as shown in Fig. A8.3c. Thus, our problem is now completely specified, and we can solve the eight homogeneous equations by numerically finding the roots of the 8×8 determinant.

The general procedure for obtaining $E_1(k_t)$ is then as follows: (1) find the $E_{1n}(0)$ of a particular HH or LH band edge energy level using the conventional method of Appendix 1; (2) increment k_t and guess at the new energy of the state; (3) find $k_{hh,lh}$ from the two $E_1(\mathbf{k})$ relations given in Eq. (A8.26) within each material (each material having its own V_0 and its own set of Luttinger parameters); (4) evaluate the Δ's within each material from their definitions in Eqs. (A8.29) and (A8.30); (5) evaluate the 8×8 coefficient determinant; (6) if it is not equal to zero, use Newton's method to repeat the process until the energy root is found for that given k_t; and (7) increment k_t and repeat the entire process to find the new energy root, using an educated initial guess. The entire $E_{1n}(k_t)$ can then be traced in this way. The rate of convergence is very good for this type of problem. For example, for

each k_t, the energy root can be found in typically 2–3 iterations (Fig. A8.4 was generated in less than 30 seconds on a Macintosh computer from the early '90ies.).

Our description of the valence subband structure is now complete, and we can move on to an example. The one-dimensional quantum confinement in the conduction band gives rise to a set of parabolic subbands in the plane of the well. In the valence band, coupling between the HH and LH subbands changes the situation drastically, giving rise to a much more interesting band structure. Figure A8.4 shows the valence subband structure for an 80 Å $GaAs/Al_xGa_{1-x}As$ quantum well with $x = 0.2$ in the barrier regions, calculated using the procedure outlined earlier. The subband structure is seen to be far from parabolic, and in some regions, the band curvature is even inverted, leading to a negative "local" hole mass.

With respect to predicting the optical gain achievable in a material with the subband structure shown in Fig. A8.4, we are particularly interested in the density of states of the subbands, also shown in Fig. A8.4. The density of states is important because it determines the relationship between carrier density and the quasi-Fermi level of the band as discussed in Appendix 2. From Fig. A8.4, we see that ρ_v is roughly $2.5\rho_c$ near the band edge, but rapidly becomes very large as mixing between the bands starts to become significant. The mismatch between ρ_c and ρ_v as well as the overall large ρ_v reduces the performance of the quantum well, increasing transparency levels and reducing the differential gain (ideally the DOS curve in Fig. A8.4 would be a straight line of magnitude one).

Because the energy bands are different along the ⟨110⟩ and ⟨100⟩ directions within the quantum well, the density of states at any given energy should in

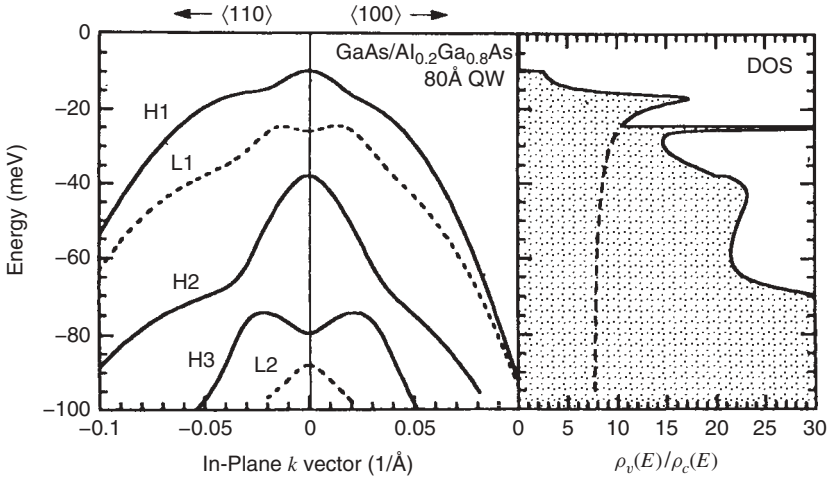


FIGURE A8.4: Plotted on the left is the valence subband structure of an 80 Å $GaAs/Al_{0.2}Ga_{0.8}As$ quantum well ($V_0 \approx 95$ meV). On the right is the total (solid curve) and H1 subband (dashed curve) density of states plotted relative to the density of states in the first conduction (C1) subband.

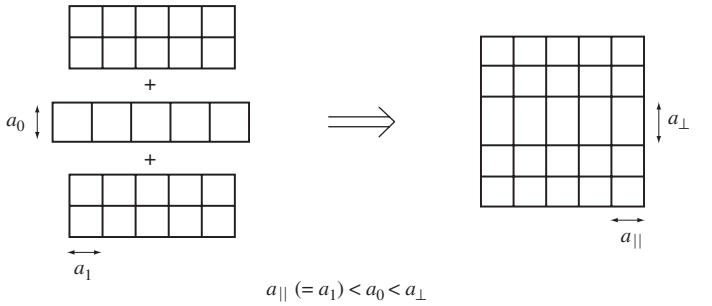
principle be calculated by averaging over all in-plane k-vector directions. Fortunately, this procedure can be approximated by calculating the density of states for just one set of energy bands, which is some appropriate average between the $\langle 100 \rangle$ and $\langle 110 \rangle$ dispersion curves shown in Fig. A8.4. These average energy bands can be found by using a coupling term of the form $W_{ave} = (W_{100} + W_{110})/2$ in the band calculation. This approach is commonly known as the *axial* approximation [3]. We use it to calculate the density of states in Figs. A8.4 and A8.6 as well as for all quantum-well gain calculations presented in Chapter 4. For the "bulk" barrier regions, the energy bands (the HH and LH effective masses) are found using a "spherical average" of the effective masses along all directions in the crystal. It basically involves replacing γ_2 with $(2\gamma_2 + 3\gamma_3)/5$ in Eq. (A8.21), and is known as the *spherical* approximation [19].

In the next section, we consider ways in which ρ_v can be modified, in beneficial ways, by the introduction of strain.

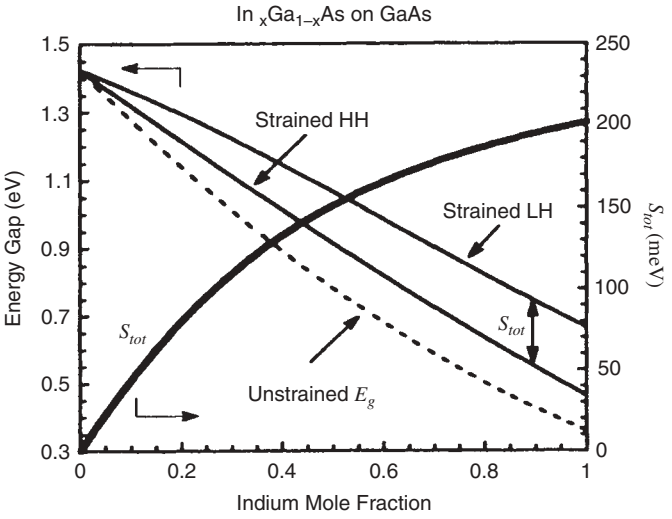
A8.4.3 Strained Quantum Wells

It was originally suggested by Yablonovich and Kane [20] and independently by Adams [21] that the introduction of compressive strain into the crystal lattice of a semiconductor could lead to enhanced performance in semiconductor lasers. To understand why, we need to examine the effects of strain, particularly in a quantum well. Introducing compressive strain into a quantum well configuration is particularly simple; just grow the well layer out of a material with a *larger* native lattice constant than the barrier layers. Because the quantum-well layer is typically very thin, instead of forming misfit dislocations, the lattice actually compresses in the plane of the well to match that of the barrier layers. In addition, the lattice constant in the direction normal to the plane becomes elongated (in an effort to keep the volume of each unit cell the same), as shown in Fig. A8.5(a).

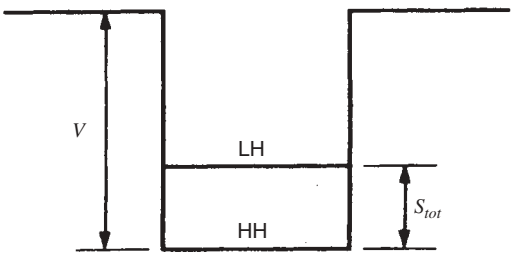
Because the energy gap of a semiconductor is related to its lattice spacing, we might expect that distortions in the crystal lattice should lead to alterations in the bandgap of the strained layer (putting aside for the moment the changes created simply by quantum confinement). In fact, there are two types of modifications that occur as discussed in more detail in Appendix 3. The first effect produces an upward shift in the conduction band as well as a downward shift in both valence bands, increasing the overall bandgap by an amount, H (that is positive for compressive strain and negative for tensile strain). The H indicates that this shift originates from the *hydrostatic* component of the strain. The second more important effect separates the HH and LH bands, each being pushed in opposite directions from the center by an amount, S. The S indicates that this shift originates from the *shear* component of the strain. Thus, the band edge degeneracy of the two valence bands is removed and two energy gaps must now be defined. The total strained bandgap can be written as $E_g + H \pm S$, where the upper sign refers to the C−LH bandgap, E_g(LH), and the lower sign refers to the C−HH bandgap, E_g(HH). In Fig. A8.5(b) we have plotted the unstrained bulk bandgap as well as the two compressively strained bulk bandgaps of InGaAs on a GaAs substrate (which has a smaller lattice



(*a*)



(*b*)



(*c*)

FIGURE A8.5: (a) Crystal lattice deformation under compressive strain. (b) Bulk band-gap of $In_xGa_{1-x}As$, when the in-plane lattice constant is compressed to that of GaAs. (c) The modified potential profile in the valence band of a quantum well, when the well material is under compressive strain.

constant than InGaAs). Note that S as defined here is positive for compressive strain, since $E_g(\text{LH}) > E_g(\text{HH})$. For tensile strain, S would be negative and the bandgap ordering would be reversed.

The two energy shifts, H and S, increase linearly with the lattice constant mismatch, which in turn increases linearly with indium mole fraction. We should expect then that the bandgap difference, $E_g(\text{LH}) - E_g(\text{HH})$, defined in the plot as S_{tot}, should be a linearly increasing function of indium mole fraction since from the above discussion, $S_{tot} = 2S$. For small indium mole fractions, we see from the plot that this is true. However, as the indium mole fraction increases, S_{tot} begins to saturate. This is a result of the interaction between the LH band and the SO band. When taken into account, this interaction introduces a correction term into the expression for the strained LH bandgap such that to second order, $S_{tot} = 2S(1 - S/\Delta)$, where Δ is the spin–orbit splitting energy. A more detailed discussion of strained bandgaps is given in Appendix 3.

In a quantum well, the splitting of the HH and LH bands can have dramatic consequences, since the large nonparabolicity of the subband structure in Fig. A8.4 is a direct result of the HH and LH band mixing. If we place the strained bandgaps shown in Fig. A8.5b into a quantum well, the situation becomes as shown in Fig. A8.5c, where the depth of the quantum well as seen by light holes is reduced by the splitting energy, S_{tot}. To predict the valence subband structure of the strained quantum well in Fig. A8.5c, we simply need to add a potential offset to the effective mass equation describing the LH band in the well. Eq. (A8.25) in the previous section now simply becomes

$$\begin{bmatrix} H_{hh} + V & W \\ W^{\dagger} & H_{lh} + V + S_{tot} \end{bmatrix} \begin{bmatrix} F_{hh} \\ F_{lh} \end{bmatrix} = E_1(\mathbf{k}) \begin{bmatrix} F_{hh} \\ F_{lh} \end{bmatrix}, \tag{A8.39}$$

where V is zero inside the well and S_{tot} is zero outside the well. The procedure for solving (A8.39) in a strained quantum well is entirely analogous to the procedure presented in Section A8.4.2 for an unstrained quantum well. Thus, we can immediately turn to an example calculation using Eq. (A8.39).

For direct comparison, we take the GaAs/AlGaAs 80 Å quantum well used in the example of Section A8.4.2 and simply add a bit of indium to the well layer. In GaAs has a larger native lattice constant than GaAs, thus sandwiched between two AlGaAs layers, it will be compressed in the plane of the well. For an indium mole fraction of 20%, the resulting HH–LH splitting energy, S_{tot}, is approximately 80 meV, or 1.4% compressive strain.

Figure A8.6 shows both the valence subband structure and the density of states of the 80 Å $In_{0.2}Ga_{0.8}As/Al_{0.2}Ga_{0.8}As$ strained quantum well calculated from Eq. (A8.39) [22]. Immediately we see that the LH bands have been pushed deep into the band (the full depth of the well is not shown). As a result, the band warping has been greatly reduced. Comparison with Fig. A8.4 reveals that the density of states in the strained quantum well is reduced significantly and matching between ρ_c and ρ_v is greatly improved. Both of these features translate into lower transparency levels and higher differential gain as calculations of gain presented in Chapter 4 reveal.

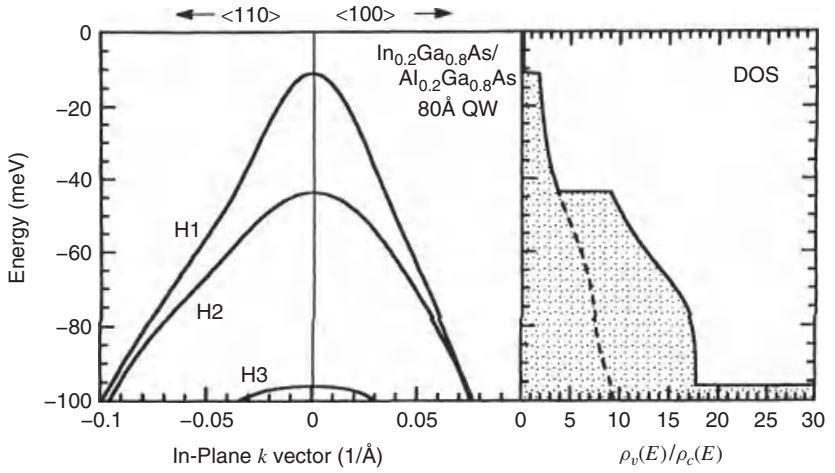


FIGURE A8.6: Plotted on the left is the valence subband structure of an 80 Å $In_{0.2}Ga_{0.8}As/Al_{0.2}Ga_{0.8}As$ strained quantum well ($V_0 \approx 175$ meV). The LH bands have been pushed further out of the well as a result of the strain and cannot be seen with the energy scale shown. On the right is the total (solid curve) and H1 subband (dashed curve) density of states plotted relative to the density of states in the first conduction (C1) subband.

Eq. (A8.39) not only applies to quantum wells, but can also be applied to bulk strained material. The $E(k)$ relations in bulk material are obtained by finding the eigenenergies of Eq. (A8.39) as we did for Eq. (A8.25) in the last section (see Eq. (A8.26) and (A8.27)) (the definitions for H and W given in Section A8.4.2 apply here as well). The general bulk solutions for the eigenenergies of (A8.39) can be obtained in closed form, but are somewhat messier than Eq. (A8.27). We leave it as an exercise for the reader to show that to first order in $\gamma_2 k^2/S$ (i.e., in the large-strain regime where $S \gg \gamma_2 k^2$), the eigenenergies of (A8.39) can be expressed as [23]

$$E_1(k_\parallel) = (\gamma_1 \mp \gamma_2)k_t^2 \pm S_{tot}/2, \quad (k_z = 0) \tag{A8.40}$$

$$E_1(k_\perp) = (\gamma_1 \pm 2\gamma_2)k_z^2 \pm S_{tot}/2. \quad (k_t = 0) \tag{A8.41}$$

The upper signs refer to the LH band solutions while the lower signs refer to the HH band solutions (we have thrown away any common shifts in the band edges to concentrate on the *difference* between the LH and HH bandgap energies). The parallel (in-plane) and perpendicular (normal to the plane) k-vectors refer to the notation used in Fig. A8.5a. Note that the above relations are also obtained by setting $W = 0$ (as seen by combining Eqs. (A8.19) and (A8.20) with Eq. (A8.39)). This makes sense because in the large-strain regime we are basically saying that $S \gg W$, allowing us to neglect the coupling between the bands altogether. The modified band structure given by (A8.40) and (A8.41) is shown to scale in Fig. A8.7. Thus, even in bulk material, strain serves to reduce the effective mass of the HH band

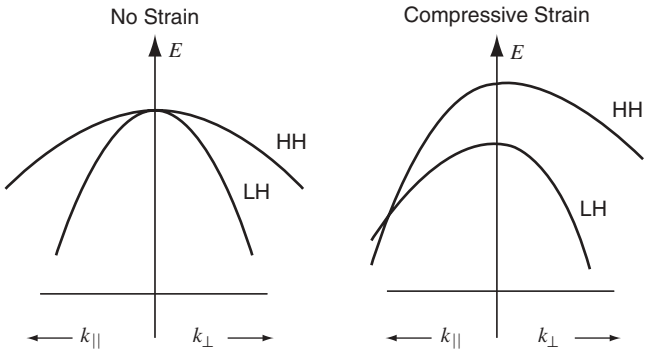


FIGURE A8.7: Effects of compressive strain on the bulk valence band structure. The relative band curvatures are drawn according to Eqs. (A8.40) and (A8.41), which are derived under the large-strain limit (the true dispersion curves would not cross each other as suggested in the figure).

dramatically within the "plane of compression." Perpendicular to the plane, it is interesting to note, however, that apart from the splitting of the HH and LH bands, the dispersion relation Eq. (A8.41) remains unchanged compared to Eq. (A8.28).

We have thus far not mentioned the effects of strain on the conduction band. The reason for this is that due to its relative isolation from other bands, the conduction band curvature remains relatively unaffected by the shifting energy gaps. Eq. (A8.13) would suggest that the increase in the bandgap should increase the conduction band effective mass slightly. However, we must be careful here because the third parameter that ties into this equation, $|M|^2$, does not necessarily remain constant (for example, it is conceivable that $|M|^2$ could increase proportionally with the bandgap, leaving m^* unaffected). In any case, the change in the conduction band curvature should be slight.

REFERENCES

1. Long D. *Energy bands in semiconductors*. New York: Interscience; 1968.
2. Kane EO. Energy band structure in p-type germanium and silicon. *J Phys Chem Solids* 1956;**1**: 82.
3. Kane EO. Band structure of indium antimonide. *J Phys Chem Solids* 1957;**1**: 249.
4. Luttinger JM, Kohn W. Motion of electrons and holes in perturbed periodic fields. *Phys Rev* 1955;**97**: 869.
5. Broido DA, Sham LJ. Effective masses of holes at GaAs-AlGaAs heterojunctions. *Phys Rev B* 1985;**31**: 888.
6. Casey HC Jr, Panish MB. *Heterostructure lasers part A: fundamental principles*. Orlando, FL: Academic Press; 1978.
7. Blakemore JS. Semiconducting and other major properties of gallium arsenide. *J Appl Phys* 1982;**53**: R123.

8. Asada M, Kameyama A, Suematsu Y. Gain and intervalence band absorption in quantum-well lasers. *IEEE J Quantum Electron* 1984;**20**: 745.
9. Chadi DJ, Clark AH, Burnhan RD. Γ_1 conduction electron g factor and matrix elements in GaAs and $Al_xGa_{1-x}As$ alloys. *Phys Rev B* 1976;**13**: 4466.
10. Hermann C, Weisbuch C. Perturbation theory in III-V compounds and alloys: a reexamination. *Phys Rev B* 1977;**15**: 823.
11. Hermann C, Wiesbuch C. *Modern problems in condensed mattersciences*. Volume 8, Optical orientation, ed. Agranovich VM, Maradudin AA. Amsterdam: North-Holland; 1984. p 463–508.
12. El Jani B, Gilbart P, Portal JC, Aulombard RL. Effective masses in Sn-doped $Ga_{1-x}Al_xAs$ ($x<0.33$) determined by the Shubnikov–de Haas effect. *J Appl Phys* 1985;**58**: 3481.
13. Nicholas RJ, Portal JC, Houlbert C, Perrier P, Pearsall TP. An experimental determination of the effective masses for $Ga_xIn_{1-x}As_yP_{1-y}$ alloys grown on InP. *Appl Phys Lett* 1979;**34**: 492.
14. Yan RH, Corzine SW, Coldren LA, Suemune I. Corrections to the expression for gain in GaAs. *IEEE J Quantum Electron* 1990;**26**: 213.
15. Eppenga R, Schuurmans MFH, Colak S. New k ·p theory for GaAs/$Ga_{1-x}A1_xAs$-type quantum wells. *Phys Rev B* 1987;**36**: 1554.
16. Colak S, Eppenga R, Schuurmans MFH. Band mixing effects on quantum well gain. *IEEE J Quantum Electron* 1987;**23**: 960.
17. Luttinger JM. Quantum theory of cyclotron resonance in semiconductors: general theory. *Phys Rev* 1956;**102**: 1030.
18. Chuang SL. Theory of hole refractions from heterojunctions. *Phys Rev. B* 1989;**40**:10 379.
19. Andreani LC, Pasquarello A, Bassani F. Hole subbands in strained GaAs-$Ga_{1-x}Al_xAs$ quantum wells: exact solution of the effective-mass equation. *Phys Rev. B* 1987;**36**: 5887.
20. Yablonovitch E, Kane EO. Reduction of lasing threshold current density by the lowering of valence band effective mass. *IEEE J Lightwave Technol* 1986;**4**: 504.
21. Adams AR. Band-structure engineering for low-threshold high-efficiency semiconductor lasers. *ElectronLett* 1986;**22**: 249.
22. Corzine SW, Yan RH, Coldren LA. Theoretical gain in strained InGaAs/AlGaAs quantum wells including valence-band mixing effects. *Appl Phys Lett* 1990;**57**: 2835.
23. Chong TC, Fonstad CG. Theoretical gain of strained-layer semiconductor lasers in the large strain regime. *IEEE J Quantum Electron* 1989;**25**: 171.

READING LIST

Ahn D, Chuang SL, Chang YC. Valence-band mixing effects on the gain and the refractive index change of quantum-well lasers. *J Appl Phys* 1988;**64**: 4056.

Ahn D, Chuang SL. Optical gain and gain suppression of quantum-well lasers with valence band mixing. *IEEE J Quantum Electron* 1990;**26**: 13.

APPENDIX NINE

Fermi's Golden Rule

A9.1 INTRODUCTION

This appendix derives a general expression for the rate of decay from an initial quantum mechanical state into a continuum of final states in the presence of a harmonic perturbation. The resulting expression for the transition rate is referred to as *Fermi's Golden Rule*, since it is a general result applicable to many quantum mechanical systems. Here we concentrate on the interaction between light and matter.

In physical terms, radiative transitions occur because the oscillating field of the photon alters the oscillating phase of the electron wavefunction in such a way that it becomes similar to the oscillating phase of another electron wavefunction typically in a different energy band. This phase-matching in time results in a strong coupling between the two electron states, analogous to the coupling of waveguide modes considered in Chapter 6. In the latter case, the coupling causes the energy initially in one mode to be transferred back and forth between the two modes. If only two electron wavefunctions were involved in the coupling, the electron would also oscillate between the two states (known as *Rabi* oscillations in the quantum world). However in most cases, a *density* of states are coupled to the initial electron wavefunction, and as a result, the electron transforms with an exponential decay from its initial wavefunction to one of many resonant wavefunctions. This electromagnetically induced phase-matching and subsequent transformation of an electron from one state to another is the fundamental mechanism of radiative transitions. The following treatment quantifies these physical arguments.

Diode Lasers and Photonic Integrated Circuits, Second Edition.
Larry A. Coldren, Scott W. Corzine, and Milan L. Mašanović.
© 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

A9.2 SEMICLASSICAL DERIVATION OF THE TRANSITION RATE

To characterize the electronic system and its interaction with light, we make use of two Hamiltonians: one that describes the electronic system in isolation, H_0, and a second that describes a classical perturbation created by the electromagnetic field, $H'(t)$. The wavefunction in Dirac notation, $|\Psi(t)\rangle$, provides a description of the state of the electron. Its evolution in time is governed by Schrödinger's equation:

$$i\hbar\frac{d}{dt}|\Psi(t)\rangle = \{H_0 + H'(t)\}|\Psi(t)\rangle. \tag{A9.1}$$

The problem we wish to solve using Eq. (A9.1) is the process of absorption (stimulated emission is analogous to absorption and hence leads to the same result, while spontaneous emission is treated separately in Chapter 4). To model an absorption event, we assume the electron initially occupies some ground state of the system, $|\psi_0\rangle$. The presence of the time-varying field excites a change in the state of the electron. Under the appropriate conditions, the electron can be excited to any number of higher energy final states, $|\psi_s\rangle$, where the sth state has energy $\hbar\omega_s$ above the ground state. This situation is depicted in Fig. A9.1.

Which state the electron eventually occupies is unknown, so we hypothesize a time-dependent superposition of the initial and possible final states of the electron:

$$|\Psi(t)\rangle = c_0(t)|\psi_0\rangle + \sum_s c_s(t)e^{-i\omega_s t}|\psi_s\rangle, \tag{A9.2}$$

with the initial conditions:

$$c_0(0) = 1 \quad \text{and} \quad c_s(0) = 0. \tag{A9.3}$$

The time-dependence of the unperturbed wavefunctions, $e^{-i\omega_s t}$, is included explicitly to remove the rapidly oscillating phase from the time dependence of the expansion coefficients ($|\psi_0\rangle$ and $|\psi_s\rangle$ are assumed to be independent of time). Note that Eq. (A9.2) satisfies Eq. (A9.1) with no perturbation present since $H_0|\psi_s\rangle = \hbar\omega_s|\psi_s\rangle$ and $H_0|\psi_0\rangle = 0$.

To obtain a set of useful equations, we begin by substituting Eq. (A9.2) into Eq. (A9.1). To move forward we recognize that all states are orthogonal to each other such that $\langle\psi_0|\psi_s\rangle = 0$ and $\langle\psi_{s'\neq s}|\psi_s\rangle = 0$. Furthermore, we have $\langle\psi_0|\psi_0\rangle = 1$ and $\langle\psi_s|\psi_s\rangle = 1$. If we multiply Eq. (A9.1) from the left with $\langle\psi_0|$, we obtain one equation for the coefficient c_0. Multiplying Eq. (A9.1) from the left with $\langle\psi_s|$, we obtain an additional set of equations, one for each c_s:

$$\begin{aligned} i\hbar\frac{dc_0}{dt} &= H'_{00}c_0 + \sum_s H'_{0s}c_s e^{-i\omega_s t}, \\ i\hbar\frac{dc_s}{dt} &= H'_{s0}c_0 e^{i\omega_s t} + \sum_{s'} H'_{ss'}c_{s'}e^{-i(\omega_{s'}-\omega_s)t}, \end{aligned} \tag{A9.4}$$

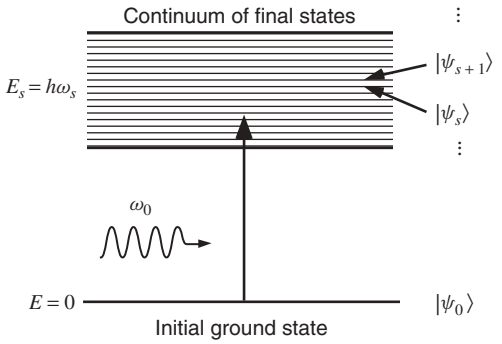


FIGURE A9.1: Model assumed for the process of absorption.

where the following shorthand notation is used:

$$H'_{s0} \equiv \langle \psi_s | H' | \psi_0 \rangle. \tag{A9.5}$$

To simplify Eq. (A9.4) further, we need to express the time dependence of the perturbation explicitly. For a harmonic perturbation, we can set

$$H'(t) \rightarrow H'(e^{i\omega_0 t} + e^{-i\omega_0 t}), \tag{A9.6}$$

where ω_0 is the oscillation frequency of the incident electromagnetic wave. The amplitude of the wave is contained in H'; however, we will not need to know the explicit details of H' for the purposes of this appendix.

The set of equations contained in Eq. (A9.4) is exact. The first approximation we make is known as the *rotating wave approximation*, which involves ignoring all terms that oscillate at frequencies comparable to or greater than either ω_s or ω_0. The reasoning is that these terms will oscillate positive and negative much faster than the time scales involved with changes in *c(t)* and hence, will average out to zero net contribution as time increases. Placing Eq. (A9.6) into Eq. (A9.4), the following combinations of angular frequencies appear: $\omega_0, \omega_s \pm \omega_0, (\omega_{s'} - \omega_s) \pm \omega_0$. If we assume $\omega_s \sim \omega_0$ and $\omega_{s'} \sim \omega_s$, then all combinations except $\omega_s - \omega_0$ are comparable to or greater than either ω_s or ω_0.

Ignoring all terms in Eq. (A9.4) except those that contain the difference frequency between the electronic and electromagnetic frequencies, $\omega_s - \omega_0$, we obtain the central equations of motion:

$$\begin{aligned} i\hbar \frac{dc_0}{dt} &= \sum_s H'_{0s} c_s e^{-i(\omega_s - \omega_0)t}, \\ i\hbar \frac{dc_s}{dt} &= H'_{s0} c_0 e^{i(\omega_s - \omega_0)t}. \end{aligned} \tag{A9.7}$$

These equations are known as the Wigner-Weisskopf equations within the rotating wave approximation.

To solve Eq. (A9.7) we integrate the second equation from 0 to t, using the initial conditions expressed in Eq. (A9.3) to obtain

$$c_s(t) = -\frac{i}{\hbar} H'_{s0} \int_0^t c_0(t') e^{i(\omega_s - \omega_0)t'} dt'. \tag{A9.8}$$

Inserting Eq. (A9.8) into the first equation of Eq. (A9.7), we obtain

$$\frac{dc_0(t)}{dt} = -\frac{1}{\hbar^2} \sum_s |H'_{s0}|^2 \int_0^t c_0(t') e^{-i(\omega_s - \omega_0)(t-t')} dt', \tag{A9.9}$$

where we have made use of the fact that $H'_{0s} = \left(H'_{s0}\right)^*$. We can convert the sum over final states to an integral with the assumption that the states are so closely spaced that they form a continuous distribution characterized by a density of states function in energy:

$$\sum_s |H'_{s0}|^2 \rightarrow \int |H'(E_s)|^2 \rho_f(E_s) \hbar \, d\omega_s. \tag{A9.10}$$

The matrix element $|H'(E_s)|^2$ represents an average of $|H'_{s0}|^2$ over all final s states existing with energies close to E_s above the ground state. Also, the final density of states function, ρ_f, represents the total *number* of states per unit energy.

Substituting Eq. (A9.10) into Eq. (A9.9) and reversing the order of integration, we can identify one portion of the equation as the inverse Fourier transform of the matrix element–density of states product:

$$f(t - t') = \int_{-\infty}^{\infty} \frac{1}{\hbar} |H'(E_s)|^2 \rho_f(E_s) e^{-i(\omega_s - \omega_0)(t-t')} \, d\omega_s. \tag{A9.11}$$

With this definition, Eq. (A9.9) becomes simply

$$\frac{dc_0(t)}{dt} = -\int_0^t c_0(t') f(t - t') \, dt'. \tag{A9.12}$$

Under certain circumstances, this integro-differential equation can be solved exactly. Its solution depends critically on the Fourier transform of the product $|H'(E_s)|^2 \rho_f(E_s)$, represented by the time-domain response function, $f(t - t')$. We will examine three such solutions here, the first of these leading to the famed Fermi's Golden Rule.

A9.2.1 Case I: The Matrix Element-Density of Final States Product is a Constant

If the density of final states is distributed evenly over a large energy range, and if the matrix element does not vary over this range, we can set $|H'(E_s)|^2 \rho_f(E_s)$

to a constant. Taking the Fourier transform we find that $f(t-t')$ becomes a delta function in time, as is evident from the following relation:

$$\int_{-\infty}^{\infty} e^{-i(\omega_s-\omega_0)(t-t')}\,d\omega_s = 2\pi\delta(t-t'). \tag{A9.13}$$

The integro-differential equation for this case reduces to a simple first-order differential equation

$$\begin{aligned} |H'(E_s)|^2\rho_f(E_s) &\to |H'|^2\rho_f \\ f(t-t') &= \frac{1}{\hbar}|H'|^2\rho_f \cdot 2\pi\delta(t-t') \\ \frac{dc_0(t)}{dt} &= -\frac{2\pi}{\hbar}|H'|^2\rho_f \cdot \int_0^t c_0(t')\delta(t-t')\,dt' = -\frac{\pi}{\hbar}|H'|^2\rho_f \cdot c_0(t). \end{aligned} \tag{A9.14}$$

The factor of 2π representing the total area under the delta function is reduced to π in the last equality because the time integration terminates in the center of the delta function such that only half of the area is included. The solution to Eq. (A9.14) is a decaying exponential:

$$c_0(t) = e^{-Wt/2} \to |c_0(t)|^2 = e^{-Wt}, \tag{A9.15}$$

where the decay rate is given by

$$W = \frac{2\pi}{\hbar}|H'|^2\rho_f. \tag{A9.16}$$

We can also solve for the final state probability coefficients to determine where the electron ends up after making the transition. Using Eq. (A9.15) in Eq. (A9.8), we obtain

$$|c_s(t\to\infty)|^2 = \frac{|H'_{s0}|^2}{(\hbar W/2)^2 + (E_s-\hbar\omega_0)^2}. \tag{A9.17}$$

If we sum this probability over all final states (i.e., multiply Eq. (A9.17) by $\rho_f dE_s$ and integrate over all energies), the total probability will equal one. Thus, the electron will eventually appear somewhere in the continuum, we just don't know exactly where. The peak probability occurs at the state with energy $E_s = \hbar\omega_0$. The electron can appear at other states as well; however, the probability of this happening declines away from the peak with a distribution characterized by a Lorentzian with $\Delta E_{FWHM} = \hbar W$.

Equation (A9.17) also tells us that the electron only interacts with final states clustered around $E_s = \hbar\omega_0$. In other words, states that are not within $\hbar W/2$ of the resonant energy effectively do not participate in the transition. Therefore, our earlier assumption that $|H'|^2\rho_f$ must be constant over a large energy range

can be refined to state that as long as $|H'(E_s)|^2\rho_f(E_s)$ is flat over energies for which Eq. (A9.17) has significant amplitude, the solutions of Case I can be applied.

This translates into evaluating $|H'(E_s)|^2\rho_f(E_s)$ at $E_s = \hbar\omega_0$, and requiring that it be constant for energies $\gg \hbar W$. In semiconductor applications, $|H'|^2\rho_f$ is usually a smooth enough function of E_s to easily meet this requirement. Therefore, interactions between light and semiconductor materials will typically induce exponentially decaying solutions for the initial state of the system.

Fermi's Golden Rule. Our ultimate objective is to obtain a transition rate that can be used in a rate equation. The preceding analysis allows us to determine how quickly photons are being removed from the electromagnetic field. For example, each transition event decays with the decay rate, W. If we resupply the ground state with a new electron every $1/W$ seconds, then the ground state will remain filled and every $1/W$ seconds on average, a photon will be absorbed. We can therefore view W as the rate at which photons are absorbed in the active region, or the transition rate. If N_p is the photon density and V_p is the mode volume, then N_pV_p photons exist in a given mode. The rate at which these photons disappear can be expressed as

$$\frac{dN_pV_p}{dt} = -W. \tag{A9.18}$$

Dividing by the mode volume and introducing the confinement factor, the photon *density* rate equation becomes

$$\frac{dN_p}{dt} = -\frac{V}{V_p}\frac{W}{V} \equiv -\Gamma R_r, \tag{A9.19}$$

where $\Gamma(= V/V_p)$ is the optical confinement factor, and $R_r(= W/V)$ is defined as the radiative transition rate *per unit volume* of active material. If we absorb the $1/V$ into ρ_f, such that the final density of states is interpreted as a density per unit energy *and* volume (as is customary), then the transition rate per unit volume becomes

$$R_r = \frac{2\pi}{\hbar}|H'(E_s)|^2\rho_f(E_s)|_{E_s=\hbar\omega_0}. \tag{A9.20}$$

This result and its interpretation as a transition rate is known as *Fermi's Golden Rule*. The evaluator bar reminds us that the product $|H'|^2\rho_f$ must be evaluated at $E_s = \hbar\omega_0$ for reasons discussed in reference to Eq. (A9.17).

The derivation presented here assumes the electron is making an upward transition to model absorption. However, we could just as well assume the electron is making a downward transition by setting $\omega_s \to -\omega_s$, allowing us to model stimulated emission. This substitution requires that we use the opposite time harmonic in the perturbation Hamiltonian Eq. (A9.6), however, aside from this simple change the rest of the derivation is identical. Therefore, R_r has the same form for both

absorption and stimulated emission. The difference between the upward and downward transition rates appears only when we include the probability of finding an electron in the initial state and no electron in the final state. The theory of gain given in Chapter 4 elaborates on this issue in some detail since the balance between absorption and stimulated emission provides the key to understanding optical gain in any material.

A9.2.2 Case II: The Matrix Element-Density of Final States Product is a Delta Function

Assume that only a single final state exists as opposed to a continuous density of final states, as might exist in an atomic transition from one energy level to another. We can model this situation by replacing the density of final states with a delta function in energy to represent the solitary energy level, E_f. To simplify matters, we will assume that $\omega_f = \omega_0$ such that perfect resonance between the electronic and electromagnetic systems exists (see Problem A9.2 for the more general solution). Taking the Fourier transform of the delta function, we find that $f(t-t')$ is a constant independent of time, which allows us to again solve Eq. (A9.12) exactly. Summarizing the results, we have

$$\begin{aligned} \rho_f(E_s) &\to \frac{1}{\hbar}\delta(\omega_s - \omega_f) \\ f(t-t') &= \frac{1}{\hbar^2}|H'|^2 \quad \text{(assuming } E_f = \hbar\omega_0\text{)} \\ \frac{dc_0(t)}{dt} &= -\frac{1}{\hbar^2}|H'|^2 \int_0^t c_0(t')\,dt' \to c_0(t) = \cos\Omega t, \quad \Omega^2 = \frac{1}{\hbar^2}|H'|^2. \end{aligned} \tag{A9.21}$$

Using Eq. (A9.8) with $\omega_f = \omega_0$, the final state probability amplitude is given by $c_f(t) = \sin\Omega t$. In words, when the electron interacts with only one possible final state, the probability amplitude oscillates back and forth sinusoidally between the initial and final states. The oscillation frequency depends on the magnitude of interaction, $|H'(E_f)|^2$. Physically, the energy is continually shifting back and forth between the electromagnetic field and the electron.

In a more general sense, we can view $f(t-t')$ as a *memory* function that represents how strongly previous events are coupled to current changes in $c_0(t)$. With only one final state to interact with, the system has infinite memory ($f(t-t')$ is independent of time), and the interaction continues indefinitely in the same periodic fashion. This periodic exchange of energy between the electromagnetic field and the electron is characteristic of *Rabi* oscillations, which occur whenever the coupled system exhibits a strong memory of interaction.

In contrast, the system considered in Case I (Fermi's Golden Rule) has no memory of previous events ($f(t-t')$ is a delta function in time). The existence of numerous *equally probable* transition pathways destroys the system memory, such that the interaction only depends on the present state of the electron. From another point of view, the probability amplitudes of the various pathways

combine in such a way as to make the process irreversible causing the initial state to decay exponentially toward a final state (rather than oscillating back and forth indefinitely). The next case bridges the gap between Case I and Case II, producing a system with finite memory and oscillatory decaying exponential solutions.

A9.2.3 Case III: The Matrix Element-Density of Final States Product is a Lorentzian

This situation can occur when analyzing the spontaneous emission process of a two-level system placed in a highly resonant cavity. In general, the analysis of spontaneous emission leads to an equation identical to Eq. (A9.12), with the exception that the density of final electronic states is interpreted as the density of free-space optical modes, and the field strength within the matrix element is set equal to the vacuum-field strength of the free-space mode at E_s. When the two-level system is placed in a cavity, the vacuum-field strengths of the free-space optical modes become enhanced near the cavity resonances, which in turn enhances the matrix element. For a highly resonant cavity, this enhancement has a Lorentzian lineshape about each resonance.

With $|H'(E_s)|^2\rho_f(E_s)$ set to a Lorentzian, its Fourier transform, $f(t-t')$, becomes a decaying exponential in time. In other words, the memory of the system decays exponentially as it recedes away from the present. One way of viewing this is that the spontaneously emitted photon only has a finite lifetime in the cavity before it escapes, limiting the memory of the system to a finite duration. In any case, Eq. (A9.12) can again be solved exactly; however, the solution is not as obvious as it was in Case I or II (see Problem A9.3). Summarizing, we have

$$|H'(E_s)|^2\rho_f(E_s) \rightarrow |H'|^2\rho_f K \cdot \frac{1}{1+(2\Delta\omega\tau_p)^2}$$

$$f(t-t') = \frac{KW_0}{4\tau_p}e^{-|t-t'|/2\tau_p} \tag{A9.22}$$

$$\frac{dc_0(t)}{dt} = -\frac{KW_0}{4\tau_p}\int_0^t c_0(t')e^{-|t-t'|/2\tau_p}dt', \quad \text{where} \quad W_0 = \frac{2\pi}{\hbar}|H'|^2\rho_f.$$

In the preceding equations, $\Delta\omega \equiv \omega_s - \omega_0$, τ_p is the photon lifetime of the cavity (note that $\Delta\omega_{FWHM}\tau_p = 1$), K is the enhancement of the vacuum-field strength within the cavity at the resonance peak ($K \propto \tau_p$), and W_0 is the decay rate found in Case I (with ρ_f in this case representing the density of free-space optical modes). Using the initial conditions expressed in Eq. (A9.3), the function satisfying Eq. (A9.22) can be written as

$$c_0(t) = \frac{p_+e^{p-t} - p_-e^{p+t}}{p_+ - p_-}, \quad \text{where} \quad p_\pm = -\frac{1}{4\tau_p}[1 \pm (1-4\tau_p KW_0)^{1/2}]. \tag{A9.23}$$

If the memory of the system characterized by the photon lifetime, τ_p, is very short such that $4\tau_p KW_0 \ll 1$, the solution reduces to

$$p_+ \approx -\frac{1}{2\tau_p},\ p_- \approx -\tfrac{1}{2}KW_0, \quad \text{and} \quad |c_0(t)|^2 \approx e^{-KW_0 t}. \tag{A9.24}$$

The last equality is valid for all but very short times, which can be verified from Eq. (A9.23) and the fact that $p_+ \gg p_-$. In words, when the photon escapes much quicker than the time it takes to make a transition (represented by the inequality $4\tau_p KW_0 \ll 1$), the initial state decays exponentially as was found in Case I for a system with no memory. From another point of view, the inequality $4\tau_p KW_0 \ll 1$ allows us to treat $|H'|^2\rho_f$ as a constant for energy ranges $\gg \hbar KW_0$, making this case equivalent to Case I. However in the present case, the decay rate is enhanced by the vacuum-field strength enhancement inside the cavity, K. This enhancement of the decay or emission rate was first postulated by Purcell many years ago.

If we go to the other extreme and assume that the system has a strong memory (the photon lifetime in the cavity is very long) such that $4\tau_p KW_0 \gg 1$, the roots become

$$p_\pm \approx -\frac{1}{4\tau_p} \pm i\left(\frac{KW_0}{4\tau_p}\right)^{1/2}. \tag{A9.25}$$

Complex roots suggest that $|c_0(t)|^2$ exhibits damped oscillatory behavior. Rabi oscillations of this nature are again the result of the system having a strong memory. In other words, the lifetime of the fields in the cavity is long enough to induce additional absorption and re-emission of electromagnetic energy. However, contrary to Case II, $|c_0(t)|^2$ does eventually reduce to zero at a decay rate $\approx 1/2\tau_p$ (which corresponds to the decay rate of the field *amplitude* within the cavity).

Case III is perhaps the most interesting theoretically because it bridges the gap between the first two cases showing clearly how the oscillating solution transforms to the exponentially decaying solution as the photon lifetime of the cavity is adjusted from infinity to zero, respectively.

READING LIST

Kroemer H. *Quantum mechanics*. Englewood Cliffs, NJ: Prentice Hall; 1994. Chapter 19.

Lai HM, Leung PT, Young K. Electromagnetic decay into a narrow resonance in an optical cavity. *Phys Rev A* 1988;**37**: 1597.

Purcell EM, Torrey HC, and Pound* RV, Resonance Absorption by Nuclear Magnetic Moments in a Solid, Radiation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, *Phys. Rev.* 1946; **69**: 37–38.

PROBLEMS

1. Determine $f(t-t')$ in Case I assuming $|H'|^2\rho_f$ is constant for energies within $E_s \pm \Delta E/2$, but *zero* everywhere else. Using this expression for $f(t-t')$, describe qualitatively the range over which the solution for $c_0(t)$ is expected to behave like a decaying exponential. In your description, be sure to consider
   (i) ΔE in relation to W, and
   (ii) E_s in relation to $\hbar\omega_0$.
   Qualitative plots of $f(t-t')$ superimposed on the expected dependence of $c_0(t')$ may prove useful in developing your answer.
2. Solve for $c_0(t)$ and $c_s(t)$ in Case II without assuming that $E_f = \hbar\omega_0$. Plot both as a function of time, assuming the energy difference $(E_f - \hbar\omega_0)^2$ is equal to
   (a) $|H'|^2$ and
   (b) $12\,|H'|^2$.
   Can you make a comment about the significance of tuning E_f to $\hbar\omega_0$?
3. Using the initial conditions in Eq. (A9.3), derive Eq. (A9.23) in Case III assuming the solution is the sum of two exponentials with constant coefficients. Numerically plot $c_0(t)$ as a function of t/τ_p for $4\tau_p K W_0$ equal to
   (a) 100,
   (b) 1, and
   (c) 0.01.
   Plot any other cases that may seem interesting to you. Can you explain this behavior?

APPENDIX TEN

Transition Matrix Element

A10.1 GENERAL DERIVATION

As derived in Chapter 4 the matrix element $|H'_{21}|^2$ can be written in terms of a *transition matrix element* $|M_T|^2$. We begin by repeating this equation, Eq. (4.22),

$$|H'_{21}|^2 = \left(\frac{q\mathscr{A}_0}{2m_0}\right)^2 |M_T|^2, \quad \text{where} \quad |M_T|^2 \equiv |\langle u_c|\hat{\mathbf{e}}\cdot\mathbf{p}|u_v\rangle|^2 |\langle F_2|F_1\rangle|^2. \quad \text{(A10.1)}$$

As discussed in Appendix 8, the momentum matrix element, $|M|^2$ can be estimated from experiment. We now need to determine $|M_T|^2$ in terms of $|M|^2$. The difference between the two matrix elements is that $|M|^2$ determines the transition probability between u_s and the basis functions (u_x, u_y, u_z, or collectively u_i), whereas $|M_T|^2$ determines the transition probability between $u_c (= u_s)$ and the valence band Bloch functions (u_{hh}, u_{lh}, u_{so}, or collectively u_v). By expanding the u_v in terms of the u_i using Eq. (A8.11), we can express $|M_T|^2$ in terms of $|M|^2$. Before we do this, however, we need to discuss spin degeneracy and how to include it here.

In Appendices 1 and 8, a simple factor of 2 for spin degeneracy was included in the definition of the density of states function. However, there are subtleties involved, which are often overlooked when we simply include a factor of 2 in our equations. For example, to include the spin degeneracy in our evaluation of $|M_T|^2$ we must obviously sum over both $u_c \rightarrow u_v$ and $\bar{u}_c \rightarrow \bar{u}_v$ transitions. However, what is not so obvious is that in our sum we must also include $\bar{u}_c \rightarrow u_v$ and $u_c \rightarrow \bar{u}_v$ transitions! This is necessary because the LH and SO valence band Bloch functions are made up of both spin-up *and* spin-down basis functions, as seen from their

Diode Lasers and Photonic Integrated Circuits, Second Edition.
Larry A. Coldren, Scott W. Corzine, and Milan L. Mašanović.
© 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

definitions in Eq. (A8.11). Therefore, a total of *four* transitions must be considered, as shown in Fig. A10.1. Because spin is already accounted for in the density of states function, we will sum over these transitions and then divide by 2 to remove the spin degeneracy. Thus, the reader should be aware that spin degeneracy has been removed from all expressions for the transition matrix element $|M_T|^2$ derived in this appendix.

Summing over the four transitions shown in Fig. A10.1, the transition matrix element defined in Eq. (A10.1) becomes

$$|M_T|_v^2 = \frac{1}{2} \sum_{u_c, \bar{u}_c} \sum_{u_v, \bar{u}_v} |\langle u_c | \hat{\mathbf{e}} \cdot \mathbf{p} | u_v \rangle|^2, \tag{A10.2}$$

where the factor of 1/2 is to remove the spin degeneracy. We have set the envelope function overlap integral, $|\langle F_2 | F_1 \rangle|^2$, equal to unity because for the moment we will be interested in transitions between two bulk plane wave electron states. Later on, we will return to a more general form of Eq. (A10.2) which does include the envelope function overlap integrals.

To simplify Eq. (A10.2), we can first of all replace the dot product between the unit polarization vector and the electron momentum operator, $\hat{\mathbf{e}} \cdot \mathbf{p}$, with the expansion, $e_x p_x + e_y p_y + e_z p_z$. Then by using the selection rules given in Eqs. (A8.8) through (A8.10), in combination with the expansions of the valence band Bloch functions given in Eqs. (A8.11), we can reduce the expression for $|M_T|^2$ to a very simple form. To aid the reader in following the derivation, we give here the intermediate step in simplifying $|M_T|^2$ for the three valence band transitions:

$$\begin{aligned}
|M_T|_{hh}^2 &= \tfrac{1}{4}|M|^2\{|-e_x - ie_y|^2 + 0 + 0 + |e_x - ie_y|^2\}, \\
|M_T|_{lh}^2 &= \tfrac{1}{12}|M|^2\{|2e_z|^2 + |e_x - ie_y|^2 + |-e_x - ie_y|^2 + |2e_z|^2\}, \\
|M_T|_{so}^2 &= \tfrac{1}{6}|M|^2\{|-e_z|^2 + |e_x - ie_y|^2 + |-e_x - ie_y|^2 + |-e_z|^2\}.
\end{aligned} \tag{A10.3}$$

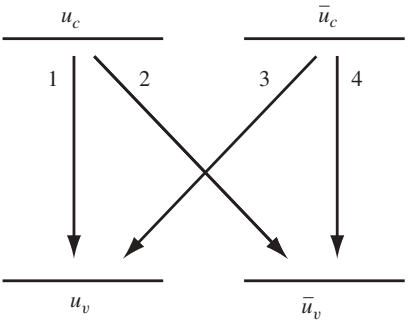


FIGURE A10.1: The four possible transitions between the spin-degenerate C and V bands, which must be considered when estimating the transition matrix element.

Each of the terms within brackets corresponds to one of the four spin-degenerate transitions (the ordering of terms from left to right corresponds to the numbering shown in Fig. A10.1). Note that in every case, the first term is equal to the fourth and the second term is equal to the third (leading to the standard factor of 2 for spin degeneracy).

To make the final expression as general as possible, we first replace every occurrence of $e_x^2 + e_y^2$ with the equivalent expression, $1 - e_z^2$ (since $\hat{\mathbf{e}}$ is a unit vector). This substitution places everything in terms of e_z. We can then interpret e_z as the component of $\hat{\mathbf{e}}$ which is parallel to the electron k-vector, since $\mathbf{k}$ is directed along z (an assumption made in defining Eq. (A8.11). In other words, we can set $e_z = \hat{\mathbf{k}} \cdot \hat{\mathbf{e}}$, where $\hat{\mathbf{k}}$ is a unit vector directed along $\mathbf{k}$. Using these substitutions, we find that

$$|M_T|_v^2/|M|^2 = \begin{cases} \frac{1}{2}(e_x^2 + e_y^2) = \frac{1}{2}(1 - |\hat{\mathbf{k}} \cdot \hat{\mathbf{e}}|^2) & \text{for HH band,} \quad \text{(A10.4)} \\ \frac{1}{6}(e_x^2 + e_y^2 + 4e_z^2) = \frac{1}{2}(\frac{1}{3} + |\hat{\mathbf{k}} \cdot \hat{\mathbf{e}}|^2) & \text{for LH band,} \quad \text{(A10.5)} \\ \frac{1}{3}(e_x^2 + e_y^2 + e_z^2) = \frac{1}{3} & \text{for SO band.} \quad \text{(A10.6)} \end{cases}$$

The *relative* transition strengths given in Eqs. (A10.4) through (A10.6) allow us to relate the transition matrix element, $|M_T|^2$, needed in our gain calculations to the experimentally measurable matrix element, $|M|^2$. Note that the use of a dot product has allowed us to drop any reference to a coordinate system and, hence, drop the constraint that the electron k-vector be directed along z (in other words, the physics does not lie in the coordinate system we choose, but in the relative orientation between the field polarization $\hat{\mathbf{e}}$ and the electron k-vector).

To examine the dependence of Eqs. (A10.4) through (A10.6) on the field polarization in a more visual fashion, we have plotted the relative transition strengths for C–HH and C–LH transitions in Fig. A10.2 as a function of the angle between the electron k-vector and the electric field polarization, $\hat{\mathbf{e}}$. These three-dimensional renderings reveal that the strength of interaction between each electron plane wave state and photon is highly polarization dependent. However, the striking features of Fig. A10.2 do not reveal themselves in bulk material because photons of a given polarization interact with a great number of electrons, all with k-vectors pointing in different directions. The average, over all these interactions transforms the interesting shapes in Fig. A10.2 into uniform spheres. In fact, the average of $|\hat{\mathbf{k}} \cdot \hat{\mathbf{e}}|^2$ for $\hat{\mathbf{k}}$ sweeping over all three dimensions is equal to 1/3. Thus, for all three valence band transitions, the bulk material transition matrix element is just equal to $1/3 \times |M|^2$ (spin excluded) for any electric field polarization.

A10.2 POLARIZATION-DEPENDENT EFFECTS

The derivation presented here assumed plane wave states for the envelope functions, which then led to the polarization dependence illustrated in Fig. A10.2. In quantum-confined structures, the envelope functions are typically constructed from

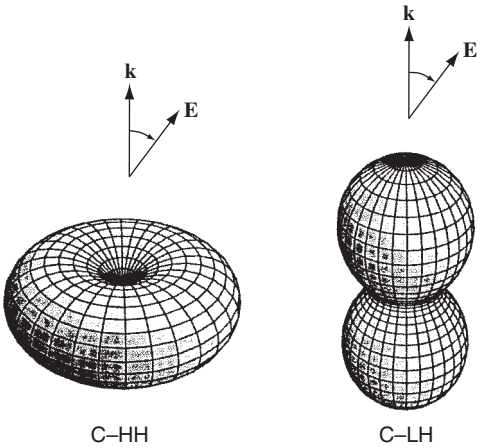


FIGURE A10.2: Dependence of the transition strength, $|M_T|^2$, on angle between the electron's k-vector and the incident electric field vector, **E**, for C–HH and C–LH transitions (C–SO transitions are independent of angle). For C–HH transitions, $|M_T|^2$ is zero when $\mathbf{E} \parallel \mathbf{k}$ and becomes a maximum of $\frac{1}{2} \times |M|^2$ when $\mathbf{E} \perp \mathbf{k}$. For C–LH transitions, when $\mathbf{E} \parallel \mathbf{k}$, $|M_T|^2$ has a peak value of $\frac{2}{3} \times |M|^2$ and is reduced to $\frac{1}{6} \times |M|^2$ when $\mathbf{E} \perp \mathbf{k}$.

two (or more) plane wave states. The magnitude squared of the transition matrix element in (A10.2) will, in general, then contain cross-terms between the various plane waves that make up the confined state. In the following discussion we will ignore these cross-terms, making the analysis simpler. Later, we will see that in quantum wells, the conclusions derived here are consistent with the band-mixing model for transitions near the band edge.

Neglect of the cross-terms in (A10.2) implies that we can treat the plane waves that make up the confined states as independent from each other. In this approximation, each plane wave's k-vector direction will then have a corresponding polarization dependence similar to that derived in Section A10.1 and shown in Fig. A10.2. Near the band edge in quantum-confined structures, the k-vectors are quantized along certain directions, and the situation will be as shown in Fig. A10.3 for a typical quantum well and quantum wire. In the quantum well, all k-vectors point along the same axis and the polarization dependence is simply proportional to Fig. A10.2. However, in the quantum-wire case, we must average over the polarization dependence of each plane wave (as indicated by the dashed curve in the lower right side of Fig. A10.3). To quantify the *average* polarization dependence, we choose our coordinate system along the confinement axis (axes) of the structure. It is then possible to evaluate the average transition strength along the three orthogonal field polarizations, $\hat{\mathbf{e}}_x$, $\hat{\mathbf{e}}_y$, and $\hat{\mathbf{e}}_z$, simply by replacing $\hat{\mathbf{k}}$ in Eqs. (A10.4) through (A10.6) with some appropriate *average* k-vector direction, $\hat{\mathbf{k}}_{ave}$.

We leave it to the reader to justify that $\hat{\mathbf{k}}_{ave}$ is obtained simply by finding the average direction of all allowed k-vectors within the *first octant* of our coordinate

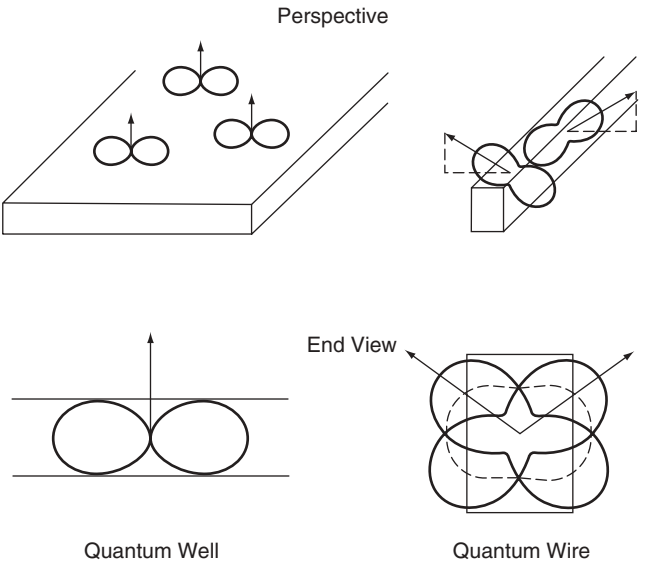


FIGURE A10.3: Illustration of how quantum confinement in a quantum well and a quantum wire serves to "polarize" the momentum of band edge electrons along certain directions. The C–HH transition strength is superimposed on each electron's k-vector in the quantum-well case. The end view of the quantum well suggests that C–HH interactions with light are strongest when the light is polarized in the plane of the well. In the quantum-wire case, the C–LH transition strength is used. The end view shows two possible k-vector directions for electrons. The average C–LH transition strength is indicated by the dashed curve.

system (if we included all octants in our average, $\hat{\mathbf{k}}_{ave}$ would always be zero!). Next we list $\hat{\mathbf{k}}_{ave}$ for bulk and various quantum-confined structures for band edge states (where "band edge" implies that the total k-vectors are simply equal to the quantized k-vectors):

$$\begin{aligned}
&\hat{\mathbf{k}}_{ave} = (1/\sqrt{3})(\hat{\mathbf{k}}_x + \hat{\mathbf{k}}_y + \hat{\mathbf{k}}_z), && \text{bulk} \\
&\hat{\mathbf{k}}_{ave} = \hat{\mathbf{k}}_z, && \text{quantum well } (L_z) \\
&\hat{\mathbf{k}}_{ave} = (1/\sqrt{2})(\hat{\mathbf{k}}_x + \hat{\mathbf{k}}_z), && \text{quantum wire } (L_x = L_z) \\
&\hat{\mathbf{k}}_{ave} = (1/\sqrt{6})(\hat{\mathbf{k}}_x + \hat{\mathbf{k}}_y + 2\hat{\mathbf{k}}_z). && \text{quantum dot } (L_x = L_y = 2L_z)
\end{aligned} \tag{A10.7}$$

The prefactors are normalization constants since $\hat{\mathbf{k}}_{ave}$ is a unit vector. The last equation was obtained assuming $k \propto 1/L$ (which is exactly true only for an infinitely deep well).

Figure A10.4 illustrates the band edge transition strengths for the three orthogonal polarizations in the four structures listed above, obtained by substituting the $\hat{\mathbf{k}}_{ave}$ defined for each structure into Eqs. (A10.4) through (A10.6). When multiplied by $|M|^2$, the numbers in Fig. A10.4 give the magnitude of the transition matrix element, $|M_T|^2$, for each particular case (spin excluded). Note that the sum of

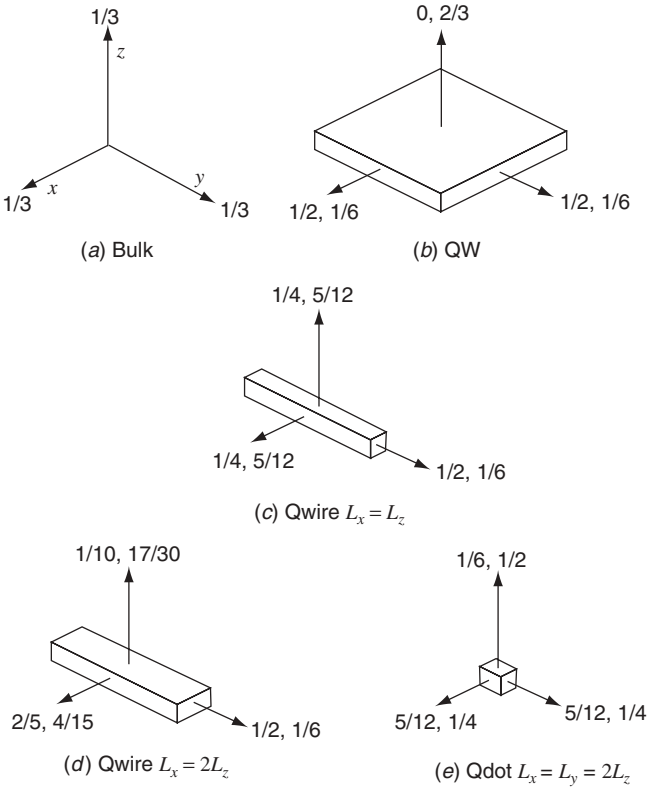


FIGURE A10.4: Relative band edge transition strengths for various quantum confinement structures. The coordinates referred to in the text are indicated in (a). The magnitude of the transition matrix element is found by multiplying the relative numbers by $|M|^2$. For example, with light polarized along the wire direction of (c), the band edge C–LH transition strength $|M_T|^2 = \frac{1}{6} \times |M|^2$.

the transition strengths over the three polarizations for each type of transition is always equal to $|M|^2$, as is the case in bulk material. Thus, the "total" transition strength for band edge transitions is always conserved. The difference is that in quantum-confined structures, a *redistribution* of the transition strength among the three polarizations occurs due to the nonuniform distribution of k-vector directions.

To treat arbitrary polarizations, we can always break the field up into the three orthogonal components shown in Fig. A10.4. The transition matrix element is then simply given by the trigonometric sum of the three components, or

$$|M_T|_v^2 = |M|^2 \sum_i e_i^2 S_i^v, \quad i = x, y, z \tag{A10.8}$$

where the S_i^v are the transition strengths determined from Fig. A10.4. As an example use of Eq. (A10.8), we examine the polarization-dependent characteristics

of a quantum-wire structure. From Fig. A10.4 (c), the *ratio* of the transition matrix element between the C–LH and C–HH transitions is $\frac{1}{3}$ when $\hat{\mathbf{e}}$ is parallel to the wire and $\frac{5}{3}$ when $\hat{\mathbf{e}}$ is perpendicular to the wire. In general, from Fig. A10.4 (c) and Eq. (A10.8), we have

$$\frac{|M_T|_{lh}^2}{|M_T|_{hh}^2} = \frac{\frac{1}{3}\cos^2\theta + \frac{5}{6}\sin^2\theta}{\cos^2\theta + \frac{1}{2}\sin^2\theta}, \tag{A10.9}$$

where θ is the angle between $\hat{\mathbf{e}}$ and the axis of the wire.

A10.3 INCLUSION OF ENVELOPE FUNCTIONS IN QUANTUM WELLS

As shown in Appendix 8 quantum confinement changes the valence band structure through the interaction and mixing of the envelope functions. That is, the general valence band wavefunction in Eq. (A8.31) consists of both HH and LH envelope function components. Thus, the transition matrix element in Eq. (A10.1) must be modified and is now expressed as

$$|M_T|_v^2 = \sum_{u_c,\bar{u}_c} |\langle u_c|\hat{\mathbf{e}}\cdot\mathbf{p}|u_A\rangle\langle F_2|F_{hh}\rangle + \langle u_c|\hat{\mathbf{e}}\cdot\mathbf{p}|u_B\rangle\langle F_2|F_{lh}\rangle|^2. \tag{A10.10}$$

Equivalent terms summing over the spin degenerate counterparts of u_A and u_B (the u_C and u_D Bloch functions) increase the sum by a factor of two. This is accounted for by the removal of the factor of $\frac{1}{2}$ appearing in Eq. (A10.2) (throughout this appendix, the spin degeneracy is removed from the transition matrix element and included in the reduced density of states function in Chapter 4).

The transition matrix element in Eq. (A10.10) can be placed in a more elegant form by following a procedure similar to that outlined in Section A10.1. However, in the present case we must average the transition matrix element over all in-plane k-vector directions to remove the cross-term which results from squaring Eq. (A10.10). With the help of Eqs. (A8.32), (A8.33), and (A8.11), the transition matrix element reduces to

$$|M_T|_v^2 = |M|^2[\tfrac{2}{3}|\langle F_2|F_{lh}\rangle|^2], \qquad \text{TM } (\hat{\mathbf{e}} \parallel \hat{\mathbf{z}}) \tag{A10.11}$$

$$|M_T|_v^2 = \frac{|M|^2}{2}[|\langle F_2|F_{hh}\rangle|^2 + \tfrac{1}{3}|\langle F_2|F_{lh}\rangle|^2], \quad \text{TE } (\hat{\mathbf{e}} \perp \hat{\mathbf{z}}) \tag{A10.12}$$

where z is assumed to be the quantization direction. At the band edge (where $F_{lh} = 0$ for HH states, and $F_{hh} = 0$ for LH states), the preceding expressions give the same transition strengths as those in Fig. A10.4 assuming the overlap integrals are close to unity (which is true to within 5% to 10% typically, and would be true exactly if the effective mass in the C and V bands were identical). However, as we move away from the band edge, band-mixing occurs such that both $F_{hh,lh}$ are present in any one wavefunction, altering the transition strengths from those shown

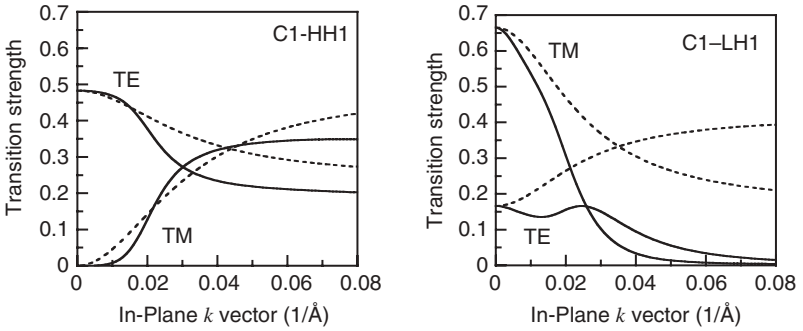


FIGURE A10.5: Relative transition strengths for both TE and TM light polarization for the two lowest subband transitions in an unstrained GaAs/$Al_{0.2}Ga_{0.8}As$ 80 Å QW. The dashed curves represent what one would calculate assuming parabolic subbands. The transition strength as plotted here is defined as $|M_T|^2/|M|^2$ (bulk value is 1/3).

in Fig. A10.4. Sample calculations of the transition strength, which illustrate this effect are plotted as a function of transverse k-vector, k_t in Fig. A10.5.

In finding the envelope functions and evaluating the overlap integrals numerically, we must make sure they are properly normalized. Normalization of the wavefunctions is obtained through the following relations

$$F_i(\text{norm}) = \frac{F_i}{\sqrt{N_i}}, \quad \text{where } N_{hh,lh} = \langle F_{hh}|F_{hh}\rangle + \langle F_{lh}|F_{lh}\rangle, \qquad N_2 = \langle F_2|F_2\rangle. \tag{A10.13}$$

The preceding envelope functions refer to the functions along the confinement direction, and hence the brackets indicate integration along z. The in-plane envelope functions are simple plane waves. Thus, we require **k** conservation in the plane of the well, which then yields an in-plane overlap integral of unity, justifying our conversion of a volume integral into an integral along z only.

READING LIST

Corzine SW, Yan R-H, Coldren LA. Optical gain in III-V bulk and quantum well semiconductors. In Zory PS Jr, editor. *Quantum well lasers*. San Diego, CA: Academic Press; 1993. Chapter 1.

APPENDIX ELEVEN

Strained Bandgaps

In this appendix, the details of how strain affects the bandgap of III–V semiconductors is considered. To provide some background, we will begin by reviewing concepts of stress and strain in a crystal lattice [1].

A11.1 GENERAL DEFINITIONS OF STRESS AND STRAIN

A crystal lattice that feels an external force will react by distorting in some fashion. The force per unit area, or stress, is usually defined by a *stress* tensor, σ_{ij}, as depicted in Fig. A11.1. *Shear* components of stress ($i \neq j$) will cause the crystal to rotate unless equal and opposite components exist (for example, if $\sigma_{23} = \sigma_{32}$). If equal and opposite components do exist, the shear stress will deform a cubic lattice into a nonrectangular shape (i.e., the crystal axes will become nonorthogonal). In typical semiconductor applications, these types of deformations are rare, and for the present purposes, we will assume that $\sigma_{ij} = 0$ for $i \neq j$. *Normal* components of stress ($i = j$) will cause the crystal to expand or contract along the crystal axes, but in contrast to shear stress, the deformed lattice remains rectangular. The six faces of a cubic crystal can be acted on by three normal forces: σ_{11}, σ_{22}, and σ_{33}. Because we are only considering normal components, the notation can be abbreviated to σ_1, σ_2, and σ_3. As suggested in Fig. A11.1, the σ_i are defined as positive for outward directed forces.

The next consideration involves the mathematical description of the lattice distortion. The strained state of the lattice is usually defined by a *strain* tensor, ε_{ij}. Each component of the strain tensor defines some aspect of the distortion of the

Diode Lasers and Photonic Integrated Circuits, Second Edition.
Larry A. Coldren, Scott W. Corzine, and Milan L. Mašanović.
© 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

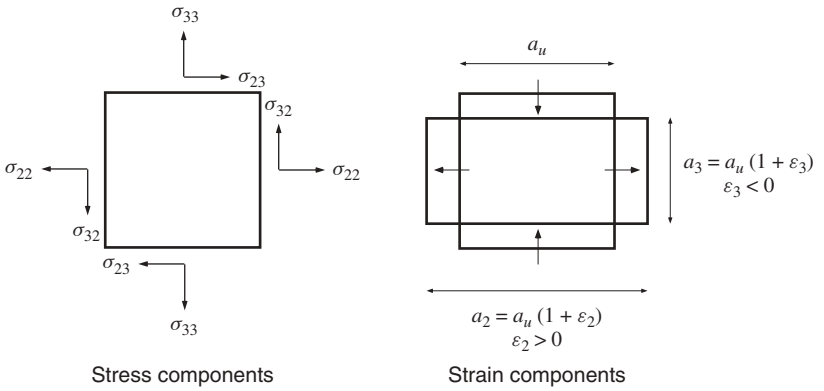


FIGURE A11.1: Components of stress tensor, σ_{ij}, and examples of the abbreviated strain tensor, ε_i, within the $y-z$ plane.

lattice away from its unstrained shape. Restricting our attention to the effects of normal forces on the lattice, we only need to consider the three diagonal components of the strain tensor: $\varepsilon_{11}, \varepsilon_{22}$, and ε_{33} (since the crystal axes remain orthogonal to each other). Again using the abbreviated notation, these can be written as $\varepsilon_1, \varepsilon_2$, and ε_3. The ε_i measure the fractional increase ($\varepsilon_i > 0$) or decrease ($\varepsilon_i < 0$) of the crystal lattice along the ith axis as illustrated in Fig. A11.1. For example, if the description of the lattice distortion is given as $\varepsilon_1 = 0, \varepsilon_2 = 0.01$, and $\varepsilon_3 = -0.01$, then we know that the lattice is undistorted along x, is larger by 1% along y, and is smaller by 1% along z, as qualitatively illustrated in the figure within the $y-z$ plane.

With the stress and strain tensors defined, we now need to relate them to predict how a particular stress or set of external forces leads to a particular strain or lattice deformation. In a uniform material, the strain is proportional to the magnitude of the applied stress, as long as we remain within the elastic limits of the material. Hooke's law in an isotropic medium expresses this scalar relation as $\sigma = C\varepsilon$, where the constant, C, is Young's modulus. In a crystal, one type of stress may lead to more than one type of strain, implying that the different components of both stress and strain tensors are potentially related. The more general form of Hooke's law excluding shear components of stress can be written as

$$\begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{12} \\ C_{12} & C_{11} & C_{12} \\ C_{12} & C_{12} & C_{11} \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}. \tag{A11.1}$$

The C_{ij} are referred to as the *elastic stiffness coefficients* or the *elastic moduli*. The expression written here assumes the crystal has cubic symmetry such that all off-diagonal elements are equal, and all diagonal elements are equal, leaving us with only two elastic moduli, denoted C_{11} and C_{12}. For crystals with less symmetry,

more C_{ij} components may need to be specified (if we include shear stress, an additional C_{44} is required to complete the description of stress and strain in cubic crystals). In common semiconductors, $C_{11}, C_{12} > 0$ with $C_{11} > C_{12}$, and both are usually described in units of 10^{11}dyn/cm^2.

Some relevant examples using Eq. (A11.1) are shown in Fig. A11.2. In the first two examples shown on the left, stress is applied to all four x and y faces of the cube such that $\sigma_1 = \sigma_2$, and no stress is applied to the z faces, such that $\sigma_3 = 0$. For both of these cases, the crystal is under *biaxial* strain. When the stress is directed outward ($\sigma_1 > 0$), the resulting strain is referred to as biaxial *tensile* strain, while inward stress ($\sigma_1 < 0$) gives rise to biaxial *compressive* strain. The resulting lattice deformation can be calculated using Eq. (A11.1). By symmetry, with $\sigma_1 = \sigma_2$, the strain in both x and y directions must be equal and we can set $\varepsilon_1 = \varepsilon_2$. With $\sigma_3 = 0$ and $\varepsilon_1 = \varepsilon_2$, the first and third equations in Eq. (A11.1) reduce to

$$\begin{aligned} \sigma_1 &= C_{11}\varepsilon_1 + C_{12}\varepsilon_1 + C_{12}\varepsilon_3, \\ 0 &= C_{12}\varepsilon_1 + C_{12}\varepsilon_1 + C_{11}\varepsilon_3. \end{aligned} \tag{A11.2}$$

These equations are valid for both tensile and compressive biaxial strain. The relationship between the strain components is immediately obtained from the lower equation:

$$\varepsilon_3 = -\frac{2C_{12}}{C_{11}}\varepsilon_1. \tag{A11.3}$$

Since $C_{11}, C_{12} > 0$, we conclude that under biaxial strain, the lattice deformation along z will be opposite to the deformation along either x or y, as depicted in

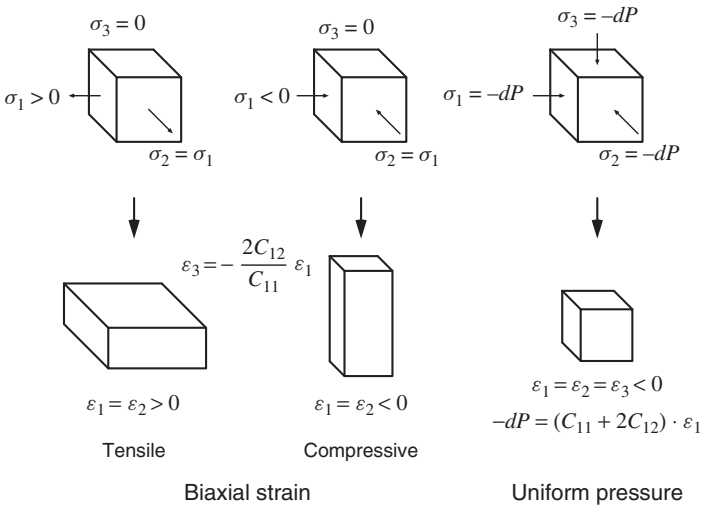


FIGURE A11.2: Examples of the relationship between stress and strain.

Fig. A11.2. This effect is similar to squeezing a balloon—as the sides are compressed, the top and bottom expand as the balloon attempts to maintain the same volume of air. Using Eq. (A11.3) in the first equation of Eq. (A11.2), we then find

$$\sigma_1 = C_{11}\varepsilon_1[1 + C_{12}/C_{11} - 2(C_{12}/C_{11})^2]. \tag{A11.4}$$

This equation gives us the absolute measure of the stress required to achieve a given strain. Note that if $C_{12} = 0$, the strain perpendicular to the stress plane (i.e., along z) reduces to zero, and Hooke's law reduces to the scalar relation: $\sigma_1 = C_{11}\varepsilon_1$.

The example shown on the right side of Fig. A11.2 corresponds to a uniform stress applied equally to all sides of the crystal. If we define this inward-directed stress as a differential pressure change, dP, surrounding the crystal, then we can set $\sigma_1 = \sigma_2 = \sigma_3 = -dP$. Adding up all three equations in Eq. (A11.1), we immediately find

$$-3\,dP = (C_{11} + 2C_{12})(\varepsilon_1 + \varepsilon_2 + \varepsilon_3). \tag{A11.5}$$

Neglecting cross-terms between the ε_i, the fractional change in the *volume* of the crystal lattice is given by

$$dV/V \approx \varepsilon_1 + \varepsilon_2 + \varepsilon_3. \tag{A11.6}$$

Therefore, Eq. (A11.5) allows us to determine the change in the crystal volume in response to a change in the uniform pressure surrounding the crystal. This relationship will be used later to relate the experimentally measured pressure dependence of the bandgap to the strain of the lattice.

A11.2 RELATIONSHIP BETWEEN STRAIN AND BANDGAP

With an understanding of how a particular stress leads to a particular strain, we are left with the task of determining how a particular strain affects the bandgap of the semiconductor, the main topic of this appendix. Pikus and Bir [2] in 1959 provided the fundamental theory necessary to describe how lattice deformations affect the Hamiltonian, which describes the interaction between the three valence bands: the heavy-hole (HH), light-hole (LH), and split-off (SO) bands. Their analysis involved transforming the coordinate system describing the crystal potential and the hole wavefunctions into a new deformed coordinate system, making use of the strain tensor. To first order in the strain, it was shown that a Hamiltonian identical in form to the Luttinger–Kohn (LK) valence band Hamiltonian described in Appendix 8 must be added to account for the strain. This new Pikus and Bir or *strain* Hamiltonian can be found directly from the LK Hamiltonian using the

following substitutions:

$$\begin{aligned} k_i k_j &\rightarrow \varepsilon_{ij}, \\ \gamma_1 &\rightarrow a, \\ -2\gamma_2 &\rightarrow b, \\ -2\sqrt{3}\gamma_3 &\rightarrow d. \end{aligned} \tag{A11.7}$$

In these substitutions, the k_i are the various wavevectors appearing in the LK Hamiltonian and the γ_i are the Luttinger parameters, both of which are discussed in Appendix 8. The ε_{ij} are the components of the strain tensor describing the lattice deformation and *a, b*, and *d* are *lattice deformation potentials*, which relate shifts in the valence bands to the strain tensor. An additional strain-dependent spin–orbit interaction Hamiltonian [3, 4] also exists; however, the deformation potentials associated with this interaction are generally more than an order of magnitude smaller [3] than those associated with the Pikus and Bir Hamiltonian and will therefore be neglected here.

In this appendix, only normal components of the stress tensor are considered, and we can set $\varepsilon_{ij} = 0$ for $i \neq j$. Furthermore, if we concentrate on biaxial strain then we can also set $\varepsilon_1 = \varepsilon_2$. Under these conditions, the 6×6 strain Hamiltonian [4, 5] relating the three twofold spin-degenerate valence bands is block-diagonalized into two identical 3×3 Hamiltonians (the interaction between states of opposite spin is removed). We can write the 3×3 strain Hamiltonian as

$$\begin{array}{c} \begin{array}{ccc} HH & LH & SO \end{array} \\ \begin{bmatrix} H-S & 0 & 0 \\ 0 & H+S & \sqrt{2}S \\ 0 & \sqrt{2}S & H+\Delta \end{bmatrix} \begin{array}{c} HH \\ LH \\ SO \end{array} \end{array} \tag{A11.8}$$

where

$$H = a(\varepsilon_1 + \varepsilon_2 + \varepsilon_3), \tag{A11.9}$$

$$S = b(\tfrac{1}{2}(\varepsilon_1 + \varepsilon_2) - \varepsilon_3), \tag{A11.10}$$

and Δ is the spin–orbit energy, which separates the SO band from the HH and LH bands. The energies, H and S, are related to the strain tensor through the a and b deformation potentials (the deformation potential, d, appears only when shear components of stress are considered).

Equation (A11.9) reveals that H is proportional to the change in volume of the crystal lattice created by the strain (compare with Eq. (A11.6)). It is referred to as the *hydrostatic* component of the strain. Because H appears with the same sign in every diagonal term of the Hamiltonian Eq. (A11.8), we conclude that changes

in the crystal volume result in a rigid shift of all three valence bands either up or down in energy, depending on whether the volume change is positive or negative.

Equation (A11.10) defines the *shear* component of the strain, S. The shear strain energy is proportional to the asymmetry in the strain parallel and perpendicular to the stress plane (the shear *strain* should not be confused with shear *stress* which is zero in this case). The dependence of the Hamiltonian on S is more complex than its dependence on H. The HH band is isolated from the other two bands; however, the LH and SO bands are coupled through the $\sqrt{2}S$ term. If $S \ll \Delta$, the LH–SO band coupling can be neglected, and the dominant contribution from S shows up in the HH and LH diagonal terms, where it is seen to split the HH and LH bands in opposite directions.

The strain Hamiltonian is written in terms of the hole energy, such that positive energy moves further into the valence band. For biaxial compressive strain, H and S are both positive, indicating an increase in the bandgap. For biaxial tensile strain, H and S are both negative. Figure A11.3 illustrates the band edge shifts of the three valence bands for both types of biaxial strain. The conduction band experiences a shift due to the hydrostatic component of strain only. In fact, we can define a separate H' with a corresponding deformation potential a' associated with the movement of the conduction band. However, it can be difficult to experimentally separate H' from the total bandgap shift [6]. A common approach is to interpret H in the strain Hamiltonian (and deformation potential a) as the total energy shift of the bandgap resulting from hydrostatic strain, and not worry about how the total shift is divided up between the conduction and valence bands. In fact, to estimate strained bandgaps, the division of H into H' and $H - H'$ is irrelevant. For heterointerface applications (including strained QWs), the division of H into H' and $H - H'$ can indicate how the conduction band offset, Q_c, is affected by strain.

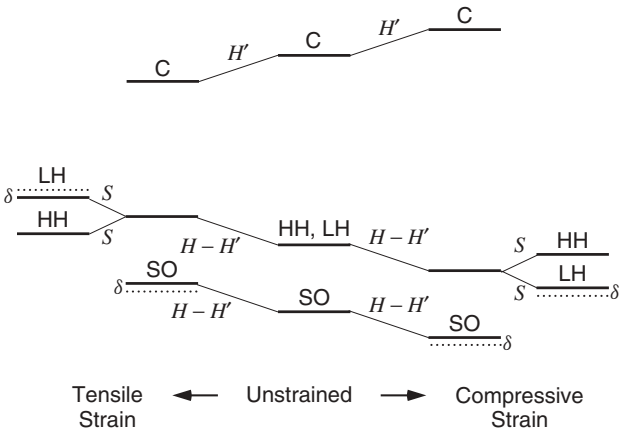


FIGURE A11.3: Qualitative band energy shifts of the conduction band and three valence bands for biaxial compressive and tensile strain. The magnitude of the energy shift is indicated next to each shift. The gray energies include LH–SO band coupling.

However, in practice Q_c is measured experimentally by other means, implying that again we do not need H'. As a result, the strained bandgap and its lineup with other material bandgaps is typically defined using H and measured values of Q_c.

To quantify the energy shifts depicted in Fig. A11.3, we need to evaluate H and S. The strain components within the stress plane, ε_1 and ε_2, are easily identified as the fractional change in the in-plane lattice constant of the strained material. The strain perpendicular to the stress plane, ε_3, was discussed earlier in reference to Fig. A11.2, and can be related to ε_1 and ε_2 using Eq. (A11.3). If we assume the in-plane lattice constant, $a_\parallel$, of the strained material is strained to match the substrate lattice constant such that $a_\parallel = a_{sub}$, we can define the strain as

$$\varepsilon \equiv \frac{a_{native} - a_{sub}}{a_{native}}, \tag{A11.11}$$

where a_{native} is the native unstrained lattice constant of the strained material. The various strain components are related to this definition as follows:

$$\varepsilon_1 = \varepsilon_2 = -\varepsilon \quad \text{and} \quad \varepsilon_3 = \frac{2C_{12}}{C_{11}}\varepsilon. \tag{A11.12}$$

The *lattice-mismatch parameter*, ε, is defined negative to the in-plane strain components by convention such that $\varepsilon > 0$ for compressive strain, and $\varepsilon < 0$ for tensile strain. Plugging these into the definitions of H and S, we find

$$H = (-a) \cdot 2\frac{C_{11} - C_{12}}{C_{11}}\varepsilon, \tag{A11.13}$$

$$S = (-b) \cdot \frac{C_{11} + 2C_{12}}{C_{11}}\varepsilon. \tag{A11.14}$$

The parentheses around the deformation potentials are used because in common semiconductors, $a, b < 0$. The parentheses therefore enclose positive numbers, and the sign of ε determines the sign of both H and S.

The change in the HH, LH, and SO bandgaps for $S \ll \Delta$ can be approximated by the diagonal terms of Eq. (A11.8). However, for larger strains, the matrix must be diagonalized. Since the HH band is isolated, the problem is reduced to finding the eigenvalues of the 2×2 submatrix which couples the LH and SO bands. Performing this procedure, the bandgap shifts become

$$\begin{aligned} \Delta E_{HH} &= H - S, \\ \Delta E_{LH} &= H + S - \delta, \\ \Delta E_{SO} &= H + \Delta + \delta, \\ \delta &= \tfrac{1}{2}\Delta\{[1 - 2(S/\Delta) + 9(S/\Delta)^2]^{1/2} - (1 - S/\Delta)\} \approx 2S^2/\Delta. \end{aligned} \tag{A11.15}$$

The energy, δ, represents an additional repulsion between the LH and SO bands, which increases approximately quadratically with S and hence is positive for both

tensile and compressive strain. This LH–SO band-coupling energy shift is indicated by the gray energy levels in Fig. A11.3. The band edge energy shifts in Eq. (A11.15) are all relative to the unstrained direct bandgap of the material, E_{g0}. Once E_{g0} is added to the energy shifts, the strained bandgaps are completely defined. The material parameters that we must know to predict the strained bandgaps (aside from the unstrained bandgap and lattice-mismatch parameter) are a, b, C_{11}, C_{12}, and Δ.

Values for all of the required strain parameters can be found in the literature. However, the hydrostatic deformation potential, a, is a bit more difficult to track down. The division of a between the valence and conduction band is not well standardized. As a result, a sometimes refers to the conduction band shift and sometimes the valence band shift, or sometimes the sum of these two (which is what we are interested in). Unfortunately, it is often difficult to know which definition is being used in any given citation. For this reason it is common to estimate a based on measurements of the pressure dependence of the bandgap. The strain example given earlier in reference to Fig. A11.2 showed that a uniform pressure is related to the volume change as follows:

$$-3\,dP = (C_{11} + 2C_{12})\,(\varepsilon_1 + \varepsilon_2 + \varepsilon_3). \tag{A11.16}$$

Furthermore, for a uniform pressure we can set $S = 0$, implying that the shift in the HH and LH bandgaps is caused solely by H. Thus, the differential change in bandgap in response to a differential change in pressure is simply

$$dE = a(\varepsilon_1 + \varepsilon_2 + \varepsilon_3). \tag{A11.17}$$

Dividing Eq. (A11.17) by Eq. (A11.16), we find

$$a = -\tfrac{1}{3}(C_{11} + 2C_{12})\frac{dE}{dP}. \tag{A11.18}$$

Because $dE/dP > 0$, we conclude that $a < 0$, as stated earlier. This formula allows us to determine a for any bandgap once the pressure dependence of the gap is known. The pressure dependence of the energy gap is commonly quoted in units of 10^{-6} eV/bar or sometimes 10^{-6} eVcm^2/kg (1 bar = 1.019 kg/cm^2). In defining a this way, a useful conversion factor to keep handy is

$$10^{-6}\ \text{eV/bar} \times 10^{11}\ \text{dyn/cm}^2 = 0.1\ \text{eV}. \tag{A11.19}$$

Strain-related parameters of III–V semiconductor materials are summarized in Table A11.1. The strain parameters related to shear stress, d and C_{44}, are included for completeness but are not required in the calculation of biaxially strained bandgaps. The hydrostatic deformation potential, a, was calculated from other parameters in the table using Eq. (A11.18); the values of *dE/dP* in the table correspond to the direct bandgap change with pressure. The first five

TABLE A11.1: Strain Parameters in III–V Semiconductors

	Lattice Constant	Deformation Potentials (eV)			Elastic Moduli (10^{11}dyn/cm^2)			(10^{-6} eV/bar)	
Material	a(Å)	a	b	d	C_{11}	C_{12}	C_{44}	dE/dP	Δ(eV)
GaAs	5.6533	−8.68	−1.7	−4.55	11.88	5.38	5.94	11.5	0.34
InAs	6.0583	−5.79	−1.8	−3.6	8.329	4.526	3.959	10.0	0.371
AlAs*	5.6611	−7.96	−1.5	−3.4	12.02	5.70	5.89	10.2	0.30
GaP*	5.4512	−9.76	−1.5	−4.6	14.12	6.253	7.047	11.0	0.10
InP	5.8688	−6.16	−2.0	−5.0	10.22	5.76	4.60	8.5	0.10
AlP*	5.4635	−8.38	−1.75	−4.8	13.2	6.3	6.15	9.75	0.10
GaSb	6.0959	−8.28	−1.8	−4.6	8.842	4.026	4.322	14.7	0.8
InSb	6.4794	−7.57	−2.0	−4.8	6.47	3.65	3.02	16.5	0.98
AlSb*	6.1355	2.04	−1.35	−4.3	8.769	4.341	4.076	−3.5	0.75

*Indirect gap.

entries were taken primarily from Adachi [7, 8] while the rest were taken from Landolt–Bornstein [9]. Complete data for AlP could not be found. Thus, blank entries in this row were filled with the average value of the other two phosphides.

If we wanted to know, for example, the strained bandgap of InGaAs grown on GaAs as a function of indium mole fraction, we could determine the strain parameters for the InGaAs ternary by linear interpolation between the GaAs and InAs values listed in Table A11.1. For most ternaries this procedure works well. For quaternaries, Vegard's law [7] can be applied to provide reasonable estimates.

A11.3 RELATIONSHIP BETWEEN STRAIN AND BAND STRUCTURE

The energy shifts discussed in this appendix refer to the band edge shifts exclusively. To predict the band structure away from the band edge, the strain Hamiltonian must be added to the LK Hamiltonian, and the combination must be diagonalized. In Appendix 8, this combined Hamiltonian uses the band edge of the HH band as the zero-energy reference. Thus, $H - S$ should be subtracted from all diagonal terms of the strain Hamiltonian. The net result once the strain Hamiltonian is diagonalized is that the difference between the HH and LH band edges should be added to the LH diagonal term in the LK Hamiltonian. In Appendix 8, this total splitting energy between the HH and LH bands was defined as S_{tot}. Using Eq. (A11.15), we can write

$$S_{tot} \equiv E_{LH} - E_{HH} = 2S - \delta \approx 2S(1 - S/\Delta), \tag{A11.20}$$

where the latter approximation expands δ to second order in S.

In Appendix 8, the LK Hamiltonian used for the subband structure calculations ignores any coupling to the SO band. Chao and Chuang [5] have analyzed the consequences of this approximation and found that significant differences do exist between the subband structure with and without SO band coupling. However, their calculations without SO band coupling assume $S_{tot} = 2S$ (i.e., they ignore the SO band coupling entirely). By assuming $S_{tot} = 2S - \delta$, the SO band coupling can at least be partially included since we obtain more accurate estimates of the LH subband energies at the zone center (i.e., at the subband edges). This procedure should improve the accuracy of strained subband structure calculations which ignore SO band coupling in the LK Hamiltonian.

REFERENCES

1. Lovett DR. *Tensor properties of crystals*. New York: Adam Hilger; 1989. Chapter 4.
2. Pikus GE, Bir GL. *Sov Phys-Solid State* 1960;**1**: 1502.
3. Laude LD, Pollak FH, Cardona M. Effects of uniaxial stress on the indirect exciton spectrum of silicon. *Phys Rev B* 1971;**3**: 2623.
4. Bahder TB. Eight-band k·p model of strained zinc-blende crystals. *Phys Rev B* 1990;**46**: 4110.
5. Chao CYP, Chuang SL. Spin-orbit-coupling effects on the valence-band structure of strained semiconductor quantum wells. *Phys Rev B* 1992;**46**: 4110.
6. Nolte DD, Walukiewicz W, Haller EE. Band-edge hydrostatic deformation potentials in III-V semiconductors. *Phys Rev Lett* 1987;**59**: 501.
7. Adachi S. Material parameters of $In_{1-x}Ga_xAs_yP_{1-y}$ and related binaries. *J Appl Phys* 1982;**53**: 8775.
8. Adachi S. GaAs, AlAs, and $Al_xGa_{1-x}As$: Material parameters for use in research and device applications. *J Appl Phys* 1985;**58**: R1.
9. Landolt-Börnstein. In: Madelung O, editor. *Numerical data and functional relationships in science and technology*. Berlin: Springer-Verlag; 1982.

APPENDIX TWELVE

Threshold Energy for Auger Processes

A12.1 CCCH PROCESS

For the CCCH process in Fig. 4.21, we can write the momentum and energy conservation laws (i.e., initial = final) as

$$\mathbf{k}_1 + \mathbf{k}_2 = \mathbf{k}_3 + \mathbf{k}_4, \tag{A12.1a}$$

$$\Delta E_1 + \Delta E_2 = -(E_g + \Delta E_3) + \Delta E_4, \tag{A12.1b}$$

where the **k**'s are vectors in k-space, and E_c has been used as the energy reference level in the latter equation. Rearranging the energy conservation law, we find

$$\Delta E_4 - E_g = \Delta E_1 + \Delta E_2 + \Delta E_3. \tag{A12.2}$$

From Eq. (4.107b), it follows that the most probable transition corresponds to the minimum possible energy of state 4. This minimum value for ΔE_4 is referred to as the *threshold* energy, E_T, of the Auger process.

To determine the threshold energy, we need to minimize Eq. (A12.2). If we assume that all bands are parabolic, we can set $\Delta E_i \propto \mathbf{k}_i \cdot \mathbf{k}_i / m_C$ for the three conduction band states and $\Delta E_3 \propto \mathbf{k}_3 \cdot \mathbf{k}_3 / m_H = \mu \mathbf{k}_3 \cdot \mathbf{k}_3 / m_C$, for the valence band where $\mu = m_C / m_H$. At the minimum, ΔE_4 will be independent of variations

Diode Lasers and Photonic Integrated Circuits, Second Edition.
Larry A. Coldren, Scott W. Corzine, and Milan L. Mašanović.
© 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

in any of the k-vectors. The differential of any one term is $d(\mathbf{k}_i \cdot \mathbf{k}_i) = \mathbf{k}_i \cdot d\mathbf{k}_i + d\mathbf{k}_i \cdot \mathbf{k}_i$. Taking the total differential of Eq. (A12.2) and setting it to zero, we obtain

$$\mathbf{k}_4 \cdot d\mathbf{k}_4 = \mathbf{k}_1 \cdot d\mathbf{k}_1 + \mathbf{k}_2 \cdot d\mathbf{k}_2 + \mu\mathbf{k}_3 \cdot d\mathbf{k}_3 = 0, \tag{A12.3}$$

$$d\mathbf{k}_3 = d\mathbf{k}_1 + d\mathbf{k}_2 - d\mathbf{k}_4. \tag{A12.4}$$

The second equation follows from taking the differential of the constraint Eq. (A12.1a). Replacing $d\mathbf{k}_3$ in Eq. (A12.3) using Eq. (A12.4), we find

$$(\mathbf{k}_1 + \mu\mathbf{k}_3) \cdot d\mathbf{k}_1 + (\mathbf{k}_2 + \mu\mathbf{k}_3) \cdot d\mathbf{k}_2 - \mu\mathbf{k}_3 \cdot d\mathbf{k}_4 = 0. \tag{A12.5}$$

The last term is zero when $\mathbf{k}_3 \| \mathbf{k}_4$, since $\mathbf{k}_4 \cdot d\mathbf{k}_4 = 0$ from Eq. (A12.3). The remaining two terms are zero for any $d\mathbf{k}_1$ and $d\mathbf{k}_2$ when $\mathbf{k}_1 = \mathbf{k}_2 = -\mu\mathbf{k}_3$. Substituting this into Eq. (A12.1a), we can determine $\mathbf{k}_4$. To summarize, we have

$$\mathbf{k}_1 = \mathbf{k}_2 = -\mu\mathbf{k}_3, \tag{A12.6}$$

$$\mathbf{k}_4 = -(1 + 2\mu)\mathbf{k}_3, \tag{A12.7}$$

where $\mu = m_C/m_H$. Thus, the most probable Auger transition occurs when all four k-vectors are colinear, with $\mathbf{k}_3$ pointing in the opposite direction to the rest. The Auger transitions in Fig. 4.20 reflect this conclusion. Using Eqs. (A12.6) and (A12.7) in Eq. (A12.1b) and solving for k_3^2, we obtain

$$k_3^2 = \frac{1}{(1 + 2\mu)(1 + \mu)} k_g^2, \tag{A12.8}$$

where k_g corresponds to $E_g = \hbar^2 k_g^2/2m_C$. The energies of the four states involved are then easily found using Eq. (A12.8):

$$\Delta E_1 = \Delta E_2 = \mu \Delta E_3 = \frac{\mu^2}{(1 + 2\mu)(1 + \mu)} E_g, \tag{A12.9}$$

$$\Delta E_4 = \frac{1 + 2\mu}{1 + \mu} E_g. \tag{A12.10}$$

The threshold energy for the CCCH process is therefore given by

$$E_T \equiv \Delta E_4 = \frac{1 + 2\mu}{1 + \mu} E_g = \frac{2m_C + m_H}{m_C + m_H} E_g. \tag{A12.11}$$

The most probable CCCH Auger transition is now completely defined.

A12.2 CHHS AND CHHL PROCESSES

For the CHHS process in Fig. 4.20, we again start by writing the momentum and energy conservation laws:

$$\mathbf{k}_1 + \mathbf{k}_2 = \mathbf{k}_3 + \mathbf{k}_4, \tag{A12.12a}$$

$$-(\Delta E_1 + \Delta E_2) = E_g + \Delta E_3 - (\Delta_{so} + \Delta E_4). \tag{A12.12b}$$

In this case, E_v has been used as the energy reference level in the latter equation. Rearranging the energy conservation law, we find

$$\Delta E_4 - (E_g - \Delta_{so}) = \Delta E_1 + \Delta E_2 + \Delta E_3. \tag{A12.13}$$

Again, we wish to minimize ΔE_4 to maximize the transition probability. Using the same procedure as outlined earlier, the minimum k-vectors are found to be

$$\mathbf{k}_1 = \mathbf{k}_2 = -\mathbf{k}_3/\mu_H, \tag{A12.14}$$

$$\mathbf{k}_4 = -(1 + 2/\mu_H)\mathbf{k}_3, \tag{A12.15}$$

where $\mu_H = m_C/m_H$. Again, all four k-vectors are colinear, with $\mathbf{k}_3$ pointing in the opposite direction to the rest. The energies of the four states involved are found to be

$$\begin{aligned} \mu_S \Delta E_1 = \mu_S \Delta E_2 &= \Delta E_3 \\ &= \frac{1/\mu_S}{(1 + 2/\mu_H)(1 + 2/\mu_H - 1/\mu_S)}(E_g - \Delta_{so}), \end{aligned} \tag{A12.16}$$

$$\Delta E_4 = \frac{(1 + 2/\mu_H)}{(1 + 2/\mu_H - 1/\mu_S)}(E_g - \Delta_{so}), \tag{A12.17}$$

where $\mu_S = m_C/m_S$. The threshold energy for the CHHS process is therefore given by

$$E_T \equiv \Delta E_4 = \frac{2m_H + m_C}{2m_H + m_C - m_S}(E_g - \Delta_{so}). \tag{A12.18}$$

The equations for the CHHL process are found by setting $\Delta_{so} \to 0$ and $m_S \to m_L$.

APPENDIX THIRTEEN

Langevin Noise

This appendix gives a more detailed account of the Langevin noise sources used in Chapter 5. It is divided into three main sections. The first section considers basic properties of Langevin noise sources and covers the general definition and evaluation of the correlation strength between two Langevin noise sources. The second section considers the specific Langevin noise correlations between: (1) the photon density and carrier density Langevin noise sources, (2) the photon density and output power Langevin noise sources, and (3) the photon density and phase Langevin noise sources. The final section makes use of the specific correlation strengths to evaluate the noise spectral densities of the photon density, output power, and carrier density, using the formulas developed in Chapter 5. Practical approximations to these lengthy expressions are also discussed.

A13.1 PROPERTIES OF LANGEVIN NOISE SOURCES

A13.1.1 Correlation Functions and Spectral Densities

One of the defining characteristics of a Langevin noise source, $F(t)$, is its completely random nature. In fact, the best analogy to $F(t)$ is a random number generator that generates a new number between $\pm\infty$ every Δt seconds in the limit of $\Delta t \to 0$. In mathematical terms, this characteristic is described as a *memoryless* process, which means that the value of $F(t)$ at time t has absolutely no correlation with any previous value $F(t-\tau)$ including $\tau \to 0^+$. Now because $F(t)$ is just as often positive as it is negative, the average value over time is zero:

Diode Lasers and Photonic Integrated Circuits, Second Edition.
Larry A. Coldren, Scott W. Corzine, and Milan L. Mašanović.
© 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

$\langle F(t)\rangle = 0$.[1] Furthermore, since $F(t)$ and $F(t-\tau)$ fluctuate randomly relative to each other, they have the same sign just as often as they have opposite signs over time. As a result, the average of the *product* of the two over time t is also zero: $\langle F(t)F(t-\tau)\rangle = 0$. The only exception is when $\tau = 0$, in which case the product is always positive: $\langle F(t)F(t)\rangle = \langle F(t)^2\rangle = \infty$ (the magnitude is infinite because $F(t)^2$ can take on any value between 0 and ∞, which when averaged is infinite). As a function of τ then, the *correlation function* $\langle F(t)F(t-\tau)\rangle$ displays a delta function-like behavior. Generalizing this result, we can define the correlation function between any two (memoryless) Langevin noise sources as

$$\langle F_i(t)F_j(t-\tau)^*\rangle = S_{ij}\cdot\delta(\tau). \tag{A13.1}$$

The proportionality constant S_{ij} defines the *correlation strength* between the two noise sources (it has units of (seconds) $\times$ (fluctuating variable units)^2). When $i = j, S_{ij}$ defines the *auto*correlation strength. When $i \neq j, S_{ij}$ defines the *cross*-correlation strength, which is nonzero only if the fluctuations of one noise source are in some way correlated with the fluctuations of the other noise source. The complex conjugate is included in the definition of the correlation function to account for possibly complex Langevin noise sources (such that at $\tau = 0$, the autocorrelation function reduces to $\langle |F_i(t)|^2\rangle$).

It is interesting to point out that according to Eq. (A13.1), a Langevin noise source has an infinite mean-square value: $\langle |F_i(t)|^2\rangle = \infty$. However, this would only be observable by a detection system that had infinite bandwidth. In practice, the fluctuations observed on an oscilloscope are never infinite because they are limited by the system rise time or measurement bandwidth. The actual *measured* mean-square noise is found from the overlap of the measurement bandwidth with the *spectral density* of the noise source as worked out in Eqs. (5.112)–(5.114) in Section 5.5.

We can relate the spectral density to the correlation function by examining the *frequency domain* correlation function as follows:

$$\begin{aligned}
\langle F_i(\omega)F_j(\omega')^*\rangle &= \left\langle \int F_i(\tau)e^{-j\omega\tau}\,d\tau \cdot \int F_j(t)^* e^{j\omega' t}\,dt \right\rangle \\
&= \iint \langle F_i(t+\tau)F_j(t)^*\rangle e^{-j\omega\tau} e^{-j(\omega-\omega')t}\,dt\,d\tau \\
&= \int \langle F_i(t+\tau)F_j(t)^*\rangle e^{-j\omega\tau}\,d\tau \cdot \int e^{-j(\omega-\omega')t}\,dt \\
&= \int \langle F_i(t)F_j(t-\tau)^*\rangle e^{-j\omega\tau}\,d\tau \cdot 2\pi\delta(\omega-\omega') \\
&\equiv S_{ij}(\omega)\cdot 2\pi\delta(\omega-\omega').
\end{aligned} \tag{A13.2}$$

[1]The brackets $\langle\ \rangle$ actually refer to a statistical average over many similarly prepared systems at the same time t. However, we can obtain the same result by averaging a single system over extended time intervals if the statistical processes involved are both *stationary* and *ergodic*, which we assume to be the

The first step in this derivation uses Eq. (5.111) to transform to the time domain. The second step sets $\tau \to t + \tau$ (with $d\tau = d\tau$) and regroups, confining the statistical average (see footnote 1) to the statistically varying processes. The third and fourth steps assume that $\langle F(t+\tau)F(t)^* \rangle$ only depends on the *relative* time delay τ between the two functions and not on the absolute time t. This characteristic is described as a *stationary* process and is not limited to Langevin noise sources but applies to all statistical processes considered in this book. For stationary processes then, we can separate out the integration over t in the third step and shift the time origin to $t \to t - \tau$ in the fourth step. The fourth step also recognizes the integration over t to be a delta function of strength 2π.

The final step in Eq. (A13.2) defines the correlation strength of the frequency domain delta function as the spectral density $S_{ij}(\omega)$ (compare with Eq. (5.113)), where

$$S_{ij}(\omega) = \int \langle F_i(t) F_j(t-\tau)^* \rangle e^{-j\omega\tau}\, d\tau. \tag{A13.3}$$

In words, the spectral density is just the Fourier transform of the correlation function. This fundamental relation is known as the *Wiener–Khinchin* theorem and it applies to all types of stationary processes (not just Langevin noise). For example, when discussing linewidth in Section 5.5.6, we stated that the power spectrum was equal to the Fourier transform of the electric field autocorrelation function. Equation (A13.2) is the proof of that statement. However, here we are interested in applying this result to Langevin noise.

Plugging the Langevin noise correlation Eq. (A13.1) into Eq. (A13.3), we find

$$S_{ij}(\omega) = S_{ij} \int \delta(\tau) e^{-j\omega\tau}\, d\tau = S_{ij}. \tag{A13.4}$$

Thus, the Langevin noise spectral density is simply equal to the correlation strength and is independent of frequency—it is a "white" noise source (this is in fact true for any memoryless, stationary process). Since the Langevin noise spectral density and correlation strength can be used interchangeably, we adopt a common notation for both:

$$\langle F_i F_j \rangle \equiv S_{ij}(\omega) = S_{ij}. \tag{A13.5}$$

In some cases it is more meaningful to interpret $\langle F_i F_j \rangle$ as the noise spectral density while in other cases the correlation aspect is more relevant. In this appendix, we will generally refer to $\langle F_i F_j \rangle$ as the correlation strength. Note that both the

case here. Thus, whether we define $\langle\ \rangle$ as a statistical average or time average is a matter of conceptual convenience. In the time domain, it is often convenient to think in terms of a time average, whereas in the frequency domain, a statistical average is usually more appropriate.

spectral density and the correlation strength have units of (seconds) × (fluctuating *timedomain* variable units)[2].

A13.1.2 Evaluation of Langevin Noise Correlation Strengths

The Langevin noise sources discussed here and in Chapter 5 are based on a shot noise model advanced by McCumber [1] and others such as Lax [2] as a method of simplifying the rigorous quantum description of noise in lasers. Within this model, the laser noise is assumed to originate from shot noise associated with the discrete random flow of particles into and out of the carrier and photon reservoirs. It can be shown that the spectral density of shot noise is constant and *proportional to the average rate of particle flow*. With this in mind, consider a reservoir where particles are flowing into and out of the reservoir by a host of discrete random processes. In the Langevin formalism, each of these discrete processes contributes shot noise to the overall noise in the reservoir. Therefore, to determine the total Langevin noise spectral density or correlation strength $\langle F_i F_i \rangle$, we simply sum over all shot noise contributions—or over all rates of particle flow into and out of reservoir i. To determine the cross-correlation strength $\langle F_i F_j \rangle$ between two reservoirs i and j, we sum only over particle flows that affect *both* reservoirs simultaneously. However in this case, when one reservoir gains a particle ($F_i > 0$), the other loses a particle ($F_j < 0$). Hence the product $\langle F_i F_j \rangle$ is always negative, and the noise between the two reservoirs is said to be *negatively correlated*.

Within the shot noise Langevin model then, the correlation strengths between the various Langevin noise sources are found by simple inspection of the rates into and out of the various reservoirs:

$$\langle F_i F_i \rangle = \sum R_i^+ + \sum R_i^-, \tag{A13.6}$$

$$\langle F_i F_j \rangle = -\left[\sum R_{ij} + \sum R_{ji} \right]. \tag{A13.7}$$

The reservoir Langevin noise sources, F_i and F_j, as well as the rates of particle flow into, R_i^+, out of, R_i^-, and between the two reservoirs, R_{ij} and R_{ji} are all in units of numbers per unit time. For single-sided spectral densities, an additional factor of 2 would be required on the RHS of these definitions.

As an example use of Eq. (A13.6), let's consider a simple and perhaps familiar example: shot noise in a detector. The shot noise associated with the current generated by discrete random absorption events is given by: $\langle i_N(t)^2 \rangle = 2qI\,\Delta f$, where I is the average current, and Δf is the bandwidth of the detection system. We can arrive at this same result by defining the current noise in terms of a Langevin noise source: $i(t) = I + F_I(t)$. After converting to pure numbers per unit time (i.e., I/q

[2]If the absorbed photons do not arrive in an entirely random fashion, it is possible to generate a noise current smaller than the standard shot noise level. Such a nonrandom, or sub-shot-noise-limited photon stream can be generated by semiconductor lasers (see Section 5.5.4 for details).

and F_I/q), we can use Eq. (A13.6) to determine the spectral density of the noise or the correlation strength $\langle F_I F_I \rangle$:

$$\frac{1}{q^2}\langle F_I F_I \rangle = \frac{I}{q} \rightarrow \langle F_I F_I \rangle = qI. \tag{A13.8}$$

Multiplying this double-sided spectral density by $2\Delta f$ (see Eq. (5.115)) gives the shot noise expression for $\langle i_N(t)^2 \rangle$. The main point of this exercise is to show the importance of first converting to numbers per unit time to get the correct proportionality constant (in this case, q).

A13.2 SPECIFIC LANGEVIN NOISE CORRELATIONS

A13.2.1 Photon Density and Carrier Density Langevin Noise Correlations

Figure 5.1 displays all rates into and out of both carrier and photon reservoirs. The associated density Langevin noise sources for each reservoir, F_N and F_P, when converted to units of numbers per unit time are VF_N and $V_p F_P$. Using Fig. 5.1 in combination with Eqs. (A13.6) and (A13.7), we immediately obtain

$$V_p^2 \langle F_P F_P \rangle = N_p V_p/\tau_p + (R_{21} + R_{12} + R'_{sp})V, \tag{A13.9}$$

$$V^2 \langle F_N F_N \rangle = \eta_i I/q + (R_{sp} + R_{nr} + R_{21} + R_{12})V, \tag{A13.10}$$

$$V_p V \langle F_P F_N \rangle = -(R_{21} + R_{12} + R'_{sp})V. \tag{A13.11}$$

The current term $\eta_i I/q$ in (A13.10) may be smaller or larger depending on the noise characteristics of the current pumping source [3].

We can simplify the correlation strengths using the following substitutions. From the steady-state relations, we can set $N_p V_p/\tau_p = (R_{21} - R_{12} + R'_{sp})V$ in Eq. (A13.9). From the fundamental relations Eqs. 4.32 and 4.48, we can also set $R_{12} = R_{21} - v_g g N_p$ and $R_{21} = R'_{sp} N_p V_p$. Finally with $R_{sp} - R'_{sp} + R_{nr} = \eta_i I_{th}/qV$ from Eq. (5.18), we obtain

$$\langle F_P F_P \rangle = 2\Gamma R'_{sp} N_p \left[1 + \frac{1}{N_p V_p}\right], \tag{A13.12}$$

$$\langle F_N F_N \rangle = \frac{2R'_{sp} N_p}{\Gamma}\left[1 + \frac{1}{2N_p V_p}\right] - \frac{v_g g N_p}{V} + \frac{\eta_i (I + I_{th})}{qV^2}, \tag{A13.13}$$

$$\langle F_P F_N \rangle = -2R'_{sp} N_p \left[1 + \frac{1}{2N_p V_p}\right] + \frac{v_g g N_p}{V_p}. \tag{A13.14}$$

For lasers putting out milliwatt power levels above threshold, the photon number in the cavity is typically $\sim 10^5$ or more. Thus it is generally safe to assume $N_p V_p \gg 1$. For this reason, Eqs. 5.136 through 5.138 are written without the terms in square brackets.

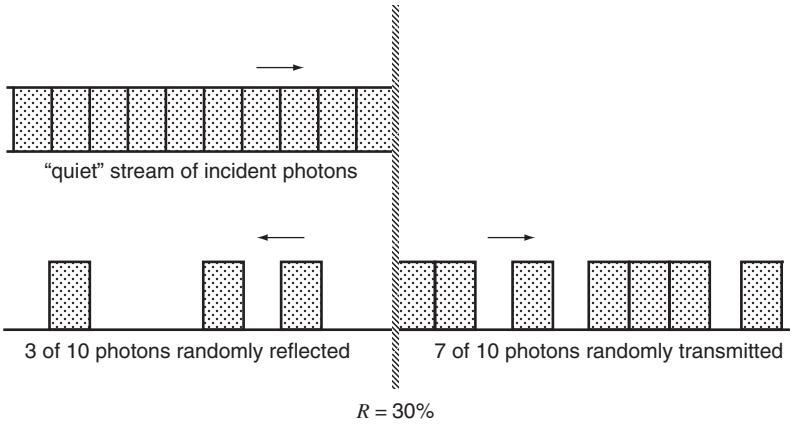


FIGURE A13.1: Illustration of partition noise created by a partially reflecting mirror. The key point is that a negative correlation exists between the partition noise reflected and transmitted at the mirror facet.

A13.2.2 Photon Density and Output Power Langevin Noise Correlations

In Chapters 2 and 5, we simply used $P_0 = (h\nu V_p v_g \alpha_m F)N_p$ or equivalently $P_0 = (\eta_0 h\nu V_p/\tau_p)N_p$ to convert from photon density to output power. However, when considering the *noise* inside and outside the cavity we must be careful. To understand why, consider a perfectly uniform stream of photons incident on a partially reflecting mirror as sketched in Fig. A13.1. While on average 30% of the photons are reflected, each individual photon must either be completely transmitted or completely reflected. This random division into reflected and transmitted photons leads to *partition* noise in the stream of both reflected and transmitted photons.

The partition noise reflected back into the cavity is accounted for in $\langle F_P F_P \rangle$ by the term $N_p V_p/\tau_p$ in Eq. (A13.9), which is equivalently the shot noise created by photons escaping the cavity. The partition noise transmitted outside the cavity however is different from that reflected back in. In fact, it is the exact inverse of the reflected noise, as Fig. A13.1 reveals. When added to the other noise contributions, the overall noise associated with the power outside the cavity is not the same as the overall noise associated with the photon density inside the cavity. This perhaps subtle point was first brought out by Yamamoto [4] using more fundamental quantum mechanical arguments.[3]

[3]Yamamoto describes the partition noise as vacuum-field fluctuations incident on the mirror facet from outside the cavity. In this case, the vacuum fields transmitted into the cavity have a negative phase relationship to those reflected off the mirror facet, and this is what provides the negative correlation between the facet noise inside and outside the cavity. Despite this difference in interpretation, Yamamoto's derivation produces the same result for output power noise considered in this appendix (however, Yamamoto's derivation requires a high mirror reflectivity resonator).

We can again use the Langevin method to determine the effect of this *negativity correlated* partition noise by treating the stream of output photons as another reservoir with its own associated Langevin noise source (similar to the detector current considered initially). Adapting Eq. (5.127), we have for the output power fluctuations:

$$\delta P(t) = (\eta_0 h\nu V_p/\tau_p)N_{p1}(t) + F_0(t). \tag{A13.15}$$

As we did in deriving Eqs. (A13.28) and (A13.26), we convert to the frequency domain, multiply both sides by $\delta P(\omega')^*$, take the time average, and integrate over ω' to obtain

$$S_{\delta P}(\omega) = (\eta_0 h\nu V_p/\tau_p)^2 S_{N_p}(\omega) + 2\ \mathrm{Re}\{(\eta_0 h\nu V_p/\tau_p)\langle N_{p1}F_0\rangle\} + \langle F_0F_0\rangle. \tag{A13.16}$$

The first term in this equation is what we would naively expect the relationship to be. However, the partition noise at the mirror facet creates two additional noise contributions which are important to consider, particularly at high output powers.

Using Eq. (5.130) for $N_{p1}(\omega)$, we can set

$$\langle N_{p1}F_0\rangle = \frac{H(\omega)}{\omega_R^2}[(\gamma_{NN} + j\omega)\langle F_PF_0\rangle + \gamma_{PN}\langle F_NF_0\rangle]. \tag{A13.17}$$

Thus, to evaluate (A13.16) we need to know the various correlations between the three noise sources, F_0, F_P, and F_N. First of all, there is no correlation between the carrier noise and the phenomenon depicted in Fig. A13.1. Therefore, we can immediately set $\langle F_NF_0\rangle = 0$. For the other two correlation strengths we can apply Eqs. (A13.6) and (A13.7) using $F_0/h\nu$ and V_pF_P to obtain

$$\langle F_0F_0\rangle = \eta_0 N_p V_p/\tau_p \cdot (h\nu)^2 = h\nu P_0, \tag{A13.18}$$

$$\langle F_PF_0\rangle = -\eta_0 N_p V_p/\tau_p \cdot (h\nu/V_p) = -P_0/V_p, \tag{A13.19}$$

$$\langle F_NF_0\rangle = 0. \tag{A13.20}$$

Using these in combination with Eqs. (A13.12) through (A13.14), we can evaluate $S_{N_p}(\omega)$ Eq. (5.125) and $\langle N_{p1}F_0\rangle$ Eq. (A13.17) from which we can ultimately obtain the output power spectral density function $S_{\delta P}(\omega)$ Eq. (A13.16). The complete expression is deferred to the next section, where we evaluate all three photon density, output power, and carrier density noise spectral density functions.

A13.2.3 Photon Density and Phase Langevin Noise Correlations

The phase Langevin noise source $F_\phi(t)$ and its related correlation strengths are found by studying the electric field in the laser cavity. To begin, we assume the noise on the electric field contains fluctuations in-phase and out-of-phase with

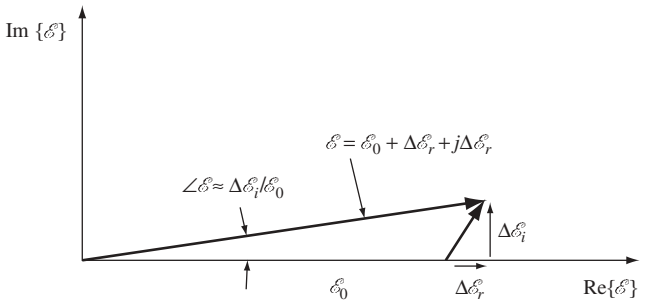


FIGURE A13.2: Vector illustration of the relationship between the instantaneous field magnitude and the quadrature (in-phase and out-of-phase) noise components.

the average field as shown in Fig. A13.2. Defining the in-phase and out-of-phase random fields as $\Delta\mathscr{E}_r(t)$ and $\Delta\mathscr{E}_i(t)$, the total instantaneous field and associated *complex* Langevin noise source become

$$\begin{aligned} \mathscr{E}(t) &= \mathscr{E}_0 + \Delta\mathscr{E}_r(t) + j\,\Delta\mathscr{E}_i(t), \\ \tilde{F}(t) &= F_r(t) + jF_i(t). \end{aligned} \tag{A13.21}$$

Here $F_r(t)$ models the in-phase field fluctuations, and $F_i(t)$ models the out-of-phase field fluctuations.

The power and phase fluctuations and their associated Langevin noise sources can be related to the in-phase and out-of-phase noise components as follows:

$$\begin{aligned} \mathscr{E}^*\mathscr{E} &\approx \mathscr{E}_0^2 + 2\mathscr{E}_0\Delta\mathscr{E}_r \qquad \text{and} \qquad \angle\mathscr{E} \approx \Delta\mathscr{E}_i/\mathscr{E}_0, \\ F_P(t) &\approx 2\sqrt{N_p}F_r(t) \qquad \text{and} \qquad F_\phi(t) \approx F_i(t)/\sqrt{N_p}. \end{aligned} \tag{A13.22}$$

The approximate equalities are valid for $\Delta\mathscr{E}_r, \Delta\mathscr{E}_i \ll \mathscr{E}_0$. For the latter relations, the fluctuating portion of $\mathscr{E}^*\mathscr{E}$ translates into photon density fluctuations, $F_P(t)$, while $\mathscr{E}_0^2$ translates into average photon density, N_p (and hence, $\mathscr{E}_0$ translates into $\sqrt{N_p}$). Now if we assume the spectral density of the in-phase and out-of-phase components of the field noise are equal in magnitude (i.e., $\langle F_iF_i\rangle = \langle F_rF_r\rangle$), we obtain

$$\langle F_rF_r\rangle = \frac{1}{4N_p}\langle F_PF_P\rangle = \frac{\Gamma R'_{sp}}{2}, \tag{A13.23}$$

$$\langle F_\phi F_\phi\rangle = \frac{1}{N_p}\langle F_iF_i\rangle = \frac{1}{N_p}\langle F_rF_r\rangle = \frac{\Gamma R'_{sp}}{2N_p}. \tag{A13.24}$$

The first equation uses Eq. (A13.12) to set $\langle F_PF_P\rangle = 2\Gamma R'_{sp}N_p$ assuming $N_pV_p \gg 1$.

To determine the cross-correlations, we note that $\langle F_\phi F_P\rangle = 2\langle F_iF_r\rangle$ from Eq. (A13.22). Assuming $\Delta\mathscr{E}_i(t)$ and $\Delta\mathscr{E}_r(t)$ are completely uncorrelated, we can

set $\langle F_i F_r \rangle = 0$. As for $\langle F_\phi F_N \rangle$, we note that phase fluctuations due to changes in carrier density are accounted for separately with the linewidth enhancement factor, α (see Eq. (5.147)). Thus, the inherent phase fluctuations that exist even when $\alpha = 0$ are unrelated to carrier noise. From these arguments, we conclude that

$$\langle F_\phi F_P \rangle = \langle F_\phi F_N \rangle = 0. \tag{A13.25}$$

Finally it should be noted that a difference appears between the phase fluctuations inside the cavity and outside the cavity as discussed by Yamamoto [4]. However, the difference in this case is relatively minor and will not be considered here (it is only important to consider at frequencies comparable to or greater than $1/\tau_p$).

A13.3 EVALUATION OF NOISE SPECTRAL DENSITIES

A13.3.1 Photon Noise Spectral Density

For reference, the spectral density of the photon density noise derived in Chapter 5 is given by

$$S_{N_p}(\omega) = \frac{|H(\omega)|^2}{\omega_R^4}[(\gamma_{NN}^2 + \omega^2)\langle F_P F_P \rangle + 2\gamma_{NN}\gamma_{PN}\langle F_P F_N \rangle + \gamma_{PN}^2 \langle F_N F_N \rangle]. \tag{A13.26}$$

Using the rate coefficients defined in Eq. (5.35) and the Langevin noise correlation strengths defined in Eqs. (A13.12) through (A13.14), the photon density noise becomes

$$S_{N_p}(\omega) = \frac{N_P}{V_p}\tau_p . \frac{a_1' + a_2'\,\omega^2}{\omega_R^4}|H(\omega)|^2, \tag{A13.27}$$

where

$$a_1' = \frac{a_2'}{\tau_{\Delta N}^2} + \omega_R^2 \frac{2}{\tau_{\Delta N}\tau_p} + \omega_R^4 \left[\frac{\eta_i (I + I_{th})}{I_{st}} + 1\right],$$

$$a_2' = \frac{8\pi (\Delta\nu)_{ST} N_p V_p}{\tau_p}\left[1 + \frac{1}{N_P V_p}\right],$$

and $(\Delta\nu)_{ST} = \Gamma R_{sp}'/4\pi N_p, I_{st} = qN_p V_p/\tau_p$. To simplify the expression, we have neglected the dependence of the single-mode spontaneous emission rate on N by setting $1/\tau_{\Delta N}' \to 0$. Also, in defining a_1' we have set $v_g a N_p/\tau_p = \omega_R^2$. Thus within a_1', the resonance frequency is *by definition*: $\omega_R^2 \equiv v_g a N_p/\tau_p$.

The three terms comprising a_1' form a power series in ω_R^2, or equivalently in N_p. At very low powers the first term dominates, while at moderate-to-high powers the last ω_R^4 term dominates. In fact at very high powers, the ω_R^4 term in a_1' dominates the entire numerator of Eq. (A13.26) and its coefficient in square brackets reduces to 2 (for a shot noise-limited current source). At very high powers, we can also set

$|H(\omega)|^2 \approx (1+\omega^2\tau_p^2)^{-1}$ if we neglect gain compression (with gain compression, τ_p is replaced by γ/ω_R^2). In this limit then, $S_{N_p}(\omega) \approx (2N_p/V_p)\cdot\tau_p/(1+\omega^2\tau_p^2)$. Using Eq. (5.114), we can estimate the mean-square photon number fluctuation. Integrating $S_{N_p}(\omega)V_p^2$ over all frequencies, we obtain $\langle|\delta N_p(t)V_p|^2\rangle = N_pV_p$. This result reveals that in the limit of high powers, the statistics of the photon number inside the cavity converges toward a Poisson distribution, indicative of a *coherent* state (in the language of quantum optics) [4].

A13.3.2 Output Power Noise Spectral Density

For reference, the spectral density of the output power noise derived earlier is given by

$$S_{\delta P}(\omega) = (\eta_0 h\nu V_p/\tau_p)^2 S_{N_p}(\omega) + 2\,\mathrm{Re}\{(\eta_0 h\nu V_p/\tau_p)\langle N_{p1}F_0\rangle\} + \langle F_0F_0\rangle. \tag{A13.28}$$

Using the expression for $\langle N_{p1}F_0\rangle$ given in Eq. (A13.17) and the Langevin noise correlation strengths defined in Eqs. (A13.18) through (A13.20), the output power noise becomes

$$S_{\delta P}(\omega) = h\nu P_0\cdot\left[\frac{a_1+a_2\,\omega^2}{\omega_R^4}|H(\omega)|^2+1\right], \tag{A13.27}$$

where

$$a_1 = \frac{8\pi(\Delta\nu)_{ST}P_0}{h\nu}\left[1+\frac{1}{N_pV_p}\right]\frac{1}{\tau_{\Delta N}^2} + \eta_0\omega_R^4\left[\frac{\eta_i(I+I_{th})}{I_{st}}-1\right]$$
$$- 2\eta_0\omega_\Delta^2\left[\omega_R^2+\frac{1}{\tau_{\Delta N}\tau_p}\right],$$
$$a_2 = \frac{8\pi(\Delta\nu)_{ST}P_0}{h\nu}\left[1+\frac{1}{N_pV_p}\right] - 2\eta_0\omega_R^2\left[\frac{\Gamma a_p}{a}+\frac{4\pi(\Delta\nu)_{ST}}{\omega_R^2\tau_p}\right],$$

and $(\Delta\nu)_{ST} = \Gamma R'_{sp}/4\pi N_p$, $I_{st} = qP_0/\eta_0 h\nu$. To simplify the expression, we have neglected the dependence of the single-mode spontaneous emission rate on N by setting $1/\tau'_{\Delta N}\to 0$. Also, in defining a_1 and a_2, we have set $v_g aN_p/\tau_p = \omega_R^2$. Thus within a_1 and a_2, the resonance frequency is *by definition*: $\omega_R^2 \equiv v_g aN_p/\tau_p$. The remainder of terms comprising the exact expression for the resonance frequency are sectioned off into the variable: $\omega_\Delta^2 \equiv \gamma_{NP}\gamma_{PN} + \gamma_{NN}\gamma_{PP} - v_g aN_p/\tau_p$.

Within a_1, the first term is independent of power and dominates at very low powers. The second term is $\propto P_0^2$ and dominates at moderate-to-high powers even though $\eta_i(I+I_{th})/I_{st}$ approaches one at high powers (for a shot noise-limited current source). The third ω_Δ^2 term can be neglected for all but very small power levels since the difference between $v_g aN_p/\tau_p$ and the actual ω_R^2 is usually negligible. Within a_2, the first term is independent of power, but nevertheless dominates

for all but very high powers. The second term is $\propto P_0$, so it eventually becomes comparable to the first term. However, gain compression limits ω_R^2 to a maximum value, such that the second term never actually grows larger than the first term. Within the square brackets of the second term, $\Gamma a_p/a$ dominates for all but very small powers since $4\pi(\Delta\nu)_{ST}/\omega_R^2\tau_p$ is $\propto 1/P_0^2$.

In comparing $S_{\delta P}(\omega)$ to $S_{N_p}(\omega)$, note that in addition to the slightly more complex frequency coefficients, a new "+1" factor appears in $S_{\delta P}(\omega)$. This factor ensures that the output power spectral density never drops below the shot noise limit of $h\nu P_0$ for all frequencies (unless the current source has sub-shot noise characteristics). Because the shot noise is included implicitly in $S_{\delta P}(\omega)$, it is in principle not necessary to explicitly add a shot noise term to the noise current in a photodetector [4]. However in practice, the partition noise created by the random loss of photons in getting from the laser to the detector contributes a shot noise-like term $\propto (1-\eta_{det})$ that should be added to the detector current noise (see Eq. (5.146a) and (5.146b) for details).

The expression for $S_{\delta P}(\omega)$ given in Eq. (5.139) neglects the ω_Δ^2 term of a_1, and the second term within square brackets of the ω_R^2 term in a_2. The first set of square brackets in both a_1 and a_2 are also set equal to one, assuming $N_pV_p \gg 1$ above threshold.

A13.3.3 Carrier Noise Spectral Density

For reference, the spectral density of the carrier density noise derived in Chapter 5 is given by

$$S_N(\omega) = \frac{|H(\omega)|^2}{\omega_R^4}[\gamma_{NP}^2\langle F_P F_P\rangle - 2\gamma_{PP}\gamma_{NP}\langle F_P F_N\rangle + (\gamma_{PP}^2+\omega^2)\langle F_N F_N\rangle]. \tag{A13.28}$$

Using the rate coefficients defined in Eq. (5.35) and the Langevin noise correlation functions defined in Eqs. (A13.12) through (A13.14), the carrier density noise becomes

$$S_N(\omega) = \frac{8\pi N_P^2}{\Gamma^2\tau_p^2\omega_R^4}(\Delta\nu)_{ST}\cdot(1+\delta)|H(\omega)|^2, \tag{A13.29}$$

where

$$\delta = -\frac{\varepsilon N_p/n_{sp}}{(1+\varepsilon N_p)^2}\left[1+\varepsilon N_p - \frac{\varepsilon N_p}{2}\left\{\frac{\eta_i(I+I_{th})}{I_{st}}+1\right\}\right] + \omega^2\tau_p^2\left[1+\frac{1}{2n_{sp}}\left\{\frac{\eta_i(I+I_{th})}{I_{st}}-1\right\}\right],$$

and $(\Delta\nu)_{ST} = \Gamma R'_{sp}/4\pi N_p, I_{st} = qN_pV_p/\tau_p$. To simplify the expression for this case, we have set the gain equal to the loss ($\Gamma v_g g = 1/\tau_p$) and assumed $N_pV_p \gg 1$. We have also written out a_p explicitly using Eq. (5.32) and set $R'_{sp}V\tau_p \to n_{sp}$. For

practical uses, it turns out that δ does not need to be considered. For example, at low powers, the entire first term reduces to $-\varepsilon N_p/n_{sp}$ which is negligible in comparison to one ($n_{sp} \sim 1$ above threshold). At very high powers, it reaches a maximum of $-1/4n_{sp}$ when $\varepsilon N_p = 1\,(\eta_i(I + I_{th})/I_{st} \to 1$ at high powers for a shot noise-limited current source). Thus, gain compression can reduce the low frequency carrier noise by as much as 25% at very high powers, however, we neglect this contribution in Chapter 5. The second term is small in comparison to one for $\omega \ll 1/\tau_p$, so unless we are interested in very high frequency carrier noise, we can neglect it as well.

The expression for $S_N(\omega)$ given in Eq. (A13.28) neglects δ in Eq. (A13.28) and sets $\omega_R^2 = v_g a N_p/\tau_p$.

REFERENCES

1. McCumber DE. Intensity fluctuations in the output of cw laser oscillators. I. *Phys Rev* 1966;**141**: 306. See also Harder C, Katz J, Margalit S, Shacham J, Yariv A. Noise equivalent circuit of a semiconductor laser diode. *IEEE J Quantum Electron* 1982;**18**: 333.
2. Lax M. Classical noise. V. noise in self-sustained oscillators. *Phys Rev* 1967;**160**: 290.
3. Yamamoto Y, Machida S. High-impedance suppression of pump fluctuation and amplitude squeezing in semiconductor lasers. *Phys Rev A* 1987;**35**: 5114.
4. Yamamoto Y, Imoto N. Internal and external field fluctuations of a laser oscillator: Part I—Quantum mechanical Langevin treatment. *IEEE J Quantum Electron* 1986;**22**: 2032.

PROBLEMS

1. Assume the output power of a laser with a given noise spectral density, S_i, is split by a beam splitter and fed into two photodetectors. Use the Langevin method to determine the detected noise spectral densities. Start by defining reservoirs for the incident and detected photon streams, with associated Langevin noise sources (F_i, F_{d1}, and F_{d2}). For the incident photon reservoir, assume an in-flow from the laser, I, and two out-flows, ηI and $(1 - \eta)I$. For the detector reservoirs assume only in-flows from the incident photon reservoir (a flowchart is useful here). Next, define all auto- and cross-correlation strengths between the three Langevin noise sources, considering the in-flow correlation term in the incident photon reservoir to be S_i instead of shot noise, I (assume photon number flow rates in all reservoirs for convenience). Finally, using a small-signal analysis, show that the spectral density of the first detected photon stream $\langle I_{d1} I_{d1}\rangle$ is equal to the product $\langle(\eta F_i + F_{d1})(\eta F_i + F_{d1})\rangle$, and then evaluate the product for both detectors. Interpret how this result applies to Eqs. 5.146a and 5.146b. How and why does the beam splitter affect the detected noise?
2. For the arrangement described in Problem A13.1, show that the spectral density of the *sum* of the detected photon streams reproduces the noise of

the incident photon stream, independent of η. What spectral density does the *difference* between the two detected photon streams yield in general, and for $\eta = 0.5$? Can you use the results for the sum and difference spectral densities to suggest a measurement technique that calibrates the measured intensity noise of a laser to the shot noise floor? This type of configuration is in fact known as a *balanced detector pair* and is commonly employed to determine whether or not a laser is operating below the standard quantum limit for noise.

APPENDIX FOURTEEN

Derivation Details for Perturbation Formulas

In Chapter 6 several calculations involve deriving an approximate solution to a perturbation on a known waveguide problem. This typically involves inserting a trial field into the wave equation, dropping out second-order terms, multiplying through by a complex conjugate of a transverse-mode eigenfunction of the unperturbed problem, integrating over the cross section, and using modal orthogonality as well as other arguments to drop additional terms, so that a simple analytic formula can be derived.

During the course of these calculations we repeatedly come on a collection of terms composed of the transverse-mode perturbation acted upon by the wave equation. The first such instance is given in Eq. (A14.1), which is the result just after multiplying times U^* and integrating over the cross section. That is,

$$2\beta\Delta\beta\int|U|^2\,dA=\int\Delta\varepsilon k_0^2|U|^2\,dA+\int[(\nabla_T^2\Delta U)U^*+\varepsilon k_0^2\Delta UU^*-\beta^2\Delta UU^*]\,dA. \quad \text{(A14.1)}$$

The terms in question are contained in the second integral on the right side of the equation (second line). We need to show that the integral is zero or negligible in the cases of interest.

Diode Lasers and Photonic Integrated Circuits, Second Edition.
Larry A. Coldren, Scott W. Corzine, and Milan L. Mašanović.
© 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

First, we multiply the complex conjugate of the transverse wave equation, Eq. (6.5), by ΔU to get

$$\Delta U(\nabla_T^2 U^*) + \Delta U[\varepsilon^*(x,y,z)k_0^2 - \beta^{*2}]U^* = 0. \tag{A14.1}$$

Thus, the last two terms in the last integral in Eq. (A14.1) can be replaced by $-\Delta U(\nabla_T^2 U^*) + (\varepsilon - \varepsilon^*)k_0^2 U^* - (\beta^2 - \beta^{*2})U^*$. If ε and β were real, then we would only need $-\Delta U(\nabla_T^2 U^*)$. Making the replacement,

$$\begin{aligned}\int [(\nabla_T^2 \Delta U)U^* + \varepsilon k_0^2 \Delta U U^* - \beta^2 \Delta U U^*]\, dA \\ = \int [(\nabla_T^2 \Delta U)U^* - \Delta U(\nabla_T^2 U^*)]\, dA \\ + \int [(\varepsilon - \varepsilon^*)k_0^2 \Delta U U^* - (\beta^2 - \beta^{*2})\Delta U U^*]\, dA. \end{aligned} \tag{A14.2}$$

Now, the first integral on the right is identically zero because

$$\begin{aligned}\int [(\nabla_T^2 \Delta U)U^* - \Delta U(\nabla_T^2 U^*)]\, dA &= \int \nabla_T \cdot [(\nabla_T \Delta U)U^* - \Delta U(\nabla_T U^*)]\, dA \\ &= \oint_\infty \hat{\mathbf{e}}_n \cdot [(\nabla_T \Delta U)U^* - \Delta U(\nabla_T U^*)]\, ds \equiv 0. \end{aligned} \tag{A14.3}$$

The latter equality uses Green's Theorem to convert the integral over the cross-sectional area to a line integral around the perimeter of the cross section (in this case, at infinity). The vector $\hat{\mathbf{e}}_n$ is the unit vector normal to the contour of integration. The contour integral at infinity is zero because both U and ΔU must vanish at infinity for any guided mode. The second integral on the right side of Eq. (A14.2) also is identically zero for ε and β real. In fact, even for complex ε and β, it tends to be negligible in comparison to the first integral on the right side of Eq. (A14.1) in most cases, since $\Delta U \ll U$, and because the loss or gain of the unperturbed problem can usually be chosen to be sufficiently small. Only in the extreme case where the gain provides most of the waveguiding effect (as in gain-guided lasers) will this term be nonnegligible. In this rare case, the use of the $\Delta\beta$ formula is questionable.

READING LIST

Haus HA. *Waves and fields in optoelectronics*. Englewood Cliffs, NJ: Prentice Hall; 1984.

APPENDIX FIFTEEN

Multimode Interference

Modal excitation of multimode waveguides can be used to realize important photonic functions of light splitting and combining through the effect of multimode interference (MMI).

A15.1 MULTIMODE INTERFERENCE-BASED COUPLERS

Multimode interference couplers are based on the principle of self-imaging. This is a property of the multimode waveguides by which an input field profile is periodically reproduced as a single or multiple images while propagating through the multimode waveguide, as illustrated in Fig. 6.24. To create an MMI effect, it is crucial to have a waveguide that supports a large number of modes. As will be explained later, the larger the number of supported modes, the better the imaging properties of the MMI components. However, the larger the number of modes, the wider the multimode waveguide needs to be, which increases the length of the MMI component, and its insertion losses. At the end of this section, we will show that the optimum MMI component length is proportional to the MMI width squared.

In Section 6.5, we have analyzed the case of waveguide excitation at the interface of two waveguides. Here, we expand that analysis to examine the input field exciting multiple modes of a multimode waveguide.

We will be analyzing a step-index multimode waveguide of width d_m, which supports M lateral modes with numbers $m = 0, 1, \ldots (M - 1)$ at a free-space wavelength λ_0. Defining the waveguide transverse effective index as n_l, the lateral wave

Diode Lasers and Photonic Integrated Circuits, Second Edition.
Larry A. Coldren, Scott W. Corzine, and Milan L. Mašanović.
© 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

vector k_{xm} and the propagation constant β_m are related to the waveguide transverse effective index (which serves as the lateral plane wave index here) by the standard relation

$$k_{xm}^2 + \beta_m^2 = k_0^2 n_l^2. \tag{A15.1}$$

For each of the supported modes, the effective waveguide width d_{effm} takes into account the polarization dependent lateral penetration depth of each mode field due to its exponential decay, as discussed in Appendix 3 and Chapter 7. The effective waveguide width can generally be approximated by the effective width of the fundamental mode, d_{eff}. The lateral component of the wave vector, k_{xm}, has to fulfill the transverse resonance condition,

$$k_{xm} = \frac{(m+1)\pi}{d_{effm}} \approx \frac{(m+1)\pi}{d_{eff}}. \tag{A15.2}$$

The propagation constant β_m can be expressed from Eqs. (A15.1) and (A15.2) as

$$\beta_m = k_0 n_l \sqrt{1 - \frac{k_{xm}^2}{k_0^2 n_l^2}} \approx k_0 n_l - \frac{(m+1)^2 \pi \lambda_0}{4 n_l d_{eff}^2}, \tag{A15.3}$$

where we have used the binomial expansion, enabled by the fact that $k_{xm}^2 \ll k_0^2 n_l^2$.

The propagation constants in a step-index multimode waveguide therefore show a nearly quadratic dependence with respect to the mode number m.

We now define the self-imaging length as the beat length between the lateral plane wave and the fundamental lateral mode in the MMI region,

$$L_{si} \equiv \frac{2\pi}{k_0 n_l - \beta_0} = \frac{8 n_l d_{eff}^2}{\lambda_0}, \tag{A15.4a}$$

which will turn out to be a quantity useful for describing the imaging lengths of key MMI structures. The propagation constant β_m can now be expressed as

$$\beta_m = k_0 n_l - \frac{2\pi (m+1)^2}{L_{si}}. \tag{A15.4b}$$

A15.2 GUIDED-MODE PROPAGATION ANALYSIS

Using the principles of the guided-mode propagation analysis, an input field profile $E(x,0)$ at a distance $z = 0$ and contained within a finite width W will, in general, excite all the guided modes of a multimode waveguide. These modes form a basis set of functions needed to describe the input field,

$$E(x,0) = \sum_{m=0}^{M-1} a_m U_m(x). \tag{A15.5}$$

The field profile at a propagation distance z from the interface will be given by the superposition of the individual modes of the MMI waveguide,

$$E(x,z) = \sum_{m=0}^{M-1} a_m U_m(x) \cdot e^{-j\beta_m z}, \tag{A15.6}$$

where the exponential time dependence of the field is implicit. Using the expression for the β_m, Eq. (A15.4b), and taking the phase of the lateral plane wave out of the sum, the MMI field profile equation is transformed into

$$E(x,z) = e^{-jk_0 n_l z} \cdot \sum_{m=0}^{M-1} a_m U_m(x) \cdot e^{j2\pi(m+1)^2(z/L_{si})}. \tag{A15.7}$$

Analyzing this expression, we notice that the types of images formed will be determined by the modal excitation factors a_m and by the properties of the mode phase factor

$$e^{j2\pi(m+1)^2(z/L_{si})}. \tag{A15.8}$$

Per example, with $z = L_{si}$, it becomes obvious that this phase factor equals 1 for all m, and that the sum of all the field components will produce an image identical to the input field - hence the name "self-imaging" length L_{si} that we have chosen. Note also that the image is delayed in phase by the phase delay of the lateral plane wave across that distance.

A15.2.1 General Interference

General interference effects in multimode waveguides occur when all the modes of the multimode waveguides have non-zero overlap with the input field, and thus get excited at the MMI input.

The general interference can yield both single and multiple images of the input field profile, based on the multimode waveguide length. This is illustrated in Figure 6.25, where calculated electric fields are superimposed over the waveguide geometry used.

Inspecting Eq. (A15.8), we notice that for $z = L_{si}, L_{si}/2$ and $L_{si}/4$, we have

$$\begin{aligned}
z = L_{si}: &\quad e^{j2\pi(m+1)^2} = 1 \quad (\text{for all m}) && \text{self-image} \\
z = \frac{L_{si}}{2}: &\quad e^{j\pi(m+1)^2} = \begin{cases} -1 & (\text{even m}) \\ 1 & (\text{odd m}) \end{cases} && \text{mirrored image} \\
z = \frac{L_{si}}{4}: &\quad e^{j(\pi/2)(m+1)^2} = \begin{cases} j & (\text{even m}) \\ 1 & (\text{odd m}) \end{cases} && \text{self-and mirrored image}
\end{aligned} \tag{A15.9}$$

The first case from Eq. (A15.9) is pretty straightforward, as based on Eq. (A15.7), we have that

$$E(x, L_{si}) = e^{-jk_0 n_l L_{si}} \sum_{m=0}^{M-1} a_m U_m(x) = e^{-jk_0 n_l L_{si}} E(x, 0), \tag{A15.10}$$

which is a self image with a phase delay of the lateral plane wave across the distance $z = L_{si}$.

For the latter two cases, we need to break the sum of modes from Eq. (A15.7) into even and odd modes and associate those with the even and odd field profiles at $z = 0$. For the mirrored image case, $z = L_{si}/2$, we have that

$$E(x, L_{si}/2) = e^{-jk_0 n_l L_{si}/2} \cdot \left(\sum_{m=even} a_m U_m(x)(-1) + \sum_{m=odd} a_m U_m(x)(1) \right) \tag{A15.11}$$

The even and odd mode sums can be expressed in terms of the even and odd input field profiles as follows

$$\begin{aligned} \sum_{m=even} a_m U_m(x) &= \frac{1}{2}(E(x, 0) + E(-x, 0)), \\ \sum_{m=odd} a_m U_m(x) &= \frac{1}{2}(E(x, 0) - E(-x, 0)). \end{aligned} \tag{A15.12}$$

Substituting Eq. (A15.12) into Eq. (A15.11), we arrive at the mirror image

$$E(x, L_{si}/2) = -e^{-jk_0 n_l L_{si}/2} \cdot E(-x, 0). \tag{A15.13}$$

Using the same procedure, for $z = L_{si}/4$ we get

$$E(x, L_{si}/4) = e^{-jk_0 n_l L_{si}/4} \left(\frac{1+j}{2} E(x, 0) - \frac{1-j}{2} E(-x, 0) \right), \tag{A15.14}$$

which represents two images, a self-image and a mirror image that are 90° out of phase with each other. Hence, two symmetric input and output waveguides and an MMI component of this length can be used to produce an efficient 2-by-2 coupler.

The expressions can be generalized into a compact expression with an integer p stepping through the four unique $z = p \cdot L_{si}/4$ values,

$$E(x, pL_{si}/4) = \frac{j^p}{2}[E(x, 0) + E(-x, 0)] + \frac{1}{2}[E(x, 0) - E(-x, 0)], \tag{A15.15}$$

where we have excluded the plane wave phase term. The term in the first bracket is generated by the symmetric or even components of the input field, while the term in the second bracket is generated by the antisymmetric or odd components.

As we can see, the pattern predicted by Eq. (A15.15) is cyclic with length, and it varies as p modulo 4 in cases where z exceeds L_{si}.

Analytical expressions for both the image z and x positions and phases relative to the input field can be obtained using Fourier analysis, and are given by

$$\begin{aligned} L &= \frac{p}{N}\left(\frac{L_{si}}{2}\right) \\ x_q &= p(2q-N)\frac{d_{eff}}{N} \\ \phi_q &= p(N-q)\frac{q\pi}{N} \end{aligned} \tag{A15.16}$$

where p indicates the imaging periodicity along z, q refers to each of the N images along x, and p and N have no common divisors. It is interesting to note that the N images are generally not equally spaced.

A15.2.2 Restricted Multimode Interference

Unlike general interference, restricted interference effects in multimode waveguides occur when the input excitation (waveguide position and field) are such that some of the modes of the multimode waveguides are not excited. This is due to the fact that the overlap integral of these modes with the input field is equal to zero. A special case of restricted interference that is of particular interest is symmetric interference, when the input waveguide is laterally centered relative to the multimode waveguide. In all symmetric interference components, only even modes of the multimode waveguide get excited, since all odd modes have a null at the center of the waveguide.

Starting from Eq. (A15.15), for symmetric excitation, the second term will disappear, and the expression for the field at distance $z = L_{si}/4$ will reduce to

$$E(x, L_{si}/4) = \frac{j}{2}[E(x,0) + E(-x,0)] = jE(x,0), \tag{A15.17}$$

since the input field is symmetric. Therefore, the self-image now appears at a distance that is 1/4 that of the general interference case. Likewise, the length periodicity of multifold images will also be reduced by a factor of 4, and N-fold images for symmetric inteference will be obtained at distances

$$L = \frac{p}{N}\left(\frac{L_{si}}{8}\right) \tag{A15.18}$$

and they will be located symmetrically along the x-axis, with equal spacings of d_{eff}/N.

Symmetric interference can be used to produce $1 \times N$ light splitters in a very length-efficient way. A 1×2 light splitter is illustrated in Fig. 6.26. Using symmetric interference components whenever possible will minimize the device size, which helps with the propagation losses and chip area utilization.

A15.3 MMI PHYSICAL PROPERTIES

A15.3.1 Fabrication

Multimode interference based components are easier to fabricate than directional couplers discussed in Chapter 6. Directional couplers require a submicron gap between waveguides, and even minute variations of the gap dimensions will greatly affect the coupler performance. On the other hand, MMIs consist of wide waveguides, and they are less sensitive to small width variations (even though their optimum length changes as square of the component width).

A15.3.2 Imaging Quality

The quadratic dependence of the propagation constants with the mode number is an approximation—therefore, the guided modes will accumulate small deviations in phase at the imaging distances, which will result in blurring of the reconstructed image. Some aleviation of this problem is achieved by using numerical modelling (such as beam propagation method, discussed later in the book) to optimize the component's length. Another effect that affects the MMI imaging properties is the scattering of the light at the edges of the MMI waveguide. The amount of this scattering will be mode dependent, thereby further degrading the imaging properties of the component. Consequently, it is of interest to use the shortest MMI components possible.

The imaging resolution of a MMI component is in principle determined by the highest supported multimode waveguide mode and can be estimated as equal to the cosine-like lobe width of that highest supported lobe, $\rho = \frac{d_{eff}}{m}$ [1].

A15.3.3 Inherent Loss and Optical Bandwidth

The inherent loss of an MMI component depends on the quality of the imaging of the input of the MMI onto the output. Basically, the inherent loss mechanisms will be the diffraction of the input field in the multimode waveguide section and the coupling losses between the imaged field and the output access waveguide. This loss is generally low due to good imaging properties and resolution of the MMIs.

Additional losses common to all optical waveguides, absorption and scattering, dictate that the length of the components be as short as possible.

Finally, the imaging length of the MMI needs to be matched for the material/waveguide structure chosen. Any departure from the optimum MMI component length will lead to the increase in the MMI inherent loss.

Optical bandwidth of the MMI components is limited by the changing mode propagation constants and refractive index values as a function of wavelength. The optimum imaging length and the field patterns vary slowly with wavelength, degrading the image at the output and reducing the power coupled into the output access waveguides. The optical bandwidth of an MMI component will be inversely proportional to the width squared of the MMI waveguide. Since the length of the

component has a square dependence on the MMI width, shorter MMI components are also expected to have higher optical bandwidth. Both 1×2 and 2×2 MMI couplers exhibit 3 dB bandwidths far in excess of 100 nm [1]. Therefore, the bandwidth does not pose any practical limitations to the MMI component design and choice.

A15.3.4 Polarization Dependence

Polarization dependence of the MMI components stems from the polarization dependence of the material system and the waveguide structure used for their fabrication. The level of polarization dependence is low, just as that of the passive waveguides. It is possible to make polarization insensitive MMI components by properly choosing the material and geometrical properties of the waveguide [1].

A15.3.5 Reflection Properties

MMI-based interconnecting components can be a source of major reflections. This is especially true for MMI 1×2 and 2×2 light splitters and combiners that are fabricated using strongly guided structures. We distinguish between reflection back into the input waveguides and internal resonance modes due to the occurrence of simultaneous self-images [1]. Because of self-imaging, reflection can be extremely efficient, even in the case of MMI devices with optimized transmission. Various techniques can be employed to suppress back reflections in the MMIs. They allow for sucessfull integration of MMI components with lasers as part of a single PIC.

REFERENCE

1. Soldano LB, Pennings ECM. Optical munti-mode interference defices based on selfimaging: principles and applications. *Journal of Lightwave Technology* 1995;**13**(4): 615–627.

APPENDIX SIXTEEN

The Electro-Optic Effect

For many tunable lasers and photonic integrated circuits it is desirable to change the index of refraction by the application of a dc (or rf) electric field. In certain crystals that do not possess inversion symmetry this is possible through what is known as the electro-optic effect. Because the index changes virtually instantaneously with the field, this effect can be used in very high modulation bandwidth modulators and tunable filters.

The electro-optic effect is somewhat more complex than one might at first expect because the change in index is generally different for different polarizations of the optical field for a given applied DC field orientation, and moreover, it is usually associated with anisotropic crystals that have different indexes for different optical polarizations initially [1]. Thus, it is necessary to use the full dielectric tensor in relating the displacement, $\mathbf{D}$, to the electric field, $\mathscr{E}$, of the lightwave. That is, the displacement might not be parallel to the electric field for all orientations. Also, the applied DC or rf field generally has still another orientation, so we have three orientations to keep track of in this process.

The displacement field is related to the optical electric field via the dielectric tensor, $\mathbf{D} = \varepsilon\mathscr{E}$, or

$$\begin{bmatrix} D_x \\ D_y \\ D_z \end{bmatrix} = \begin{bmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{yx} & \varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{zx} & \varepsilon_{zy} & \varepsilon_{zz} \end{bmatrix} \begin{bmatrix} \mathscr{E}_x \\ \mathscr{E}_y \\ \mathscr{E}_z \end{bmatrix}. \tag{A16.1}$$

If we choose axes to diagonalize the matrix, then we have found the principle dielectric axes of the medium. Switching to the optical index, we then

Diode Lasers and Photonic Integrated Circuits, Second Edition.
Larry A. Coldren, Scott W. Corzine, and Milan L. Mašanović.
© 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

have

$$\begin{bmatrix} D_x \\ D_y \\ D_z \end{bmatrix} = \varepsilon_0 \begin{bmatrix} n_{xx}^2 & 0 & 0 \\ 0 & n_{yy}^2 & 0 \\ 0 & 0 & n_{zz}^2 \end{bmatrix} \begin{bmatrix} \mathscr{E}_x \\ \mathscr{E}_y \\ \mathscr{E}_z \end{bmatrix}. \tag{A16.2}$$

We continue to allow for anisotropy of the index, but we have defined an index ellipsoid with axes along the material's principle axes. When the field is polarized along one of these directions, **D** will be parallel to $\boldsymbol{\mathscr{E}}$. In most materials at least two of these indexes are equal. These are called the *ordinary* index, while the third is called the *extraordinary* index. Rays with their electric field aligned along these directions are also referred to as the ordinary or extraordinary rays. By convention we label n_{zz} as the extraordinary index, and the z-axis is called the *optic* axis. Materials with two such indexes are called *birefringent*. In the III–V semiconductor materials considered in this book, the index is isotropic when no additional DC or rf field is applied. Thus, all the indexes are the same with no applied field. However, with the application of a field, these materials become birefringent also.

The general index ellipsoid from Eq. (A16.2) is

$$\frac{x^2}{n_{xx}^2} + \frac{y^2}{n_{yy}^2} + \frac{z^2}{n_{zz}^2} = 1. \tag{A16.3}$$

The index of any propagating wave can be found by constructing a plane perpendicular to its k-vector that passes through the origin of the ellipsoid. The wave's electric field direction is then constructed as a line from the origin on this plane. The index is given by the distance to the surface of the ellipsoid. Figure A16.1 illustrates this construction. Note that for GaAs or InP this ellipsoid is just a sphere in the absence of any applied dc or rf field.

With the application of a dc or rf field to an electro-optic material this ellipsoid is distorted, so that generally off-diagonal terms appear in Eq. (A16.2). The electro-optic tensors are defined from the perturbation to each of the terms in the index ellipsoid. That is,

$$\Delta\left(\frac{1}{n^2}\right)_{ij} = \sum_k r_{ijk}\mathscr{E}_k, \tag{A16.4}$$

where **r** is the linear electro-optic tensor and $\mathscr{E}_k$ is the applied dc electric field. Three subscripts are necessary to account for the relative orientation of the crystal axis, the optical electric field, and the dc electric field. It is also possible that quadratic effects are important, so we can also define a similar equation where $\boldsymbol{\mathscr{E}}$ is replaced by $\boldsymbol{\mathscr{E}}^2$ and **r** is replaced by **s**, the quadratic electro-optic tensor. Quadratic effects are important for wavelengths near the absorption edge in bulk and MQW waveguides, where an electric field can move the effective absorption edge [2,3]. Also, in depletion regions of *pn*-junctions quadratic

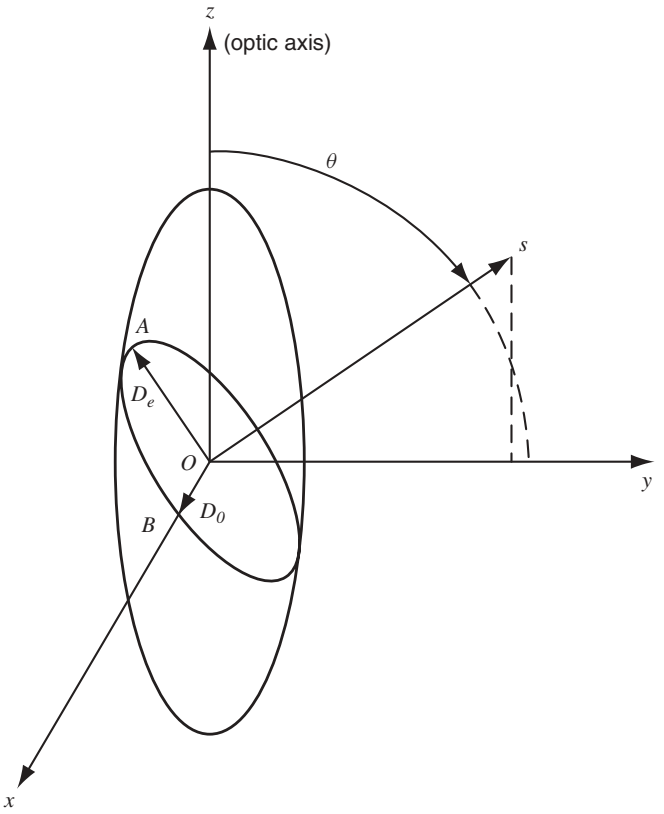


FIGURE A16.1: Plot of index ellipsoid [1] (From *Optical Waves in Crystals*, A. Yariv and P. Yeh, Copyright © 1984 John Wiley & Sons, Inc. Reprinted by permission of John Wiley and Sons, Inc.)

effects become dominant if a significant carrier density is being depleted at the same time [4]. We will come back to this topic after completing the linear effect.

To simplify the notation somewhat, an abbreviated subscript notation is usually introduced, so that the three-dimensional tensor can be expressed in two dimensions. The notation involves using numbers from 1 to 6 for the first two subscripts, where $1 \equiv xx, 2 \equiv yy, 3 \equiv zz, 4 \equiv yz, 5 \equiv xz,$ and $6 \equiv xy$. The final subscript, which refers to the applied dc or rf electric field orientation also is given a numerical subscript, where $1 \equiv x, 2 \equiv y,$ and $3 \equiv z$. Thus, Eq. (A16.4) becomes

$$\Delta\left(\frac{1}{n^2}\right)_i = \sum_{j=1}^{3} r_{ij}\mathscr{E}_j, \qquad i = 1, 2, \ldots, 6 \tag{A16.5}$$

With the application of a dc or rf field, the distorted index ellipsoid takes on the form,

$$\left(\frac{1}{n^2}\right)_1 x^2 + \left(\frac{1}{n^2}\right)_2 y^2 + \left(\frac{1}{n^2}\right)_3 z^2 + 2\left(\frac{1}{n^2}\right)_4 yz + 2\left(\frac{1}{n^2}\right)_5 xz + 2\left(\frac{1}{n^2}\right)_6 xy = 1, \tag{A16.6}$$

where the additional terms are solely due to (A15.5) and the original terms may also be changed if nonzero perturbation terms are generated by Equation. To determine the change in index for a propagating optical wave, we again must diagonalize the index matrix, or equivalently, find the new principle axes for the distorted index ellipsoid. This generally involves a rotation in space of the original axes. Once this new principle axis set (x', y', z') is found, the index ellipsoid will again have the form of (A15.3). The coefficient of x'^2 will give the reciprocal of the square of the perturbed index, $n_{x'}$, for an optical field polarized along the x'-axis, and similarly for the other components.

In the case of GaAs, InP, or their alloys, the process of finding the perturbed principle axes and the related indexes is somewhat simplified, since for crystals of this zinc blende class (*cubic*-$\bar{4}3m$) only r_{41}, r_{52}, and r_{63} are nonzero for x, y, and z aligned with the crystal axes. Also, these all have the same value, so we set them all equal to r_{41}. Thus, for a dc field in the z-direction, or [001], Eq. (A16.6) becomes

$$\frac{x^2 + y^2 + z^2}{n_0^2} + 2r_{41}\mathscr{E}_z xy = 1, \tag{A16.7}$$

where n_0 is the unperturbed index for the isotropic semiconductor in all directions. That is, in the $x - y$ plane, the index ellipsoid is somewhat squashed in the first and third quadrants and somewhat stretched in the second and fourth quadrants. To put Eq. (A16.7) in the form of Eq. (A16.3), it can be shown that a rotation of the coordinate system about the z-axis by 45° is required. That is, $z = z'$, and

$$\begin{aligned} x &= x' \cos 45^\circ + y' \sin 45^\circ, \\ y &= -x' \sin 45^\circ + y' \cos 45^\circ, \end{aligned} \tag{A16.8}$$

so that the new (primed) axes become $x' = [1\bar{1}0]$ and $y' = [110]$. Plugging this into Eq. (A16.7), we obtain

$$\left(\frac{1}{n_0^2} - r_{41}\mathscr{E}_z\right) x'^2 + \left(\frac{1}{n_0^2} + r_{41}\mathscr{E}_z\right) y'^2 + \left(\frac{1}{n_0^2}\right) z^2 = 1. \tag{A16.9}$$

Thus,

$$\frac{1}{n_{x'}^2} = \frac{1}{n_0^2} - r_{41}\mathscr{E}_z \qquad \text{and} \qquad \frac{1}{n_{y'}^2} = \frac{1}{n_0^2} + r_{41}\mathscr{E}_z. \tag{A16.10}$$

For $r_{41}\mathscr{E}_z \ll 1/n_0^2$, we can approximately solve for the desired perturbed indices,

$$n_{x'} = n_0 + \frac{n_0^3}{2} r_{41}\mathscr{E}_z,$$
$$n_{y'} = n_0 - \frac{n_0^3}{2} r_{41}\mathscr{E}_z. \tag{A16.11}$$

The result of the above calculation is that the index for an optical field polarized along the $[1\bar{1}0]$ direction (i.e., x'-direction) will be increased by $(n_0^3/2)r_{41}\mathscr{E}_z$, while the index for an optical field polarized along the [110] direction (i.e., y'-direction) will be decreased by $(n_0^3/2)r_{41}\mathscr{E}_z$. There will be no change in index for a component of optical field along the direction of the applied field [001]. Thus, for a waveguide on a (001) wafer aligned perpendicular to the natural (110) cleavage planes in these III–V materials, the application of a surface-normal field will cause the index of a TE mode to change, but it will not change for a TM mode.

In GaAs, r_{41} varies from 1.1 to 1.5×10^{-12} m/V in the 0.9 to 1.3 μm wavelength range. In other materials the electro-optic coefficient can be much larger than this. Of course, it is really the product $n^3 r_{ij}$, which gives the best measure of the index change for a given applied field. In GaAs, $n^3 r_{41} \sim 55\times10^{-12}$ m/V. Thus, for fields ~400 kV/cm, obtainable in *pin* undoped regions (as well as in the depletion regions of *pn*-junctions), $\Delta n/n \sim 0.001$. Outside the semiconductor regime, the most popular material for photonic integrated circuits is lithium niobate, $LiNbO_3$. In this *trigonal*-3*m* material several terms in the electro-optic tensor are nonzero. At 1.3 μm, $r_{13} = -r_{23} \sim 7$; $r_{22} = -r_{12} = -r_{61} \sim 3.3$; $r_{33} \sim 30$; and $r_{51} = r_{42} \sim 26\times10^{-12}$ m/V. The ordinary and extraordinary indices of refraction are 2.22 and 2.14, respectively. Also, $n_0^3 r_{33} \sim 328\times10^{-12}$ m/V.

In certain III–V device configurations the quadratic electro-optic effect is larger than the linear effect. This can happen if the wavelength of the lightwave is close to the absorption edge in bulk or quantum-well material. When a field is applied the absorption edge moves to longer wavelengths via either the Franz–Keldysh (FK) in bulk or quantum-confined Stark effect (QCSE) in MQWs. Since the index of refraction is decreasing roughly as $1/\lambda$ above the absorption edge, the application of a field also increases the index at some wavelength in this region. The combined effect is nearly quadratic with electric field [4]. Also, if free carriers are depleted when a field is applied, e.g., in the depletion region of a *pn*-junction, the index will increase due to the removal of the index reduction associated with the existence of free carriers as well as the shift in absorption edge due to a removal of band filling. This effect again is approximately quadratic with the applied field.

Analogous to the linear effect, the quadratic index change in all cases can be written as

$$\Delta n_q = \frac{n_0^3}{2} s\mathscr{E}_z^2. \tag{A16.12}$$

The difficulty in these cases is that the quadratic coefficient, s, also decreases with increasing distance away from the absorption edge, and therefore, it is difficult to

effectively parameterize the problem as with the linear electro-optic effect. With the QCSE, associated with absorption due to excitons in MQWs, the index shift per unit field change can be larger than in bulk GaAs, but the effect also decays more rapidly with increasing wavelength. For example, the effect becomes less than the linear effect ~40 nm from the absorption edge, whereas in bulk materials the quadratic effect due to the absorption edge alone is larger than the linear effect 100 nm away. The depletion of charge also only works effectively in bulk materials, since the charge will tend to screen the excitons in quantum wells. Thus, it is unclear whether or not quantum wells can offer a significant practical advantage in electro-optic modulators.

Figure A16.2 plots the phase modulation available in a doped GaAs waveguide under reverse bias. This device has a *pn*-junction in the center of the doped waveguide, and the waveguide is oriented along the [011] direction so that the quadratic effects add to the linear effect. A separate curve is plotted for each physical effect. The net phase shift $\Delta\phi$, in a device of length L, is obtained from the plotted phase-shift efficiency, η_{ps}, from

$$\Delta\phi = \eta_{ps} L(\Delta V), \tag{A16.13}$$

where ΔV is the applied voltage shift.

In all cases the proximity of the operating wavelength to the absorption edge is limited by loss. That is, although the amount of index change available increases

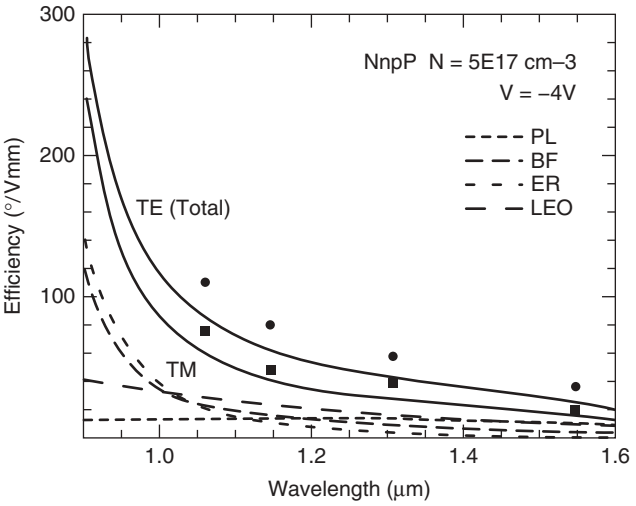


FIGURE A16.2: Calculated and measured (data) phase-shift efficiency versus wavelength with $L_m = 740$ μm, and equal *n*- and *p*-type doping to the center of the 0.25 μm-thick GaAs waveguide [5]. The AlGaAs cladding material contains 40% Al. PL, BF, ER and LEO refer to the plasma, band-filling, electro-refractive, and linear electro-optic effect respectively. All except LEO are approximately quadratic with applied reverse bias in this configuration.

as the absorption edge is approached, so does the loss and the loss change. As a consequence, it is important to limit the residual loss and loss change to some values when comparing different kinds of index modulation. As defined in Chapter 2, the ratio of the changes in the real to the imaginary parts of the refractive index is called the chirp parameter, α. Thus for good phase modulation with low loss this parameter should be large. As a rule of thumb, we prefer $\alpha > 10$ for reasonably low-loss phase modulation. Figure A16.3 shows the calculated and measured phase modulation

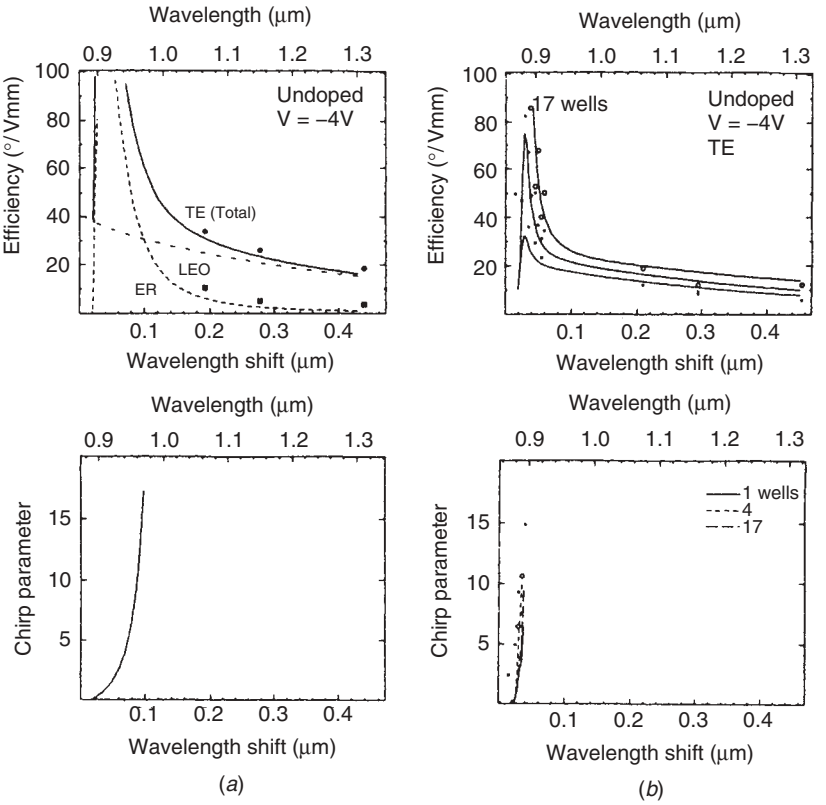


FIGURE A16.3: Calculated and measured phase-shift efficiency (top) and chirp (bottom) for undoped material versus wavelength deviation from the zero-bias absorption edges at 0.87 μm for bulk and 0.855 μm for the MQW waveguides [5, 6]. The waveguide regions (MQW separate-confinement region) are 0.25 μm wide clad by $Al_{0.4}Ga_{0.6}As$. The bulk guide is undoped GaAs; the MQW region has either 1, 4, or 17 GaAs wells, 10-nm-thick, separated by $Al_{0.2}Ga_{0.8}As$ barriers. In all cases the phase-shift efficiency and chirp are calculated between 0 and −4V. For the bulk case (a) the TE mode measurements are given by circles; the TM mode by squares. The calculated net quadratic electro-refractive (ER) effect is shown with the short dashed line, the net linear electro-optic (LEO) effect is shown with the long dashed line, and the total with a solid line. For the MQW case (b) only the TE mode is measured.

efficiency and the chirp parameter for undoped bulk and MQW waveguides for comparison.

REFERENCES

1. Yariv A, Yeh P. *Optical waves in crystals*. New York: Wiley; 1984. Chapters 4 and 7.
2. Alping A, Coldren LA. Electrorefraction in GaAs and InGaAsP and its application to phase modulators. *J Appl Phys* 1987;**61**: 2430.
3. Miller DAB, Chemla DS, Schmitt-Rink S. Relation between electroabsorption in bulk semiconductors and in quantum wells: the quantum-confined Franz-Keldysh effect. *Phys Rev B* 1985;**33**: 6976
4. Mendoza-Alvarez JG, Coldren LA, Alping A, Yan RH, Hausken T, Lee K, Pedrotti K. Analysis of depletion edge translation lightwave modulators. *IEEE J Lightwave Technol* 1988;**6**: 793.
5. Hausken T. [dissertation]. ECE Technical Report #90-18. Santa Barbara (CA): University of California Santa Barbara; 1990. Chapter 2.
6. Hausken T. [dissertation]. ECE Technical Report #90-18. Santa Barbara (CA): University of California Santa Barbara; 1990. Chapter 4.

READING LIST

Yariv A. *Optical electronics*. 4th ed. Philadelphia: Saunders College Publishing; 1991. Chapter 9.

Yariv A, Yeh P. *Optical waves in crystals*. New York: Wiley; 1984. Chapter 7.

APPENDIX SEVENTEEN

Solution of Finite Difference Problems

A17.1 MATRIX FORMALISM

In Section 7.3.6, the finite-difference technique was introduced to solve for the effective index and field profiles of an arbitrary channel waveguide structure. The scalar wave equation was discretized to provide the following linear matrix equation:

$$\frac{U_j^{i-1}}{\Delta X^2} + \frac{U_{j-1}^{i}}{\Delta Y^2} - \left(\frac{2}{\Delta X^2} + \frac{2}{\Delta Y^2} - \left(n_j^i\right)^2 \right) U_j^i + \frac{U_{j+1}^{i}}{\Delta Y^2} + \frac{U_j^{i+1}}{\Delta X^2} = \bar{n}^2 U_j^i. \quad \text{(A17.1)}$$

where U_j^i and n_j^i are the normalized electric field and refractive index at the grid point (i,j) for $i = 0, 1, 2, \ldots, I+1$ and $j = 0, 1, 2, \ldots, J+1$. Also, $\Delta X = k_0 \Delta x$ and $\Delta Y = k_0 \Delta y$ are the normalized coordinate steps between grid points, and $\bar{n}$ is the effective index of the waveguide mode.

To complete the problem we need to specify boundary conditions. Although different choices exist, the simplest choice is simply to set the fields to zero around the border of the computational window. This approximation is valid as long as we are far enough away from the guiding layers that the true field solutions are essentially zero. Experimentation with the size of the computational window can determine the validity of this approximation. So our boundary conditions are $U_j^0 = U_j^{I+1} = U_0^i = U_{J+1}^i = 0$, and we are left solving for the fields within the window. In other words, we need to solve all equations for $i = 1$ to I and $j = 1$ to J, or a total of $I \times J$ coupled equations.

Diode Lasers and Photonic Integrated Circuits, Second Edition.
Larry A. Coldren, Scott W. Corzine, and Milan L. Mašanović.
© 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

Defining coefficients for terms with common j indices, Eq. (A17.1) becomes

$$\frac{U_j^{i-1}}{\Delta X^2} + \left[bU_{j-1}^i + a_j^i U_j^i + bU_{j+1}^i\right] + \frac{U_j^{i+1}}{\Delta X^2} = \bar{n}^2 U_j^i, \tag{A17.2}$$

where

$$a_j^i = \left(n_j^i\right)^2 - \frac{2}{\Delta X^2} - \frac{2}{\Delta Y^2},$$
$$b = \frac{1}{\Delta Y^2}.$$

We can compact the y-direction into matrix notation by defining a vector that encompasses y for each x position, i:

$$\mathbf{U}^i = \begin{bmatrix} U_1^i \\ \vdots \\ U_J^i \end{bmatrix}. \tag{A17.3}$$

Then, by vertically listing all J Eqs. (A17.2) for a given i, we can group common i indices into a matrix-difference equation along x:

$$\mathbf{B}\mathbf{U}^{i-1} + \mathbf{A}^i\mathbf{U}^i + \mathbf{B}\mathbf{U}^{i+1} = \bar{n}^2\mathbf{U}^i, \tag{A17.4}$$

where

$$\mathbf{A}^i = \begin{bmatrix} a_1^i & b & 0 & \cdots & 0 & 0 \\ b & a_2^i & b & 0 & & 0 \\ 0 & b & a_3^i & b & 0 & \vdots \\ \vdots & 0 & & \ddots & & 0 \\ 0 & & 0 & b & a_{J-1}^i & b \\ 0 & 0 & \cdots & 0 & b & a_J^i \end{bmatrix},$$

$$\mathbf{B} = \frac{1}{\Delta X^2}\mathbf{I},$$

and $\mathbf{I}$ is the $J \times J$ identity matrix. The second b does not appear in the top and bottom rows of the $\mathbf{A}^i$ matrix because $U_0^i = 0$ and $U_{J+1}^i = 0$ (this is in fact the reason we need to set the fields to zero at the boundaries).

Now writing all I Eqs. (A17.4) along x in matrix form, we obtain

$$\begin{bmatrix} \mathbf{A}^1 & \mathbf{B} & 0 & \cdots & 0 & 0 \\ \mathbf{B} & \mathbf{A}^2 & \mathbf{B} & 0 & & 0 \\ 0 & \mathbf{B} & \mathbf{A}^3 & \mathbf{B} & 0 & \vdots \\ \vdots & 0 & & \ddots & & 0 \\ 0 & & 0 & \mathbf{B} & \mathbf{A}^{I-1} & \mathbf{B} \\ 0 & 0 & \cdots & 0 & \mathbf{B} & \mathbf{A}^I \end{bmatrix} \begin{bmatrix} \mathbf{U}^1 \\ \mathbf{U}^2 \\ \mathbf{U}^3 \\ \vdots \\ \mathbf{U}^{I-1} \\ \mathbf{U}^I \end{bmatrix} = \bar{n}^2 \begin{bmatrix} \mathbf{U}^1 \\ \mathbf{U}^2 \\ \mathbf{U}^3 \\ \vdots \\ \mathbf{U}^{I-1} \\ \mathbf{U}^I \end{bmatrix}, \tag{A17.5}$$

where again the second **B** does not appear in the top and bottom rows because $\mathbf{U}^0 = 0$ and $\mathbf{U}^{I+1} = 0$. We can write this equation symbolically as

$$\mathbf{AU} = \bar{n}^2\mathbf{U} \quad \text{with} \quad \mathbf{U} = \begin{bmatrix} \mathbf{U}^1 \\ \vdots \\ \mathbf{U}^I \end{bmatrix}. \tag{A17.6}$$

This is the matrix equation to be solved for the eigenvalue, $\bar{n}^2$, and eigenvector, **U**. In this compact form, **A** represents an $I \times I$ matrix with elements that are themselves $J \times J$ matrices, and **U** is an I length vector with components that are themselves J length vectors. To solve the equation, we expand each element of **A** into a $J \times J$ block, so that **A** becomes an $(IJ) \times (IJ)$ matrix with scalar elements. Likewise, we expand the component vectors of **U**, so that **U** becomes an (IJ) length vector with scalar components. Various matrix methods are then available for determining the eigenvalues and eigenvectors.

A17.2 ONE-DIMENSIONAL DIELECTRIC SLAB EXAMPLE

To illustrate the basic numerical technique, we consider the most simple slab waveguide problem. In this case the scalar wave equation correctly describes the TE mode with its electric field polarized along the y-direction as illustrated in Fig. A17.1.

In this case the eigenvalue Eq. (A17.5) reduces to

$$\begin{bmatrix} \left(n_1^2 - \frac{2}{\Delta X^2}\right) & \frac{1}{\Delta X^2} & 0 & 0 \\ \frac{1}{\Delta X^2} & \left(n_2^2 - \frac{2}{\Delta X^2}\right) & \frac{1}{\Delta X^2} & 0 \\ 0 & & \ddots & 0 \\ 0 & 0 & \frac{1}{\Delta X^2} & \left(n_I^2 - \frac{2}{\Delta X^2}\right) \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_I \end{bmatrix} = \bar{n}^2 \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_I \end{bmatrix}, \tag{A17.7}$$

where we have simplified the notation in this one-dimensional case, in which there is no variation of the index in the y-direction, by letting, $n_i = n_j^i$, and $U_i = U_j^i$, since j only takes on a single value.

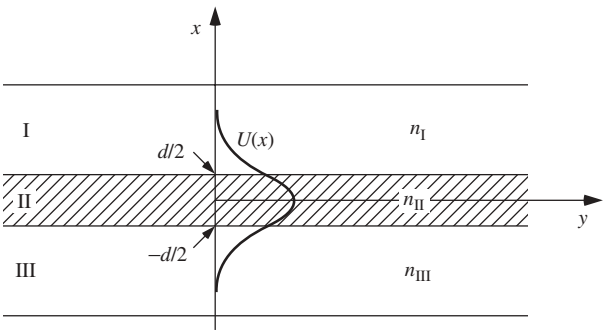


FIGURE A17.1: Schematic of dielectric slab waveguide assuming propagation in the z-direction.

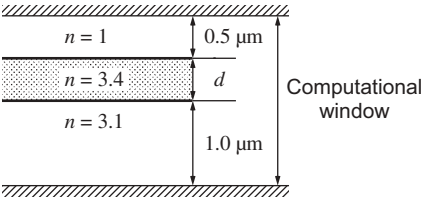


FIGURE A17.2: One-dimensional slab waveguide example.

Figure A17.2 shows an illustrative one-dimensional problem. To demonstrate the importance of the grid size, we solve the problem for several different finite-difference step sizes and two different slab thicknesses. The results are given in Tables A17.1 and A17.2.

TABLE A17.1: Results of Finite Difference Calculation of Slab Waveguide in Fig. A17.2.*

Δx (μm)	Air Grid Pts	Guide Grid Pts	Sub. Grid Pts	Matrix Size	Effective Index	Error (%)
0.25	2	2	4	8 × 8	3.315 23	1.081
0.125	4	4	8	16 × 16	3.292 13	0.376
0.0625	8	8	16	32 × 32	3.283 29	0.107
0.03125	16	16	32	64 × 64	3.280 70	0.028

*Slab waveguide thickness $d = 0.5$ μm (single transverse mode). Exact effective index is $\bar{n}_0 = 3.279\,790$ (for $\lambda_0 = 1.3$ μm).

TABLE A17.2: Results of Finite Difference Calculation of Slab Waveguide in Fig. A17.2.†

Δx (μm)	Air Grid Pts	Guide Grid Pts	Sub. Grid Pts	Matrix Size	Effective Indices	Error (%)
0.25	2	4	4	10 × 10	3.365 83	0.242
					3.272 23	1.234
0.125	4	8	8	20 × 20	3.360 46	0.082
					3.245 07	0.394
0.0625	8	16	16	40 × 40	3.358 49	0.021
					3.235 82	0.108
0.03125	16	32	32	80 × 80	3.357 92	0.006
					3.233 22	0.028

†Slab waveguide thickness $d = 1.0$ μm (two modes allowed). Exact effective indices are $\bar{n}_0 = 3.357\,718$ and $\bar{n}_1 = 3.232\,331$ (for $\lambda_0 = 1.3$ μm).

READING LIST

Flannery BP, Teukolsky SA, Vetterling WT. *Numerical recipes: art of scientific computing*. Cambridge, UK: Cambridge University Press; 1986.

Index

Absorption, 9, 158–62, 178, 241, 249, 580–592
  free carrier, 249, 437
  intervalence band, 81, 437
  modulator, 476–9
  spectrum, 185–7, 221–2, 477
    broadening of, 181–5
    including excitonic effects, 187–90
  in waveguides (*see* Dielectric Waveguides)
Acousto-Optic
  modulator, 305
Active Region(s), 218–38
  bulk, 64, 308
  definition, 46
  design of, 66–70
  double-heterostructure, 46
  number of wells to use in, 66–70
  optimum design of, 66–70
  quantum dot, 179–81
  quantum-well, 66, 308, 309, 615–23
  strained, 66, 201–2, 218–20, 623–7
  surface-to-volume ratio of, 204
  volume, 52–3, 248
Aging of Lasers, 202
Alloy Scattering, 80
Ambipolar Diffusion, 205, 244–5
Amplifiers, 481–3
  Semiconductor Optical, 481
    gain, 483
    noise figure, 483
    wavelength converter, 490
Arrayed Waveguide Grating, 383–4
  demultiplexer, 383
  dispersion relationship, 384
  free-spectral range, 385
  laser, 472
  multiplexer, 383
  number of waveguides, 387
  phase errors, 387
  router, 387
Atomic Orbitals, 6, 611–12
Attenuation Constant
  for antiguide, 439–41
  for waveguide bend, 441–5
Auger Recombination (*also see* Recombination Rate)
  coefficient, 215–17
  in quantum wells, 215–17
  phonon-assisted, 215
  strain-reduced, 218, 245
  threshold energy for, 214, 659–61
Autocorrelation Function, 303–4, 664–5
  relation to frequency spectrum, 303–5

Balanced Detector Pair, 673
Band(s), 5–10, 516–20
  filling of, 529–32
  Light- and heavy-hole, 519, 612, 650–653
  offsets, 16, 538, 652
  quantum-well, 526, 615–23
    with strain, 623–7, 647–56
  shifts due to strain, 625, 650–656
  split-off hole, 612–14, 651

Diode Lasers and Photonic Integrated Circuits, Second Edition.
Larry A. Coldren, Scott W. Corzine, and Milan L. Mašanović.

© 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

Band(s) (*Continued*)
valence and conduction, 6–7, 158–60, 172, 177, 516–21, 580–582, 611–23
Feynman model of, 516–20
Kronig-Penney model of, 516
nonparabolicity of, 216, 520–522, 611
Band-Mixing Effects, 616–27, 645–6
Bandgap Shrinkage, 187–90
Barrier States, 220, 227
Beam-Propagation Method (BPM), 432–4
Birefringent Materials, 686
Bit-Error Rate, 288–90
Blackbody Radiation, 561–2, 584–5
Bloch Functions, 162–3, 165–70, 216, 610–614, 618–23
Boltzmann Approximation, 200, 212, 214–15, 229, 532–7
Bowing Parameter, 17–18
Bragg Condition, 113, 116, 344, 345–8, 593, 595–7
Brillouin Zone, 519

Carrier (Density)
clamping of, 46, 64, 252, 255, 285, 287, 542–4
confinement of, 11, 12, 13
diffusion of, 28, 47, 204, 205–11, 248–9, 537–42
effective diffusion velocity of, 208
fluctuations, 287, 292, 295
gain vs., 64, 223–6
curve fits, 64, 224
large-signal, 279
leakage, 47–8, 75, 248–9, 537–42
drift, 541, 543
temperature dependence of, 80–83, 236–8
lifetime, 49, 257, 283–6, 591
differentia, 257, 28–6, 311
single-mode differential, 258
spontaneous, 591
stimulated, 255
recombination current vs., 48, 227–9, 229–34
curve fits, 224, 231
relation to Fermi energy, 529–37
approximate expressions for, 532–7
in bulk, 532–7
in quantum wells, 530–531
small-signal, 71–3, 261–76
thermionic emission of, 47, 249–50, 538–40
threshold, 46, 59, 65, 69–70, 250
shift due to feedback, 314
transient solution of
small-signal, 266-70
transparency, 9, 64, 218–20, 223–5
temperature dependence of, 81
tuning of lasing frequency via, 135–8
Cavity
active/passive, 56–8, 107–13, 123–41, 452–74, 571–6
cold
linewidth of, 293
response of, 293
enhancement of spontaneous emission, 636–7
external, 107–13, 307
round-trip phase of, 108, 312
four-mirror, 92, 111–13
in-plane, 25, 52, 124
lifetime (*see* Photon/lifetime)
loss, 46
optimum value for net external, 110
multisection, 96, 452–74
optical, 559–61
optimum design of, 64–70
power flow curves in a laser, 572
Q, 109, 259–60
resonance(s), 102–3, 108–11, 112
round-trip time of, 313
three-mirror, 92, 107–11, 312–14
vertical, 25, 52, 124
volume, 52,53, 248
Chaotic Behavior, 325
Charge Neutrality, 47, 177–9, 218–20, 234–8, 242–3, 585, 592
Chemical Beam Epitaxy (CBE), 23
Chirp Parameter, 693
Codirectional Coupler Filter, 370–376
Coherence
collapse, 325
length, 304
time, 304–8
Coherent State, 296, 670
Coherent Receiver (*see* Receiver)
Coherent System, 491–3
Confinement Factor, 52–3, 339, 568–73
axial, 572–3
in active/passive cavities, 573–6
standing wave enhancement of, 573
field, 339
lateral, 555
optimum, 66–70
transverse, 550, 569–70
Constelation Diagram, 492, 493
Correlation Function, 433, 661–4
Coupler
frustrated total internal reflection, 435
grating assisted, 372, 374
multimode interference, 379, 380, 683

Star, 381, 382
Coupling Constant, 121, 344–8, 457, 459, 461, 472, 602
closed-form expression for, 354
for directional couplers, 356–76
for gratings
closed-form expression for, 354
for square wave profile, 121, 354, 602
for sampled gratings, 458–63
Coupling Length, 364
Coupling of Modes, 99
codirectional, 356–70
exact interpretation of, 390
contradirectional, 342–8
exact interpretation of, 113-23
Coupled Mode Theory, 121, 342–70
Covalent Bonds, 6, 511
Critical Angle, 396, 404
Current(s)
clamping of leakage, 542–4
confinement of, 12, 24–30
gain vs., 229–32
curve fits, 231, 234
leakage, 12, 75, 85–6, 539–42
minimization of laser, 64–70
modulation bandwidth dependence on, 72, 265
spontaneous emission, 195–7, 592
stimulated emission, 255, 539–42
temperature dependence of, 31, 80–84
threshold, 45, 60, 65, 66, 70, 76, 252
nonradiative component of, 70
shift due to feedback, 313
transparency, 229–38
vs. carrier density, 48, 227–9, 231–3
curve fits, 224, 231
Curved Waveguide (*see* Dielectric Waveguide(s))

Damping Factor, 73, 262–6, 327
offset, 264, 327
with transport effects, 311
Dangling Bonds, 200, 202, 211
Deep-Level Traps, 200
Defects, 13, 17, 200, 218–20
dark-line, 202
Density of States, 159, 164, 522–7, 531, 581
for different dimensions, 172
optical, 194, 229, 559–61, 585, 567–8
in a resonator, 636
quantum-well, 172, 177–9, 524–7
reduced, 159, 170–173, 185–7, 240–241, 580–582, 588
standing wave and plane wave, 525
Detuning Parameter
for directional couplers, 362
for gratings, 116, 117, 118, 345, 353, 598
Dielectric Constant, 336, 545
Fourier decomposition of, 343, 345
Dielectric Interface
angled, 327
reflection and transmission of, 96–8
for guided modes, 96, 114
Dielectric Perturbations, 336–9, 339–41, 356–60, 675–6
periodic, 342–53, 356–63, 372
Dielectric Waveguide(s), 10, 24–30, 91–2, 395–445, 545–55
analysis of, 400–434
standing wave, 400–403
transverse resonance, 403–10, 438–9
antiguide, 438–40
ARROW, 440–441
bending losses in, 441–5
channel, 418–21, 551–5
curved, 421–4
analysis using conformal mapping, 422
transformed into straight waveguide, 423
cutoff and quasi-modes of, 404–8
five-layer, 390
losses, 437–41
absorption, 437–8
free-carrier, 437
intraband, 437
deep level, 438
radiation, 438
scattering, 438
modes
asymmetric, in curved waveguide, 424
bend, 424
characteristic equation for, 402–3, 407, 414, 547–8
orthogonality between, 336
radiation 438-9
slab, 400–438, 546–50, 551–5, 697–8
multilayer, 408–10
numerical techniques for, 424–7, 427–34, 695–8
parabolic, 414–18
power and effective width of, 336–9
power transferred from one into another, 376–8
radiation losses in, 437–45
separate-confinement heterostructure (SCH), 308–11, 341, 541–4
confinement factor in, 434–5
three-layer, 338, 353, 400–406, 546–50
normalized solutions of, 547–9
WKB method for arbitrary, 410–418
zigzag ray picture of, 338, 401, 413, 417

Differential Quantum Efficiency (*see* Efficiency)
Dipole Moment, 175
Directional Coupler(s), 356–70, 390
  filters and switches, 370–376
  four-port, 363–6
    coupling length of, 364
  grating-assisted, 372–6, 452–67, 470–474
  vertical, 371
Dispersion Curves
  codirectional coupler filter, 373
  coupled waveguide, 370
  grating, 347
  waveguide, 406, 409
Distributed
  Bragg reflector (DBR) (*see* Gratings)
    laser, 92, 123–41, 452–67
  feedback (DFB) laser, 92, 141–51, 353–6
    gain-coupled, 142, 147
    quarter-wave shifted, 142–4
    unshifted, 141–2, 144
    with cleaved mirror, 143, 147
    with HR mirror, 146, 149

Echelle grating, 389
  demultiplexer, 389
  laser, 474
  multiplexer, 389
Edge-Emitting
  cavity (*see* Cavity/in-plane)
  laser (*see* Laser/in-plane)
Effective
  index (*see* Index of Refraction)
  index technique, 418–21, 551–5
  length, 109, 121–3
  mirror (*see* Mirror)
  number of mirror periods, 117–21, 595–7
  width of mode, 336–9, 353–4
Effective Mass, 14, 16, 162, 172, 215–18, 511, 520–521, 611
  conduction band, 613–15
  equations, 615–18, 625
  reduced, 172
  with strain, 218–20, 242–3, 626–7
Efficiency
  collection, 50, 64
  differential quantum, 62–3, 109, 250, 573–6
    in active/passive cavities, 573–6
    from each facet, 106–7, 576
    methods of measuring the, 75–8
    temperature dependence of, 31, 82
  external LED quantum, 51
  internal quantum/injection, 47, 49, 233, 248, 542–4
    methods of measuring the, 75–8
    temperature dependence of, 82
  optical, 249, 250
  power conversion, 79, 274
  radiative, 49–50, 229, 250, 542
  wall-plug, 79, 274
Einstein A and B Coefficients, 162, 174, 582–4
Elastic Moduli, 648
Electro-Optic
  coefficients, 689
  effect, 110, 124, 136, 370–372, 685–92
  modulation, 370–372, 373
  switches, 370–372
  tensor, 685–92
Energy
  bands, 5–7, 516–21, 615–27
    Feynman model of, 516–28
    Kronig-Penney model of, 516
    offset, *see* Band(s) offset
  conservation of, 7, 160, 164–5, 211–18, 580–581
  density, 566, 580
  gap, 13–20
  levels, 5, 6, 7
  uncertainty, 160, 181–5, 582–4, 633
Envelope Function, 160–162, 165–70, 520, 610–611, 615–23, 645–6
  approximation, 610
Excitons, 187–90
Expected Value, 510
External Quantum Efficiency (*see* Efficiency)
Externally Modulated Lasers, 474

Fabry-Perot
  etalon, 100–104
    asymmetric, 103
    scattering matrix of, 101
    symmetric, lossless, 102
    transfer function using Mason's rule, 104–5
    transmission matrix of, 102
  laser, 105–7
  resonator, 403, 406
Far Fields, 555–7
Feedback
  dependence of linewidth on, 300–301, 321–7
  dependence of noise on, 321–7
  effects of, 311–27
  from a cleaved fiber, 311
  from substrate-air interface, 317
  in-phase, 313, 325
  level, 313
    critical, 325

method of measuring, 315
out-of-phase, 325
phase, 311–27
rate, 313
method of measuring, 315
reduction of, 327
regimes of, 321–7
Fermi (*also see* Quasi-Fermi Level)
factors, 160, 175, 192, 212, 215, 588–92
function, 9, 161, 172, 212, 529, 531, 580
levels, 161, 200, 527–37
occupation probability (*see* Fermi function)
Fermi-Dirac
distribution (*see* Fermi function)
integral, 532
Fermi's Golden Rule, 162, 164, 170–173, 174, 175, 187, 215, 629–38
Filter(s)
bandwidth, 291
codirectionally coupled filters, 356–76
directional coupler, 356–70
enhanced tunability of, 373–6 356–76
grating-assisted coupler, 372–6
low-pass, 310–311
narrowband, 289–91
passband, 356–76
tunable, 452–84
wavelength-selective, 356–76
Finite-Difference Technique, 427–32, 693–6
Franz-Keldysh (Electro-Absorption) Effect, 689–92
Frequency
Bragg (*see* Bragg Condition), 113, 116, 344
chirping, 270–276, 286–7, 692–4
noise (*see* Noise/frequency)
normalized, 355, 419–21, 444, 548–9
response (*see* Modulation Response)
shift due to carrier density, 270–271, 320
shift due to current injection, 458–63
shift due to feedback, 318–20
shift due to injection, 317–20
shift due to temperature, 273–6
small-signal modulation of, 272

Gain, 8, 45, 55, 174–90, 585–9
band edge, 180
bandwidth, 109–10, 185
characteristics of, 174–9, 185–7, 585–9
clamping of, 59, 110, 252, 255, 287
compression, 70, 72, 258, 263–6, 268, 310–311, 543, 673, 674
intermodal, 259, 278
method of measuring, 271–6
definition of material, 53, 174–5
dependence on
doping, 218–20, 234–8
temperature, 234–8
well width, 234–8
differential, 64–5, 71–4, 224–5, 258–65, 589
increase due to strain, 218–20
nominal, 258
temperature dependence of, 80
Einstein's approach to calculating, 585–9
in microcavity lasers, 587–9
including lineshape broadening, 180
margin, 143–9
modal, 57, 341, 565–77
peak, 109–10
relation between spontaneous emission, 193–7, 587–9, 590
spectrum, 185–7, 220–2, 588
broadening of, 181–5
modelling of, 277–83
redshift due to bandgap shrinkage, 189
threshold, 45, 55, 59, 76, 251
of DBR lasers, 124–5
shift due to feedback, 319
vs. carrier density, 64–5, 223–6
curve fits, 65, 224, 589
vs. current density, 75–8, 229–32
curve fits, 231, 233
Gap Factor, 112
Gaussian Spot Size, 416, 443, 555
Goos-Hanchen Shift, 434–5
Grating(s), 24, 91–2, 113–23, 342–52, 593–605
arrayed waveguide, (*see* Arrayed waveguide grating)
average loss of, 116, 571
average propagation constant of, 116
coupling constant of, 121, 344–7, 603
closed-form expression for, 354
for square wave profile, 121, 354, 603
decay constant of, 347, 603
detuning parameter, 116, 117–18, 345, 346, 598
diffraction efficiency of, 354
Echelle, (*see* Echelle grating)
effective length of, 121–3, 599
effective mirror model of, 121–3
lossless, 119
lossy, 121
passband, 117
penetration depth of (*see* Penetration Depth)
physical length of, 350
reflectivity, 119–21, 348–52, 596–7
bandwidth of, 350–351
coupled-mode limit of, 352, 602–3

Grating(s) (*Continued*)
Fourier limit of, 121, 601–2
including arbitrary input and output layers, numerical determination of, 123
phase, 349
spectrum, 119–21, 350–351, 597–600
sampled, 458–63
coupling constant of, 459-61
spectral width, 119
stopband, 117, 347, 350–351
transmission, 352
with terminations, 350–351, 603–5
GRINRODs, 418
Group Velocity, 54, 58, 174–5, 194, 249, 251, 271, 434, 567, 576–7

Hamiltonian, 164, 510
Luttinger-Kohn (LK), 615–17, 625, 656
perturbed, 630
Pikus and Bir, 650–652
Heat
diode laser performance, 84
effects of, 66, 72, 86
flow, 79–80
Hermite-Gaussian polynomials, 416, 442–4
Heterodyne Receiver (*see* Receiver)
Hooke's Law, 650
Huygen's Principle, 556

Ideality Factor, 72
Index Ellipsoid, 686–8
Index of Refraction
carrier density dependence of, 58, 136, 269, 270, 302, 691–3
change in, 10–11, 27, 35, 36, 136, 270, 340–341, 371, 374, 452
quadratic, 691–4
relative to change in gain, 270, 292
using the electro-optic effect, 687–94
complex, 270, 546
dispersion of, 58, 347, 370, 373, 406, 409
effective, 58, 97, 114, 425–7, 545–6, 553–4, 556–7
closed-form expression for, 355
change in, 136, 270, 340–341, 371, 374, 452
periodic perturbation of, 344, 358
group, 58, 59, 270, 561, 576–7
group effective, 58, 59, 351, 374, 454, 576–7
ordinary and extraordinary, 688
Injection locking, 317–20
dynamic characteristic, 320
frequency shift, 318–20
carrier induced, 319–20
injected signal caused, 318–19
total, 320
stability, 321–2
static characteristic, 317
Interference Fringe, 304
visibility, 304
Internal Modal Loss (*see* Loss)
Internal Quantum Efficiency (*see* Efficiency)
Intraband Relaxation Time, 182, 582–4, 588

Joyce-Dixon Approximation, 534, 536

K-Factor, 264–6, 325
with transport effects, 311
k-Selection Rule, 165, 166–70, 171–4
exceptions to, 170

Langevin
correlation strengths, 644–71
between photons and carriers, 667
between photons and phase, 670–671
between photons and power, 669
method, 294–6, 299–301, 666–7
noise sources, 294–6, 301–3, 663–74
negatively correlated, 296, 666, 668–9
as white noise, 295, 665
noise spectral densities, 294–6, 663–6
Laser
array, 484
arrayed waveguide grating, 472
cavity examples, 130–132
characterization of, 78-9
proton-implanted, 27–8, 34–9
dielectrically apertured, 28, 34–9
mesa confined, 28, 34–9
coupled-cavity, 111–13
diode, semiconductor, 1, 2–5
distributed Bragg reflector (DBR), 92, 123–41
multisection, 452–7
distributed feedback (DFB), 92, 141–51, 353–6
gain-coupled, 136, 143–7
quarter-wave shifted, 142
unshifted, 141–3
with cleaved mirror, 141–3, 353–6
with high reflectivity coating, 146
Echelle grating, 474
externally modulated, 474
monolithic example DFB, 478
Fabry-Perot, 92, 100–105
gain-guided., 27
gas, solid-state, YAG, HeNe, C02, 1, 2–5
heterostructure

buried (BH), 27, 29
double (DH), 10–13, 308, 539, 541
separate confinement (SCH), 12, 308–11
graded index (GRINSCH), 12
in-plane (IPL), 24
characterization of, 75–8
index-guided, 27
injection locked, 317–20
micro-cavity and thresholdless, 195, 252–3, 587–8, 592
mode locked, 279–83
oxide stripe, 27
proton-implanted, 27
ridge, 27
ring, 366–70
tunable, 92, 113, 124
DBR, 452–574
Digital supermode DBR, 464
double ring-resonator, 467–8
grating coupled sampled reflector, 470
Sampled Grating DBR, 458–61
Superstructure grating DBR, 464
Y-branch SGDBR, 467
Typical parameters for, 260–261
V-groove, removed
vertical-cavity surface-emitting (VCSEL), 28, 29, 34–9, 52, 133–41
Lasing
threshold (*see* Threshold)
transition to, 253–5
wavelength, 58
Lattice
constant, 13–20
deformation potentials, 651–3
matching, 15–20
mismatch parameter, 653
strained, 15–20, 201–2, 218–20
critical thickness of, 202, 218
Lifetime
carrier, 48, 257, 283–6, 591
differential, 257, 283–6, 311
single-mode differential, 257
spontaneous, 591
stimulated, 255
photon/cavity, 53–4, 55, 57, 71, 249–50, 292, 636–8
dependence on feedback, 313
effective, 294
Light-Emitting Diode (LED), 9, 45, 49–51
diffraction grating
shaped
Lineshape
broadening, 181–5, 195–7, 582–4, 586–92
function, 182, 195–7, 582–4, 586–92
Lorentzian, 182
Linewidth
enhancement factor, 271, 292, 307, 320, 692–3
method of measuring, 272–6
narrowing with external cavity, 327
spectral, 288, 292–4, 303–8, 663
dependence on noise, 292–4
method of measuring, 303 - 8
Schawlow-Townes, 292–4
modified, 294, 303, 307
with satellite peaks, 306, 325
Liquid-Phase Epitaxy (LPE), 20–21, 201
Local Oscillator, 491–9
Loss(es)
bend, 441–5
free carrier, 249
in gratings, 116, 122, 571
margin, 110, 128, 140
mirror (*see* Mirror)
modal, 55, 57, 249–50, 570–571
methods of measuring the, 75–8
temperature dependence of, 81
modulation, 110
optical, 45, 249–50
scattering, 98, 438
waveguide radiation, 437–45
Luttinger Parameters, 617–19

Mach-Zehnder Interferometer, 479–81
Many-Body Effects, 187–90
Mason's rule, 104–5, 366–7
Material
parameters, 14
quaternary, 13–20
semi-insulating (SI), 27–30
strained (*also see* Strain), 13–20, 200–202, 218–20
ternary, 13–20
Matrix Elements
Energy (H'_{21}), 165–6, 631–3
Momentum (M), 166–9
Transition (M_T) (*also see* Transition), 165–70, 175, 185–7, 639–46
polarization dependence of, 165, 169–70, 641–6
quantum-well, 645–6
Maxwell-Bolzmnann Distribution, 529, 561
Memoryless Process, 661
Metal-Organic Chemical Vapor Deposition (MOCVD), 20–22, 201–2
Mirror(s)
cleaved, 24, 57

Mirror(s) (*Continued*)
effective, 92, 107–8, 111–13, 121–3, 124–7, 311–13, 355–6
facet damage of, 72
grating (*see* Gratings)
loss, 57, 75, 250
in active/passive cavities, 57, 574–5
in DBR lasers, 126
dependence on feedback, 311–14
modulation of, 258–9, 261
optimum, 67
variation with wavelength, 91–2, 126–8, 140
multilayer (*see* Gratings)
total internal reflection, 435
Modal Excitation, 376–8
Modal Gain, 56–7, 341, 565–77
multiple quantum wells active region, 68
Modal Transition
adiabatic, 422
waveguide offset, 422
Mode(s)
alignment with gain spectrum, 80–81, 192, 256, 278
asymmetric in curved waveguide, 424
external cavity, 316, 319–20
multiple, 325–7
spacing of, 325-7
transition from single to multiple, 325–7
Fabry-Perot, 102–4
feedback, 319–27
hopping, 84–6, 135, 301, 324
impedance, 93, 376–7
internal cavity, 319–20
lasing, 45, 56
locking, 279–83
techniques, 282
pulse width, 282
pulse period, 282
numbers, 559–61
optical, 10–13, 24–30, 52, 55, 159, 194–5, 255–6, 277–83
density of, 194, 229, 559–61, 585, 587–8
in a cavity, 636
quantum mechanical description of, 192
pulling, 314–16
repeat, 109–10
saturation of side, 256
single, 91–2, 110, 136, 139–41, 149–51
tuning, 452–63
transition rates, 159
spacing, 58, 109, 125
suppression of side, 109–10
suppression ratio (MSR)
dynamic, 278
in grating-assisted coupler-laser, 470–472
static, 93, 139–41, 149–51
tuning of (*see* Wavelength/tuning)
waveguide (*see* Dielectric Waveguide)
Modulation
bandwidth, 72, 262–3
ways to maximize the, 72–3
doping, 170, 219, 235–8
electro-optic, 370–372, 373
index
frequency, 271–2
intensity, 272
measuring the FM-to-IM, 271–6
phase, 692–4
differential phase shift keying, 493
quadrature phase shift keying, 493
vector, 499
response, 46
comparison of FM and IM, 272
damping of, 71
FM, 270–276
laser, 70–74, 261–6
LED, 51
low-frequency roll-off of, 310–312
with *RC* parasitics, 330
transfer function, 71, 261–6
natural roots of, 267
Modulator
absorption, 476–9
amplifier, 481–3
directional coupler, 371, 374
Franz-Keldysh, 691–4
laser combined with, 488–91
length, 488–91
Mach-Zehnder, 479
phase, 488–91
Molecular Beam Epitaxy (MBE), 20–22, 201–2
metal-organic (MOMBE), 23
Momentum
conservation of, 7–9, 160, 169, 212, 215, 581, 657–8
density of states, 525
matrix element (*see* Matrix Element)
operator, 163–5, 510, 610, 612, 640
Multimode interference, 378–81, 677–84
component properties, 381
coupler, 379–80, 683
general, 380, 679–81
restricted, 683
self-imaging, 380, 680
symmetric, 379, 681

Near Fields, 555–7
Noise
  carrier, 295, 301–3, 307–8, 671–2
  correlation strengths, 295, 302, 662–69
  current source shot, 296, 298–301, 665, 671
  dependence on feedback, 321–8
  detector, 300–301, 664–5, 671
  excess intensity, 296–8
  field, 303
  figure of optical amplifiers, 483
  floor, 287, 298–301
  frequency, 292, 301–3, 305–6
  inherent quantum, 297
  Langevin (*see* Langevin)
  mode partition, 301
  negative correlation between reflected and transmitted, 295–6
  output power, 296, 670–671
  partition, 298–301, 666
  photon, 295, 689–90
  quadrature components of, 293
  relative intensity (RIN), 288–308
    low frequency, 297
    low power, 297
    multimode, 301
    peak, 298
    per unit bandwidth, 298–9
    spectrum, 288–308
  shot, 296, 299–301, 664–5
  sources, 258–9, 288
  sub-shot, 299–301, 664
  thermal, 299
Noise-Free Operation, 299–301

Obliquity Factor, 556
Operator, 509
Optic Axis, 688
Optical
  amplifier, 481–3
  efficiency, 249, 250
  heterodyne receiver, 496
  isolator, 304, 327
  loss, 45, 249–50
  modes (*see* Modes)
  receiver (*see* Receiver)
Optimum Laser Designs
  for high speeds, 70–74
  for low current, 64–70
Organometallic Vapor-Phase Epitaxy (OMVPE), 20–22
Orthogonality
  of electromagnetic waves, 336
  of particle waves, 169, 512–13
Output Power (*see* Power)

Parameters
  internal laser, 75–9
  material, 14
Pauli Exclusion Principle, 6, 560
Penetration Depth, 118, 351–2
  energy, 123, 435, 599, 605–7
  loss, 600
  phase, 123, 599–600
Perturbation Theory, 339–41, 675–6
Phase
  change in, 315
  noise (*see* Noise)
  round-trip, 313, 315
    change in, 315
Photodiode
  edge coupled, 486
  noise, 487
  quantum efficiency, 486
  responsivity, 485
  uni-traveling carrier, 487–9
  waveguide, 485
Photon (Density)
  confinement of, 10–13, 24–30
  fluctuations, 287, 295
  in terms of the field strength, 175
  large-signal, 279
  lifetime, 53–5, 57, 71, 251, 292, 636–7
    dependence on feedback, 311–13
    effective, 293
  modulation bandwidth dependence on, 71
  recycling, 50
  relationship to output power, 62
  small-signal, 70–71, 261–6, 308–10
  steady-state
    above-threshold, 62, 139, 149–50, 252
    below-threshold, 63, 251
  transient solution of
    small-signal, 268-70
Photonic Integrated Circuits (PICs), 451–99
Piezoelectricity, 190
Population Inversion, 158, 177–9, 185, 590
  factor, 192–3, 590
    curve fit, 253, 256
Potential Well (1-D), 511–16
Power
  dependence of linewidth on, 292–3
  dependence of RIN on, 294–8, 671
  dissipation, 79, 83
  flow curves in a laser cavity, 572
  flowing into a port, 93
  fluctuations, 290
  fraction from each facet, 62, 66, 106–7, 250
    in an active/passive cavity, 576
    in a DBR laser, 126

Power (*Continued*)
modulation bandwidth dependence on, 72
normalized flow of, 92–5
output of laser, 62, 250, 252
propagation in a waveguide, 336–9
saturation in amplifiers, 482–3
shift due to feedback, 314
small-signal, 71, 310–311
spectrum of laser, 307, 665
spontaneous (LED), 49–51, 250, 252
saturation of, 256
vs. current, 60–64, 253–5
derivative analysis of, 84–6
experimental characterization of, 75–8
linearity of, 63, 84–6, 254
Prepatterned substrate, 37–9
Principle Dielectric Axes, 685–6
Propagation
constant, 55,
discrete, 117
grating-induced replicas of, 346–8
numerical determination of, 424–7, 432–4
perturbation of, 339–41
vs. frequency, 347, 370, 373
delay, 99
loss (*see* Attenuation Constant)
vector components, 396–400, 413
Purcell effect, 198

Quantum-Confined
regime, 11, 523
Stark Effect (QCSE), 689–92
Quantum Dots, 179
excitonis effects, 189
gain, 180
Quantum Efficiency (*see* Efficiency)
Quantum Noise Floor, 296
Quantum Numbers, 513–16, 522–3, 526–7
Quantum Well (QW), 12, 511–16
absorption spectrum, 187–8, 221–2
band structure, 615–23
method of calculating, 621–2
characteristic equation for, 512–16
dependence of gain on width of, 234–8
dependence of modal gain on number, 68
gain, 176–9, 185–7, 218–38, 588–9
strained, 15–20, 201–2, 218, 219–20, 623–7
subbands, 169–70, 176–9, 185
Quantum Wires, 163, 169–70, 172, 641–5
Quarter-Wave Dielectric Stacks (*see* Gratings)
Quasi-Fermi Level(s), 161, 176–9, 530, 538–40, 580–582, 585–7, 591–2
clamping of, 543
effect of doping on, 218–20, 235–8
effect of strain on, 218–20
separation, 79, 161, 176–9
required to achieve gain, 161, 176–9, 185–7, 587

Rabi Oscillations, 629, 635, 637
Radiation
spectral density, 174, 582
equivalent for spontaneous emission, 561–2, 585
Rate Coefficients, 258, 259
with transport effects, 309
Rate Equations, 45, 54, 249, 325
carrier, 45, 46–9, 284
above-threshold, 62
differential analysis cif, 257–76
including transport effects, 308–9
large-signal, 277–9
multimode
large-signal, 277–9
small-signal, 259
steady-state, 255-6
numerical solutions to, 277–9
phase, 302
photon, 45, 52–5
small-signal, 45, 257, 258–9, 261–2, 308–9
with Langevin noise driving terms, 294
steady-state, 49, 62, 250–256
Receiver
array, 484
bandwidth, 491–9
coherent
In-phase and quadrature, 493, 497
optical heterodyne, 491–9
tunable, 498
polarization diversity, 498
Recombination Rate
Auger, 9, 55, 70, 211–18, 228–32, 284, 543, 657–59
bandgap dependence of, 213–14
temperature dependence of 58, 215–16
bimolecular, 48, 65, 227–9, 284, 591
clamping of, 62, 64, 252, 255
defect and impurity, 8–9, 55, 199–202
in SCH regions, 541–3
interface, 8–9, 202–11
leakage, 48, 542–4
linear, 200–204
nonradiative, 8–9, 23, 45–6, 48, 70, 199–218, 248–9
radiative, 8–9, 48, 162–5, 192–7, 227–9, 580–583
surface, 8–9

Reference Planes, 52–3, 96, 98, 100, 107–8, 141, 348–52
Reflection
  coefficient, 57, 95, 397, 414
    facet, 327
    of gratings (*see* Gratings)
  of plane waves, 396–400
  phase angle, 398–400, 435
  supression, 327
  total internal, 435–6
    mirrors, 435
    frustrated, 435
Reflectivity
  grating (*see* Gratings)
  mean mirror, 57, 75
  of Fabry-Perot etalon, 102–3
Regrowth, 33–8, 123, 202, 208
Relaxation Resonance
  description of, 70–72
  frequency, 70–74, 224–5, 261–6, 293–4, 305–8
    ways to maximize the, 72–4
    with transport effects, 310
Reservoir
  analogy, 47, 61
  model, 248, 664–5
Resistance
  series, 79, 83, 85
  shunt, 85
Resonant Cavity, 10–12, 56, 292
Resonance Condition, 56, 315
Ring
  filter, 366
  laser, 366–70
  resonator, 366
Ringing Frequency, 266–8
Rotating Wave Approximation, 631
Rowland circle, 384
  radius in an AWG, 386

Scattering
  coefficients, 92–105
    relations between, 94–6
  junctions, 93, 95
    dielectric interface, 96–8
    Fabry-Perot etalon, 100–104
    transmission line, 98–100
  loss, 98, 438
  matrices, 91, 94
    four-port, 363–5
  theory, 92–5
Schrödinger's Equation, 162–3, 510, 609, 630
  normal mode expansion of, 516–20
  time-dependent expansion of, 630–632
Section
  active, 57
  passive, 57
Self-Heterodyne Technique, 304
Self-Imaging Length, 680
Self-Pulsations, 322
Semiconductor Amplifier, 481–3
Shockley-Read-Hall Recombination Theory, 200–201, 203
Shot Noise Floor, 295–301
Signal-to-Noise Ratio (SNR), 288–9
Single Frequency Operation, 92, 123, 128, 139, 141, 149, 452–75
Snell's Law, 396–7
Spectral Density
  carrier, 295, 302
  current, 297
  definition of, 290
  frequency, 306
  Langevin noise (*see* Langevin)
  output power, 296
  photon, 295
  single-sided and double-sided, 290
  units of, 291
Spin Degeneracy, 6, 172, 524, 560, 639–41, 645
Spin-Orbit
  energy, 612–13, 649, 659
  interaction, 613
Splitter
  frustrated total internal reflection, 437
  multimode interference, 379
Spontaneous Emission, 8, 45, 48–50, 55, 158–62, 192–9, 227–9, 248–9, 579–92
  amplified, 293–4
  as a noise source, 159–60, 292–3, 301–2, 307
  Einstein's approach to calculating, 589–92
  factor, 53, 139–41, 149–51, 198, 248, 250, 252–3, 562–3
  in microcavities, 195, 587–8, 636–7
  lifetime, 591
  relation between gain and, 193–7, 587–8, 590
  relation between stimulated emission and, 161–2, 192, 585
  single-mode, 159–62, 192–3, 250, 253
    curve fit, 253, 261
  spectrum, 195–7, 227, 562–3
    broadening of, 182, 195–7, 590
  using the metal box assumption, 194–5
Standard Quantum Limit, 296, 298–301
Standing Wave Effects, 58, 567–73
Stationary Process, 665
Step Response (*see* Transient Response)

Stimulated Emission, 8, 45, 48–50, 54, 58–9, 70–71, 158–62, 248–9, 292–3, 579–92
  net rate of, 161
  relation between gain and, 54, 174–6
  relation between spontaneous emission and, 161, 192, 585
Stimulated Rate Constant, 161, 174, 582–5
Strain, 16–20, 64, 201–2, 218, 219–20, 623–7, 647–56
  band edge shifts created by, 623–5, 650–656
  biaxial, 649
  compressive and tensile, 16–20, 218–20, 221–30, 623–7, 649
  critical thickness of, 16–20, 202
  effect on effective masses, 218–20, 242–4, 623–7
  effect on performance of active material, 218–38
  hydrostatic and shear components of, 623–5, 651–3
  tensor, 647–50
Stress
  normal and shear components of, 647–50
  tensor, 647–50
Surface
  passivation, 210–211
  recombination (*see* Recombination Rate)
    equivalent velocity for diffusion, 208–10, 243–5
    velocity, 202–10, 243–5
  states, 202

Table of
  approximations for N and E_F, 533, 534
  density of states for different dimensions, 172
  gain vs. carrier density curve fits, 224
  gain vs. current density curve fits, 231
  material parameters, 14
  strain parameters, 655
  transmission and scattering matrices, 96, 97, 100
  typical laser parameter values, 260–261
Temperature
  characteristic, 81, 237–8
  dependence of laser parameters on, 80–84, 237–8
  rise of, 79
  tuning, 271–6
Thermal
  conductivity, 79
  effect on lasing frequency, 271–6
  impedance, 79–80, 274
  time constant, 274
Thermionic Emission, 47, 249, 538–41
Threshold
  carrier density, 45, 59, 65–6, 70, 251
    shift due to feedback, 314
  characteristic equation defining, 57, 355–6
    for DFB lasers, 141–3, 353–4, 355–6
    for three-mirror cavity, 108
  condition in active/passive cavities, 57, 574–6
  current, 45, 60–62, 65, 66, 70, 252
    nonradiative component of, 70
    shift due to feedback, 314
  definition of lasing, 56, 106
    in DBR lasers, 125
    in DFB lasers, 141–3
  gain, 45, 57, 59, 75–6, 251
  numerical determination of, 106
    in DFB lasers, 141–3, 353–4, 355–6
Threshold Energy (*see* Auger Recombination)
Transient Response
  small-signal, 266–70
Transition(s)
  allowed and forbidden, 165–70, 188
  Auger, 213
  matrix element (*see* Matrix Element)
  nonradiative, 199–218
  radiative, 8, 158–73
    band-to-band, 112, 123, 158, 170
    band-to-bound and bound-to-bound, 170
    band-to-exciton, 187
    rate of, 162–5
    relations between, 161
Transmission
  coefficient, 98, 390
    for Fabry-Perot etalon, 100–104
  line, 98–100
  matrices, 91, 94–105
    applied to gratings., 115–21
    multiplying, 94, 100, 102, 113–15
    relations between coefficients, 96, 593–7
Transparency
  carrier density, 9, 55–6, 64, 219, 220, 223–5
    temperature dependence of, 80–81
  condition, 185–7
  current, 229–38
Transport
  effects, 70, 72, 308–11
  factor, 310
Transverse
  electric (TE) and magnetic (TM) modes, 166
    coupling between, 359–62
    reflection coefficient for, 397–8
Transceiver, 488–91

Triplexer, 491–2
Tunable
  filters, 370–376, 452–7
  Lasers (*see* Lasers)
Turn-On Delay, 278, 283–7
Two-Port Networks, 93–5
  cascading, 95, 99, 101–2, 114–15, 117
  properties of, 97
  reciprocal, 97

Uncertainty Principle, 509
Unger Approximation, 533–7

Vacuum Fields, 158–62, 192–3, 636–7
Vector Modulation, 494
Vector Potential, 164
  in terms of the electric field, 175
Vegard's Law, 16–20
Vertical Cavity
  examples, 130–132
  geometry of, 25, 28, 52
Voltage
  ideal diode, 79–80
  series, 79–80
  threshold, 79–80, 274
    vs. current derivative analysis of, 84–6

Wall-Plug Efficiency (*see* Efficiency)
Wave Equation, 6
  electromagnetic, 336, 545
  electron, 510
  numerical solutions of, 424–7
  scalar, 411, 426, 695
  transverse, 336, 411, 677
Wavefunction, 162–3, 164, 509, 609–11
  quantum-well, 169
Waveguide (*see* Dielectric Waveguides)
Waveguide losses (*see* Dielectric Waveguide Losses)
Wavelength
  Bragg (*see* Bragg Condition)
  chirp, 270–276, 286–7
    maximum transmission distance, 475
  division multiplexing (WDM), 484
  fine tuning and channel changing of, 452–72
  modal, 58
  spacing between modes (*see* Mode)
  tuning, 92, 110, 113,124, 127–8, 135–5, 270–276, 452–72
WKB approximation, 411–16
Wiener-Khinchin Theorem, 665
Widely Tunable Lasers (*see* Lasers)
Wigner-Weisskopf Equations, 631

WILEY SERIES IN MICROWAVE AND OPTICAL ENGINEERING

KAI CHANG, Editor
Texas A&M University

FIBER-OPTIC COMMUNICATION SYSTEMS, Fourth Edition • *Govind P. Agrawal*

ASYMMETRIC PASSIVE COMPONENTS IN MICROWAVE INTEGRATED CIRCUITS • *Hee-Ran Ahn*

COHERENT OPTICAL COMMUNICATIONS SYSTEMS • *Silvello Betti, Giancarlo De Marchis, and Eugenio Iannone*

PHASED ARRAY ANTENNAS: FLOQUET ANALYSIS, SYNTHESIS, BFNs, AND ACTIVE ARRAY SYSTEMS • *Arun K. Bhattacharyya*

HIGH-FREQUENCY ELECTROMAGNETIC TECHNIQUES: RECENT ADVANCES AND APPLICATIONS • *Asoke K. Bhattacharyya*

RADIO PROPAGATION AND ADAPTIVE ANTENNAS FOR WIRELESS COMMUNICATION LINKS: TERRESTRIAL, ATMOSPHERIC, AND IONOSPHERIC • *Nathan Blaunstein and Christos G. Christodoulou*

COMPUTATIONAL METHODS FOR ELECTROMAGNETICS AND MICROWAVES • *Richard C. Booton, Jr.*

ELECTROMAGNETIC SHIELDING • *Salvatore Celozzi, Rodolfo Araneo, and Giampiero Lovat*

MICROWAVE RING CIRCUITS AND ANTENNAS • *Kai Chang*

MICROWAVE SOLID-STATE CIRCUITS AND APPLICATIONS • *Kai Chang*

RF AND MICROWAVE WIRELESS SYSTEMS • *Kai Chang*

RF AND MICROWAVE CIRCUIT AND COMPONENT DESIGN FOR WIRELESS SYSTEMS • *Kai Chang, Inder Bahl, and Vijay Nair*

MICROWAVE RING CIRCUITS AND RELATED STRUCTURES, Second Edition • *Kai Chang and Lung-Hwa Hsieh*

MULTIRESOLUTION TIME DOMAIN SCHEME FOR ELECTROMAGNETIC ENGINEERING • *Yinchao Chen, Qunsheng Cao, and Raj Mittra*

DIODE LASERS AND PHOTONIC INTEGRATED CIRCUITS, Second Edition • *Larry Coldren, Scott Corzine, and Milan Masanovic*

EM DETECTION OF CONCEALED TARGETS • *David J. Daniels*

RADIO FREQUENCY CIRCUIT DESIGN • *W. Alan Davis and Krishna Agarwal*

RADIO FREQUENCY CIRCUIT DESIGN, Second Edition • *W. Alan Davis*

MULTICONDUCTOR TRANSMISSION-LINE STRUCTURES: MODAL ANALYSIS TECHNIQUES • *J. A. Brandão Faria*

PHASED ARRAY-BASED SYSTEMS AND APPLICATIONS • *Nick Fourikis*

SOLAR CELLS AND THEIR APPLICATIONS, Second Edition • *Lewis M. Fraas and Larry D. Partain*

FUNDAMENTALS OF MICROWAVE TRANSMISSION LINES • *Jon C. Freeman*

OPTICAL SEMICONDUCTOR DEVICES • *Mitsuo Fukuda*

MICROSTRIP CIRCUITS • *Fred Gardiol*

HIGH-SPEED VLSI INTERCONNECTIONS, Second Edition • *Ashok K. Goel*

FUNDAMENTALS OF WAVELETS: THEORY, ALGORITHMS, AND APPLICATIONS, Second Edition • *Jaideva C. Goswami and Andrew K. Chan*

HIGH-FREQUENCY ANALOG INTEGRATED CIRCUIT DESIGN • *Ravender Goyal (ed.)*

RF AND MICROWAVE TRANSMITTER DESIGN • *Andrei Grebennikov*

ANALYSIS AND DESIGN OF INTEGRATED CIRCUIT ANTENNA MODULES • *K. C. Gupta and Peter S. Hall*

PHASED ARRAY ANTENNAS, Second Edition • *R. C. Hansen*

STRIPLINE CIRCULATORS • *Joseph Helszajn*

THE STRIPLINE CIRCULATOR: THEORY AND PRACTICE • *Joseph Helszajn*

LOCALIZED WAVES • *Hugo E. Hernández-Figueroa, Michel Zamboni-Rached, and Erasmo Recami (eds.)*

MICROSTRIP FILTERS FOR RF/MICROWAVE APPLICATIONS, Second Edition • *Jia-Sheng Hong*

MICROWAVE APPROACH TO HIGHLY IRREGULAR FIBER OPTICS • *Huang Hung-Chia*

NONLINEAR OPTICAL COMMUNICATION NETWORKS • *Eugenio Iannone, Francesco Matera, Antonio Mecozzi, and Marina Settembre*

FINITE ELEMENT SOFTWARE FOR MICROWAVE ENGINEERING • *Tatsuo Itoh, Giuseppe Pelosi, and Peter P. Silvester (eds.)*

INFRARED TECHNOLOGY: APPLICATIONS TO ELECTROOPTICS, PHOTONIC DEVICES, AND SENSORS • *A. R. Jha*

SUPERCONDUCTOR TECHNOLOGY: APPLICATIONS TO MICROWAVE, ELECTRO-OPTICS, ELECTRICAL MACHINES, AND PROPULSION SYSTEMS • *A. R. Jha*

TIME AND FREQUENCY DOMAIN SOLUTIONS OF EM PROBLEMS USING INTEGTRAL EQUATIONS AND A HYBRID METHODOLOGY • *B. H. Jung, T. K. Sarkar, S. W. Ting, Y. Zhang, Z. Mei, Z. Ji, M. Yuan, A. De, M. Salazar-Palma, and S. M. Rao*

OPTICAL COMPUTING: AN INTRODUCTION • *M. A. Karim and A. S. S. Awwal*

INTRODUCTION TO ELECTROMAGNETIC AND MICROWAVE ENGINEERING • *Paul R. Karmel, Gabriel D. Colef, and Raymond L. Camisa*

MILLIMETER WAVE OPTICAL DIELECTRIC INTEGRATED GUIDES AND CIRCUITS • *Shiban K. Koul*

ADVANCED INTEGRATED COMMUNICATION MICROSYSTEMS • *Joy Laskar, Sudipto Chakraborty, Manos Tentzeris, Franklin Bien, and Anh-Vu Pham*

MICROWAVE DEVICES, CIRCUITS AND THEIR INTERACTION • *Charles A. Lee and G. Conrad Dalman*

ADVANCES IN MICROSTRIP AND PRINTED ANTENNAS • *Kai-Fong Lee and Wei Chen (eds.)*

SPHEROIDAL WAVE FUNCTIONS IN ELECTROMAGNETIC THEORY • *Le-Wei Li, Xiao-Kang Kang, and Mook-Seng Leong*

ARITHMETIC AND LOGIC IN COMPUTER SYSTEMS • *Mi Lu*

OPTICAL FILTER DESIGN AND ANALYSIS: A SIGNAL PROCESSING APPROACH • *Christi K. Madsen and Jian H. Zhao*

THEORY AND PRACTICE OF INFRARED TECHNOLOGY FOR NONDESTRUCTIVE TESTING • *Xavier P. V. Maldague*

METAMATERIALS WITH NEGATIVE PARAMETERS: THEORY, DESIGN, AND MICROWAVE APPLICATIONS • Ricardo Marqués, Ferran Martín, and Mario Sorolla

OPTOELECTRONIC PACKAGING • *A. R. Mickelson, N. R. Basavanhally, and Y. C. Lee (eds.)*

OPTICAL CHARACTER RECOGNITION • *Shunji Mori, Hirobumi Nishida, and Hiromitsu Yamada*

ANTENNAS FOR RADAR AND COMMUNICATIONS: A POLARIMETRIC APPROACH • *Harold Mott*

INTEGRATED ACTIVE ANTENNAS AND SPATIAL POWER COMBINING • *Julio A. Navarro and Kai Chang*

ANALYSIS METHODS FOR RF, MICROWAVE, AND MILLIMETER-WAVE PLANAR TRANSMISSION LINE STRUCTURES • *Cam Nguyen*

LASER DIODES AND THEIR APPLICATIONS TO COMMUNICATIONS AND INFORMATION PROCESSING • *Takahiro Numai*

FREQUENCY CONTROL OF SEMICONDUCTOR LASERS • *Motoichi Ohtsu (ed.)*

INVERSE SYNTHETIC APERTURE RADAR IMAGING WITH MATLAB ALGORITHMS • *Caner Özdemir*

SILICA OPTICAL FIBER TECHNOLOGY FOR DEVICE AND COMPONENTS: DESIGN, FABRICATION, AND INTERNATIONAL STANDARDS • *Un-Chul Paek and Kyunghwan Oh*

WAVELETS IN ELECTROMAGNETICS AND DEVICE MODELING • *George W. Pan*

OPTICAL SWITCHING • Georgios Papadimitriou, Chrisoula Papazoglou, and Andreas S. Pomportsis

MICROWAVE IMAGING • *Matteo Pastorino*

ANALYSIS OF MULTICONDUCTOR TRANSMISSION LINES • *Clayton R. Paul*

INTRODUCTION TO ELECTROMAGNETIC COMPATIBILITY, Second Edition • *Clayton R. Paul*

ADAPTIVE OPTICS FOR VISION SCIENCE: PRINCIPLES, PRACTICES, DESIGN AND APPLICATIONS • *Jason Porter, Hope Queener, Julianna Lin, Karen Thorn, and Abdul Awwal (eds.)*

ELECTROMAGNETIC OPTIMIZATION BY GENETIC ALGORITHMS • *Yahya Rahmat-Samii and Eric Michielssen (eds.)*

INTRODUCTION TO HIGH-SPEED ELECTRONICS AND OPTOELECTRONICS • *Leonard M. Riaziat*

NEW FRONTIERS IN MEDICAL DEVICE TECHNOLOGY • *Arye Rosen and Harel Rosen (eds.)*

ELECTROMAGNETIC PROPAGATION IN MULTI-MODE RANDOM MEDIA • *Harrison E. Rowe*

ELECTROMAGNETIC PROPAGATION IN ONE-DIMENSIONAL RANDOM MEDIA • *Harrison E. Rowe*

HISTORY OF WIRELESS • *Tapan K. Sarkar, Robert J. Mailloux, Arthur A. Oliner, Magdalena Salazar-Palma, and Dipak L. Sengupta*

PHYSICS OF MULTIANTENNA SYSTEMS AND BROADBAND PROCESSING • *Tapan K. Sarkar, Magdalena Salazar-Palma, and Eric L. Mokole*

SMART ANTENNAS • *Tapan K. Sarkar, Michael C. Wicks, Magdalena Salazar-Palma, and Robert J. Bonneau*

NONLINEAR OPTICS • *E. G. Sauter*

APPLIED ELECTROMAGNETICS AND ELECTROMAGNETIC COMPATIBILITY • *Dipak L. Sengupta and Valdis V. Liepa*

COPLANAR WAVEGUIDE CIRCUITS, COMPONENTS, AND SYSTEMS • *Rainee N. Simons*

ELECTROMAGNETIC FIELDS IN UNCONVENTIONAL MATERIALS AND STRUCTURES • *Onkar N. Singh and Akhlesh Lakhtakia (eds.)*

ANALYSIS AND DESIGN OF AUTONOMOUS MICROWAVE CIRCUITS • *Almudena Suárez*

ELECTRON BEAMS AND MICROWAVE VACUUM ELECTRONICS • *Shulim E. Tsimring*

FUNDAMENTALS OF GLOBAL POSITIONING SYSTEM RECEIVERS: A SOFTWARE APPROACH, Second Edition • *James Bao-yen Tsui*

SUBSURFACE SENSING • *Ahmet S. Turk, A. Koksal Hocaoglu, and Alexey A. Vertiy (eds.)*

RF/MICROWAVE INTERACTION WITH BIOLOGICAL TISSUES • *André Vander Vorst, Arye Rosen, and Youji Kotsuka*

InP-BASED MATERIALS AND DEVICES: PHYSICS AND TECHNOLOGY • *Osamu Wada and Hideki Hasegawa (eds.)*

COMPACT AND BROADBAND MICROSTRIP ANTENNAS • *Kin-Lu Wong*

DESIGN OF NONPLANAR MICROSTRIP ANTENNAS AND TRANSMISSION LINES • *Kin-Lu Wong*

PLANAR ANTENNAS FOR WIRELESS COMMUNICATIONS • *Kin-Lu Wong*

FREQUENCY SELECTIVE SURFACE AND GRID ARRAY • *T. K. Wu (ed.)*

ACTIVE AND QUASI-OPTICAL ARRAYS FOR SOLID-STATE POWER COMBINING • *Robert A. York and Zoya B. Popovic (eds.)*

OPTICAL SIGNAL PROCESSING, COMPUTING AND NEURAL NETWORKS • *Francis T. S. Yu and Suganda Jutamulia*

ELECTROMAGNETIC SIMULATION TECHNIQUES BASED ON THE FDTD METHOD • *Wenhua Yu, Xiaoling Yang, Yongjun Liu, and Raj Mittra*

SiGe, GaAs, AND InP HETEROJUNCTION BIPOLAR TRANSISTORS • *Jiann Yuan*

PARALLEL SOLUTION OF INTEGRAL EQUATION-BASED EM PROBLEMS • *Yu Zhang and Tapan K. Sarkar*

ELECTRODYNAMICS OF SOLIDS AND MICROWAVE SUPERCONDUCTIVITY • *Shu-Ang Zhou*

MICROWAVE BANDPASS FILTERS FOR WIDEBAND COMMUNICATIONS • *Lei Zhu, Sheng Sun, and Rui Li*