Semiconductor Physics and Devices

Basic Principles

Third Edition

Donald A. Neamen
University of New Mexico



Boston Burr Ridge. IL Dubuque. IA Madison WI New York San Francisco St Louis
Bangkok Bogota Caracas Kuala Lumpur Lisbon London Madrid Mexico City
Milan Montreal New Delhi Santiago Seoul Singapore Sydney Taipei Toronto

Page 1 of 566 INNOSCIENCE EXHIBIT 1015

McGraw-Hill Higher- Education
A Division of The *McGraw-Hill* Companies

SEMICONDUCTOR PHYSICS AND DEVICES: BASIC PRINCIPLES
THIRD EDITION

Published by McGraw-Hill, a business unit of The McGraw-Hill Companies, Inc., 1221 Avenue of the Americas, New York, NY 10020. Copyright © 2003, 1997, 1992 by The McGraw-Hill Companies, Inc. All rights reserved. No part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written consent of The McGraw-Hill Companies, Inc., including, but not limited to, in any network or other electronic storage or transmission, or broadcast for distance learning.

Some ancillaries, including electronic and print components, may not be available to customers outside the United States.

This book is printed on acid-free paper.

International 1 2 3 4 5 6 7 8 9 0 DOC/DOC 0 9 8 7 6 5 4 3 2
Domestic 2 3 4 5 6 7 8 9 0 DOC/DOC 0 9 8 7 6 5 4 3

ISBN 0–07–232107–5
ISBN 0–07–119862–8 (ISE)

Publisher: *Elizabeth A. Jones*
Senior developmental editor: *Kelley Butcher*
Executive marketing manager: *John Wannemacher*
Project manager: *Joyce Waiters*
Production supervisor: *Sherry L. Kane*
Designer: *David W. Hash*
Cover designer: *Rokusek Design*
Cover image: *©Eyewire, Inc.*
Media project manager: *Sandra M. Schnee*
Media technology senior producer: *Phillip Meek*
Compositor: *Interactive Composition Corporation*
Typeface: *10/12 Times Roman*
Printer: *R. R. Donnelley/Crawfordsville, IN*

Library of Congress Cataloging-in-Publication Data

Neamen, Donald A.
Semiconductor physics and devices : basic principles / Donald A. Neamen. — 3rd ed.
p. cm.
Includes bibliographical references and index.
ISBN 0–07–232107–5 (acid-free paper)
I. Semiconductors. I. Title.

QC611 .N39 2003
537.6'22—dc21 2002019681
CIP

INTERNATIONAL EDITION ISBN 0–07–119862–8
Copyright © 2003. Exclusive rights by The McGraw-Hill Companies, Inc., for manufacture and export. This book cannot be re-exported from the country to which it is sold by McGraw-Hill. The International Edition is not available in North America.

www.mhhe.com

ABOUT THE AUTHOR

Donald A. Neamen is a professor emeritus in the Department of Electrical and Computer Engineering at the University of New Mexico where he taught for more than 25 years. He received his Ph.D. from the University of New Mexico and then became an electronics engineer at the Solid State Sciences Laboratory at Hanscom Air Force Base. In 1976, he joined the faculty in the EECE department at the University of New Mexico, where he specialized in teaching semiconductor physics and devices courses and electronic circuits courses. He is still a part-time instructor in the department.

In 1980, Professor Neamen received the Outstanding Teacher Award for the University of New Mexico. In 1983 and 1985, he was recognized as Outstanding Teacher in the College of Engineering by Tau Beta Pi. In 1990, and each year from 1994 through 2001, he received the Faculty Recognition Award, presented by graduating EECE students. He was also honored with the Teaching Excellence Award in the College of Engineering in 1994.

In addition to his teaching, Professor Neamen served as Associate Chair of the EECE department for several years and has also worked in industry with Martin Marietta, Sandia National Laboratories, and Raytheon Company. He has published many papers and is the author of *Electronic Circuit Analysis and Design*, 2nd edition.

CONTENTS IN BRIEF

Preface xi

Chapter 1 The Crystal Structure of Solids 1

Chapter 2 Introduction to Quantum Mechanics 24

Chapter 3 Introduction to the Quantum Theory of Solids 56

Chapter 4 The Semiconductor in Equilibrium 103

Chapter 5 Carrier Transport Phenomena 154

Chapter 6 Nonequilibrium Excess Carriers in Semiconductors 189

Chapter 7 The pn Junction 238

Chapter 8 The pn Junction Diode 268

Chapter 9 Metal–Semiconductor and Semiconductor Heterojunctions 326

Chapter 10 The Bipolar Transistor 367

Chapter 11 Fundamentals of the Metal–Oxide–Semiconductor Field-Effect Transistor 449

Chapter 12 Metal–Oxide–SemiconductorField-Effect Transistor: Additional Concepts 523

Chapter 13 The Junction Field-Effect Transistor 570

Chapter 14 Optical Devices 617

Chapter 15 Semiconductor Power Devices 668

Appendix A Selected List of Symbols 703

Appendix B System of Units, Conversion Factors, and General Constants 711

Appendix C The Periodic Table 715

Appendix D The Error Function 717

Appendix E "Derivation" of Schrodinger's Wave Equation 719

Appendix F Unit of Energy—The Electron-Volt 721

Appendix G Answers to Selected Problems 723

Index 731

CONTENTS

Preface xi

CHAPTER 1
The Crystal Structure of Solids 1

Preview 1
1.1 Semiconductor Materials 1
1.2 Types of Solids 2
1.3 Space Lattices 3
1.3.1 Primitive and Unit Cell 3
1.3.2 Basic Crystal Structures 4
1.3.3 Crystal Planes and Miller Indices 5
1.3.4 The Diamond Structure 9
1.4 Atomic Bonding 11
***1.5** Imperfections and Impurities in Solids 13*
1.5.1 Imperfections in Solids 13
1.5.2 Impurities in Solids 15
***1.6** Growth of Semiconductor Materials 16*
1.6.1 Growth from a Melt 16
1.6.2 Epitaxial Growth 18
1.7 Summary 19
Problems 21

CHAPTER 2
Introduction to Quantum Mechanics 24

Preview 24
2.1 Principles of Quantum Mechanics 25
2.1.1 Energy Quanta 25
2.1.2 Wave–Particle Duality 26
2.1.3 The Uncertainty Principle 29
2.2 Schrodinger's Wave Equation 30
2.2.1 The Wave Equation 30
2.2.2 Physical Meaning of the Wave Function 32
2.2.3 Boundary Conditions 32
2.3 Applications of Schrodinger's Wave Equation 33
2.3.1 Electron in Free Space 33
2.3.2 The Infinite Potential Well 34
2.3.3 The Step Potential Function 38
2.3.4 The Potential Barrier 42
***2.4** Extensions of the Wave Theory to Atoms 45*
2.4.1 The One-Electron Atom 45
2.4.2 The Periodic Table 48
2.5 Summary 50
Problems 51

CHAPTER 3
Introduction to the Quantum Theory of Solids *56*

Preview 56
3.1 Allowed and Forbidden Energy Bands 57
3.1.1 Formation of Energy Bands 57
**3.1.2 The Kronig–Penney Model 61*
3.1.3 The k-Space Diagram 66
3.2 Electrical Conduction in Solids 70
3.2.1 The Energy Band and the Bond Model 70
3.2.2 Drift Current 72
3.2.3 Electron Effective Mass 73
3.2.4 Concept of the Hole 76
3.2.5 Metals, Insulators, and Semiconductors 78
3.3 Extension to Three Dimensions 80
3.3.1 The k-Space Diagrams of Si and GaAs 81
3.3.2 Additional Effective Mass Concepts 82
3.4 Density of States Function 83
3.4.1 Mathematical Derivation 83
3.4.2 Extension to Semiconductors 86
3.5 Statistical Mechanics 88
3.5.1 Statistical Laws 88

3.5.2 The Fermi–Dirac Probability Function 89
3.5.3 The Distribution Function and the Fermi Energy 91
3.6 Summary 96
Problems 98

CHAPTER 4
The Semiconductor in Equilibrium 103

Preview 103
4.1 Charge Carriers in Semiconductors 104
4.1.1 Equilibrium Distribution of Electrons and Holes 104
4.1.2 The n_0 and p, Equations 106
4.1.3 The Intrinsic Carrier Concentration 110
4.1.4 The Intrinsic Fermi-Level Position 113
4.2 Dopant Atoms and Energy Levels 115
4.2.1 Qualitative Description 115
4.2.2 Ionization Energy 117
4.2.3 Group III–V Semiconductors 119
4.3 The Extrinsic Semiconductor 120
4.3.1 Equilibrium Distribution of Electrons and Holes 121
4.3.2 The n_0p_0 Product 124
**4.3.3 The Fermi–Dirac Integral 125*
4.3.4 Degenerate and Nondegenerate Semiconductors 127
4.4 Statistics of Donors and Acceptors 128
4.4.1 Probability Function 128
4.4.2 Complete Ionization and Freeze-Out 129
4.5 Charge Neutrality 132
4.5.1 compensated Semiconductors 133
4.5.2 Equilibrium Electron and Hole Concentrations 133
4.6 Position of Fermi Energy Level 139
4.6.1 Mathematical Derivation 139
4.6.2 Variation of E_F with Doping Concentration and Temperature 142
4.6.3 Relevance of the Fermi Energy 144
4.7 Summary 145
Problems 148

CHAPTER 5
Carrier Transport Phenomena 154

Preview 154
5.1 Carrier Drift 154
5.1.1 Drift Current Density 155
5.1.2 Mobility Effects 157
5.1.3 Conductivity 162
5.1.4 Velocity Saturation 167
5.2 Carrier Diffusion 169
5.2.1 Diffusion Current Density 170
5.2.2 Total Current Density 173
5.3 Graded Impurity Distribution 173
5.3.1 Induced Electric Field 174
5.3.2 The Einstein Relation 176
***5.4** The Hall Effect 177
5.5 Summary 180
Problems 182

CHAPTER 6
Nonequilibrium Excess Carriers in Semiconductors 189

Preview 189
6.1 Carrier Generation and Recombination 190
6.1.1 The Semiconductor in Equilibrium 190
6.1.2 Excess Carrier Generation and Recombination 191
6.2 Characteristics of Excess Carriers 194
6.2.1 Continuity Equations 195
6.2.2 Time-Dependent Diffusion Equations 196
6.3 Ambipolar Transport 197
6.3.1 Derivation of the Ambipolar Transport Equation 198
6.3.2 Limits of Extrinsic Doping and Low Injection 200
6.3.3 Applications of the Ambipolar Transport Equation 203
6.3.4 Dielectric Relaxation Time Constant 211
**6.3.5 Haynes–Shockley Experiment 213*
6.4 Quasi-Fermi Energy Levels 216

*6.5 Excess-Carrier Lifetime 218
6.5.1 *Shockley–Read–Hall Theory of Recombination* 219
6.5.2 *Limits of Extrinsic Doping and Low Injection* 222
*6.6 Surface Effects 224
6.6.1 *Surface States* 224
6.6.2 *Surface Recombination Velocity* 226
6.7 Summary 229
Problems 231

CHAPTER 7
The pn Junction 238

Preview 238
7.1 Basic Structure of the pn Junction 238
7.2 Zero Applied Bias 240
7.2.1 *Built-in Potential Barrier* 240
7.2.2 *Electric Field* 242
7.2.3 *Space Charge Width* 246
7.3 Reverse Applied Bias 247
7.3.1 *Space Charge Width and Electric Field* 248
7.3.2 *Junction Capacitance* 251
7.3.3 *One-Sided Junctions* 253
*7.4 Nonuniformly Doped Junctions 255
7.4.1 *Linearly Graded Junction* 255
7.4.2 *Hyperabrupt Junctions* 258
7.5 Summary 260
Problems 262

CHAPTER 8
The pn Junction Diode 268

Preview 268
8.1 pn Junction Current 269
8.1.1 *Qualitative Description of Charge Flow in a pn Junction* 269
8.1.2 *Ideal Current–Voltage Relationship* 270
8.1.3 *Boundary Conditions* 271
8.1.4 *Minority Carrier Distribution* 275
8.1.5 *Ideal pn Junction Current* 277
8.1.6 *Summary of Physics* 281
8.1.7 *Temperature Effects* 284
8.1.8 *The "Short" Diode* 284
8.2 Small-Signal Model of the pn Junction 286
8.2.1 *Diffusion Resistance* 286
8.2.2 *Small-Signal Admittance* 288
8.2.3 *Equivalent Circuit* 295
8.3 Generation–Recombination Currents 297
8.3.1 *Reverse-Bias Generation Current* 297
8.3.2 *Forward-Bias Recombination Current* 300
8.3.3 *Total Forward-Bias Current* 303
8.4 Junction Breakdown 305
*8.5 Charge Storage and Diode Transients 309
8.5.1 *The Turn-off Transient* 309
8.5.2 *The Turn-on Transient* 312
*8.6 The Tunnel Diode 313
8.7 Summary 316
Problems 318

CHAPTER 9
Metal–Semiconductor and Semiconductor Heterojunctions 326

Preview 326
9.1 The Schottky Barrier Diode 326
9.1.1 *Qualitative Characteristics* 327
9.1.2 *Ideal Junction Properties* 329
9.1.3 *Nonideal Effects on the Barrier Height* 333
9.1.4 *Current–Voltage Relationship* 337
9.1.5 *Comparison of the Schottky Barrier Diode and the pn Junction Diode* 341
9.2 Metal–Semiconductor Ohmic Contacts 344
9.2.1 *Ideal Nonrectifying Barriers* 345
9.2.2 *Tunneling Barrier* 346
9.2.3 *Specific Contact Resistance* 348
9.3 Heterojunctions 349
9.3.1 *Heterojunction Materials* 350
9.3.2 *Energy-Band Diagrams* 350
9.3.3 *Two-Dimensional Electron Gas* 351
*9.3.4 *Equilibrium Electrostatics* 354
*9.3.5 *Current–Voltage Characteristics* 359
9.4 Summary 359
Problems 361

CHAPTER 10
The Bipolar Transistor 367

Preview 367
10.1 The Bipolar Transistor Action 368
10.1.1 The Basic Principle of Operation 369
10.1.2 Simplified Transistor Current Relations 370
10.1.3 The Modes of Operation 374
10.1.4 Amplification with Bipolar Transistors 376
10.2 Minority Carrier Distribution 377
10.2.1 Forward-Active Mode 378
10.2.2 Other Modes of Operation 384
10.3 Low-Frequency Common-Base Current Gain 385
10.3.1 Contributing Factors 386
10.3.2 Mathematical Derivation of Current Gain Factors 388
10.3.3 Summary 392
10.3.4 Example Calculations of the Gain Factors 393
10.4 Nonideal Effects 397
10.4.1 Base Width Modulation 397
10.4.2 High Injection 401
10.4.3 Emitter Bandgap Narrowing 403
10.4.4 Current Crowding 405
**10.4.5* Nonuniform Base Doping 406*
10.4.6 Breakdown Voltage 408
10.5 Equivalent Circuit Models 413
**10.5.1* Ebers–Moll Model 414*
10.5.2 Gummel–Poon Model 416
10.5.3 Hybrid-Pi Model 418
10.6 Frequency Limitations 422
10.6.1 Time-Delay Factors 422
10.6.2 Transistor Cutoff Frequency 424
10.7 Large-Signal Switching 427
10.7.1 Switching Characteristics 427
10.7.2 The Schottky-Clamped Transistor 429
***10.8** Other Bipolar Transistor Structures 430
10.8.1 Polysilicon Emitter BJT 430
10.8.2 Silicon–Germanium Base Transistor 431
10.8.3 Heterojunction Bipolar Transistors 434
10.9 Summary 435
Problems 438

CHAPTER 11
Fundamentals of the Metal–Oxide–Semiconductor Field-Effect Transistor 449

Preview 449
11.1 The Two-Terminal MOS Structure 450
11.1.1 Energy-Band Diagrams 450
11.1.2 Depletion Layer Thickness 455
11.1.3 Work Function Differences 458
11.1.4 Flat-Band Voltage 462
11.1.5 Threshold Voltage 465
11.1.6 Charge Distribution 471
11.2 Capacitance–VoltageCharacteristics 474
11.2.1 Ideal C-V Characteristics 474
11.2.2 Frequency Effects 479
11.2.3 Fixed Oxide and Interface Charge Effects 480
11.3 The Basic MOSFET Operation 483
11.3.1 MOSFET Structures 483
11.3.2 Current–Voltage Relationship—Concepts 486
**11.3.3 Current–Voltage Relationship—Mathematical Derivation 490*
11.3.4 Transconductance 498
11.3.5 Substrate Bias Effects 499
11.4 Frequency Limitations 502
11.4.1 Small-Signal Equivalent Circuit 502
11.4.2 Frequency Limitation Factors and Cutoff Frequency 504
***11.5** The CMOS Technology 507
11.6 Summary 509
Problems 513

CHAPTER 12
Metal–Oxide–SemiconductorField-Effect Transistor: Additional Concepts *523*

Preview 523
12.1 Nonideal Effects 524
12.1.1 Subthreshold Conduction 524

12.1.2 Channel Length Modulation 526
12.1.3 Mobility Variation 530
12.1.4 Velocity Saturation 532
12.1.5 Ballistic Transport 534
12.2 MOSFET Scaling 534
12.2.1 Constant-Field Scaling 534
12.2.2 Threshold Voltage—First Approximations 535
12.2.3 Generalized Scaling 536
12.3 Threshold Voltage Modifications 537
12.3.1 Short-Channel Effects 537
12.3.2 Narrow-Channel Effects 541
12.4 Additional Electrical Characteristics 543
12.4.1 Breakdown Voltage 544
**12.4.2 The Lightly Doped Drain Transistor 550*
12.4.3 Threshold Adjustment by Ion Implantation 551
***12.5** Radiation and Hot-Electron Effects 554
12.5.1 Radiation-Induced Oxide Charge 555
12.5.2 Radiation-Induced Interface States 558
12.5.3 Hot-Electron Charging Effects 560
12.6 Summary 561
Problems 563

CHAPTER **13**
The Junction Field-Effect Transistor *570*

Preview 570
13.1 JFET Concepts 571
13.1.1 Basic pn JFET Operation 571
13.1.2 Basic MESFET Operation 575
13.2 The Device Characteristics 577
13.2.1 Internal Pinchoff Voltage, Pinchoff Voltage, and Drain-to-Source Saturation Voltage 577
13.2.2 Ideal DC Current-Voltage Relationship—Depletion Mode JFET 582
13.2.3 Transconductance 587
13.2.4 The MESFET 588
***13.3** Nonideal Effects 593
13.3.1 Channel Length Modulation 594
13.3.2 Velocitj Saturation Efects 596
13.3.3 Subthreshold and Gate Current Effects 596
***13.4** Equivalent Circuit and Frequency Limitations 598
13.4.1 Small-Signal Equivalent Circuit 598
13.4.2 Frequency Limitation Factors and Cutoff Frequency 600
***13.5** High Electron Mobility Transistor 602
13.5.1 Quantum Well Structures 603
13.5.2 Transistor Performance 604
13.6 Summary 609
Problems 611

CHAPTER **14**
Optical Devices 617

Preview 617
14.1 Optical Absorption 618
14.1.1 Photon Absorption Coefficient 618
14.1.2 Electron–Hole Pair Generation Rate 621
14.2 Solar Cells 623
14.2.1 The pn Junction Solar Cell 623
14.2.2 Conversion Ejjiciency and Solar Concentration 626
14.2.3 Nonuniform Absorption Efects 628
14.2.4 The Heterojunction Solar Cell 628
14.2.5 Amorphous Silicon Solar Cells 630
14.3 Photodetectors 631
14.3.1 Photoconductor 632
14.3.2 Photodiode 634
14.3.3 PIN Photodiode 639
14.3.4 Avalanche Photodiode 640
14.3.5 Phototransistor 641
14.4 Photoluminescence and Electroluminescence 642
14.4.1 Basic Transitions 643
14.4.2 Luminescent Efficiency 645
14.4.3 Materials 645
14.5 Light Emitting Diodes 647
14.5.1 Generation of Light 648

PREFACE

PHILOSOPHY AND GOALS

The purpose of the third edition of this book is to provide a basis for understanding the characteristics, operation, and limitations of semiconductor devices. In order to gain this understanding, it is essential to have a thorough knowledge of the physics of the semiconductor material. The goal of this book is to bring together quantum mechanics, the quantum theory of solids, semiconductor material physics. and semiconductor device physics. All of these components are vital to the understanding of both the operation of present day devices and any future development in the field.

The amount of physics presented in this text is greater than what is covered in many introductory semiconductor device books. Although this coverage is more extensive, the author has found that once the basic introductory and material physics have been thoroughly covered. the physics of the semiconductor device follows quite naturally and can he covered fairly quickly and efficiently. The emphasis on the underlying physics will also be a benefit in understanding and perhaps in developing new semiconductor devices.

Since the objective of this text is to provide an introduction to the theory of semiconductor devices, there is a great deal of advanced theory that is not considered. In addition. fabrication processes are not described in detail. There are a few references and general discussions about processing techniques such as diffusion and ion implantation, but only where the results of this processing have direct impact on device characteristics.

PREREQUISITES

This book is intended for junior and senior undergraduates. The prerequisites for understanding the material are college mathematics. up to and including differential equations, and college physics, including an introduction to modern physics and electrostatics. Prior completion of an introductory course in electronic circuits is helpful, but not essential.

ORGANIZATION

The text begins with the introductory physics, moves on to the semiconductor material physics, and then covers the physics of semiconductor devices. Chapter 1 presents an introduction to the crystal structure of solids, leading to the ideal single-crystal semiconductor material. Chapters 2 and 3 introduce quantum mechanics and the quantum theory of solids, which together provide the necessary basic physics.

Chapters 4 through6 cover the semiconductor material physics. Chapter4 presents the physics of the semiconductor in thermal equilibrium; Chapter 5 treats the transport

phenomena of the charge carriers in a semiconductor. The nonequilibrium excess carrier characteristics are then developed in Chaptcr 6. Understanding the behavior of excess carriers in a semiconductor is vital to the goal of understanding the device physics.

The physics of the basic semiconductor devices is developed in Chapters 7 through 13. Chaptcr 7 treats the electrostatics of the basic pn junction. and Chapter 8 covers the current-voltage characteristics of the pn junction. Metal-semiconductorjunctions, both rectifying and nonrectifying, and semiconductor heterojunctions are considered in Chapter 9, while Chapter 10 treats the bipolar transistor. The physics of the metal-oxide-semiconductor field-effect transistor is presented in Chapters 11 and 12. and Chapter 13 covers the junction field-effect transistor. Once the physics of the pn junction is developed, the chapters dealing with the three basic transistors may be covered in any order—these chapters are written so as not to depend on one another. Chapter 14 considers optical devices and finally Chapter 15 covers power semiconductor devices.

USE OF THE BOOK

The text is intended for a one-semester course at the junior or senior level. As with most textbooks, there is more material than can be conveniently covered in one semester; this allows each instructor some flexibility in designing the course to hislher own specific needs. Two possible orders of presentation are discussed later in a separate section in this preface. However, the text is not an encyclopedia. Sections in each chapter that can be skipped without loss of continuity are identified by an asterisk in both the table of contents and in the chapter itself. These sections, althoughimportant to the development of semiconductor device physics, can he postponed to a later time.

The material in the text has been used extensively in a course that is required for junior-level electrical engineering students at the University of New Mexico. Slightly less than half of the semester is devoted to the first six chapters; the remainder of the semester is devoted to the pn junction, the bipolar transistor. and the metal-oxide-semiconductor field-effect transistor. A few other special topics may be briefly considered near the end of the semester.

Although the bipolar transistor is discussed in Chapter 10 before the MOSFET or JFET, each chapter dealing with one of the three basic types of transistors is written to stand alone. Any one of the transistor types may be covered first.

NOTES TO THE READER

This book introduces the physics of semiconductor materials and devices. Although many electrical engineering students are more comfortable building electronic circuits or writing computer programs than studying the underlying principles of semiconductor devices, the material presented here is vital to an understanding of the limitations of electronic devices, such as the microprocessor.

Mathematics is used extensively throughout the hook. This may at times seem tedious, but the end result is an understanding that will not otherwise occur. Although some of the mathematical models used to describe physical processes may seem abstract, they have withstood the test of time in their ability to describe and predict these physical processes.

The reader is encouraged to continually refer to the preview sections so that the objective of the chapter and the purposes of each topic can be kept in mind. This constant review is especially important in the first five chapters, dealing with basic physics.

The reader must keep in mind that, although some sections may be skipped without loss of continuity, many instructors will choose to cover these topics. The fact that sections are marked with an asterisk does not minimize the importance of these subjects.

It is also important that the reader keep in mind that there may be questions still unanswered at the end of a course. Although the author dislikes the phrase. "it can be shown that. . .," there are some concepts used here that rely on derivations beyond the scope of the text. This hook is intended as an introduction to the subject. Those questions remaining unanswered at the end of the course, the reader is encouraged to keep "in a desk drawer." Then, during the next course in this area of concentration, the reader can take out these questions and search for the answers.

ORDER OF PRESENTATION

Each instructor has a personal preference for the order in which the course material is presented. Listed below are two possible scenarios. The first case, called the classical approach, covers the bipolar transistor before the MOS transistor. However, because the MOS transistor topic is left until the end of the semester. time constraints may shortchange the amount of class time devoted to this important topic.

The second method of presentation listed, called the nonclassical approach, discusses the MOS transistor before the bipolar transistor. Two advantages to this approach are that the MOS transistor will not get shortchanged in terms of time devoted to the topic and, since a "real device" is discussed earlier in the semester, the reader may have more motivation to continue studying this course material. A possible disadvantage to this approach is that the reader may be somewhat intimidated by jumping from Chapter 7 to Chapter I I. However. the material in Chapters 1 I and I? is written so that this jump can be made.

Unfortunately, because of time constraints, every topic in every chapter cannot be covered in a one-semester course. The remaining topics must be left for a second-semester course or for further study by the reader.

Classical **approach**	
Chapter 1	Crystal structure
Chapters 2, 3	Selected topics from quantum mechanics and theory of solids
Chapter 4	Semiconductor physics
Chapter 5	Transport phenomena
Chapter 6	Selected topics from nonequilibrium characteristics
Chapters 7, 8	The pn junction and diode
Chapter 9	A brief discussion of the Schottky diode
Chapter 10	The bipolar transistor
Chapters 11, 12	The MOS transistor

	Nonclassical approach
Chapter 1	Crystal structure
Chapters 2, 3	Selected topics from quantum mechanics and theory of solids
Chapter 4	Semiconductor physics
Chapter 5	Transport phenomena
Chapter 7	The pn junction
Chapters 11, 12	The MOS transistor
Chapter 6	Selected topics from nonequilibrium characteristics
Chapter 8	The pn junction diode
Chapter 9	A brief discussion of the Schottky diode
Chapter 10	The bipolar transistor

FEATURES OF THE THIRD EDITION

- ***Preview section:*** A preview section introduces each chapter. This preview links the chapter to previous chapters and states the chapter's goals, i.e., what the reader should gain from the chapter.
- ***Exumples:*** An extensive number of worked examples are used throughout the text to reinforce the theoretical concepts being developed. These examples contain all the details of the analysis or design, so the reader does not have to fill in missing steps.
- ***Test your understanding:*** Exercise or drill problems are included throughout each chapter. These problems are generally placed immediately after an example problem, rather than at the end of a long section. so that readers can immediately test their understanding of the material just covered. Answers are given for each drill problem so readers do not have to search for an answer at the end of the book. These exercise problems will reinforce readers' grasp of the material before they move on to the next section.
- *Summary* ***section:*** A summary section, in bullet form, follows the text of each chapter. This section summarizes the overall results derived in the chapter and reviews the basic concepts developed.
- ***Glossary of important terms:*** A glossary of important terms follows the Summary section of each chapter. This section defines and summarizes the most important terms discussed in the chapter.
- ***Checkpoint:*** A checkpoint section follows the Glossary section. This section states the goals that should have been met and states the abilities the reader should have gained. The Checkpoints will help assess progress before moving on to the next chapter.
- ***Review questions:*** A list of review questions is included at the end of each chapter. These questions serve as a self-test to help the reader determine how well the concepts developed in the chapter have been mastered.
- ***End-of-chupterproblems*** A large number of problems are given at the end of each chapter, organized according to the subject of each section in the chapter

body. A larger number of problems have been included than in the second edition. Design-oriented or open-ended problems are included at the end in a Summary and Review section.

- *Computer simulation:* Computer simulation problems are included in many end-of-chapter problems. Computer simulation has not been directly incorporated into the text. However, a website has been established that considers computer simulation using MATLAB. This website contains computer simulations of material considered in most chapters. These computer simulations enhance the theoretical material presented. There also are exercise or drill problems that a reader may consider.
- ***Reading list:*** A reading list finishes up each chapter. The references, that are at an advanced level compared with that of this text, are indicated by an asterisk.
- ***Answers** to selected problems:* Answers to selected problems are given in the last appendix. Knowing the answer to a problem is an aid and a reinforcement in problem solving.

ICONS



Computer Simulations



Design Problems and Examples

SUPPLEMENTS

This hook is supported by the following supplements:

- Solutions Manual available to instructors in paper form and on the website.
- Power Point slides of important figures are available on the website.
- Computer simulations are available on the website.

ACKNOWLEDGMENTS

lamindebtedtothe many students I have hadoverthe years who have helped in the evolution of the third edition as well as the first and second editions of this text. I am grateful for their enthusiasm and constructive criticism. The University of New Mexico has my appreciationfor providing an atmosphere conducive to writing this hook.

I want to thank the many people at McGraw-Hill, for their tremendous support. Aspecial thanks to Kelley Butcher, senior developmental editor. Her attention to details and her enthusiasm throughout the project are especially recognized and appreciated. I also appreciate the efforts of Joyce Watters, project manager, who guided the work through its final phase toward publication.

The following reviewers deserve thanks for their constructive criticism and suggestions for the third edition of this text.

Thomas Mantei, *University of Cincinnati*
Cheng Hsiao Wu, *University of Missouri—Rolla*
Kamtoshi Najita, *University of Hawaii at Manoa*
John Naber, *University of Louisville*
Gerald Oleszek, *University of Colorado—Colorado Springs*
Marc Cahay, *University of Cincinnati*

The following reviewers deserve thanks for their constructive criticism and suggestions for the second edition:

Jon M. Meese, *University of Missouri—Columbia*
Jacob *B*. Khurgin, *Johns Hopkins University*
Hong Koo Kim, *University of Pittsburgh*
Gerald M. Oleszek, *University of Colorado — Colorado Springs*
Ronald J. Roedel, *Arizona Stare University*
Leon McCaughan, *University of Wisconsin*
A. Anil Kumar, *Prairie View A & M University*

Since the third edition is an outgrowth of the first edition of the text, the following reviewers of the first edition deserve my continued thanks for their thorough reviews and valuable suggestions:

Timothy J. Drummond, *Sandia Laboratories*
J. *L*. Davidson, *Vanderbilt University*
Robert Jackson, *University of Massachusetts—Amherst*
C. H. *Wu, University of Missouri—Rolla*
D. K. Reinhard, *Michigan State University*
Len Trombetta, *University of Houston*
Dan Moore, *Virginia Polytechnic Institute and State University*
Bruce *P*. Johnson, *University ofNevada — Reno*
William Wilson, *Rice University*
Dennis Polla, *university of Minnesota*
G. E. Stillman, *University of Illinois—Urbana-Champaign*
Richard C. Jaeger, *Auburn University*
Anand Kulkarni, *Michigan Technological University*
Ronald *D*. Schrimpf, *University of Arizona*

I appreciate the many fine and thorough reviews — your suggestions have made this a better book.

Donald A. Neamen

Semiconductors and the Integrated Circuit

PREVIEW

We often hear that we are living in the information age. Large amounts of information can be obtained via the Internet, for example, and can also be obtained quickly over long distances via satellite communication systems. The development of the transistor and the integrated circuit (IC) has lead to these remarkable capabilities. The IC permeates almost every facet of our daily lives, including such things as the compact disk player, the fax machine, laser scanners at the grocery store, and the cellular telephone. One of the most dramatic examples of IC technology is the digital computer—a relatively small laptop computer today has more computing capability than the equipment used to send a man to the moon a few years ago. The semiconductor electronics field continues to be a fast-changing one. with thousands of technical papers published each year. ■

HISTORY

The semiconductor device has a fairly long history, although the greatest explosion of IC technology has occured during the last two or three decades.' The metal–semiconductor contact dates back to the early work of Rraun in 1874, who discovered the asymmetric nature of electrical conduction between metal contacts and semiconductors, such as copper, iron, and lead sulfide. These devices were used as

'This brief introduction is intended to give a flavor of the history of the semiconductor device and integrated circuit. Thousands of engineers and scientists have made significant contributions to the development of semiconductor electronics-the few events and names mentioned here are not meant to imply that these are the only significant events or people involved in the semiconductor history.

detectors in early experiments on radio. In 1906, Pickard took out a patent for a point contact detector using silicon and, in 1907. Pierce published rectification characteristics of diodes made by sputtering metals onto a variety of semiconductors.

By 1935. selenium rectifiers and silicon point contact diodes were available for use as radio detectors. With the development of radar. the need for detector diodes and mixers increased. Methods of achieving high-purity silicon and germanium were developed during this time. A significant advance in our understanding of the metal–semiconductor contacr was aided by developments in the semiconductor physics. Perhaps most important during this period was Bethe's thermionic-emission theory in 1942, according to which the current is determined by the process of emission of electrons into the metal rather than by drift or diffusion.

Another big breakthrough came in December 1947 when the first transistor was constructed and tested at Bell Telephone Laboratories by William Shockley, John Bardeen, and Walter Brattain. This tirst transistor was a point contacr device and used polycrystalline germanium. The transistor effect was soon demonstrated in silicon as well. A significant improvement occurred at the end of 1949 when single-crystal material was used rather than rhe polycrystalline material. The single crystal yields uniform and improved properties throughout the whole semiconductor material.

The next significant step in the development of the transistor was the use of the diffusion process to form the necessary junctions. This process allowed better control of the transistor characteristics and yielded higher-frequency devices. The diffused mesa transistor was commercially available in germanium in 1957 and in silicon in 1958. The diffusion process also allowed many transistors to be fabricated on a single silicon slice. so the cost of these devices decreased.

THE INTEGRATED CIRCUIT (IC)

Up to this point, each component in an electronic circuit had ro he individually connected by wires. In September 1958. Jack Kilby of Texas Instruments demonstrated the first integrated circuit, which was fabricated in germanium. At about the same time, Robert Noyce of Fairchild Semiconductor introduced the integrated circuit in silicon using a planar technology. The tirst circuit used bipolar transistors. Practical MOS transistors were then developed in the mid-'60s. The MOS technologies, especially CMOS, have become a major focus for IC design and development. Silicon is the main semiconductor material. Gallium arsenide and other compound semiconductors are used for special applications requiring very high frequency devices and for optical devices.

Since that first IC, circuit design has become more sophisticated, and the integrated circuit more complex. A single silicon chip may be on the order of 1 square centimeter and contain over a million transistors. Some ICs may have more than a hundred terminals, while an individual transistor has only three. An IC can contain the arithmetic, logic. and memory functions on a single semiconductor chip—the primary example of this type of IC is the microprocessor. Intense research on silicon processing and increased automation in design and manufacturing have led to lower costs and higher fabrication yields.

FABRICATION

The integrated circuit is a direct result of the development of various processing techniques needed to fabricate the transistor and interconnect lines on the single chip. The total collection of these processes for making an IC is called a *technology*. The following few paragraphs provide an introduction to a few of these processes. This introduction is intended to provide the reader with some of the basic terminology used in processing.

Thermal Oxidation A major reason for the success of silicon ICs is the fact that an excellent native oxide, SiO_2, can be formed on the surface of silicon. This oxide is used as a gate insulator in the MOSFET and is also used as an insulator, known as the field oxide, between devices. Metal interconnect lines that connect various devices can be placed on top of the field oxide. Most other semiconductors do not form native oxides that are of sufficient quality to be used in device fabrication.

Silicon will oxidize at room temperature in air forming a thin native oxide of approximately 25 A thick. However, most oxidations are done at elevated temperatures since the basic process requires that oxygen diffuse through the existing oxide to the silicon surface where a reaction can occur A schematic of the oxidation process is shown in Figure 0.1. Oxygen diffuses across a stagnant gas layer directly adjacent to the oxide surface and then diffuses through the existing oxide layer to the silicon surface where the reaction between O_2 and Si forms SiO_2. Because of this reaction, silicon is actually consumed from the surface of the silicon. The amount of silicon consumed is approximately 44 percent of the thickness of the final oxide.

Photomasks and Photolithography The actual circuitry on each chip is created through the use of photomasks and photolithography. The photomask is a physical representation of a device or a portion of a device. Opaque regions on the mask are made of an ultraviolet-light-absorbing material. A photosensitive layer, called photoresist, is first spread over the surface of the semiconductor. The photoresist is an

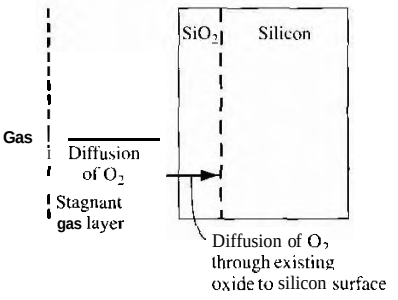


Figure 0.1 | Schematic of the oxidation process.

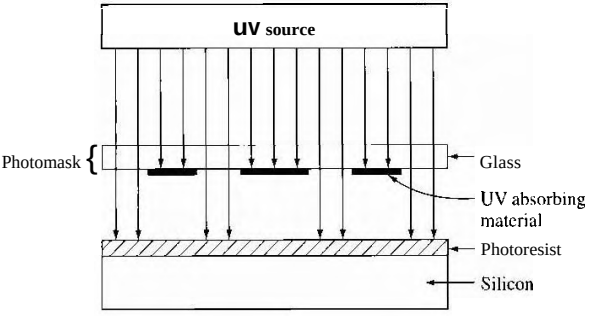


Figure 0.2 | Schematic showing the use of a photomask

organic polymer that undergoes chemical change when exposed to ultraviolet light. The photoresist is exposed to ultraviolet light through the photomask as indicated in Figure 0.2. The photoresist is thcn developed in a chemical solution. Thc developer is used to remove the unwanted portions of the photoresist and generate the appropriate patterns on the silicon. The photomasks and photolithography process is critical in that it determines how small the devices can he made. Instead of using ultraviolet light, electrons and x-rays can also be used to expose the photoresist.

Etching After the photoresist pattern is formed, the remaining photoresist can be used as a mask, so that the material not covered hy the photoresist can be etched. Plasma etching is now the standard process used in IC fabrication. Typically. an etch gas such as chlorofluorocarbons are injected into a low-pressure chamber. A plasma is created by applying a radio-frequency voltage between cathode and anode terminals. The silicon wafer is placed on the cathode. Positively charged ions in the plasma are accelerated toward the cathode and bombard the wafer normal to the surface. The actual chemical and physical reaction at the surface is complex. but the net result is that silicon can he etched anisotropically in very selected regions of the wafer. If photoresist is applied on the surface of silicon dioxide. then the silicon dioxide can also be etched in a similar way.

Diffusion A thermal process that is used extensively in IC fabrication is diffusion. Diffusion is the process hy which specific types of "impurity" atoms can be introduced into the silicon material. This doping process changes the conductivity type of the silicon so that pn junctions can be formed. (The pn junction is a basic building block of semiconductor devices.) Silicon wafers are oxidized to form a layer of silicon dioxide and windows are opened in thc oxide in selected areas using photolithography and etching as just described.

The wafers are thcn placed in a high-temperature furnace (about 1100 C) and dopant atoms such as boron or phosphorus are introduced. The dopant atoms gradually diffuse or move into the silicon duc to a density gradient. Since the diffusion process requires a gradient in the concentration of atoms, the final concentration of

diffused atoms is nonlinear. as shown in Figure 0.3. When the wafer is removed from the furnace and the wafer temperature return:, to room temperature, the diffusion coefficient of the dopant atoms is essentially zero so that the dopant atoms are then fixed in the silicon material.

Ion Implantation A fabrication process that is an alternative to high-temperature diffusion is ion implantation. A beam of dopant ions is accelerated to a high energy and is directed at the surface of a semiconductor. As the ions enter the silicon, they collide with silicon atoms and lose energy and finally come to rest at some depth within the crystal. Since the collision process is statistical in nature, there is a distribution in the depth of penetration of the dopant ions. Figure 0.4 shows such an example of the implantation of boron into silicon at a particular energy.

Two advantages of the ion implantation process compared to diffusion are (1) the ion implantation process is a low temperature process and (2) very well defined doping layers can be achieved. Photoresist layers or layers of oxide can he used to block the penetration of dopant atoms so that ion implantation can occur in very selected regions of the silicon.

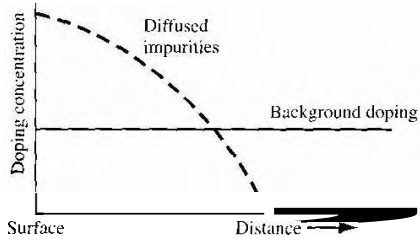


Figure 0.3 | Final concentration of diffused impurities into the surface of a semiconductor.

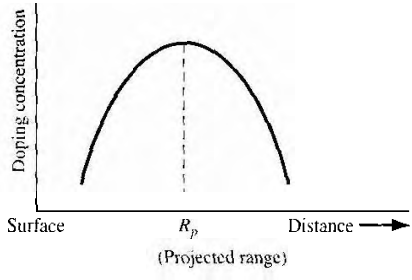


Figure 0.4 | Final concentration of ion-implanted boron into silicon.

One disadvantage of ion implantation is that the silicon crystal is damaged by the penetrating dopant atoms because of collisions between the incident dopant atoms and the host silicon atoms. However, most of the damage can he removed by thermal annealing the silicon at an elevated temperature. The thermal annealing temperature, however, is normally much less that the diffusion process temperature.

Metallization, Bonding, and Packaging After the semiconductor devices have been fabricated by the processing steps discussed. they need to be connected to each other to form the circuit. Metal films are generally deposited by a vapor deposition technique and the actual interconnect lines are formed using photolithography and etching. In general, a protective layer of silicon nitride is finally deposited over the entire chip.

The individual integrated circuit chips are separated by scribing and breaking the wafer. The integratedcircuit chip is then tnounted in apackage. Lead bonders are finally used to attach gold or aluminum wires between the chip and package terminals.

Summary: Simplified **Fabrication of** a pn **Junction** Figure 0.5 shows the basic steps in fomung a pn junction. These steps involve some of the processing described in the previous paragraphs.

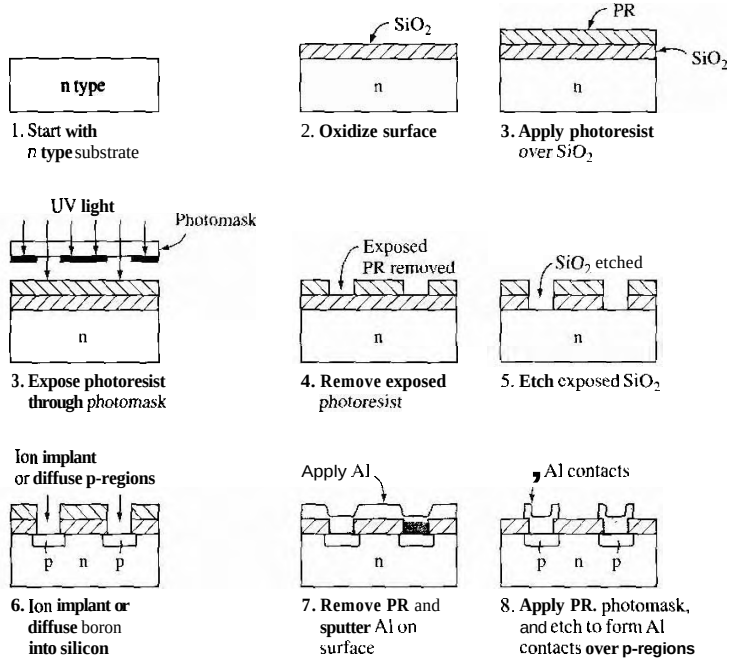


Figure 0.5 / The **basic** steps **in** *forming* **a** pn **junction.**

CHAPTER 1

The Crystal Structure of Solids

PREVIEW

This text deals with the electrical properties and characteristics of semiconductor materials and devices. The electrical properties of solids are therefore of primary interest. The semiconductor is in general a single-crystal material. The electrical properties of a single-crystal material are determined not only by the chemical composition but also by the arrangement of atoms in the solid; this being true, a brief study of the crystal structure of solids is warranted. The formation, or growth, of the single-crystal material is an important part of semiconductor technology. A short discussion of several growth techniques is included in this chapter to provide the reader with some of the terminology that describes semiconductor device structures. This introductory chapter provides the necessary background in single-crystal materials and crystal growth for the basic understanding of the electrical properties of semiconductor materials and devices. ■

1.1 | SEMICONDUCTOR MATERIALS

Semiconductors are a group of materials having conductivities between those of metals and insulators. Two general classifications of semiconductors are the elemental semiconductor materials, found in group IV of the periodic table, and the compound semiconductor materials, most of which are formed from special combinations of group III and group V elements. Table 1.1 shows a portion of the periodic table in which the more common semiconductors are found and Table 1.2 lists a few of the semiconductor materials. (Semiconductors can also be formed from combinations of group II and group VI elements. but in general these will not beconsidered in this text.)

The elemental materials, those that are composed of single species of atoms, are silicon and germanium. Silicon is by far the most common semiconductor used in integrated circuits and will be emphasized to a great extent.

Table 1.1 | A portion of the periodic table

III	IV	V
B	C	
Al	Si	P
Ga	Ge	As
In		Sb

Table 1.2 | A list of some semiconductor materials

Elemental semiconductors	
Si	Silicon
Ge	Germanium
Compound semiconductors	
AlP	Aluminum phosphide
AlAs	Aluminum arsenide
GaP	Gallium phosphidr
GaAs	Gallium arsenide
InP	Indium phosphide

The two-element, or *binary,* compounds such as gallium arsenide or gallium phosphide are formed by combining one group III and one group V element. Gallium arsenide is one of the more common of the compound semiconductors. Its good optical properties make it useful in optical devices. GaAs is also used in specialized applications in which, for example, high speed is required.

We can also form a three-element, or *ternary,* compound semiconductor. An example is $Al_xGa_{1-x}As$, in which the subscript x indicates the fraction of the lower atomic number element component. More complex semiconductors can also be formed that provide flexibility when choosing material properties.

1.2 | TYPES OF SOLIDS

Amorphous, polycrystalline, and single crystal are the three general types of solids. Each type is characterized by the size of an ordered region within the material. An ordered region is a spatial volume in which atoms or molecules have a regular geometric arrangement or periodicity. Amorphous materials have order only within a few atomic or molecular dimensions, while polycrystalline materials have a high degree

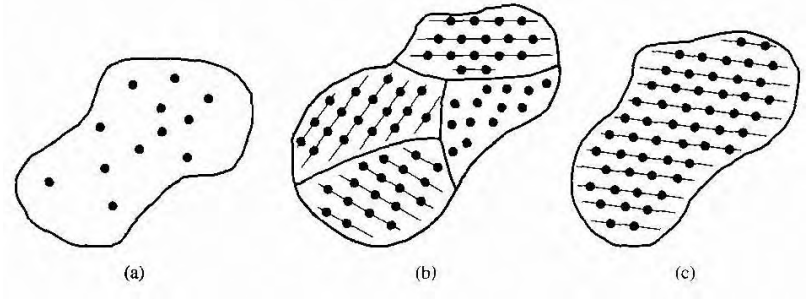


Figure 1.1 | Schematics of three general types of crystals: (a) amorphous, (b) polycrystalline, (c) single crystal.

of order over many atomic or molecular dimensions. These ordered regions. or single-crystal regions, vary in size and orientation with respect to one another. The single-crystal regions are called grains and are separated from one another by grain boundaries. Single-crystal materials, ideally, have a high degree of order, or regular geometric periodicity, throughout the entire volume of the material. The advantage of a single-crystal material is that. in general, its electrical properties are superior to those of a nonsingle-crystal material, since grain boundaries tend to degrade the electrical characteristics. Two-dimensional representations of amorphous, polycrystalline, and single-crystal materials are shown in Figure 1.1.

1.3 | SPACE LATTICES

Our primary concern will be the single crystal with its regular geometric periodicity in the atomic arrangement. A representative unit, or group of atoms, is repeated at regular intervals in each of the three dimensions to form the single crystal. The periodic arrangement of atoms in the crystal is called the *lattice*.

1.3.1 Primitive and Unit Cell

We can represent a particular atomic array by a dot that is called a *lattice point*. Figure 1.2 shows an infinite two-dimensional array of lattice points. The simplest means of repeating an atomic array is by translation. Each lattice point in Figure 1.2 can be translated a distance a_1 in one direction and a distance b_1 in a second noncolinear direction to generate the two-dimensiunal lattice. A third noncolinear translation will produce the three-dimensional lattice. The translation directions need not be perpendicular.

Since the three-dimensional lattice is a periodic repetition of a group of atoms, we do not need to consider the entire lattice, but only a fundamental unit that is being repeated. *A unit cell* is a small volume of the crystal that can be used to reproduce the entire crystal. Aunit cell is not a unique entity. Figure 1.3 shows several possible unit cells in a two-dimensional lattice.

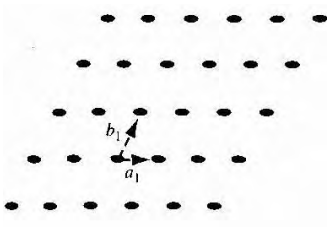


Figure 1.2 | Two-dimensional representation of a single-crystal lattice.

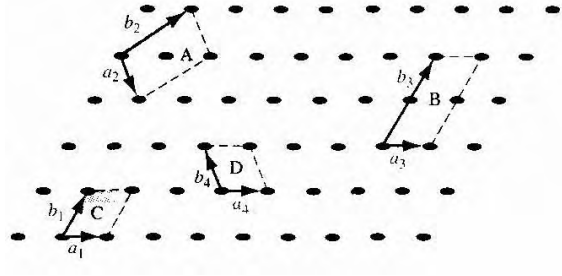


Figure 1.3 | Two-dimensional representation of a single-crystal lattice showing various possible unit cells.

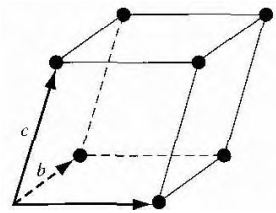


Figure 1.4 | A generalized primitive unit cell.

The unit cell A can be translated in directions a_2 and b_2, the unit cell B can be translated in directions a_3 and b_3, and the entire two-dimensional lattice can be constructed by the translations of either of these unit cells. The unit cells C and D in Figure 1.3 can also be used to construct the entire lattice by using the appropriate translations. This discussion of two-dimensional unit cells can easily be extended to three dimensions to describe a real single-crystal material.

Aprirnitive cell is the smallest unit cell that can be repeated to form the lattice. In many cases, it is more convenient to use a unit cell that is not a primitive cell. Unit cells may be chosen that have orthogonal sides, for example, whereas the sides of a primitive cell may be nonorthogonal.

A generalized three-dimensional unit cell is shown in Figure 1.4. The relationship between this cell and the lattice is characterized by three vectors $\bar{a}$, 6, and ?, which need not be perpendicular and which may or may not be equal in length. Every equivalent lattice point in the three-dimensional crystal can he found using the vector

$$\bar{r} = p\bar{a} + q\bar{b} + s\bar{c} \tag{1.1}$$

wherep, q, and s are integers. Since the location of the origin is arbitrary, we will let p, q, and s be positive integers for simplicity.

1.3.2 Basic Crystal Structures

Before we discuss the semiconductor crystal, let us consider three crystal structures and determine some of the basic characteristics of these crystals. Figure 1.5 shows the simple cubic, body-centered cubic, and face-centered cubic structures. For these simple structures, we may choose unit cells such that the general vectors $\bar{a}$, 6, and $\bar{c}$ are perpendicular to each other and the lengths are equal. The *simple* ***cubic*** (sc) structure has an atom located at each corner: the ***hod?-centered*** *cubic* (bcc) structure has an additional atom at the center of the cube; and the *face-centered cubic* (fcc) structure has additional atoms on each face plane.

By knowing the crystal structure of a material and its lattice dimensions, we can determine several characteristics of the crystal. For example, we can determine the volume density of atoms.

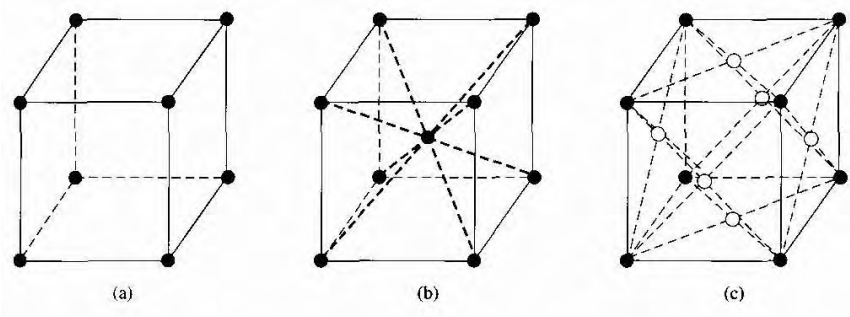


Figure 1.5 | Three lattice types: (a) simple cuhic. (b) body-centered cubic. (c) face-centered cuhic.

Objective | **EXAMPLE 1.1**

To find the volume density of atoms in a crystal.

Consider a single-crystal material that is a body-centered cuhic with a lattice constant $a = 5\,\text{Å} = 5 \times 10^{-8}$ cm. **A** corner atom is shared by eight unit cells which meet at each corner so that each comer atom effectively contributes one-eighth of its volume to each unit cell. The eight comer atoms then contribute an equivalent of one atom to the unit cell. If we add the body-centered atom to the comer atoms, each unit cell contains an equivalent of two atoms.

■ Solution

The volume density of atoms is then found as

$$\text{Density} = \frac{2\ \text{atoms}}{(5 \times 10^{-8})^3} = 1.6 \times 10^{22}\ \text{atoms per cm}^3$$

■ Comment

The volume density of atoms just calculated represents the order of magnitude of density for most materials. The actual density is a function of the crystal type and crystal structure since the packing density—number of atoms per unit cell—depends on crystal structure.

TEST **YOUR** UNDERSTANDING

E1.1 The lattice constant of a face-centered-cubic structure is 4.75 Å. Determine the volume density of atoms. ($1.77 \times 10^{22}\ \text{cm}^{-3}$)

E1.2 The volume density of atoms for a simple cubic lattice is $3 \times 10^{22}\ \text{cm}^{-3}$. Assume that the atoms are hard spheres with each atom touching its nearest neighbor. Determine the lattice constant and the radium of the atom. ($a_0 = 3.22\,\text{Å}$, $r = 1.61\,\text{Å}$)

1.3.3 Crystal Planes and Miller Indices

Since real crystals are not infinitely large, they eventually terminate at a surface. Semiconductor devices are fabricated at or near a surface, so the surface properties

may influence the device chnracteristics. We would like to be able to describe these surfaces in terms of the lattice. Surfaces, or planes through the crystal, can be described by first considering the intercepts of the plane along the $\bar{a}$, $\bar{b}$, and $\bar{c}$ axes used to describe the lattice.

EXAMPLE 1.2 Objective

To describe the plane shown in Figure 1.6. (The lattice points in Figure 1.6 are shown along the $\bar{a}$, $\bar{b}$, and? axes only.)

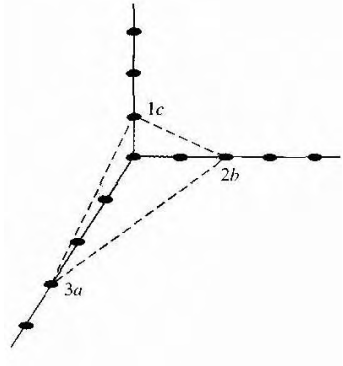


Figure **1.6** | A representative crystal lattice plane.

Solution

From Equation (1.1), the intercepts of the plane correspond to $p = 3, q = 2$, and $s = 1$. Now write the reciprocals of the intercepts, which gives

$$\left(\frac{1}{3}, \frac{1}{2}, \frac{1}{1}\right)$$

Multiply by the lowest common denominator, which in this case is 6, to obtain (2, 3, 6). The plane in Figure 1.6 is then referred to as the (236) plane. The integers are referred to as the Miller indices. We will refer to a general plane as the (*hkl*) plane.

■ **Comment**

We can show that the same three Miller indices are obtained for any plane that is parallel to the one shown in Figure 1.6. Any parallel plane is entirely equivalent to any other

Three planes that are commonly considered in a cubic crystal are shown in Figure 1.7. The plane in Figure 1.7a is parallel to the $\bar{b}$ and $\bar{c}$ axes so the intercepts are given as $p = 1, q = \infty$, and $s = \infty$. Taking the reciprocal, we obtain the Miller indices as (1, 0, 0), so the plane shown in Figure 1.7a is referred to as the (100) plane. Again, any plane parallel to the one shown in Figure 1.7a and separated by an integral

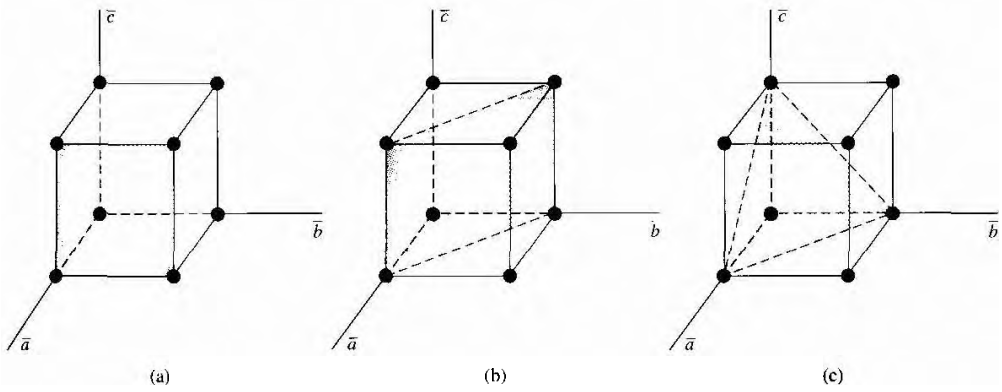


Figure 1.7 | Three lattice planes: (a)(100) plane. (b) (110) plane. (c) (111) plane.

number of lattice constants is equivalent and is referred to as the (100) plane. One advantage to taking the reciprocal of the intercepts to obtain the Miller indices is that the use of infinity is avoided when describing a plane that is parallel to an axis. If we were to describe a plane passing through the origin of our system, we would obtain infinity as one or more of the Miller indices after taking the reciprocal of the intercepts. However, the location of the origin of our system is entirely arbitrary and so, by translating the origin to another equivalent lattice point. we can avoid the use of infinity in the set of Miller indices.

For the simple cubic structure, the body-centered cubic. and the face-centered cubic, there is a high degree of symmetry. The axes can be rotated by 90" in each of the three dimensions and each lattice point can again be described by Equation (1.1) as

$$\bar{r} = p\bar{a} + q\bar{b} + s\bar{c} \qquad (1.1)$$

Each face plane of the cubic structure shown in Figure 1.7a is entirely equivalent. These planes are grouped together and are referred to as the {100} set of planes.

We may also consider the planes shown in Figures 1.7b and 1.7c. The intercepts of the plane shown in Figure 1.7b are $p = 1$, q = I, and $s = \infty$. The Miller indices are found by taking the reciprocal of these intercepts and, as a result, this plane is referred to as the (110) plane. In a similar way, the plane shown in Figure 1.7c is referred to as the (111) plane.

One characteristic of a crystal that can be determined is the distance between nearest equivalent parallel planes. Another characteristic is the surface concentration of atoms, number per square centimeter (#/cm^2), that are cut by a particular plane. Again, a single-crystal semiconductor is not infinitely large and must terminate at some surface. The surface density of atoms may be important, for example, in determining how another material, such as an insulator, will "fit" on the surface of a semiconductor material.

EXAMPLE 1.3 Objective

To calculate the surface density of atoms on a particular plane in a crystal.

Consider the body-centered cubic structure and the (110) plane shown in Figure 1.8a. Assume the atoms can be represented as hard spheres with the closest atoms touching each other. Assume the lattice constant is $a_1 = 5$ Å. Figure 1.8b shows how the atoms are cut by the (110) plane.

The atom at each corner is shared by four similar equivalent lattice planes. so each corner atom effectively contributes one-fourth of its area to this lattice plane as indicated in the figure. The four corner atoms then effectively contribute one atom to this lattice plane. The atom in the center is completely enclosed in the lattice plane. There is no other equivalent plane that cuts the center atom and the comer atoms, so the entire center atom is included in the number of atoms in the crystal plane. The lattice plane in Figure 1.8b, then. contains two atoms.

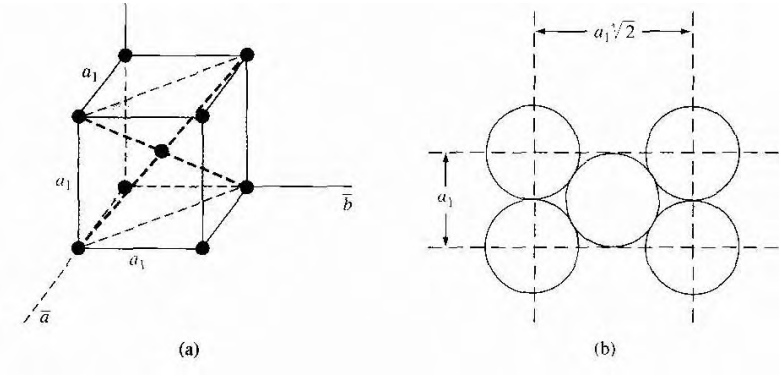


Figure 1.8 | (a) The (110) plane in a body-centered cubic and (b) the atoms cut by the (110) plane in a body-centered cubic.

■ **Solution**

We find the surface density by dividing the number of lattice atoms by the surface area, or in this case

$$\text{Surface density} = \frac{2 \text{ atoms}}{(a_1)(a_1\sqrt{2})} = \frac{2}{(5 \times 10^{-8})^2(\sqrt{2})}$$

which is

$$5.66 \times 10^{14} \text{ atoms/cm}^2$$

■ **Comment**

The surface density oi atoms is a function of the panicular crystal plane in the lattice and generally varier from one crystal plane to another

E1.3 Determine the distance between nearest (110) planes in a simple cubic lattice with a lattice constant of $a_0 = 4.83$ Å. (Ans. 3.42 Å)

E1.4 The lattice constant of a face-centered-cubic structure is 4.75 Å. Calculate the surface density of atoms for (a) a (100) plane and (b) a (110) plane.
[Ans. (a) 8.86×10^{14} cm^{-2}, (b) 6.27×10^{14} cm^{-2}]

Inaddition to describing crystal planes in a lattice, we may want todescribe a particulardirection in the crystal. The direction can be expressed as a set of three integers which are the components of a vector in that direction. For example, the body diagonal in a simple cubic lattice is composed of vector components I. 1, I. The body diagonal is then described as the [I I I]direction. The brackets are used to designate direction as distinct from the parentheses used for the crystal planes. The three basic directions and the associated crystal planes for the simple cubic structure are shown in Figure 1.9. Note that in the simple cubic lattices, the [*hkl*] direction is perpendicular to the (*hkl*) plane. This perpendicularity may not be true in noncubic lattices.

1.3.4 The Diamond Structure

As already stated, silicon is the most common semiconductor material. Silicon is referred to as a group IV element and has a diamond crystal structure. Germanium is also a group IV element and has the same diamond structure. A unit cell of the. diamond structure, shown in Figure 1.10, is more complicated than the simple cubic structures that we have considered up to this point.

We may begin to understand the diamond lattice by considering the tetrahedral structure shown in Figure 1.11. This structure is basically a body-centered cubic with

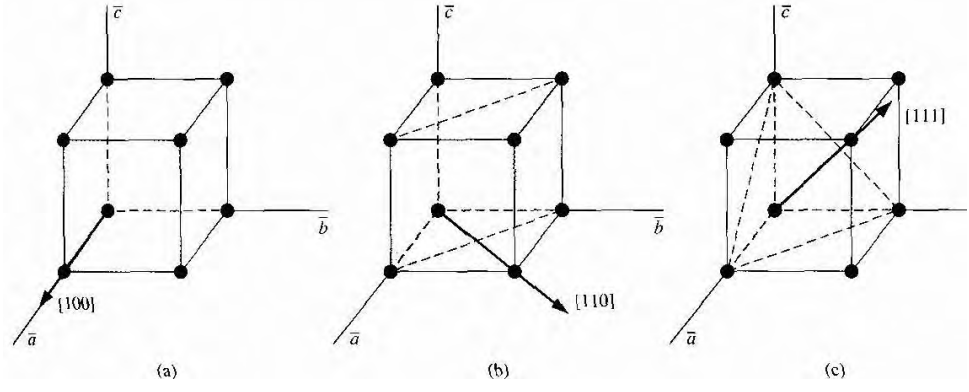


Figure 1.9 | Three lattice directions and planes: (a) (100) plane **and** [100] directiun, (b) (110) plane and [110] direction, (c) (111) plane and [111] direction.

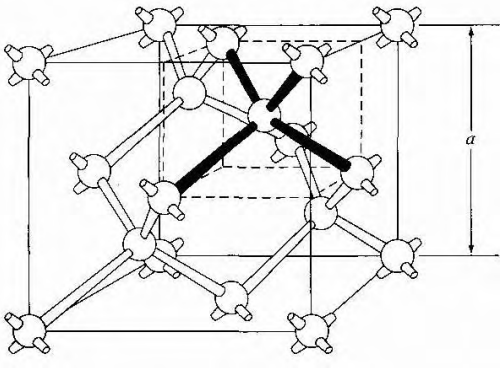


Figure 1.10 | The diamond structure.

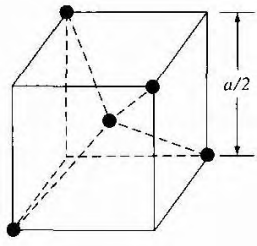


Figure 1.11 | The tetrahedral structure of closest neighbors in the diamond lattice.

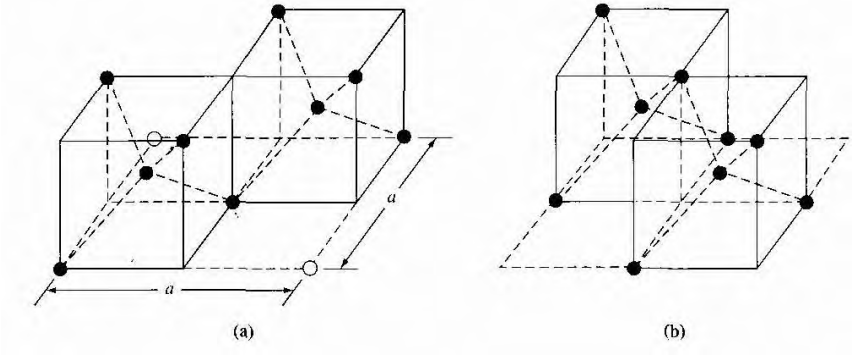


Figure 1.12 | Portions of the diamond lattice: (a) bottom half and (b) top half

four of the comer atoms missing. Every atom in the tetrahedral structure has four nearest neighbors and it is this structure which is the basic building block of the diamond lattice.

There are several ways to visualize the diamond structure. One way to gain a further understanding of the diamond lattice is by considering Figure 1.12. Figure 1.12a shows two body-centercd cubic, or tetrahedral, structures diagonally adjacent to each other. The shaded circles represent atoms in the lattice that are generated when the structure is translated to the right or left, one lattice constant, a. Figure 1.12b represents the top half of the diamond structure. The top half again consists of two tetrahedral structures joined diagonally, but which are at 90" with respect to the bottom-half diagonal. An important characteristic of the diamond lattice is that any atom within the diamond structure will have four nearest neighboring atoms. We will note this characteristic again in our discussion of atomic bonding in the next section.

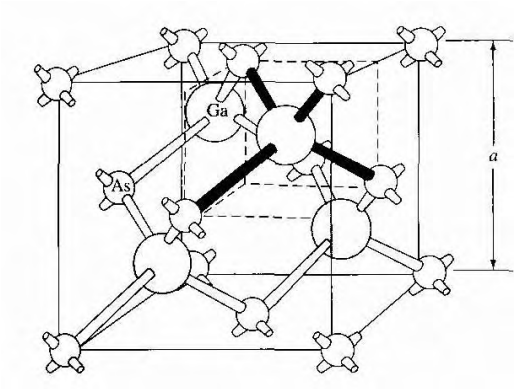


Figure 1.13 | The zincblende (sphalerite) lattice of GaAs.

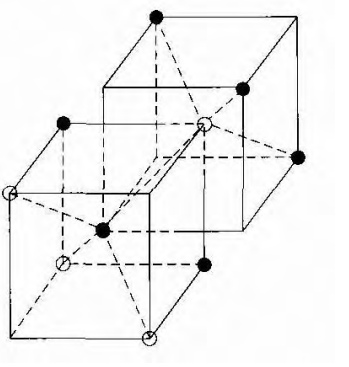


Figure 1.14 | The tetrahedral structure of closest neighbors in the zincblende lattice

The diamond structure refers to the particular lattice in which all atoms are of the same species, such as silicon or germanium. The rincblende (sphalerite) structure differsfrom the diamond structure only in that there are two different types of atoms in the lattice. Compound semiconductors, such as gallium arsenide, have the zincblende structure shown in Figure 1.13. The important feature of both the diamond and the zincblende structures is that the atoms are joined together to form a tetrahedron. Figure 1.14 shows the basic tetrahedral structure of GaAs in which each Ga atom has four nearest As neighbors and each As atom has four nearest Ga neighbors. This figurealso begins to show the interpenetration of two sublattices that can be used to generate the diamond or zincblende lattice.

TEST YOUR UNDERSTANDING

E1.5 ***The*** lattice constant of silicon is 5.43 Å. Calculate the volume density of silicon atoms. (5×10^{22} atoms/cm^3)

1.4 | ATOMIC BONDING

We have been considering various single-crystal structures. The question arises as to why one particular crystal structure is favored over another for a particular assembly of atoms. Afundamental law of nature is that the total energy of a system in thermal equilibrium tends to reach a minimum value. The interaction that occurs between atoms to form a solid and to reach the minimum total energy depends on the type of atom or atoms involved. The type of bond, or interaction, between atoms, then, depends on the particular atom or atoms in the crystal. If there is not a strong bond between atoms, they will not "stick together" to create a solid.

The interaction between atoms can be described by quantum mechanics. Although an introduction to quantum mechanics is presented in the next chapter, the quantum-mechanical description of the atomic bonding interaction is still beyond the scope of this text. We can nevertheless obtain a qualitative understanding of how various atoms interact by considering the valence, or outermost, electrons of an atom.

The atoms at the two extremes of the periodic table (excepting the inert elements) tend to lose or gain valence electrons, thus forming ions. These ions then essentially have complete outer energy shells. The elements in group I of the periodic table tend to lose their one electron and become positively charged. while the elements in group VII tend to gain an electron and become negatively charged. These oppositely charged ions then experience a coulomb attraction and form a bond referred to as an ***ionic bond.*** If the ions were to get too close, a repulsive force would become dominant, so an equilibrium distance results between these two ions. In a crystal, negatively charged ions tend to be surrounded by positively charged ions and positively charged ions tend to he surrounded by negatively charged ions, so a periodic array of the atoms is formed to create the lattice. A classic example of ionic bonding is sodium chloride.

The interaction of atoms tends to form closed valence shells such as we see in ionic bonding. Another atomic bond that tends to achieve closed-valence energy shells is ***covalent bonding,*** an example of which is found in the hydrogen molecule. A hydrogen atom has one electron and needs one more electron to complete the lowest energy shell. A schematic of two noninteracting hydrogen atoms, and the hydrogen molecule with the covalent bonding, are shown in Figure 1.15. Covalent bonding results in electrons being shared between atoms, so that in effect the valence energy shell of each atom is full.

Atoms in group IV of the periodic table, such as silicon and germanium, also tend to form covalent bonds. Each of these elements has four valence electrons and needs four more electrons to complete the valence energy shell. If a silicon atom, for example, has four nearest neighbors, with each neighbor atom contributing one valence electron to be shared. then the center atom will in effect have eight electrons in its outer shell. Figure 1.16a schematically shows live noninteracting silicon atoms with the four valence electrons around each atom. A two-dimensional representation

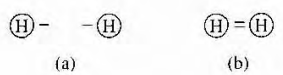


Figure 1.15 | Representation of (a) hydrogen valence electrons and (b) covalent bonding in a hydrogen molecule.

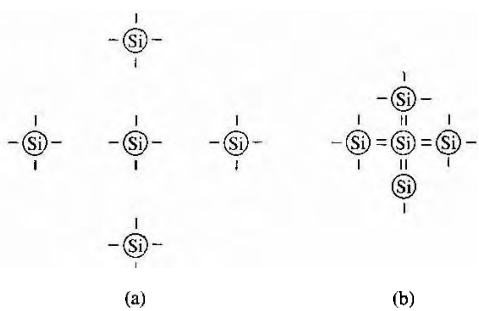


Figure 1.16 | Representation of (a) silicon valence electrons and (b) covalent bonding in the silicon crystal.

of the covalent bonding in silicon is shown in Figure 1.16b. The center atom has eight shared valence electrons.

A significant difference between the covalent bonding of hydrogen and of silicon is that, when the hydrogen molecule is formed, it has no additional electrons to form additional covalent bonds, while the outer silicon atoms always have valence electrons available for additional covalent bonding. The silicon array may then be formed into an infinite crystal, with each silicon atom having four nearest neighbors and eight shared electrons. The four nearest neighbors in silicon forming the covalent bond correspond to the tetrahedral structure and the diamond lattice, which were shown in Figures 1.11 and 1.10, respectively. Atomic bonding and crystal structure are obviously directly related.

The third major atomic bonding scheme is referred to as *metallic bonding*. Group I elements have one valence cleclron. If two sodium atoms ($Z = 1$ I), for example. are brought into close proximity, the valence electrons interact in a way similar to that in covalent bonding. When a third sodium atom is brought into close proximity with the first two, the valence electrons can also interact and continue to form a bond. Solid sodium has a body-centered cubic structure, so each atom has eight nearest neighbors with each atom sharing many valence electrons. We may think of the positive metallic ions as being surrounded by a sea of negative electrons, the solid being held together by the electrostatic forces. This description gives a qualitative picture of the metallic bond.

A fourth type of atomic bond. called the *Van der Waals* bond, is the weakest of the chemical bonds. A hydrogen fluoride (HF) molecule, for example, is formed by an ionic bond. The effective center of the positive charge of the molecule is not the same as the effective center of the negative charge. This nonsymmetry in the charge distribution results in a small electric dipole that can interact with the dipoles of other HF molecules. With these weak interactions, solids formed by the Van der Waals bonds have a relatively low melting temperature — in fact, most of these materials are in gaseous form at room temperature.

*1.5 | IMPERFECTIONS AND IMPURITIES IN SOLIDS

Up to this point, we have been considering an ideal single-crystal structure. In a real crystal, the lattice is not perfect, hut contains imperfections or defects; that is, the perfect geometric periodicity is disrupted in some manner. Imperfections tend to alter thc electrical properties of a material and, in some cases, electrical parameters can be dominated by these defects or impurities.

1.5.1 Imperfections in Solids

One type of imperfection that all crystals have in common is atomic thermal vibration. Aperfect single crystal contains atoms at particular lattice sites, the atoms separated from each other by a distance we have assumed to be constant. The atoms in a

*Indicates sections that can bc skipped without loss of continuity.

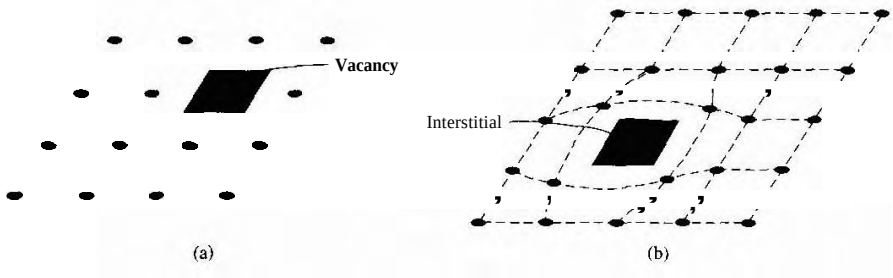


Figure 1.17 | Two-dimensional representation of a single-crystal lattice showing (a) a vacancy defect and (b) an interstitial defect.

crystal, however, have a certain thermal energy, which is a function of temperature. The thermal energy causes the atoms to vibrate in a random manner about an equilibrium lattice point. This random thermal motion causes the distance between atoms to randomly fluctuate, slightly disrupting the perfect geometric arrangement of atoms. This imperfection, called lattice vibrations, affects some electrical parameters, as we will see later in our discussion of semiconductor material characteristics.

Another type of defect is called a *point defect.* There are several of this type that we need to consider. Again, in an ideal single-crystal lattice, the atoms are arranged in a perfect periodic arrangement. However, in a real crystal, an atom may be missing from a particular lattice site. This defect is referred to as a *vacancy;* it is schematically shown in Figure 1.17a. In another situation, an atom may be located between lattice sites. This defect is referred to as an *interstitial* and is schematically shown in Figure 1.17b. In the case of vacancy and interstitial defects, not only is the perfect geometric arrangement of atoms broken, but also the ideal chemical bonding between atoms is disrupted, which tends to change the electrical properties of the material. A vacancy and interstitial may be in close enough proximity lo exhibit an interaction between the two point defects. This vacancy-interstitial defect, also known as a Frenkel defect, produces different effects than the simple vacancy or interstitial.

The point defects involve single atoms or single-atom locations. In forming single-crystal materials, more complex defects may occur. A line defect. for example, occurs when an entire row of atoms is missing from its normal lattice site. This defect is referred to as a line *dislocation* and is shown in Figure 1.18. As with a point defect, a line dislocation disrupts both the normal geometric periodicity of the lattice and the ideal atomic bonds in the crystal. This dislocation can also alter the electrical properties of the material, usually in a more unpredictable manner than the simple point defects.

Other complex dislocations can also occur in it crystal lattice. However. this introductory discussion is intended only to present a few of the basic types of defect, and to show that a real crystal is not necessarily a perfect lattice structure. The effect of these imperfections on the electrical properties of a semiconductor will be considered in later chapters.

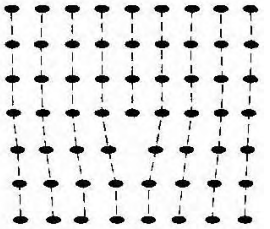


Figure 1.18 | A two-dimensional representation of a line dislocation.

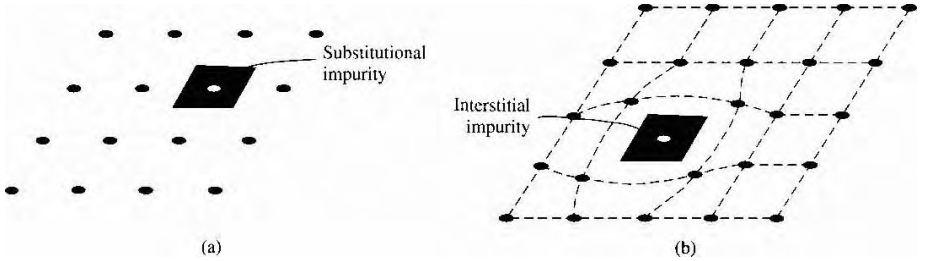


Figure 1.19 | Two-dimensional representation of a single-crystal lattice showing (a) a substitutional impurity and (b) an intersititat impurity.

1.5.2 Impurities in Solids

Foreign atoms, or impurity atoms, may be present in a crystal lattice. Impurity atoms may be located at normal lattice sites, in which case they are called *substitutional* impurities. Impurity atoms may also be located between normal sites, in which case they are called *interstitial* impurities. Both these impurities are lattice defects and are schematically shown in Figure 1.19. Some impurities, such as oxygen in silicon, tend to he essentially inert; however, other impurities, such as gold or phosphorus in silicon, can drastically alter the electrical properties of the material.

In Chapter 4 we will see that, by adding controlled amounts of particular impurity atoms, the electrical characteristics of a semiconductor material can be favorably altered. The technique of adding impurity atoms to a semiconductor material in order to change its conductivity is called *doping*. There are two general methods of doping: impurity diffusion and ion implantation.

The actual diffusion process depends to some extent on the material but, in general, impurity diffusion occurs when a semiconductor crystal is placed in a high-temperature (= 1000°C) gaseous atmosphere containing the desired impurity atom. At this high temperature, many of the crystal atoms can randomly move in and out of their single-crystal lattice sites. Vacancies may be created by this random motion so

that impurity atoms can move through the lattice by hopping from one vacancy to another. Impurity diffusion is the process by which impurity particles move from a region of high concentration near the surface, to a region of lower concentration within the crystal. When the temperature decreases, the impurity atoms become permanently frozen into the substitutional lattice sites. Diffusion of various impurities into selected regions of a semiconductor allows us to fabricate complex electronic circuits in a single semiconductor crystal.

Ion implantation generally takes place at a lower temperature than diffusion. A beam of impurity ions is accelerated to kinetic energies in the range of 50 keV or greater and then directed to the surface of the semiconductor. The high-energy impurity ions enter the crystal and come to rest at some average depth from the surface. One advantage of ion implantation is that controlled numbers of impurity atoms can be introduced into specific regions of the crystal. A disadvantage of this technique is that the incident impurity atoms collide with the crystal atoms. causing lattice-displacement damage. However, most of the lattice damage can he removed by thermal annealing, in which the temperature of the crystal is raised for a short time. Thermal annealing is a required step after implantation.

*1.6 | GROWTH OF SEMICONDUCTOR MATERIALS

The success in fabricating very large scale integrated (VLSI) circuits is a result, to a large extent, of the development of and improvement in the formation or growth of pure single-crystal semiconductor materials. Semiconductors are some of the purest materials. Silicon, for example, has concentrations of most impurities of less than 1 part in 10 billion. The high purity requirement means that extreme care is necessary in the growth and the treatment of the material at each step of the fabrication process. The mechanics and kinetics of crystal growth are extremely complex and will be described in only very general terms in this text. However, a general knowledge of the growth techniques and terminology is valuable.

1.6.1 Growth from a Melt

A common technique for growing single-crystal materials is called the ***Czochralski method.*** In this technique, a small piece of single-crystal material, known as a ***seed,*** is brought into contact with the surface of the same material in liquid phase, and then slowly pulled from the melt. As the seed is slowly pulled, solidification occurs along the plane between the solid–liquid interface. Usually the crystal is also rotated slowly as it is being pulled, to provide a slight stirring action to the melt, resulting in a more uniform temperature. Controlled amounts of specific impurity atoms, such as boron or phosphorus, may be added to the melt so that the grown semiconductor crystal is intentionally doped with the impurity atom. Figure 1.20 shows a schematic of the Czochralski growth process and a silicon ingot or boule grown by this process.

*Indicates sections that can be skipped without loss of continuity

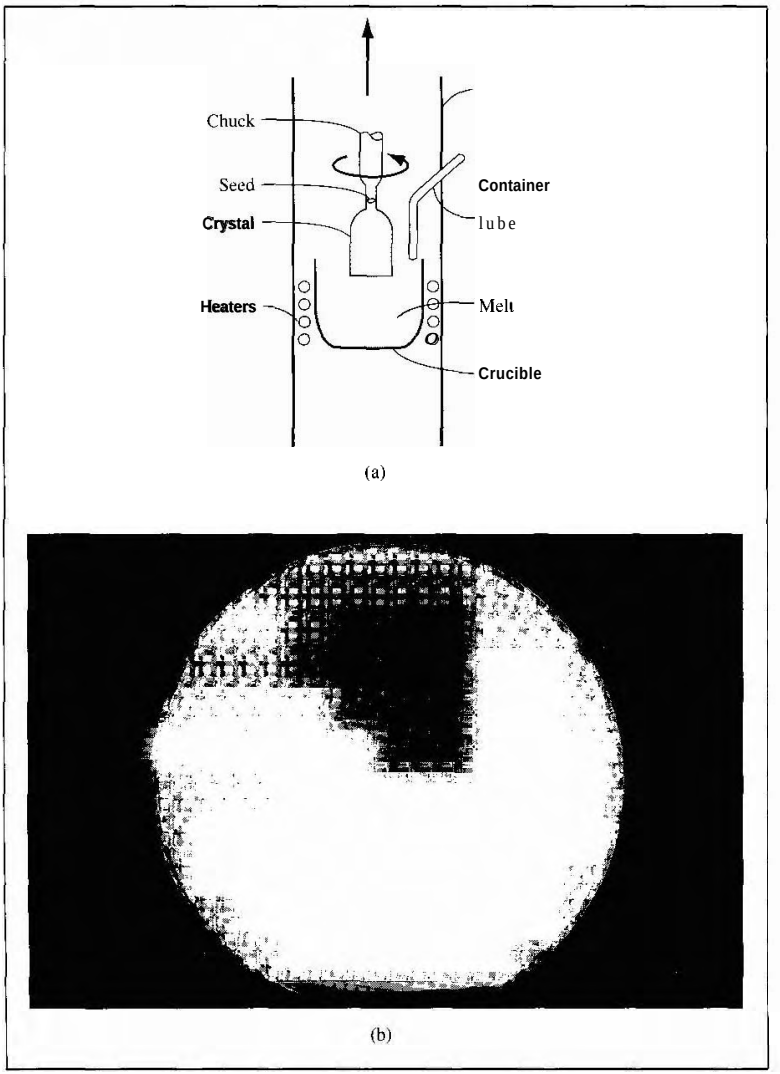


Figure 1.20 **(a)** Model of a crystal puller and (b) photograph of a silicon wafer with an may of integrated circuits. The circuits are tested on the wafer then sawed apan into chips that are mounted into packages. (Photo courtesy of Intel Corporation.)

Some impurities may be present in the ingot that are undesirable. Zone refining is a common technique for purifying material. A high-temperature coil. or r-f induction coil, is slowly passed along the length of the boule. The temperature induced by the coil is high enough so that a thin layer of liquid is formed. At the solid–liquid interface, there is a distrihution of impurities between the two phases. The parameter that describes this distrihution is called the ***segregation*** *coefficient:* the ratio of the concentration of impurities in the solid to the concentration in the liquid. If the segregation coefficient is 0.1, for example, the concentration of impurities in the liquid is a factor of 10 greater than that in the solid. As the liquid zone moves through the material. the impurities are driven along with the liquid. After several passes of the r-f coil, most impurities are at the end of the bar, which can then be cut off. The moving molten zone, or the zone-refining technique, can result in considerable purification.

After the semiconductor is grown, the boule is mechanically trimmed to the proper diameter and a Rat is ground over the entire length of the boule to denote the crystal orientation. The Rat is perpendicular to the [110] direction or indicates the (110) plane. (See Figure 1.20b.) This then allows the individual chips to be fabricated along given crystal planes so that the chips can he sawed apan more easily. The boule is then sliced into wafers. The wafer must he thick enough to mechanically support itself. A mechanical two-sided lapping operation produces a Rat wafer of uniform thickness. Since the lapping procedure can leave a surface damaged andcontaminated by the mechanical operation, the surface must be removed by chemical etching. The final step is polishing. This provides a smooth surface on which devices may be fabricated or further growth processes may be carried out. This final semiconductor wafer is called the substrate material.

1.6.2 Epitaxial Growth

A common and versatile growth technique that is used extensively in device and integrated circuit fabrication is epitaxial growth. ***Epitaxial growth*** is a process whereby a thin, single-crystal layer of material is grown on the surface of a single-crystal substrate. In the epitaxial process, the single-crystal substrate acts as the seed, although the process takes place far below the melting temperature. When an epitaxial layer is grown on a substrate of the same material, the process is termed ***homoepitaxy.*** Growing silicon on a silicon substrate is one example of a homoepitaxy process. At present, a great deal of work is being done with ***heteroepitaxy.*** In a heteroepitaxy process, although the substrate and epitaxial materials are not the same, the two crystal structures should be very similar if single-crystal growth is to he obtained and if a large number of defects are to be avoided at the epitaxial–substrate interface. Growing epitaxial layers of the ternary alloy AlGaAs on a GaAs substrate is one example of a heteroepitaxy process.

One epitaxial growth technique that has been used extensively is called ***chemical vapor-phase deposition (CVD).*** Silicon epitaxial layers, for example, are grown on silicon substrates by the controlled deposition of silicon atoms onto the surface from a chemical vapor containing silicon. In one method, silicon tetrachloride reacts with hydrogen at the surface of a heated substrate. The silicon atoms are released in

the reaction and can he deposited onto the substrate, while the other chemical reactant, HCI, is in gaseous form and is swept out of the reactor. A sharp demarcation between the impurity doping in the suhstrate and in the epitaxial layer can be achieved using the CVD process. This technique allows great flexibility in the fabrication of semiconductor devices.

***Liquid-phase** epitaxy* is another epitaxial growth technique. **A** compound of the semiconductor with another element may have a melting temperature lower than that of the semiconductor itself. The semiconductor substrate is held in the liquid compound and, since the temperature of the melt is lower than the melting temperature of the substrate, the substrate does not melt. As the solution is slowly cooled, a single-crystal semiconductor layer grows on the seed crystal. This technique, which occurs at a lower temperature than the Czochralski method, is useful in growing group III–V compound semiconductors.

A versatile technique for growing epitaxial layers is the *molecular beam epitaxy* (MBE) process. Asubstrate is held in vacuum at a temperature normally in the range of 400 to 800°C, a relatively low temperature compared with many semiconductor-processing steps. Semiconductor and dopant atoms are then evaporated onto the surface of the substrate. In this technique, the doping can he precisely controlled resulting in very complex doping profiles. Complex ternary compounds, such as AlGaAs, can be grown on substrates, such as GaAs, where abrupt changes in the crystal composition are desired. Many layers of various types of epitaxial compositions can be grown on a substrate in this manner. These structures are extremely beneficial in optical devices such as laser diodes.

1.7 | SUMMARY

- A few of the most common semiconductor (materialswere listed. Silicon is the most common semiconductor material.
- The properties of semiconductors and other materials are determined to a large extent by the single-crystal lattice structure. The unit cell is a small volume of the crystal that is used to reproduce the entire crystal. Three basic unit cells **are** the simple cuhic. hody-centered cubic. and face-centered cubic.
- Silicon has the diamond crystal structure. Atoms are formed in a tetrahedral contiguration with four nearest neighbor atoms. The binary semiconductors have a zincblende lattice, that is basically the came as the diamond lattice.
- Miller indices are used to describe planer in a crystal lattice. These planes may be used to describe the surface of a semicnnductor material. The Miller indices are also used to describe directions in a crystal.

  Imperfectionsdo exist in semiconductor materials. A few of these imperfections are vacancies, substitutiunal impurities, and interstitial impurities. Small amounts of controlled substitutional impurities can favorably alter semiconductor properties as we will see in later chapters.

  Abrief description of semiconductor growth methods was given. Bulk growth produces the starting semiconductor material or suhctrate. Epitaxial growth can be used to control the surface properties of a semiconductor Most semiconductor devices are fabricated in the epitaxial layer.

GLOSSARY OF IMPORTANT TERMS

binary semiconductor A two-element compound semiconductor, such as gallium arsenide (GaAs).

covalent bonding The bonding between atoms in which valence electrons are shared.

diamond lattice The atomic crystal structure of silicon, for example, in which each atom has four nearest neighbors in a tetrahedral configuration.

doping The process of adding specific types of atoms to a semiconductor to favorably alter the electrical characteristics.

elemental semiconductor A semiconductor composed of a single species of atom, such as silicon or germanium.

epitaxial layer A thin, single-crystal layer of material formed on the surface of a substrate.

ion implantation One particular process of doping a semiconductor.

lattice The periodic arrangement of atoms in a crystal.

Miller indices The set of integers used to describe a crystal plane.

primitive cell The smallest unit cell that can be repeated to form a lattice.

substrate A semiconductor wafer or other material used as the starling material for further semiconductor processing. such as epitaxial growth or diffusion.

ternary semiconductor A three-element compound semiconductor. such as aluminum gallium arsenide (AlGaAs).

unit cell A small volume of a crystal that can be used to reproduce the entire crystal.

zincblende lattice A lattice structure identical to the diamond lattice except that there arc two types of atoms instead of one.

CHECKPOINT

After studying this chapter, the reader should hare the ability to:

- Determine the volume density of atoms for various lattice structures.
- Determine the Miller indices of a crystal-lattice plane.
- Sketch a lattice plane riven the Miller indices.
- Determine the surface density of atoms on a given crystal-lattice plane.
- Understand and describe various defects in a single-crystal lattice.

REVIEW QUESTIONS

1. List two elemental semiconductor materials and two compound semiconductor materials.
2. Sketch three lattice structures: (*a*) simple cubic, (*b*) body-centered cubic, and (*c*) face-centered cubic.
3. Describe the procedure for finding the volume density of atoms in a crystal.
4. Describe the procedure for obtaining the Miller indices that describe a plane in a crystal.
5. What is meant by a substitutional impurity in a crystal? What is meant by an interstitial impurity'?

PROBLEMS

Section 1.3 Space Lattices

1.1 Determine the number of atoms per unit cell in a *(a)* face-centered cubic, (b) body-centered cubic, and *(c)* diamond lattice.

1.2 (a) The lattice constant of GaAs is 5.65 Å. Determine the number of Ga atoms and As atoms per cm^3. *(h)* Determine the volume density of germanium atoms in a germanium semiconductor. The lattice constant of germanium is 5.65 Å.

Assume that each atom is a hard sphere with the surface of each atom in contact with the surface of its nearest neighbor. Determine the percentage of total unit cell volume that is occupied in *(a)* a simple cubic lattice, (b) a face-centered cubic lattice, (c) a body-centered cubic lattice, and (d) a diamond lattice.

A material, with a volume of 1 cm', is composed of an fcc lattice with a lattice constant of 2.5 *mm*. The "atoms" in this material are actually coffee beans. Assume the coffee beans are hard spheres with each bean touching its nearest neighbor. Determine the volume of coffee after the coffee beans have been ground. (Assume 100 percent packing density of the ground coffee.)

1.5 If the lattice constant of silicon is 5.41 Å, calculate (a) the distance from the center of one silicon atom to the center of its nearest neighbor, *(h)* the number density of silicon atoms (#per cm^3), and *(c)* the mass density (grams per cm^3) of silicon.

1.6 A crystal is composed of two elements, A and B. The basic crystal structure is a body-centered cubic with elements A at each of the corners and element B in the center. The effective radius of element A is 1.02 Å. Assume the elements arc hard spheres with the surface of each A-type atom in contact with the surface of its nearest A-type neighhor. Calculate (a) the maximum radius of the B-type atom that will fit into this structure, and *(b)* the volume density (#/cm^3) of both the A-type atoms and the B-type atoms.

The crystal structure of sodium chloride (NaCl) is a simple cubic with the Na and Cl atoms alternating positions. Each Na atom is then surrounded by six Cl atoms and likewise each Cl atom is surrounded by six Na atoms. (a) Sketch the atoms in a (100) plane. (b) Assume the atoms arc hard spheres with nearest neighbors touching. The effective radius of Na is 1.0 Å and the effective radius of Cl is 1.8 Å. Determine the lattice constant. (c) Calculate the volume density of Na and Cl atoms. (d) Calculate the mass density of NaCl.

(a) A material is composed of two types of atoms. Atom A has an effective radius of 2.2 Å and atom B has an effective radius of 1.8 Å. The lattice is a bcc with atoms A at the comers and atom B in the center. Determine the lattice constant and the volume densities of A atoms and B atoms. (b) Repeat part (a) with atoms B at the corners and atom Ain the center *(c)* What comparison can be made of the materials in pans *(a)* and *(b)*?

Consider the materials described in Problem 1.8 in pans (a) and (b). For each case, calculate the surface density of A atoms and B atoms in the (110) plane. What corn parison can be made of the two materials?

1.10 *(a)* The crystal structure of a particular material consists of a single atom in the center of a cube. The lattice constant is a, and the diameter of the atom is a_0. Determine the volume density of atoms and the surface density of atoms in the (110) plane.
(b) Compare the results of part (a) to the results for the case of the simple cubic structure shown in Figure 1.5a with the same lattice constant.

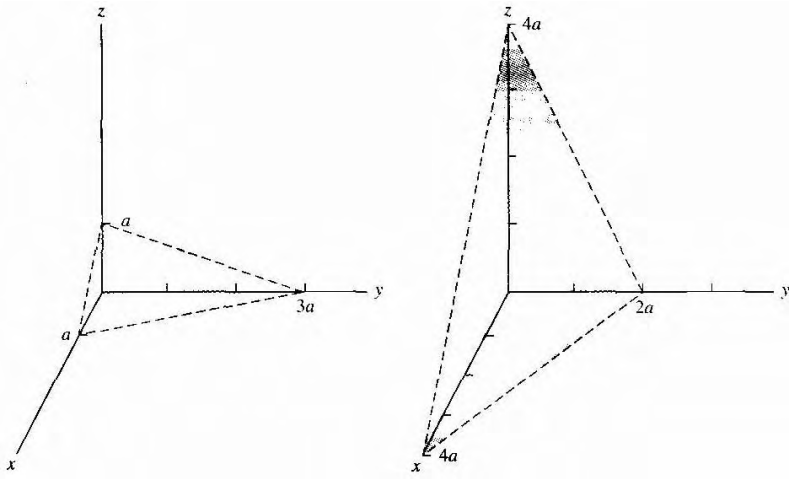


Figure 1.21 | Figure for Problem 1.12.

1.11 Consider a three-dimensional cubic lattice with a lattice constant equal loo. (a) Sketch the following planes: (*i*) (100), (*ii*) (110), (iii) (310), and (*iv*) (230). (*b*) Sketch the following directions: (i) [100], (ii) [110], (*iii*) [310], and (iv) [230].

1.12 For a simple cubic lattice, determine the Miller indices for the planes shown in Figure 1.21.

1.13 The lattice constant of a simple cubic cell is 5.63 Å. Calculate the distance between the nearest parallel (*a*) (100), (*b*) (110), and (*c*) (111) planes.

1.14 The lauice constant of a single crystal is 4.50 Å. Calculate the surface density of atoms (# per cm^2) on the following planes: (*i*) (100), (ii) (110), (*iii*) (111) for each of the following lattice structures: (*a*) simple cubic, (h) body-centered cubic, and (*c*) face-centered cubic.

1.15 Determine the surface density of atoms for silicon on the (a) (100) plane, (*b*) (110) plane, and (*c*) (111) plane.

1.16 Consider a face-centered cubic lattice. Assume the atoms are hard spheres with the surfaces of the nearest neighbors touching. Assume the radius of the atom is 2.25 Å. (*a*) Calculate the volume density of atoms in the crystal. (*b*) Calculate the distance between nearest (110) planes. (c) Calculate the surface density of atoms on the (I10) plane.

Section 1.4 Atomic Bonding

1.17 Calculate the density of valence electrons in silicon

1.18 The structure of GaAs is the zincblende lattice. The lattice constant is 5.65 Å. Calculate the density of valence electrons in GaAs.

Section 1.5 Imperfections and Impurities in Solids

1.19 (a) If 2×10^{16} boron atoms per cm' are added to silicon as a substitutional impurity, determine what percentage of the silicon atoms are displaced in the single crystal lattice. (*b*) Repeat pall (*a*) for 10^{15} boron atoms per cm^3.

1.20 (a) Phosphorus atoms, at a concentration of $5 \times 10^{16}\ cm^{-3}$, are added to a pure sample of silicon. Assume the phosphorus atoms are distributed homogeneously throughout the silicon. What is the fraction by weight of phosphorus'? (*b*) If boron atoms, at a concentration of $10^{18}\ cm^{-3}$, are added to the material in part (a), determine the fraction by weight of boron.

1.21 If 2×10^{15} gold atoms per cm^3 are added to silicon as a substitutional impurity and are distributed uniformly throughout the semiconductor, determine the distance between gold atoms in terms of the silicon lattice constant. (Assume the gold atoms are distributed in a rectangular or cubic array.)

READING LIST

1. Azaroff. L.V., and J. J. Brophy. Electronic *Processes* in *Materials*. New York: McGraw-Hill, 1963.
2. Campbell, S. A. The Science *and Engineering* of Microelectronic *Fabrication*. New York: Oxford University Press, 1996.
3. Kittel, C. *Introduction* to Solid Stare *Physics*, 7th ed. Berlin: Springer-Verlag, 1993.

*4. Li, S. S. Semiconductor Physical Electronics. New York: Plenum Press, 1993.

5. McKelvey, J. P. Solid State Physicsfor *Engineering and Materials* Science. Malabar, FL: Krieger, 1993.
6. Pierret, R. F. Semiconductor *Device* Fundamentals. Reading, MA: Addison-Wesley, 1996.
7. Runyan, W. R., and K. E. Bean. *Semiconductor Integrated* Circuit Processing and Technology Reading, MA: Addison-Wesley, 1990.
8. Singh, J. *Semiconductor Devices:* Basic *Principles*. New York: John Wiley and Sons, 2001.
9. Streetman, B. G., and S. Banerjee. *Solid* State Electronic *Devices*, 5th ed. Upper Saddle River, NJ: Prentice Hall, 2000.
10. Sze, S. M. VLSI *Technology*. New York: McGraw-Hill, 1983.

*11. Wolfe, C. M., N. Holonyak. Jr., and G. E. Stillman. Physical Properties *of Semiconductors*. Englewood Cliffs. NJ: Prentice Hall, 1989.

*Indicates references that are at an advanced level compared to this text.

CHAPTER 2

Introduction to Quantum Mechanics

PREVIEW

The goal of this text is to help readers understand the operation and characteristics of semiconductor devices. Ideally, we would like to begin discussing these devices immediately. However, in order to understand the current–voltage characteristics, we need some knowledge of the electron behavior in a crystal when the electron is subjected to various potential functions.

The motion of large objects, such as planets and satellites, can be predicted to a high degree of accuracy using classical theoretical physics based on Newton's laws of motion. But certain experimental results, involving electrons and high-frequency electromagnetic waves, appear to he inconsistent with classical physics. However, these experimental results can be predicted by the principles of quantum mechanics. The quantum mechanical wave theory is the basis for the theory of semiconductor physics.

We are ultimately interested in semiconductor materials whose electrical properties arc directly related to the behavior of electrons in the crystal lattice. The behavior and characteristics of these electrons can be described by the formulation of quantum mechanics called wave mechanics. The essential elements of this wave mechanics, using Schrodinger's wave equation, are presented in this chapter.

The goal of this chapter is to provide a brief introduction to quantum mechanics so that readers gain an understanding of and become comfortable with the analysis techniques. This introductory material forms the basis of semiconductor physics..

2.1 | PRINCIPLES OF QUANTUM MECHANICS

Before we delve into the mathematics of quantum mechanics. there are three principles we need to consider: the principle of energy quanta, the wave–particle duality principle, and the uncertainty principle.

2.1.1 Energy Quanta

One experiment that demonstrates an inconsistency between experimental results and the classical theory of light is called the photoelectric effect. If monochromatic light is incident on a clean surface of a material, then under certain conditions, electrons (photoelectrons) are emitted from the surface. According to classical physics. if the intensity of the light is large enough, the work function of the material will be overcome and an electron will be emitted from the surface independent of the incident frequency. This result is not observed. The observed effect is that, at a constant incident intensity, the maximum kinetic energy of the photoelectron varies linearly with frequency with a limiting frequency $\nu = \nu_0$, below which no photoelectron is produced. This result is shown in Figure 2.1. If the incident intensity varies at a constant frequency, the rate of photoelectron emission changes, but the maximum kinetic energy remains the same.

Planck postulated in 1900 that thermal radiation is emitted from a heated surface in discrete packets of energy called *quanta*. The energy of these quanta is given by $E = h\nu$, where ν is the frequency of the radiation and h is a constant now known as Planck's constant ($h = 6.625 \times 10^{-34}$ J-s). Then in 1905. Einstein interpreted the photoelectric results by suggesting that the energy in a light wave is also contained in discrete packets or bundles. The particle-like packet of energy is called a photon, whose energy is also given by $E = h\nu$. A photon with sufficient energy, then, can knock an electron from the surface of the material. The minimum energy required to remove an electron is called the work function of the material

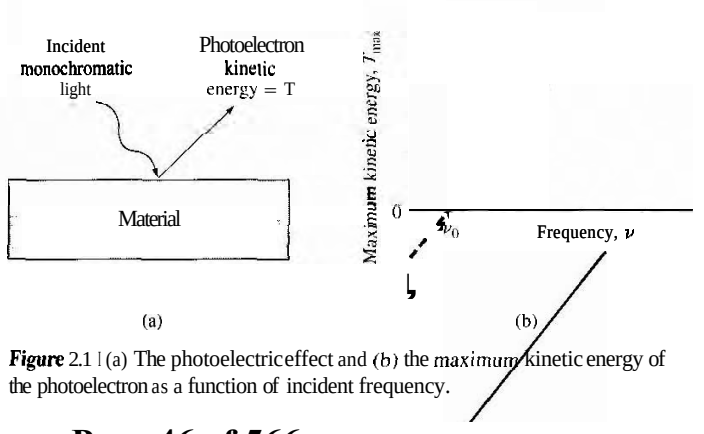


Figure 2.1 | (a) The photoelectric effect and (b) the maximum kinetic energy of the photoelectron as a function of incident frequency.

and any excess photon energy goes into the kinetic energy of the photoelectron. This result was confirmed experimentally as demonstrated in Figure 2.1. The photoelectric effect shows the discrete nature of the photon and demonstrates the particle-like behavior of the photon.

The maximum kinetic energy of the photoelectron can be written as

$$T_{max} = \frac{1}{2}mv^2 = h\nu - h\nu_0 \quad (\nu \geq \nu_0) \tag{2.1}$$

where $h\nu$ is the incident photon energy and $h\nu_0$ is the minimum energy, or work function, required to remove an electron from the surface.

EXAMPLE 2.1 Objective

To calculate the photon energy corresponding to a particular wavelength.

Consider an x-ray with a wavelength of $\lambda = 0.708 \times 10^{-8}$ cm.

■ **Solution**

The energy is

$$E = h\nu = \frac{hc}{\lambda} = \frac{(6.625 \times 10^{-34})(3 \times 10^{10})}{0.708 \times 10^{-8}} = 2.81 \times 10^{-15} \text{ J}$$

This value of energy may be given in the more common unit of electron-volt (see Appendix F). We have

$$E = \frac{2.81 \times 10^{-15}}{1.6 \times 10^{-19}} = 1.75 \times 10^{4} \text{ eV}$$

■ **Comment**

The reciprocal relation between photon energy and wavelength is demonstrated: A large energy corresponds to a short wavelength.

2.1.2 Wave–Particle Duality

We have seen in the last section that light waves, in the photoelectric effect, behave as if they are particles. The particle-like behavior of electromagnetic waves was also instrumental in the explanation of the Compton effect. In this experiment. an x-ray beam was incident on a solid. A portion of the x-ray beam was deflected and the frequency of the deflected wave had shifted compared to the incident wave. The observed change in frequency and the deflected angle corresponded exactly to the expected results of a "billiard ball" collision between an x-ray quanta, or photon, and an electron in which both energy and momentum are conserved.

In 1924. de Broglie postulated the existence of matter waves. He suggested that since waves exhibit particle-like behavior, then panicles should he expected to show wave-like properties. The hypothesis of de Broglie was the existence of a

wave–particle duality principle. The momentum of a photon is given by

$$p = \frac{h}{\lambda} \tag{2.2}$$

where A is the wavelength of the light wave. Then, de Broglie hypothesized that the wavelength of a particle can be expressed as

$$\lambda = \frac{h}{p} \tag{2.3}$$

wherep is the momentum of the particle and A is known as the *de Broglie wavelength* of the matter wave.

The wave nature of electrons has been tested in several ways. In one experiment by Davisson and Germer in 1927, electrons from a heated filament were accelerated at normal incidence onto a single crystal of nickel. A detector measured the scattered electrons as a function of angle. Figure 2.2 shows the experimental setup and Figure 2.3 shows the results. The existence of a peak in the density of scattered electrons can be explained as a constructive interference of waves scattered by the periodic atoms in the planes of the nickel crystal. The angular distribution is very similar to an interference pattern produced by light diffracted from a grating.

In order to gain some appreciation of the frequencies and wavelengths involved in the wave-particle duality principle. Figure 2.4 shows the electromagnetic frequency spectrum. We see that a wavelength of 72.7 A obtained in the next example is in the ultraviolet range. Typically, we will be considering wavelengths in the

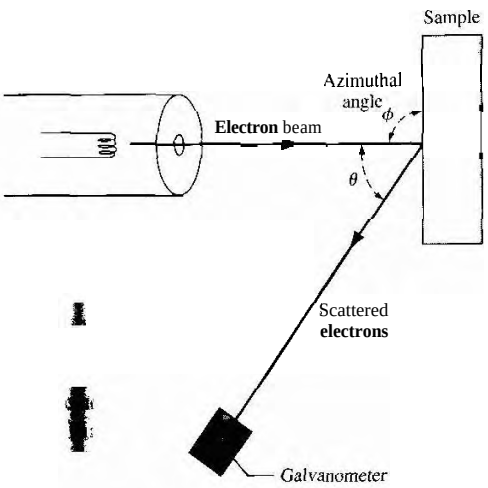


Figure 2.2 | Experimental arrangement of the Davisson–Germer experiment.

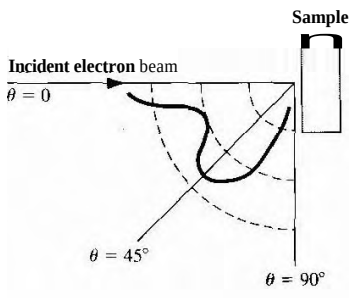


Figure 2.3 | Scattered electron flux as a function of scattering angle for the Davisson–Germer experiment.

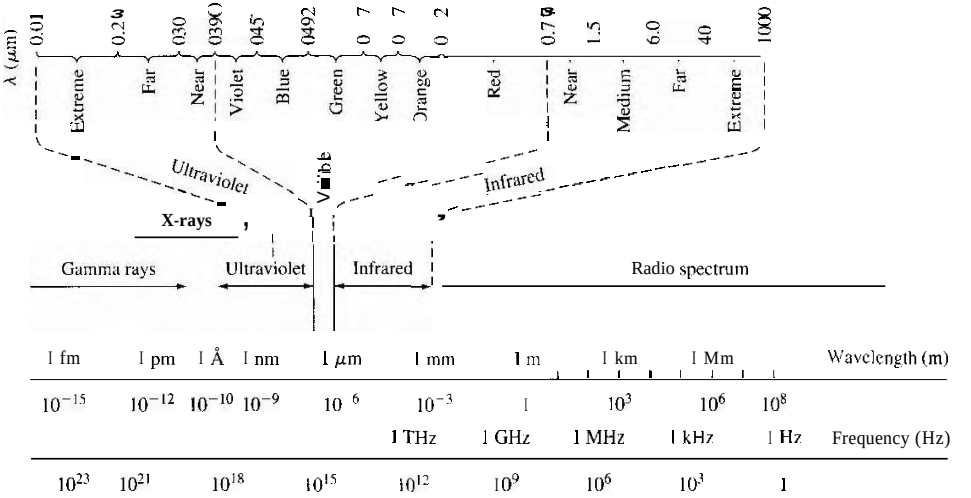


Figure 2.4 | The electromagnetic frequency spectrum

ultraviolet and visible range. These wavelengths are very short compared to the usual radio spectrum range.

EXAMPLE 2.2 | Objective

To calculate the de Broglie wavelength of a panicle.

Concidcr an electron traveling at a velocity of 10^7 cm/sec $= 10^5$ m/s.

■ **Solution**

The momentum is given by

$$p = mv = (9.11 \times 10^{-31})(10^5) = 9.11 \times 10^{-26}$$

Then, the de Broglie wavelength is

$$\lambda = \frac{h}{p} = \frac{6.625 \times 10^{-34}}{9.11 \times 10^{-26}} = 7.27 \times 10^{-9} \text{ m}$$

or

$$\lambda = 72.7 \text{ Å}$$

■ **Comment**

This calculation shows the order of magnitude of the de Broglie wavelength for a "typical" electron.

In some cases electromagnetic waves behave as if they are particles (photons) and sometimes particles behave as if they are waves. This wave–particle duality

principle of quantum mechanics applies primarily to small particles such as electrons, but it has also been shown to apply to protons and neutrons. For very large particles, we can show that the relevant equations reduce to those of classical mechanics. The wave–particle duality principle is the basis on which we will use wave theory to describe the motion and behavior of electrons in a crystal.

TEST YOUR UNDERSTANDING

E2.1 Determine the energy of a photon having wavelengths of (a) $\lambda = 10.000$ Å and (b) $\lambda = 10$ Å. [Ans. (a) 1.99×10^{-19} J or 1.24 eV; (b) 1.99×10^{-16} J or 1.24×10^{3} eV]

E2.2 (a) Find the momentum and energy of a particle with mass of 5×10^{-31} kg and a de Broglie wavelength of 180 Å. (b) An electron has a kinetic energy of 20 meV. Determine the de Broglie wavelength. [Ans. (a) $p = 3.68 \times 10^{-26}$ kg-m/s, $E = 1.35 \times 10^{-21}$ J or 8.46×10^{-3} eV; (b) $\lambda = 86.7$ Å, $p = 7.64 \times 10^{-26}$ kg-m/s]

2.1.3 The Uncertainty Principle

The Heisenberg uncertainty principle, given in 1927, also applies primarily to very small particles, and states that we cannot describe with absolute accuracy the hchavior of these subatomic particles. The uncertainty principle describes a fundamental relationship between conjugate variables, including position and momentum and also energy and time.

The first statement of the uncertainty principle is that it is impossible to simultaneously describe with absolute accuracy the position and momentum of a particle. If the uncertainty in the momentum is Δp and the uncertainty in the postion is Δx, then the uncenainty principle is stated as[1]

$$\Delta p \, \Delta x \geq \hbar \tag{2.4}$$

where $\hbar$ is defined as $\hbar = h/2\pi = 1.054 \times 10^{-34}$ J-s and is called a modified Planck's constant. This statement may be generalized to include angular position and angular momentum.

The second statement of the uncertainty principle is that it is impossible to simultaneously describe with absolute accuracy the energy of a particle and the instant of time the particle has this energy. Again. if the uncertainty in the energy is given by AE and the uncertainty in the time is given by Ar, then the uncertainty principle is stated as

$$\Delta E \, \Delta t \geq \hbar \tag{2.5}$$

One way to visualize the uncenainty principle is to consider the simultaneous measurement of position and momentum, and the simultaneous measurement of energy and time. The uncertainty principle implies that these simultaneous measurements

[1] In some texts, the uncenainty principle is stated as $\Delta p \, \Delta x \geq \hbar/2$. We are interested here in the order of magnitude and will not be concerned with small differences.

are in error to a certain extent. However, the modified Planck's constant $\hbar$ is very small; the uncertainty principle is only significant for subatomic particles. We must keep in mind nevertheless that the uncertainty principle is a fundamental statement and does not deal only with measurements.

One consequence of the uncertainty principle is that we cannot. for example. determine the exact position of an electron. We will, instead, determine the *probability* of finding an electron at a particular position. In later chapters, we will develop a *probability density function* that will allow us to determine the probability that an electron has a particular energy. So in describing electron behavior, we will he dealing with probability functions.

TEST YOUR UNDERSTANDING

E2.3 The uncertainty in position of an electron is 12 Å. Determine the minimum uncertainty in momentum and also the corresponding uncertainly in kinetic energy. (Ans. $\Delta p = 8.79 \times 10^{-26}$ kg-m/s, $\Delta E = 0.0265$ eV)

E2.4 An electron's energy is measured with an uncertainty of 1.2 eV. What is the minimum uncertainty in time over which the energy is measured? (Ans. $\Delta t = 5.49 \times 10^{-16}$ s)

2.2 | SCHRODINGER'S WAVE EQUATION

The various experimental results involving electromagnetic waves and particles. which could not be explained by classical laws of physics, showed that a revised formulation of mechanics was required. Schrodinger, in 1926. provided a formulation called *wave mechanics,* which incorporated the principles of quanta introduced by Planck, and the wave–particle duality principle introduced by de Broglie. Based on the wave–particle duality principle. we will describe the motion of electrons in a crystal by wave theory. This wave theory is described by Schrodinger's wave equation.

2.2.1 The Wave Equation

The one-dimensional, nonrelativistic Schrodinger's wave equation is given by

$$\frac{-\hbar^2}{2m} \cdot \frac{\partial^2 \Psi(x,t)}{\partial x^2} + V(x)\Psi(x,t) = j\hbar \frac{\partial \Psi(x,t)}{\partial t} \tag{2.6}$$

where $\Psi(x,t)$ is the wave function, $V(x)$ is the potential function assumed to he independent of time, m is the mass of the particle, and j is the imaginary constant $\sqrt{-1}$. There are theoretical arguments that justify the form of Schrodinger's wave equation. hut the equation is a basic postulate of quantum mechanics. The wave function $\Psi(x,t)$ will be used to describe the behavior of the system and, mathematically, $\Psi(x,t)$ can be a complex quantity.

We may determine the time-dependent portion of the wave function and the position-dependent, or time-independent, portion of the wave function by using the

technique of separation of variables. Assume that the wave function can he written in the form

$$\Psi(x,t) = \psi(x)\phi(t) \tag{2.7}$$

where $\psi(x)$ is a function of the position x only and $\phi(t)$ is a function of time t only. Substituting this form of the solution into Schrodinger's wave equation, we obtain

$$\frac{-\hbar^2}{2m}\phi(t)\frac{\partial^2\psi(x)}{\partial x^2} + V(x)\psi(x)\phi(t) = j\hbar\psi(x)\frac{\partial\phi(t)}{\partial t} \tag{2.8}$$

If we divide by the total wave function. Equation (2.8) becomes

$$\frac{-\hbar^2}{2m}\frac{1}{\psi(x)}\frac{\partial^2\psi(x)}{\partial x^2} + V(x) = j\hbar\cdot\frac{1}{\phi(t)}\cdot\frac{\partial\phi(t)}{\partial t} \tag{2.9}$$

Since the left side of Equation ***(2.9)*** is a function of position x only and the right side of the equation is a function of time t only, each side of this equation must he equal to aconstant. We will denote this separation of variables constant by η.

The time-dependent portion of Equation (2.9) is then written as

$$\eta = j\hbar\cdot\frac{1}{\phi(t)}\cdot\frac{\partial\phi(t)}{\partial t} \tag{2.10}$$

where again the parameter η is called a separation constant. The solution of Equation (2.10) can be written in the form

$$\phi(t) = e^{-j(E/\hbar)t} \tag{2.11}$$

The form of this solution is the classical exponential form of a sinusoidal wave where $\eta/\hbar$ is the radian frequency w. We have that $E = h\nu$ or $E = h\omega/2\pi$. Then $\omega = \eta/\hbar = E/\hbar$ so that the separation constant is equal to the total energy E of the particle.

The time-independent portion of Schrodinger's wave equation can now he written from Equation (2.9) as

$$\frac{-\hbar^2}{2m}\cdot\frac{1}{\psi(x)}\cdot\frac{\partial^2\psi(x)}{\partial x^2} + V(x) = E \tag{2.12}$$

where the separation constant is the total energy E of the particle. Equation (2.12) may he written as

$$\boxed{\frac{\partial^2\psi(x)}{\partial x^2} + \frac{2m}{\hbar^2}(E - V(x))\psi(x) = 0} \tag{2.13}$$

where again m is the mass of the particle, $V(x)$ is the potential experienced by the particle, and E is the total energy of the particle. This time-independent Schrodinger's wave equation can also be justitied on the basis of the classical wave equation as

shown in Appendix E. The pseudo-derivation in the appendix is a simple approach but shows the plausibility of the time-independent Schrodinger's equation.

2.2.2 Physical Meaning of the Wave Function

We are ultimately trying to use the wave function $\Psi(x,t)$ to describe the behavior of an electron in a crystal. The function $\Psi(x,t)$ is a wave function. so it is reasonable to ask what the relation is between the function and the electron. The total wave function is the product of the position-dependent, or time-independent, function and the time-dependent function. We have from Equation (2.7) that

$$\boxed{\Psi(x,t) = \psi(x)\phi(t) = \psi(x)e^{-j(E/\hbar)t}} \tag{2.14}$$

Since the total wave function $\Psi(x,t)$ is a complex function. it cannot by itself represent a real physical quantity.

Max Born postulated in 1926 that the function $|\Psi(x,t)|^2\,dx$ is the probability of finding the particle between x and $x + dx$ at a given time, or that $|\Psi(x,t)|^2$ is a probability density function. We have that

$$|\Psi(x,t)|^2 = \Psi(x,t)\cdot\Psi^*(x,t) \tag{2.15}$$

where $\Psi^*(x,t)$ is the complex conjugate function. Therefore

$$\Psi^*(x,t) = \psi^*(x)\cdot e^{+j(E/\hbar)t}$$

Then the product of the total wave function and its complex conjugate is given by

$$\Psi(x,t)\Psi^*(x,t) = \left[\psi(x)e^{-j(E/\hbar)t}\right]\left[\psi^*(x)e^{+j(E/\hbar)t}\right] = \psi(x)\psi^*(x) \tag{2.16}$$

Therefore, we have that

$$|\Psi(x,t)|^2 = \psi(x)\psi^*(x) = |\psi(x)|^2 \tag{2.17}$$

is the probability density function and is independent of time. One major difference between classical and quantum mechanics is that in classical mechanics, the position of a particle or body can be determined precisely, whereas in quantum mechanics, the position of aparticlr is found in terms o f a probability. We will determine the probability density function for several examples, and, since this property is independent of time. we will, in general, only be concerned with the time-independent wave function.

2.2.3 Boundary Conditions

Since the function $|\Psi(x,t)|^2$ represents the probability density function, then for a single particle. we must have that

$$\boxed{\int_{-\infty}^{\infty}|\psi(x)|^2\,dx = 1} \tag{2.18}$$

The probability of finding the particle somewhere is certain. Equation (2.18) allows us to normalize the wave function and is one boundary condition that is used to determine some wave function coefficients.

The remaining boundary conditions imposed on the wave function and its derivative are postulates. However. we may state the boundary conditions and present arguments that justify why they must be imposed. The wave function and its first derivative must have the following properties if the total energy E and the potential $V(x)$ are finite everywhere.

Condition 1. $\psi(x)$ must be finite, single-valued, and continuous.

Condition 2. $\partial\psi(x)/\partial x$ must be finite, single-valued, and continuous.

Since $|\psi(x)|^2$ is a probability density, then $\psi(x)$ must be finite and single-valued. If the probability density were to become infinite at some point in space, then the probability of finding the particle at this position would be certain and the uncertainty principle would be violated. If the total energy E and the potential $V(x)$ are finite everywhere, then from Equation (2.13), the second derivative must be finite, which implies that the first derivative must be continuous. The first derivative is related to the particle momentum, which must be finite and single-valued. Finally, a finite first derivative implies that the function itself must he continuous. In some of the specific examples that we will consider, the potential function will become infinite in particular regions of space. For these cases. the first derivative will not necessarily be continuous, but the remaining boundary conditions will still hold.

2.3 | APPLICATIONS OF SCHRODINGER'S WAVE EQUATION

We will now apply Schrodinger's wave equation in several examples using various potential functions. These examples will demonstrate the techniques used in the solution of Schrodinger's differential equation and the results of these examples will provide an indication of the electron behavior under these various potentials. We will utilize the resulting concepts later in the discussion of semiconductor properties.

2.3.1 Electron in Free Space

As a first example of applying the Schrodinger's wave equation, consider the motion of an electron in free space. If there is no force acting on the particle, then the potential function $V(x)$ will be constant and we must have $E > V(x)$. Assume, for simplicity, that the potential function $V(x) = 0$ for all x. Then, the time-independent wave equation can he written from Equation (2.13) as

$$\frac{\partial^2\psi(x)}{\partial x^2} + \frac{2mE}{\hbar^2}\psi(x) = 0 \tag{2.19}$$

The solution to this differential equation can be written in the form

$$\psi(x) = A\exp\left[\frac{jx\sqrt{2mE}}{\hbar}\right] + B\exp\left[\frac{-jx\sqrt{2mE}}{\hbar}\right] \tag{2.20}$$

Recall that the time-dependent portion of the solution is

$$\psi(t) = e^{-j(E/\hbar)t} \tag{2.21}$$

Then the total solution for the wave function is given by

$$\Psi(x,t) = A\exp\left[\frac{j}{\hbar}(x\sqrt{2mE} - Et)\right] + B\exp\left[\frac{-j}{\hbar}(x\sqrt{2mE} + Et)\right] \tag{2.22}$$

This wave function solution is a traveling wave, which means that a particle moving in free space is represented by a traveling wave. The first term, with the coefficient A. is a wave traveling in the $+x$ direction, while the second term, with the coefficient B. is a wave traveling in the x direction. The value of these coefficients will he determined from boundary conditions. We will again see the traveling-wave solution for an electron in a crystal or semiconductor material.

Assume, for a moment, that we have a particle traveling in the $+x$ direction. which will be described by the $+x$ traveling wave. The coefficient B = 0. We can write the traveling-wave solution in the form

$$\Psi(x,t) = A\exp[j(kx - \omega t)] \tag{2.23}$$

where k is a wave number and is

$$k = \frac{2\pi}{\lambda} \tag{2.24}$$

The parameter A is the wavelength and, comparing Equation (2.23) with Equation (2.22), the wavelength is given by

$$\lambda = \frac{h}{\sqrt{2mE}} \tag{2.25}$$

From de Broglie's wave–particle duality principle, the wavelength is also given by

$$\lambda = \frac{h}{p} \tag{2.26}$$

A free particle with a well-defined energy will also have a well-defined wavelength and momentum.

The probability density function is $\Psi(x,t)\Psi^*(x,t) = AA^*$, which is aconstanr independent of position. A free particle with a well-defined momentum can be found anywhere with equal probahility. This result is in agreement with the Heisenberg uncertainty principle in that a precise momentum implies an undefined position.

A localized free particle is defined by a wave packet, formed by a superposition of wave functions with different momentum or k values. We will not consider the wave packet here.

2.3.2 The Infinite Potential Well

The problem of a particle in the infinite potential well is a classic example of a bound particle. The potential $V(x)$ as a function of position for this problem is shown in

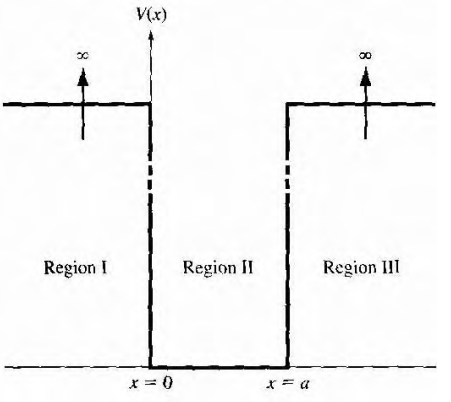


Figure 2.5 Potential function of the infinite potential well.

Figure 2.5. The particle is assumed to exist in region II so the particle is contained within a finite region of space. The time-independent Schrodinger's wave equation is again given by Equation *(2.13)* as

$$\frac{\partial^2\psi(x)}{\partial x^2} + \frac{2m}{\hbar^2}(E - V(x))\psi(x) = 0 \qquad (2.13)$$

where E is the total energy of the particle. If E is finite, the wave function must be zero, or $\psi(x) = 0$, in both regions I and III. A particle cannot penetrate these infinite potential barriers, so the probability of finding the particle in regions I and III is zero.

The time-independent Schrodinger's wave equation in region II, where $V = 0$. becomes

$$\frac{\partial^2\psi(x)}{\partial x^2} + \frac{2mE}{\hbar^2}\psi(x) = 0 \qquad (2.27)$$

A particular form of solution to this equation is given by

$$\psi(x) = A_1 \cos Kx + A_2 \sin Kx \qquad (2.28)$$

where

$$K = \sqrt{\frac{2mE}{\hbar^2}} \qquad (2.29)$$

One boundary condition is that the wave function $\psi(x)$ must be continuous so that

$$\psi(x = 0) = \psi(x = a) = 0 \qquad (2.30)$$

Applying the boundary condition at $x = 0$, we must have that $A_1 = 0$. At x = a, we have

$$\psi(x = a) = 0 = A_2 \sin Ka \tag{2.31}$$

This equation is valid if Ka $= n\pi$, where the parameter n is a positive integer, or n = 1, 2, 3, ... The parameter n is referred to as a quantum number. We can write

$$K = \frac{n\pi}{a} \tag{2.32}$$

Negative values of n simply introduce a negative sign in the wave function and yield redundant solutions for the probability density function. Wc cannot physically distinguish any difference between $+n$ and $-n$ solutions. Because of this redundancy, negative values of n are not considered.

The coefficient A_2 can be found from the normalization boundary condition that was given by Equation (2.18) as $\int_{-\infty}^{\infty} \psi(x)\psi^*(x)\,dx = 1$. If we assume that the wave function solution $\psi(x)$ is a real function, then $\psi(x) = \psi^*(x)$. Substituting the wave function into Equation (2.18), we have

$$\int_0^a A_2^2 \sin^2 Kx\,dx = 1 \tag{2.33}$$

Evaluating this integral gives[2]

$$A_2 = \sqrt{\frac{2}{a}} \tag{2.34}$$

Finally. the time-independent wave solution is given by

$$\psi(x) = \sqrt{\frac{2}{a}} \sin\left(\frac{n\pi x}{a}\right) \qquad \text{where } n = 1, 2, 3, \ldots \tag{2.35}$$

This solution represents the electron in the intinite potential well and is a standing wave solution. The free electron was represented by a traveling wave. and now the bound particle is represented by a standing wave.

The parameter K in the wave solution was defined by Equations (2.29) and (2.32). Equating these two expressions for K. we obtain

$$\frac{2mE}{\hbar^2} = \frac{n^2\pi^2}{a^2} \tag{2.36}$$

[2]A more thorough analysis shows that $|A_2|^2 = 2/a$, so solutions lor the coefficient A_2 include $+\sqrt{2/a}$, $-\sqrt{2/a}$, $+j\sqrt{2/a}$, $-j\sqrt{2/a}$, or any complex number whose magnitude is $\sqrt{2/a}$. Since the wave function itself has no physical meaning, the choice of which coefficient to use is immaterial: They all produce the same probability density function.

The total energy can then be written as

$$E = E_n = \frac{\hbar^2 n^2 \pi^2}{2ma^2} \qquad \text{where } n = 1, 2, 3, \ldots \tag{2.37}$$

For the particle in the infinite potential well. the wave function is now given by

$$\psi(x) = \sqrt{\frac{2}{a}} \sin Kx \tag{2.38}$$

where the constant K must have discrete values, implying that the total energy of the particle can only have discrete values. This result *means* that the energy *of the* particle is quantized. That is, *the energy of the* particle can only have particular discrete *values.* The quantization of the particle energy is contrary to results from classical physics, which would allow the particle to have continuous energy values. The discrete energies lead to quantum states that will be considered in more detail in this and later chapters. The quantization of the energy of a bound particle is an extremely important result.

Objective | **EXAMPLE 2.3**

Tocalculate the first three energy levels of an electron in an infinite potential well

Consider an electron in an infinite potential well of width 5 Å.

■ Solution

From Equation **(2.37)** we have

$$E_n = \frac{(\hbar^2 n^2 \pi^2)}{2ma^2} = \frac{n^2 (1.054 \times 10^{-34})^2 \pi^2}{2(9.11 \times 10^{-31})(5 \times 10^{-10})^2} = n^2 (2.41 \times 10^{-19}) \text{ J}$$

$$E_n = \frac{n^2 (2.41 \times 10^{-19})}{1.6 \times 10^{-19}} = n^2 (1.51) \text{ eV}$$

Then.

$$E_1 = 1.51 \text{ eV}, \quad E_2 = 6.04 \text{ eV}, \quad E_3 = 13.59 \text{ eV}$$

■ Comment

This calculation shows the order of magnitude *of* the energy levels of a bound electron.

Figure 2.6a shows the first four allowed energies for the particle in the infinite potential well, and Figures 2.6b and 2.6c show the corresponding wave functions and probability functions. We may note that as the energy increases, the probability of finding the particle at any given value of x becomes more uniform.

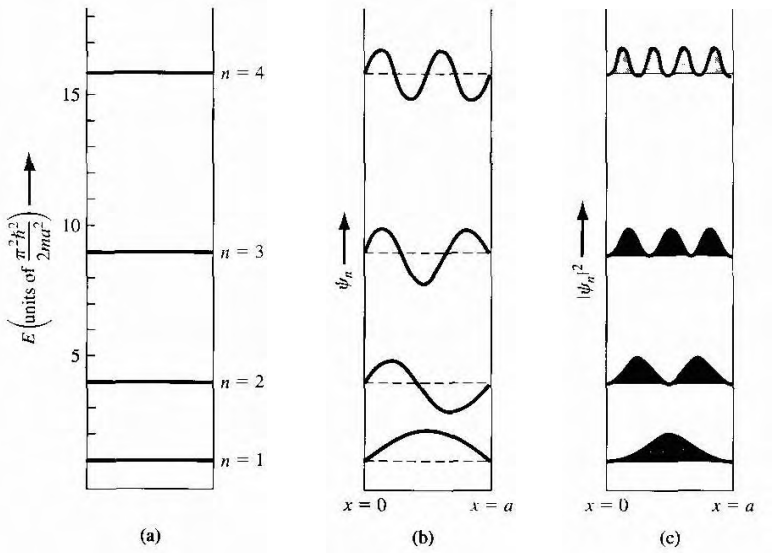


Figure 2.6 | Particle in an infinite potential well: (a) Four lowest discrete energy levels. (b) Corresponding wave functions. (c) Corresponding probability functions. *(From Pierret [9].)*

TEST YOUR UNDERSTANDING

E2.5 The width of the infinite potential well in Example 2.3 is doubled to 10 Å. Calculate the first three energy levels in terms of electron volts for an electron. (Ans. 0.376 eV, 1.50 eV, 3.38 eV)

E2.6 The lowest energy of a particle in an infinite potential well with a width of 100 Å is 0.025 eV. What is the mass of the particle? (Ans. 1.37×10^{-31} kg)

2.3.3 The Step Potential Function

Consider now a step potential function as shown in Figure 2.7. In the previous section. we considered a particle being confined between two potential barriers. In this example, we will assume that a flux of particles is incident on the potential barrier. We will assume that the particles are traveling in the $+x$ direction and that they originated at $x = -\infty$. A particularly interesting result is obtained for the case when the total energy of the particle is less than the barrier height, or $E < V_0$.

We again need to consider the time-independent wave equation in each of the two regions. This general equation was given in Equation (2.13) as $\partial^2\psi(x)/\partial x^2 + 2m/\hbar^2(E - V(x))\psi(x) = 0$. The wave equation in region I, in which $V = 0$, is

$$\frac{\partial^2\psi_1(x)}{\partial x^2} + \frac{2mE}{\hbar^2}\psi_1(x) = 0 \tag{2.39}$$

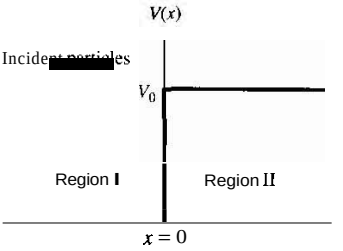


Figure 2.7 | The step potential function.

The general solution to this equation can be written in the form

$$\psi_1(x) = A_1 e^{jK_1 x} + B_1 e^{-jK_1 x} \quad (x \leq 0) \tag{2.40}$$

where the constant K_1 is

$$K_1 = \sqrt{\frac{2mE}{\hbar^2}} \tag{2.41}$$

The first term in Equation (2.40) is a traveling wave in the $+x$ direction that represents the incident wave, and the second term is a traveling wave in the $-x$ direction that represents a reflected wave. As in the case of a free particle, the incident and reflected particles are represented by traveling waves.

For the incident wave, $A_1 \cdot A;$ is the probability density function of the incident panicles. If we multiply this probability density function by the incident velocity, then $v_i \cdot A_1 \cdot A_1^*$ is the flux of incident particles in units of #/cm^2-s. Likewise, the quantity $v, . B_1 . B_1^*$ is the Hux of the reflected particles, where v_r is the velocity of the reflected wave. (The parameters v_i and v_r in these terms are actually the magnitudes of the velocity only.)

In region II, the potential is V = V_0. If we assume that E < V_0, then the differential equation describing the wave function in region 11 can be written as

$$\frac{\partial^2 \psi_2(x)}{\partial x^2} - \frac{2m}{\hbar^2}(V_0 - E)\psi_2(x) = 0 \tag{2.42}$$

The general solution may then be written in the form

$$\psi_2(x) = A_2 e^{-K_2 x} + B_2 e^{+K_2 x} \quad (x \geq 0) \tag{2.43}$$

where

$$K_2 = \sqrt{\frac{2m(V_0 - E)}{\hbar^2}} \tag{2.44}$$

One boundary condition is that the wave function $\psi_2(x)$ must remain finite, which means that the coefficient $B_2 = 0$. The wave function is now given by

$$\psi_2(x) = A_2 e^{-K_2 x} \quad (x \geq 0) \tag{2.45}$$

The wave function at x = 0 must be continuous so that

$$\psi_1(0) = \psi_2(0) \tag{2.46}$$

Then from Equations (2.40), (2.45), and (2.46), we obtain

$$A_1 + B_1 = A_2 \tag{2.47}$$

Since the potential function is everywhere finite, the first derivative of the wave function must also be continuous so that

$$\left.\frac{\partial \psi_1}{\partial x}\right|_{x=0} = \left.\frac{\partial \psi_2}{\partial x}\right|_{x=0} \tag{2.48}$$

Using Equations (2.40), (2.45), and (2.48), we obtain

$$jK_1A_1 - jK_1B_1 = -K_2A_2 \tag{2.49}$$

We can solve Equations (2.47) and (2.49) to determine the coefficients B_1 and A_2 in terms of the incident wave coefficient A_1. The results are

$$B_1 = \frac{-(K_2^2 + 2jK_1K_2 - K_1^2)A_1}{(K_2^2 + K_1^2)} \tag{2.50a}$$

and

$$A_2 = \frac{2K_1(K_1 - jK_2)A_1}{(K_2^2 + K_1^2)} \tag{2.50b}$$

The reflected probability density function is given by

$$B_1 \cdot B_1^* = \frac{(K_2^2 - K_1^2 + 2jK_1K_2)(K_2^2 - K_1^2 - 2jK_1K_2)A_1 \cdot A_1^*}{(K_2^2 + K_1^2)^2} \tag{2.51}$$

We can define a reflection coefficient, R, as the ratio of the reflected flux to the incident flux, which is written as

$$R = \frac{v_r \cdot B_1 \cdot B_1^*}{v_1 \cdot A_1 \cdot A_1^*} \tag{2.52}$$

where v_i and v_r are the incident and reflected velocities. respectively. of the particles. In region I, V = 0 so that E = T, where T is the kinetic energy of the particle. The kinetic energy is given by

$$T = \frac{1}{2}mv^2 \tag{2.53}$$

so that the constant K_1, from Equation (2.41), may be written as

$$K_1 = \sqrt{\frac{2m}{\hbar^2}\left(\frac{1}{2}mv^2\right)} = \sqrt{m^2\frac{v^2}{\hbar^2}} = \frac{mv}{\hbar} \tag{2.54}$$

The incident velocity can then be written as

$$v_i = \frac{\hbar}{m} \cdot K_1 \tag{2.55}$$

Since the reflected particle also exists in region I, the reflected velocity (magnitude) is given by

$$v_r = \frac{\hbar}{m} \cdot K_1 \tag{2.56}$$

The incident and reflected velocities (magnitudes) are equal. The reflection coefficient is then

$$R = \frac{v_r \cdot B_1 \cdot B_1^*}{v_i \cdot A_1 \cdot A_1^*} = \frac{B_1 \cdot B_1^*}{A_1 \cdot A_1^*} \tag{2.57}$$

Substituting the expression from Equation (2.51) into Equation (2.57), we obtain

$$R = \frac{B_1 \cdot B_1^*}{A_1 \cdot A_1^*} = \frac{\left(K_2^2 - K_1^2\right)^2 + 4K_1^2K_2^2}{\left(K_2^2 + K_1^2\right)^2} = 1.0 \tag{2.58}$$

The result of R = 1 implies that all of the particles incident on the potential barrier for E < V_0 are eventually reflected. Particles are not absorbed or transmitted through the potential barrier. This result is entirely consistent with classical physics and one might ask why we should consider this problem in terms of quantum mechanics. The interesting result is in terms of what happens in region II.

The wave solution in region 11 was given by Equation (2.45) as $\psi_2(x) = A_2 e^{-K_2 x}$. The coefficient A_2 from Equation (2.47) is $A_2 = A_1 + B_1$, which we derived from the boundary conditions. For the case of E < V_0, the coefficient A_2 is not zero. If A_2 is not zero, then the probability density function $\psi_2(x) \cdot \psi_2^*(x)$ of the particle being found in region II is not equal to zero. ***This result implies that there is a finite probability that the incident particle will penetrate the potential barrier and exist in region II. The probability of a particle penetrating the potentiul burrier is another difference between classical and quantum mechanics: The quantum mechanical penetration is classically not allowed.*** Although there is a finite probability that the particle may penetrate the barrier, since the reflection coefficient in region I is unity, the particle in region II must eventually turn around and move back into region I.

Objective | **EXAMPLE 2.4**

To calculate the penetration depth of a panicle impinging on a potential barrier.

Consider an incident electron that is traveling at a velocity of 1 × 10^5 m/s in region I.

Solution

With $V(x) = 0$, the total energy is also equal to the kinetic energy so that

$$E = T = \frac{1}{2}mv^2 = 4.56 \times 10^{-21}\text{J} = 2.85 \times 10^{-2}\text{ eV}$$

Now, assume that the potential barrier at $x = 0$ is twice as large as the total energy of the incident particle, or that $V_0 = 2E$. The wave function solution in region II is $\psi_2(x) = A_2 e^{-K_2 x}$, where the constant K_2 is given by $K_2 = \sqrt{2m(V_0 - E)/\hbar^2}$.

In this example, we want to determine the distance $x = d$ at which the wave function magnitude has decayed to e^{-1} of its value at $x = 0$. Then, for this case, we have $K_2 d = 1$ or

$$1 = d\sqrt{\frac{2m(2E - E)}{\hbar^2}} = d\sqrt{\frac{2mE}{\hbar^2}}$$

The distance is then given by

$$d = \sqrt{\frac{\hbar^2}{2mE}} = \frac{1.054 \times 10^{-34}}{\sqrt{2(9.11 \times 10^{-31})(4.56 \times 10^{-21})}} = 11.6 \times 10^{-10} \text{ m}$$

or

$$d = 11.6 \text{ Å}$$

■ **Comment**

This penetration distance corresponds to approximately two lattice constants of silicon. The numbers used in this example are rather arbitrary. We used a distance at which the wave function decayed to e^{-1} of its initial value. We could have arbitrarily used e^{-2}, for example, but the results give an indication of the magnitude of penetration depth.

The case when the total energy of a particle, which is incident on the potential barrier, is greater than the barrier height, or $E > V_0$, is left as an exercise at the end of the chapter.

TEST YOUR UNDERSTANDING

E2.7 The probability of finding a panicle at a distanced in region II compared to that at $x = 0$ is given by $\exp(-2K_2 d)$. Consider an electron traveling in region I at a velocity of 10' m/s incident on a potential barrier whose height is 3 times the kinetic energy of the electron. Find the probability of finding the electron at a distance d compared to $x = 0$ where d is (*a*) 10 Å and (*b*) 100 Å into the potential barrier.

[Ans. (*a*) 8.72 percent, (*b*) 2.53×10^{-9} percent]

2.3.4 The Potential Barrier

We now want to consider the potential barrier function, which is shown in Figure 2.8. The more interesting problem, again, is in the case when the total energy of an incident particle is $E < V_0$. Again assume that we have a flux of incident particles originating on the negative x axis traveling in the $+x$ direction. As before, we need to solve Schrodinger's time-independent wave equation in each of the three regions. The

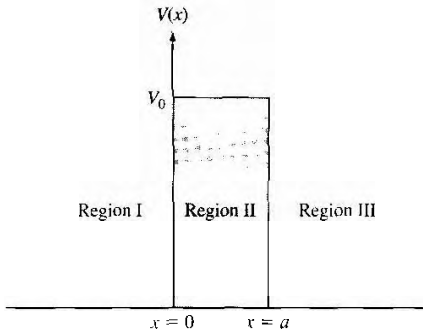


Figure **2.8** | The potential barrier function.

solutions of the wave equation in regions I, II, and III are given, respectively, as

$$\psi_1(x) = A_1 e^{jK_1 x} + B_1 e^{-jK_1 x} \tag{2.59a}$$

$$\psi_2(x) = A_2 e^{K_2 x} + B_2 e^{-K_2 x} \tag{2.59b}$$

$$\psi_3(x) = A_3 e^{jK_1 x} + B_3 e^{-jK_1 x} \tag{2.59c}$$

where

$$K_1 = \sqrt{\frac{2mE}{\hbar^2}} \tag{2.60a}$$

and

$$K_2 = \sqrt{\frac{2m}{\hbar^2}(V_0 - E)} \tag{2.60b}$$

The coefficient B_3 in Equation (2.59c) represents a negative traveling wave in region III. However, once a particle gets into region III, there are no potential changes to cause a reflection; therefore, the coefficient B_3 must be zero. We must keep both exponential terms in Equation (2.59b) since the potential barrier width is finite; that is, neither term will become unbounded. We have four boundary relations for the boundaries at $x = 0$ and $x = a$ corresponding to the wave function and its first derivative being continuous. We can solve for the four coefficients B_1, A_2, B_2, and A_3 in terms of A_1. The wave solutions in the three regions are shown in Figure 2.9.

One particular parameter of interest is the transmission coefficient, in this case defined as the ratio of the transmitted flux in region III to the incident flux in region I. Then the transmission coefficient T is

$$T = \frac{v_t \cdot A_3 \cdot A_3^*}{v_i \cdot A_1 \cdot A_1^*} = \frac{A_3 \cdot A_3^*}{A_1 \cdot A_1^*} \tag{2.61}$$

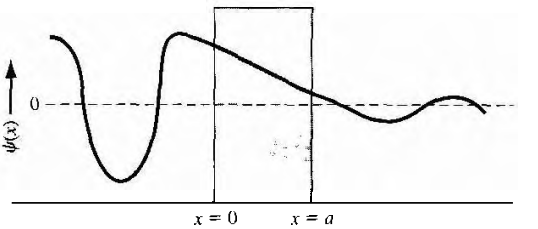


Figure 2.9 | The wave functions through the potential barrier.

where v_t and v_i are the velocities of the transmitted and incident particles, respectively. Since the potential $V = 0$ in both regions I and III, the incident and transmitted velocities are equal. The transmission coefficient may be determined by solving the boundary condition equations. For the special case when $E \ll V_0$, we find that

$$T \approx 16\left(\frac{E}{V_0}\right)\left(1 - \frac{E}{V_0}\right)\exp(-2K_2 a) \tag{2.62}$$

Equation (2.62) implies that there is a finite probability that a particle impinging a potential barrier will penetrate the barrier and will appear in region III. This phenomenon is called tunneling *and it,* roo, *contradicts classical mechanics.* We will see later how this quantum mechanical tunneling phenomenon can be applied to semiconductor device characteristics, such as in the tunnel diode.

EXAMPLE 2.5 | Objective

To calculate the probability of an electron tunneling through a potential harrier.

Consider an electron with an energy of 2 eV impinging on a potential barrier with $V_0 =$ 20 eV and a width of 3 Å.

■ Solution

Equation (2.62) is the tunneling probability. The factor K_2 is

$$K_2 = \sqrt{\frac{2m(V_0 - E)}{\hbar^2}} = \sqrt{\frac{2(9.11 \times 10^{-31})(20 - 2)(1.6 \times 10^{-19})}{(1.054 \times 10^{-34})^2}}$$

or

$$K_2 = 2.17 \times 10^{10}\ \text{m}^{-1}$$

Then

$$T = 16(0.1)(1 - 0.1)\exp[-2(2.17 \times 10^{10})(3 \times 10^{-10})]$$

and finally

$$T = 3.17 \times 10^{-6}$$

■ **Comment**

The tunneling probability may appear to be a small value, but the value is nut zero. If a large number of particles impinge on a potential barrier, a significant number can penetrate the barrier.

TEST YOUR UNDERSTANDING

E2.8 Estimate the tunneling probability of an electron tunneling through a rectangular barrier with a barrier height of $V_0 = 1\,\text{eV}$ and a barrier width of 15 Å. The electron energy is 0.20 eV. ($T = 2.76 \times 10^{-9}$)

E2.9 For a rectangular potential barrier with a height of $V_0 = 2$ eV and an electron with an energy of 0.25 eV, plot the tunneling probability versus barrier width over the range $2 \leq a \leq 20$ Å. Use a log scale for the tunneling probability.



E2.10 A certain semiconductor device requires a tunneling probability of $T = 10^{-5}$ for an electron tunneling through a rectangular barrier with a barrier height of $V_0 = 0.4$ eV. The electron energy is 0.04 eV. Determine the maximum barrier width. ($a = 19.3$ Å)

Additional applications of Schrodinger's wave cquation with various one-dimensional potential functions are found in problems at the end of the chapter. Several of these potential functions represent quantum well structures that are found in modern semiconductor devices.

*2.4 | EXTENSIONS OF THE WAVE THEORY TO ATOMS

So far in this chapter. we have considered several one-dimensional potential energy functions and solved Schrodinger's time-independent wave equation to obtain the probability function of finding a particle at various positions. Consider now the one-electron, or hydrogen, atom potential function. We will only briefly consider the mathematical details and wave function solutions, but the results are extremely interesting and important.

2.4.1 The One-Electron Atom

The nucleus is a heavy, positively charged proton and the electron is a light. negatively charged particle that, in the classical Bohr theory, is revolving around the nucleus. The potential function is due to the coulomb attraction between the proton and electron and is given by

$$V(r) = \frac{-e^2}{4\pi\epsilon_0 r} \tag{2.63}$$

where e is the magnitude of the electronic charge and ϵ_0 is the permittivity of free space. This potential function, although spherically symmetric, leads to a three-dimensional problem in spherical coordinates.

*Indicates sections that can be skipped without loss of continuity.

We may generalize the time-independent Schrodinger's wave equation to three dimensions by writing

$$\nabla^2\psi(r,\theta,\phi) + \frac{2m_0}{\hbar^2}(E - V(r))\psi(r,\theta,\phi) = 0 \tag{2.64}$$

where V^2 is the Laplacian operator and must be written in spherical coordinates for this case. The parameter m_0 is the rest mass of the electron.[3] In spherical coordinates, Schrodinger's wave equation may be written as

$$\frac{1}{r^2}\cdot\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right) + \frac{1}{r^2\sin^2\theta}\cdot\frac{\partial^2\psi}{\partial\phi^2} + \frac{1}{r^2\sin\theta}\cdot\frac{\partial}{\partial\theta}\left(\sin\theta\cdot\frac{\partial\psi}{\partial\theta}\right) + \frac{2m_0}{\hbar^2}(E - V(r))\psi = 0 \tag{2.65}$$

The solution to Equation (2.65) can be determined by the separation-of-variables technique. We will assume that the solution to the time-independent wave equation can be written in the form

$$\psi(r,\theta,\phi) = R(r)\cdot\Theta(\theta)\cdot\Phi(\phi) \tag{2.66}$$

where R, Θ, and Φ, are functions only of r, θ, and ϕ, respectively. Substituting this form of solution into Equation (2.65), we will obtain

$$\frac{\sin^2\theta}{R}\cdot\frac{\partial}{\partial r}\left(r^2\frac{\partial R}{\partial r}\right) + \frac{1}{\Phi}\cdot\frac{\partial^2\Phi}{\partial\phi^2} + \frac{\sin\theta}{\Theta}\cdot\frac{\partial}{\partial\theta}\left(\sin\theta\cdot\frac{\partial\Theta}{\partial\theta}\right) + r^2\sin^2\theta\cdot\frac{2m_0}{\hbar^2}(E - V) = 0 \tag{2.67}$$

We may note that the second term in Equation (2.67) is a function of ϕ only, while all the other terms are functions of either r or θ. We may then write that

$$\frac{1}{\Phi}\cdot\frac{\partial^2\Phi}{\partial\phi^2} = -m^2 \tag{2.68}$$

where m is a separation of variables constant? The solution to Equation (2.68) is of the form

$$\Phi = e^{jm\phi} \tag{2.69}$$

Since the wave function must be single-valued. we impose the condition that m is an integer, or

$$m = 0, \pm1, \pm2, \pm3, \ldots \tag{2.70}$$

[3]The mass should be the rest mass of the two-panicle system, hut since the proton mass is much greater than the electron mass, the equivalent mass reduces to that of the electron.

[4]Where m means the separation-of-variables constant developed historically. That meaning will be retained here even though there may be some confusion with the electron mass. In general, the mass parameter will be used in conjunction with a subscript.

Incorporating the separation-of-variables constant we can further separate the variables Band r and generate two additional separation-of-variables constants 1and n. The separation-of-variablesconstants n, l, and m are known as *quantum* ***numbers*** and are related by

$$\begin{aligned} n &= 1, 2, 3, \ldots \\ l &= n-1, n-2, n-3, \ldots, 0 \\ |m| &= l, l-1, \ldots, 0 \end{aligned} \tag{2.71}$$

Each set of quantum numbers corresponds to a quantum state which the electron may occupy.

The electron energy may be written in the form

$$\boxed{E_n = \frac{-m_0 e^4}{(4\pi\epsilon_0)^2 2\hbar^2 n^2}} \tag{2.72}$$

where n is the principal quantum number. The negative energy indicates that the electron is bound to the nucleus and we again see that the energy of the bound electron is quantized. If the energy were to become positive, then the electron would no longer he a bound panicle and the total energy would no longer be quantized. Since the parametern in Equation (2.72) is an integer, the total energy of the electron can take on only discrete values. The quantized energy is again a result of the particle being bound in a finite region of space.

TEST YOUR UNDERSTANDING

E2.11 Calculate the lowest energy (in electron volts) of an electron in a hydrogen atom. (Ans. $E_1 = -13.6$ eV)

The solution of the wave equation may be designated by ψ_{nlm}, where n, I, and mare again the various quantum numbers. For the lowest energy state, $n = 1.1 = 0$, and $m = 0$, and the wave function is given by

$$\psi_{100} = \frac{1}{\sqrt{\pi}} \cdot \left(\frac{1}{a_0}\right)^{3/2} e^{-r/a_0} \tag{2.73}$$

This function is spherically symmetric, and the parameter a_0 is given by

$$a_0 = \frac{4\pi\epsilon_0\hbar^2}{m_0 e^2} = 0.529 \text{ Å} \tag{2.74}$$

and is equal to the Bohr radius.

The radial probability density function, or the probability of finding the electron at a particular distance from the nucleus, is proportional to the product $\psi_{100} \cdot \psi_{100}^*$ and also to the differential volume of the shell around the nucleus. The probability

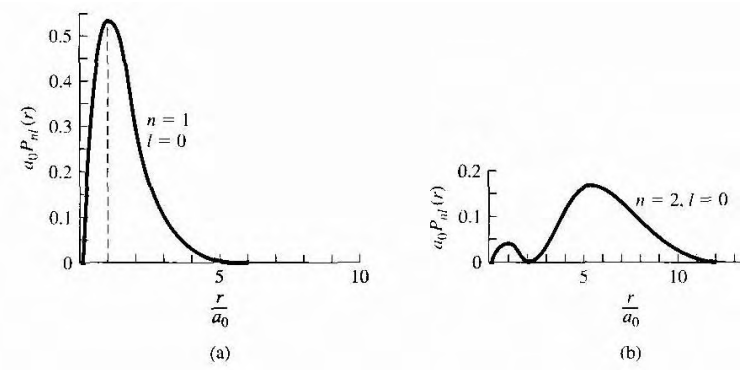


Figure 2.10 | The radial probability density function for the one-electron atom in the (a) lowest energy state and (b) next-higher energy slate.
(From Eisberg and Resnick [4].)

density function for the lowest energy state is plotted in Figure 2.10a. The most probable distance from the nucleus is at r $= a_0$, which is the same as the Bohr theory. Considering this spherically symmetric probability function, we may now begin to conceive the concept of an electron cloud, or energy shell, surrounding the nucleus rather than a discrete particle orbiting around the nucleus.

The radial probability density function for the next higher, spherically symmetric wave function, corresponding to n $= 2, l = 0.$ and $m = 0$, is shown in Figure 2.10b. This figure shows the idea of the next-higher energy shell of the electron. The second energy shell is at a greater radius from the nucleus than the first energy shell. As indicated in the figure, though, there is still a small probability that the electron will exist at the smaller radius. For the case of n $= 2$ and I $=$ I. there are three possible states corresponding to the three allowed values of the quantum numberm. These wave functions are no longer spherically symmetric.

Although we have not gone into a great deal of mathematical detail for the one-electron atom, three results are important for the further analysis of semiconductor materials. The first is the solution of Schrodinger's wave equation, which again yields electron probability functions, as it did for the simpler potential functions. In developing the physics of semiconductor materials in later chapters, we will also be considering electron probability functions. The second result is the quantization of allowed energy levels for the hound electron. The third is the concept of quantum numbers and quantum states, which evolved from the separation-of-variables technique. We will consider this concept again in the next section and in later chapters when we deal with the semiconductor material physics.

2.4.2 The Periodic Table

The initial portion of the periodic table of elements may he determined by using the results of the one-electron atom plus two additional concepts. The tirst concept

needed is that of *electron spin*. The electron has an intrinsic angular momentum, or spin, which is quantized and may take on one of two possible values. The spin is designated by a quantum number s, which has a value of $s = +\frac{1}{2}$ or $s = -\frac{1}{2}$. We now have four basic quantum numbers: n, l, m, and s.

The second concept needed is the *Pauli exclusion principle*. The Pauli exclusion principle states that, in any given system (an atom, molecule, or crystal), no two electrons may occupy the same quantum state. In an atom, the exclusion principle means that no two electrons may have the same set of quantum numbers. We will see that the exclusion principle is also an important factor in determining the distribution of electrons among available energy states in a crystal.

Table 2.1 shows the first few elements of the periodic table. For the first element, hydrogen, we have one electron in the lowest energy state corresponding to $n = 1$. From Equation (2.71) both quantum numbers l and m must be zero. However, the electron can take on either spin factor $+\frac{1}{2}$ or $-\frac{1}{2}$. For helium, two electrons may exist in the lowest energy state. For this case, $l = m = 0$, so now both electron spin states are occupied and the lowest energy shell is full. The chemical activity of an element is determined primarily by the valence, or outermost, electrons. Since the valence energy shell of helium is full, helium does not react with other elements and is an inert element.

The third element, lithium, has three electrons. The third electron must go into the second energy shell corresponding to $n = 2$. When $n = 2$, the quantum number l may be 0 or 1, and when $l = 1$, the quantum number m may be -1, 0, or $+1$. In each case, the electron spin factor may be $+\frac{1}{2}$ or $-\frac{1}{2}$. For $n = 2$, then, there are eight possible quantum states. Neon has ten electrons. Two electrons are in the $n = 1$ energy shell and eight electrons are in the $n = 2$ energy shell. The second energy shell is now full, which means that neon is also an inert element.

From the solution of Schrodinger's wave equation for the one electron atom, plus the concepts of electron spin and the Pauli exclusion principle, we can begin to build up the periodic table of elements. As the atomic numbers of the elements increase, electrons will begin to interact with each other, so that the buildup of the periodic table will deviate somewhat from the simple method.

Table 2.1 | Initial portion of the periodic table

Element	Notation	n	l	m	s
Hydrogen	$1s^1$	1	0	0	$+\frac{1}{2}$ or $-\frac{1}{2}$
Helium	$1s^2$	1	0	0	$+\frac{1}{2}$ and $-\frac{1}{2}$
Lithium	$1s^2 2s^1$	2	0	0	$+\frac{1}{2}$ or $-\frac{1}{2}$
Beryllium	$1s^2 2s^2$	2	0	0	$+\frac{1}{2}$ and $-\frac{1}{2}$
Boron	$1s^2 2s^2 2p^1$	2	1		
Carbon	$1s^2 2s^2 2p^2$	2	1		
Nitrogen	$1s^2 2s^2 2p^3$	2	1		$m = 0, -1, +1$
Oxygen	$1s^2 2s^2 2p^4$	2	1		$s = +\frac{1}{2}, -\frac{1}{2}$
Fluorine	$1s^2 2s^2 2p^5$	2	1		
Neon	$1s^2 2s^2 2p^6$	2	1		

2.5 | SUMMARY

- We considered some of the basic concepts of quantum mechanics, which can be used to describe the behavior of electrons under various potential functions. The understanding of electron behavior is crucial in understanding semiconductor physics.
- The wave–particle duality principle is an important element in quantum mechanics. Particles can have wave-like behavior and waves can have particle-like behavior.
- Schrodinger's wave equation forms the basis for describing and predicting the behavior of electrons.
- Man Born postulated that $|\psi(x)|^2$ is a probability density function.
- A result of applying Schrodinger's wave equation to a bound panicle is that the energy of the bound particle is *quantized.*
- A result of applying Schrodinger's wave equation to an electron incident on a potential barrier is that there is a tinite probability of *tunneling.*
- The basic structure of the periodic table is predicted by applying Schrodinger's wave equation to the one-electron atom.

GLOSSARY OF IMPORTANT TERMS

de Bruglie wavelength The wavelength of a particle given as the ratio of Planck's constant to momentum.

Heisenberg uncertainty principle The principle that states that we cannot describe with absolute accuracy the relationship between sets of conjugate variables that describe the brhavior of particles, such as momentum and position.

Pauli exclusion principle The principle that states that no two electrons can occupy the same quantum state.

photon The particle-like packet of electromagnetic energy.

quanta The particle-like packet of thermal radiation.

quantized energies The allowed discrete energy levels that bound panicles may occupy.

quantum numbers **A** set of numbers that describes the quantum state of a particle, such as an electron in an atom.

quantum state A particular state of an electron that may he described, for example, by a set of quantum numbers.

tunneling The quantum mechanical phenomenon by which a particle may penetrate through a thin potential barrier.

wave–particle duality The characteristic by which electromagnetic waves sometimes exhibit particle-like behavior and particles sometimes exhibit wave-like behavior.

CHECKPOINT

After studying this chapter, the reader should have the ability to:

- Discuss the principle of energy quanta, the wave–particle duality principle, and the uncertainty principle.
- Apply Schrodinger's wave equation and boundary conditions to problems with various potential functions.
- Determine quantized energy levels of bound particles.
  Determine the approximate tunneling probability of a particle incident on a potential barrier.

REVIEW QUESTIONS

1. State the wave–particle duality principle and state the relationship between momentum and wavelength.
2. What is the physical meaning of Schrodinger's wave function?
3. What is meant by a probability density function?
4. List the boundary conditions for solutions to Schrodinger's wave equation.
5. What is meant by quantized energy levels?
6. Describe the concept of tunneling.
7. List the quantum numbers of the one-electron atom and discuss how they were developed

PROBLEMS



2.1 The classical wave equation for a two-wire transmission line is given by $\partial^2 V(x,t)/\partial x^2 = \mathrm{L\,C}\cdot \partial^2 V(x,t)/\partial t^2$. One possible solution is given hy $V(x,t) = (\sin Kx)\cdot(\sin \omega t)$ where $K = n\pi/a$ and $\omega = K/\sqrt{LC}$. Sketch, on the same graph. the function $V(x,t)$ as a function of x for $0 \le x \le a$ and $n = 1$ when (i) $wr = 0$, $(ii)\ \omega t = \pi/2$, (iii) $\omega t = \pi$, (iv) $\omega t = 3\pi/2$, and $(v)\ \omega t = 2\pi$.



2.2 The function $V(x,t) = \cos(2\pi x/\lambda - \omega t)$ is also a solution to the classical wave equation. Sketch on the same graph the function $V(x,t)$ as a function of x for $0 \le x \le 3\lambda$ when: $(i)\ \omega t = 0$, $(ii)\ \omega t = 0.25\pi$. $(iii)\ \omega t = 0.5\pi$, (iv) $\omega t = 0.75\pi$, and $(v)\ \omega t = \pi$.



2.3 Repeat Problem 2.2 for the function $V(x,t) = \cos(2\pi x/\lambda + \mathrm{wt})$.

24 Determine the phase velocities of the traveling waves described in Problems 2.2 and **2.3.**

Section 2.1 Principles of Quantum Mechanics

2.5 The work function of a material refers to the minimum energy required to remove an electron from the material. Assume that the work function of gold is 4.90 eV and that of cesium is 1.90 eV. Calculate the maximum wavelength of light for the photoelectric emission of electrons for gold and cesium.

2.6 Calculate the de Broglie wavelength, $\mathrm{A} = h/p$, for: *(a)* An electron with kinetic energy of (i) 1.0 eV, and (ii) 100 eV. *(b)* A proton with kinetic energy oi 1.0 eV. *(c)* A singly ionized tungsten atom with kinetic energy of I.0 eV. *(d)* A 2000-kg truck traveling at 20 m/s.

2.7 According to classical physics, the average energy of an electron in **an** electron gas at thermal equilibrium is $3kT/2$. Determine, for $T = 300$ K, the average electron energy (in eV), average electron momentum, and the de Broglie wavelength.

***2.8** **An** electron and a photon have the same energy. At what value of energy (in eV) will the wavelength of the photon he 10 rimes that of the electron?

2.9 *(a)* An electron is moving with a velocity of 2 x 10^6 cm/s. Determine the electron energy (in eV), momentum, and de Broglie wavelength (in Å). *(b)* The de Broglie wavelength of an electron is 125 Å. Determine the electron energy (in eV), momentum, and velocity.

2.10 It is desired to produce x-ray radiation with a wavelength of I Å. *(a)* Through what potential voltagedifference must the electron be accelerated in vacuum so that it can,

upon colliding with a target. generate such a photon'! (Assume that all of the electron's energy is transferred to the photon.) (*b*) What is the dc Broglie wavelength of the electron in part (*a*) just before it hits the target'!

2.11 When the uncertainty principle is considered, it is not possible to locate a photon in space more precisely than about one wavelength. Consider a photon with wavelength $A = 1\ \mu m$. What is the uncertainty in the photon's (*a*) momentum and (*b*) energy?

2.12 The uncertainty in position is 12 Å for a particle of mass 5×10^{-29} kg. Determine the minimum uncertainty in (a) the momentum of the panicle and (*b*) the kinetic energy of the particle.

2.13 Repeat Problem 2.12 for a particle of mass 5×10^{-26} kg.

2.14 An automobile has a mass of 1500 kg. What is the uncertainty in the velocity (in miles per hour) when its center of mass is located with an uncertainty no greater than I cm'!

2.15 (*a*) The uncertainty in the position of an electron is no greater than I Å. Determine the minimum uncertainty in its momentum. *(b)* The electron's energy is measured with an uncertainty no greater than 1 eV. Determine the minimum uncertainty in the time aver which the measurement is made.

Section 2.2 Schrodinger's Wave Equation

2.16 Assume that $\Psi_1(x, t)$ and $\Psi_2(x, t)$ are solutions of the one-dimensional time-dependent Schrodinger's wave equation. *(n)* Show that $\Psi_1 + *>$ is a solution. (*b*) Is $\Psi_1 \cdot \Psi_2$ a solution of the Schrodinger's equation in general'! Why or why not?

2.17 Considerthe wave function $\Psi(x, t) = A(\sin \pi x)e^{-j\omega t}$ for $-1\ 5\ x \leq +1$. Determine **A** so that $\int_{-1}^{1} |\Psi(x, t)|^2 dx = 1$

2.18 Consider the wave function $\Psi(x, t) = A(\sin n\pi x)e^{-j\omega t}$ for $0 \leq x \leq$ I. Determine A so that $\int_0^1 |\Psi(x, t)|^2 dx = 1$

2.19 The solution to Schrodinger's wave equation for a particular situation is given by $\psi(x) = \sqrt{2/a_0} \cdot e^{-x/a_0}$. Determine the probability of finding the particle between the following limits: (*a*) $0 \leq x \leq a_0/4$, (*b*) $a_0/4 \leq x \leq a_0/2$, and (*c*) $0 \leq x \leq a_0$.

Section 2.3 Applications of Schrodinger's Wave Equation

2.20 An electron in free space is described by a plane wave given by $\Psi(x, t) = Ae^{j(kx-\omega t)}$ where $k = 1.5 \times 10^9\ m^{-1}$ **and** $\omega = 1.5 \times 10^{13}$ rad/s. (*a*) Determine the phase velocity of the plane wave. (b) Calculate the wavelength, momentum, and kinetic energy (in eV) of the electron.

2.21 An electron is traveling in the negative *x* direction with a kinetic energy of 0.015 eV. Write the equation of a plane wave that describes thic particle.

2.22 An electron is hound in a one-dimensional infinite potential well with a width of 100 Å. Determine the electron energy levels for $n = 1, 2, 3$.

2.23 A one-dimensional infinite potential well with a width of 12 A contains an electron. (a) Calculate the first two energy levels that the electron may occupy. (*b*) If an electron drops from the second energy level to the first, what is the wavelength of *a* photon that might be emitted"

2.24 Consider a panicle with mass of 10 mg in an infinite potential well 1.0 cm wide. (*a*) If the energy of the particle is 10 mJ, calculate the value of *n* for that state. (*b*) What is

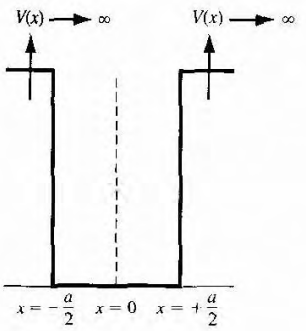


Figure 2.11 | Potential function for Problem 2.26.

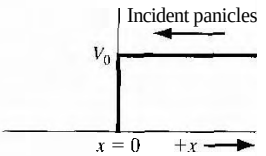


Figure 2.12 | Potential function for Problem 2.30

the kinetic energy of the $(n + 1)$ state? (*c*) Would quantum effects be observable for this particle?

2.25 Calculate the lowest energy level for a neutron in a nucleus, by treating it as if it were in an infinite potential well of width equal to 10^{-14} m. Compare this with the lowest energy level for an electron in the same infinite potential well.

2.26 Consider the particle in the infinite potential well as shown in Figure 2.11. Derive and sketch the wave functions corresponding to the four lowest energy levels. (Do not normalize the wave functions.)

***2.27** Consider a three-dimensional infinite potential well. The potential function is given by $V(x) = 0$ for $0 < x < a, 0 < y < a, 0 < z < a$, and $V(x) = \infty$ elsewhere. Start with Schrodinger's wave equation, use the separation of variables technique, and show that the energy is quantized and is given by

$$E_{n_x n_y n_z} = \frac{\hbar^2 \pi^2}{2ma^2}\left(n_x^2 + n_y^2 + n_z^2\right)$$

where $n_x = 1, 2, 3, \ldots, n_y = 1, 2, 3, \ldots, n_z = 1, 2, 3, \ldots$.

2.28 Consider a **free** electron bound within a two-dimensional infinite potential well defined by $V = 0$ for $0 < x < 25$ Å, $0 < y < 50$ Å, and $V = \infty$ elsewhere. Determine the expression for the allowed electron energies.

Describe any similarities and any differences to the results of the one-dimensional infinite potential well.

2.29 Consider a proton in a one-dimensional inhnite potential well shown in Figure 2.5. (a) Derive the expression for the allowed energy states of the proton. (*b*) Calculate the energy difference (in units of eV) between the lowest possible energy and the next higher energy state fur (i) $a = 4$ Å, and (*ii*) $a = 0.5$ cm.

2.30 For the step potential function shown in Figure 2.12, assume that $E > V_0$ and that particles are incident from the $+x$ direction traveling in the $-x$ direction. (*a*) Write the wave solutions for each region. (*b*) Derive expressions for the transmission and reflection coefficients.

2.31 Consider the penetration of a step potential function of height 2.4 eV by an electron whose energy is 2.1 eV. Determine the relative probability of finding the electron at

the distance (*a*) 12 Å beyond the barrier, and ***(h)***48 Å beyond the barrier, compared to the probability of finding the incident panicle at the barrier edge.

2.32 Evaluate the transmission coefficient for an electron of energy 2.2 eV impinging on a potential barrier of height 6.0 eV and thickness 10^{-10} m. Repeat the calculation for a barrier thickness of 10^{-9} m. Assume that Equation (2.62) is valid.

2.33 (*a*) Estimate the tunneling probability of a particle with **an** effective mass of $0.067 m_0$ (an electron in gallium arsenide), where m_0 is the mass of an electron, tunneling through a rectangular potential barrier of height $V_0 = 0.8$ eV and width 15 Å. The panicle kinetic energy is 0.20 eV. (*b*) Repeat pan ***(a)***if the effective mass of the panicle is $1.08 m_0$ (an electron in silicon).

2.34 A proton attempts to penetrate a rectangular potential barrier of height 10 MeV and thickness 10^{-14} m. The panicle has a total energy of 3 MeV. Calculate the probability that the particle will penetrate the potential barrier. Assume that Equation (2.62) is valid.

***2.35** **An** electron with energy E is incident on a rectangular potential barrier as shown in Figure 2.8. The potential harrier is of width a and height $V_0 \gg E$. (*a*) Write the form of the wave function in each of the three regions. (b) For this geometry, determine what coefficient in the wave function solutions is zero. (r)Derive the expression for the transmission coefficient for the electron (tunneling probability). (d) Sketch the wave function for the electron in each region.

***2.36** **A** potential function is shown in Figure 2.13 with incident particles coming from $-\infty$ with a total energy E > V_2. The constants k are defined as

$$k_1 = \sqrt{\frac{2mE}{\hbar^2}} \qquad k_2 = \sqrt{\frac{2m}{\hbar^2}(E - V_1)} \qquad k_3 = \sqrt{\frac{2m}{\hbar^2}(E - V_2)}$$

Assume a special case for which $k_2 a = 2n\pi$, $n = 1, 2.3, \ldots$. Derive the expression, in terms of the conslants, k_1, k_2, and k_3, for the transmission coefficient. The transmission coefficient is defined as the ratio of the flux of particles in region III to the incident flux in region I.

***2.37** Consider the one-dimensional potential function shown in Figure 2.14. Assume the total energy of an electron is E < V_0. (a) Write the wave solutions that apply in each

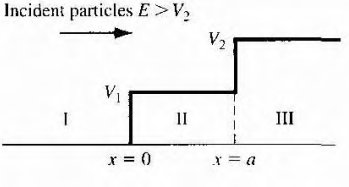


Figure 2.13| Potential function for Problem 2.36.

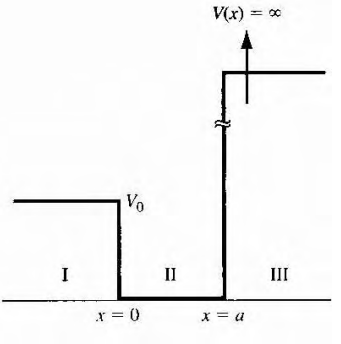


Figure 2.14| Potential function for Problem 2.37.

region. (*b*) Write the set of equations that result from applying the boundary conditions. (*c*) Show explicitly why, or why not. the energy levels of theelectron are quantized.

Section 2.4 Extensions of the Wave Theory to Atoms

2.38 Calculate the energy of the electron in the hydrogen atom (in units of eV) for the first four allowed energy levels.

2.39 Show that the most probable value of the radius r for the 1s electron in a hydrogen atom is equal to the Bohr radius a_0.

2.40 Show that the wave function for ψ_{100} given by Equation (2.73) is a solution to the differential equation given by Equation (2.64).

2.41 What property do H. Li, Na, and K have in common?

READING LIST

*1. Datta, S. *Quantum Phenomena*. Vol. 8 of ***Modular Series** *on Solid State* **Devices.** Reading. ***Mass.:*** Addison–Wesley, 1989.

*2. deCogan, D. ***Solid State Devices: A Quantum Physics Approach.*** New York: Springer-Verlag, 1987.

3. Eisberg, R. M. *Fundamentals* ***of*** *Modern* **Physics.** New York: Wiley, 1961

4. Eisberg, R., and R. Resnick. ***Qunntum*** *Physics of Atoms,* ***Molecules, Solids, Nuclei, arid Particles.*** New York: Wiley, 1974.

5. Kano, K. *Semiconductor* **Devices.** Upper Saddle River, NJ: Prentice Hall, 1998.

6. Kittel, C. *Introduction to* **Solid State Physics,** 7th ed. Berlin: Springer-Verlag, 1993.

7. McKelvey, J. P. ***Solid State Physics for*** *Engineering and Materials* **Science.** Malabar, FL.: Krieger Publishing, 1993.

8. Pauling, L., and E. B. Wilson. *Introduction to Quantum Mechanics.* New York: McGraw-Hill, 1935.

9. Pierret, R. F. *Semiconductor Device Fundamentals*. Reading, MA. Addison-Wesley Publishing Co., 1996.

10. Pohl, H. A. ***Qunntum Mechunics for Science*** *and Engineering*. Englewood Cliffs, N.J.: Prentice Hall, 1967.

11. Schiff, L. I. *Quantum Mechanics.* New York: McGraw-Hill, 1955.

12. Shur, M. ***Inrroduction*** *to* ***Electronic*** *Devices.* New York: John Wiley and Sons, 1996.

CHAPTER 3

Introduction to the Quantum Theory of Solids

PREVIEW

In the last chapter, we applied quantum mechanics and Schrodinger's wave equation to determine the behavior of electrons in the presence of various potential functions. We found that one important characteristic of an electron bound to an atom or bound within a finite space is that the electron can take on only discrete values of energy; that is, the energies are quantized. We also discussed the Pauli exclusion principle, which stated that only one electron is allowed to occupy any given quantum state. In this chapter, we will generalize these concepts to the electron in a crystal lattice.

One of our goals is to determine the electrical properties of a semiconductor material, which we will then use to develop the current-voltage characteristics of semiconductor devices. Toward this end, we have two tasks in this chapter: to determine the properties of electrons in a crystal lattice, and to determine the statistical characteristics of the very large number of electrons in a crystal.

To start, we will expand the concept of discrete allowed electron energies that occur in a single atom to a band of allowed electron energies in a single-crystal solid. First we will qualitatively discuss the feasibility of the allowed energy bands in a crystal and then we will develop a more rigorous mathematical derivation of this theory using Schrodinger's wave equation. This energy band theory is a basic principle of semiconductor material physics and can also be used to explain differences in electrical characteristics between metals, insulators, and semiconductors.

Since current in a solid is due to the net flow of charge, it is important to determine the response of an electron in the crystal to an applied external force, such as an electric field. The movement of an electron in a lattice is different than that of an electron in free space. We will develop a concept allowing us to relate the quantum mechanical behavior of electrons in a crystal to classical Newtonian mechanics. This

analysis leads to a parameter called the electron effective mass. As part of this development, we will find that we can define a new particle in a semiconductor called a *hole*. The motion of both electrons and holes gives rise to currents in a semiconductor.

Because the number of electrons in a semiconductor is very large, it is impossible to follow the motion of each individual particle. We will develop the statistical behavior of electrons in a crystal, noting that the Pauli exclusion principle is an important factor in determining the statistical law the electrons must follow. The resulting probability function will determine the distribution of electrons among the available energy states. The energy band theory and the probability function will be used extensively in the next chapter, when we develop the theory of the semiconductor in equilibrium. ■

3.1 | ALLOWED AND FORBIDDEN ENERGY BANDS

In the last chapter, we treated the one-electron, or hydrogen, atom. That analysis showed that the energy of the bound electron is quantized: Only discrete values of electron energy are allowed. The radial probability density for the electron was also determined. This function gives the probability of finding the electron at a particular distance from the nucleus and shows that the electron is not localized at a given radius. We can extrapolate these single-atom results to a crystal and qualitatively derive the concepts of allowed and forbidden energy bands. We can then apply quantum mechanics and Schrodinger's wave equation to the problem of an electron in a single crystal. We find that the electronic energy states occur in hands of allowed states that are separated by forbidden energy bands.

3.1.1 Formation of Energy Bands

Figure 3.1a shows the radial probability density function for the lowest electron energy state of the single, noninteracting hydrogen atom, and Figure 3.1b shows the same probability curves for two atoms that are in close proximity to each other. The wave functions of the two atom electrons overlap, which means that the two electrons

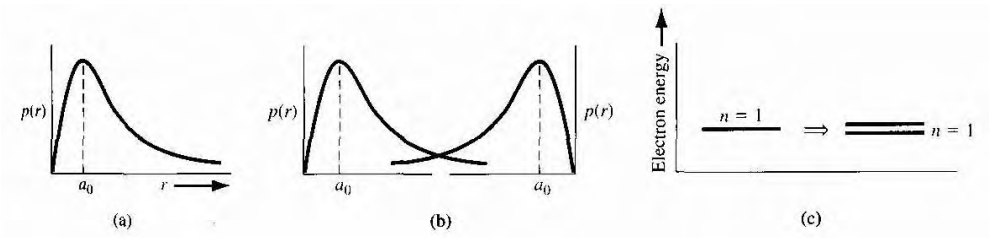


Figure 3.1 | (a) Probability density function of an isolated hydrogen atom. (b) Overlapping probability density functions of two adjacent hydrogen atoms. (c) The splitting of the $n = 1$ state.

will interact. This interaction or perturbation results in the discrete quantized energy level splitting into two discrete energy levels, schematically shown in Figure 3.1c. The splitting of the discrete state into two states is consistent with the Pauli exclusion principle.

A simple analogy of the splitting of energy levels by interacting particles is the following. Two identical race cars and drivers are far apart on a race track. There is no interaction between the cars, so they both must provide the same power to achieve a given speed. However, if one car pulls up close behind the other car, there is an interaction called ***draft.*** The second car will be pulled to an extent by the lead car. The lead car will therefore require more power to achieve the same speed, since it is pulling the second car and the second car will require less power since it is being pulled by the lead car. So there is a "splitting" of power (energy) of the two interacting race cars. (Keep in mind not to take analogies too literally.)

Now, if we somehow start with a regular periodic arrangement of hydrogen-type atoms that are initially very far apart, and begin pushing the atoms together, the initial quantized energy level will split into a band of discrete energy levels. This effect is shown schematically in Figure 3.2, where the parameter r_0 represents the equilibrium interatomic distance in the crystal. At the equilibrium interatomic distance, there is a band of allowed energies, but within the allowed band, the energies are at discrete levels. The Pauli exclusion principle states that the joining of atoms to form a system (crystal) does not alter the total number of quantum states regardless of size. However, since no two electrons can have the same quantum number. the discrete energy must split into a band of energies in order that each electron can occupy a distinct quantum state.

We have seen previously that, at any energy level, the number of allowed quantum states is relatively small. In order to accommodate all of the electrons in a crystal, then, we must have many energy levels within the allowed band. **As** an example, suppose that we have a system with 10^{19} one-electron atoms and also suppose that, at the equilibrium interatomic distance, the width of the allowed energy band is 1 eV. For simplicity, we assume that each electron in the system occupies a different energy level and, if the discrete energy states are equidistant, then the energy levels are separated by 10^{-19} eV. This energy difference is extremely small, so that for all practical purposes, we have a quasi-continuous energy distribution through the allowed

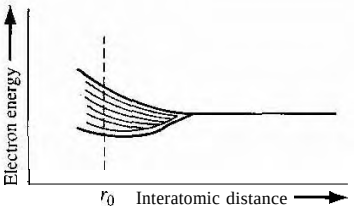


Figure 3.2 | The splitting of an energy state into a band of allowed energies.

energy band. The fact that 10^{-19} eV is a very small difference between two energy states can be seen from the following example.

Objective

EXAMPLE 3.1

To calculate the change in kinetic energy of an electron when the velocity changes by a small value.

Consider an electron traveling at a velocity of 10^7 cm/s. Assume the velocity increases by a value of 1 cm/s. The increase in kinetic energy is given by

$$\Delta E = \frac{1}{2}mv_2^2 - \frac{1}{2}mv_1^2 = \frac{1}{2}m\left(v_2^2 - v_1^2\right)$$

Let $v_2 = v_1 + \Delta v$. Then

$$v_2^2 = (v_1 + \Delta v)^2 = v_1^2 + 2v_1\Delta v + (\Delta v)^2$$

But $\Delta v \ll v_1$, so we have that

$$\Delta E \approx \frac{1}{2}m(2v_1\Delta v) = mv_1\Delta v$$

■ Solution

Substituting the number into this equation, we obtain

$$\Delta E = (9.11 \times 10^{-3})(10^5)(0.01) = 9.11 \times 10^{-28} \text{ J}$$

which may be converted to units of electron volts as

$$\Delta E = \frac{9.11 \times 10^{-28}}{1.6 \times 10^{-19}} = 5.7 \times 10^{-9} \text{ eV}$$

■ Comment

A change in velocity of 1 cm/s compared with 10^7 cm/s results in a change in energy of 5.7×10^{-9} eV, which is orders of magnitude larger than the change in energy of 10^{-19} eV between energy states in the allowed energy hand. This example serves to demonstrate that a difference in adjacent energy states of 10^{-19} eV is indeed very small, so that the discrete energies within an allowed band may be treated as a quasi-continuous distribution.

Consider again a regular periodic arrangement of atoms, in which each atom now contains more then one electron. Suppose the atom in this imaginary crystal contains electrons up through the $n = 3$ energy level. If the atoms are initially very far apart, the electrons in adjacent atoms will not interact and will occupy the discrete energy levels. If these atoms are brought closer together, the outermost electrons in the $n = 3$ energy shell will begin to interact initially, so that this discrete energy level will split into a band of allowed energies. If the atoms continue to move closer together, the electrons in the $n = 2$ shell may begin to interact and will also split into a band of allowed energies. Finally, if the atoms become sufficiently close together, the innermost electrons in the $n = 1$ level may interact. so that this energy level may also split into a band of allowed energies. The splitting of these discrete energy levels is

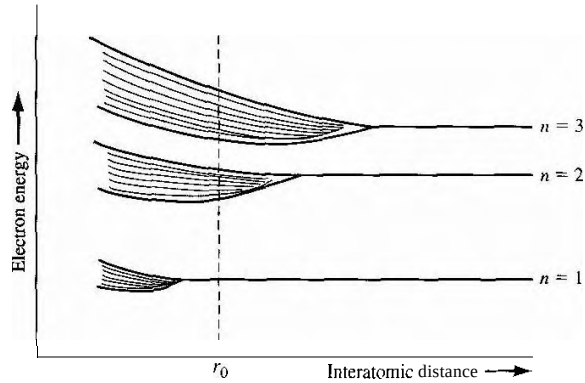


Figure 3.3 | Schematic showing the splitting of three energy states into allowed bands of energies.

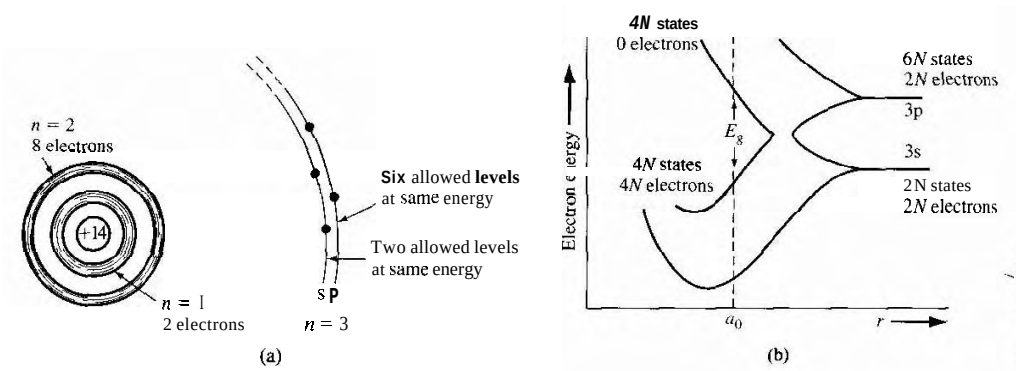


Figure 3.4 | (a) Schematic of an isolated silicon atom. (b) The splitting of the 3s and 3p states of silicon into the allowed and forbidden energy bands.
(From Shockley [51].)

qualitatively shown in Figure 3.3. If the equilibrium interatomic distance is r_0, then we have bands of allowed energies that the electrons may occupy separated by bands of forbidden energies. This energy-band splitting and the formation of allowed and forbidden bands is the energy-band theory of single-crystal materials.

The actual hand splitting in a crystal is much more complicated than indicated in Figure 3.3. A schematic representation of an isolated silicon atom is shown in Figure 3.4a. Ten of the fourteen silicon atom electrons occupy deep-lying energy levels close to the nucleus. The four remaining valence electrons are relatively weakly bound and are the electrons involved in chemical reactions. Figure 3.4b shows the hand splitting of silicon. We need only consider the n = 3 level for the valence electrons, since the first two energy shells are completely full and are tightly bound to the nucleus. The

3s state corresponds to n = 3 and l = 0 and contains two quantum states per atom. This state will contain two electrons at T = 0 K. The 3p state corresponds to n = 3 and l = 1 and contains six quantum states per atom. This state will contain the remaining two electrons in the individual silicon atom.

As the interatomic distance decreases, the 3s and 3p states interact and overlap. At the equilibrium interatomic distance, the bands have again split. hut now four quantum states per atom are in the lower band and four quantum states per atom are in the upper band. At absolute zero degrees, electrons are in the lowest energy state, so that all states in the lower band (the valence band) will he full and all states in the upper band (the conduction band) will be empty. The bandgap energy E_g between the top of the valence band and the bottom of the conduction hand is the width of the forbidden energy band.

We havediscussed qualitatively how and why bands of allowed and forbidden energies are formed in a crystal. The formation of these energy bands is directly related to the electrical characteristics of the crystal, as we will see later in our discussion.

*3.1.2 The Kronig–Penney Model

In the previous section, we discussed qualitatively the spitting of allowed electron energies as atoms are hrought together to form a crystal. The concept of allowed and forbidden energy bands can bc developed more rigorously by considering quantum mechanics and Schrodinger's wave equation. It may be easy for the reader to "get lost" in the following derivation, but the result forms the basis for the energy-band theory of semiconductors.

The potential function of a single, noninteracting, one-electron atom is shown in Figure 3.5a. Also indicated on the figure are the discrete energy levels allowed for theelectron. Figure 3.5b shows the same type of potential function for the case when several atoms are in close proximity arranged in a one-dimensional array. The potential functions of adjacent atoms overlap, and the net potential function for this case is shown in Figure 3.5c. It is this potential function we would need to use in Schrodinger's wave equation to model a one-dimensional single-crystal material.

The solution to Schrodinger's wave equation, for this one-dimensional single-crystal lattice, is made more tractable by considering a simpler potential function. Figure 3.6 is the one-dimensional Kronig–Penney model of the periodic potential function, which is used to represent a one-dimensional single-crystal lattice. We need to solve Schrodinger's wave equation in each region. As with previous quantum mechanical problems, the more interesting solution occurs for the case when $E < V_0$, which corresponds to a particle being bound within the crystal. The electrons are contained in the potential wells, but we have the possibility of tunneling between wells. The Kronig–Penney model is an idealized periodic potential representing a one-dimensional single crystal. but the results will illustrate many of the important features of the quantum behavior of electrons in a periodic lattice.

To obtain the solution to Schrodinger's wave equation, we make use of a mathematical theorem by Bloch. The theorem states that all one-electron wave functions,

*Indicates sections that can he skipped without loss of continuity

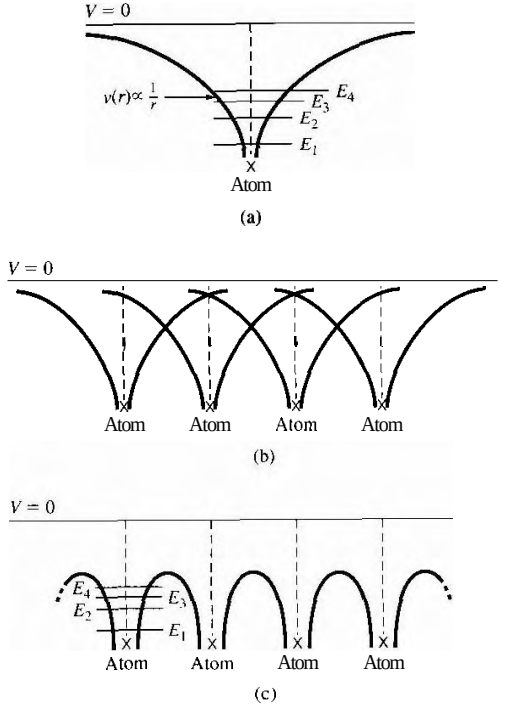


Figure 3.5 | (a) Potential function of a single isolated atom. (b) Overlapping potential functions of adjacent atoms. (c) Net potential function of a one-dimensional single crystal.

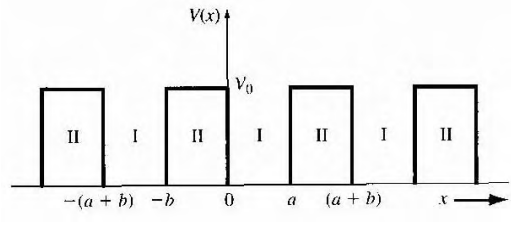


Figure 3.6 | The one-dimensional periodic potential function of the Kronig–Penney model.

for problems involving periodically varying potential energy functions, must be of the form

$$\psi(x) = u(x)e^{jkx} \tag{3.1}$$

The parameter k is called a constant of motion and will be considered in more detail as we develop the theory. The function $u(x)$ is a periodic function with period $(a + b)$.

We stated in Chapter 2 that the total solution to the wave equation is the product of the time-independentsolution and the time-dependent solution, or

$$\Psi(x, t) = \psi(x)\phi(t) = u(x)e^{jkx} \cdot e^{-j(E/\hbar)t} \tag{3.2}$$

which may be written as

$$\Psi(x, t) = u(x)e^{j(kx-(E/\hbar)t)} \tag{3.3}$$

This traveling-wave solution represents the motion of an electron in a single-crystal material. The amplitude of the traveling wave is a periodic function and the parameter k is also referred to as a wave number.

We can now begin to determine a relation between the parameter k, the total energy E, and the potential V_0. If we consider region I in Figure 3.6 ($0 < x < a$) in which $V(x) = 0$, take the second derivative of Equation (3.1), and substitute this result into the time-independent Schrodinger's wave equation given by Equation (2.13). we obtain the relation

$$\frac{d^2u_1(x)}{dx^2} + 2jk\frac{du_1(x)}{dx} - (k^2 - \alpha^2)u_1(x) = 0 \tag{3.4}$$

The function $u_1(x)$ is the amplitude of the wave function in region I and the parameter a is defined as

$$\alpha^2 = \frac{2mE}{\hbar^2} \tag{3.5}$$

Considernow a specific region II, $-b < x < 0$, in which $V(x) = V_0$, and apply Schrodinger's wave equation. We obtain the relation

$$\frac{d^2u_2(x)}{dx^2} + 2jk\frac{du_2(x)}{dx} - \left(k^2 - \alpha^2 + \frac{2mV_0}{\hbar^2}\right)u_2(x) = 0 \tag{3.6}$$

where $u_2(x)$ is the amplitude of the wave function in region II. We may define

$$\frac{2m}{\hbar^2}(E - V_0) = \alpha^2 - \frac{2mV_0}{\hbar^2} = \beta^2 \tag{3.7}$$

so that Equation (3.6) may be written as

$$\frac{d^2u_2(x)}{dx^2} + 2jk\frac{du_2(x)}{dx} - (k^2 - \beta^2)u_2(x) = 0 \tag{3.8}$$

Note that from Equation (3.7). if $E > V_0$, the parameter β is real, whereas if $E < V_0$, then β is imaginary.

The solution to Equation (3.4), for region I, is of the form

$$u_1(x) = Ae^{j(\alpha-k)x} + Be^{-j(\alpha+k)x} \quad \text{for } (0 < x < a) \tag{3.9}$$

and the solution to Equation (3.8), for region II, is of the form

$$u_2(x) = Ce^{j(\beta-k)x} + De^{-j(\beta+k)x} \quad \text{for } (-b < x < 0) \tag{3.10}$$

Since the potential function $V(x)$ is everywhere finite, both the wave function $\psi(x)$ and its first derivative $\partial\psi(x)/\partial x$ must be continuous. This continuity condition implies that the wave amplitude function $u(x)$ and its first derivative $\partial u(x)/\partial x$ must also be continuous.

If we consider the boundary at $x = 0$ and apply the continuity condition to the wave amplitude, we have

$$u_1(0) = u_2(0) \tag{3.11}$$

Substituting Equations (3.9) and (3.10) into Equation (3.11), we obtain

$$A + B - C - D = 0 \tag{3.12}$$

Now applying the condition that

$$\left.\frac{du_1}{dx}\right|_{x=0} = \left.\frac{du_2}{dx}\right|_{x=0} \tag{3.13}$$

we obtain

$$(\alpha - k)A - (\alpha + k)B - (\beta - k)C + (\beta + k)D = 0 \tag{3.14}$$

We have considered region I as $0 < x < a$ and region II as $-b < x < 0$. The periodicity and the continuity condition mean that the function u_1, as $x \to a$, is equal to the function u_2, as $x \to -b$. This condition may be written as

$$u_1(a) = u_2(-b) \tag{3.15}$$

Applying the solutions for $u_1(x)$ and $u_2(x)$ to the boundary condition in Equation (3.15) yields

$$Ae^{j(\alpha-k)a} + Be^{-j(\alpha+k)a} \quad Ce^{-j(\beta-k)b} - De^{j(\beta+k)b} = 0 \tag{3.16}$$

The last boundary condition is

$$\left.\frac{du_1}{dx}\right|_{x=a} = \left.\frac{du_2}{dx}\right|_{x=-b} \tag{3.17}$$

which gives

$$(\alpha - k)Ae^{j(\alpha-k)a} - (\alpha + k)Be^{-j(\alpha+k)a} - (\beta - k)Ce^{-j(\beta-k)b} + (\beta + k)De^{j(\beta+k)b} = 0 \tag{3.18}$$

We now have four homogeneous equations, Equations (3.12), (3.14), (3.16), and (3.18), with four unknowns as a result of applying the four boundary conditions. In a set of simultaneous, linear, homogeneous equations, there is a nontrivial solution if,

and only if, the determinant of the coefficients is zero. In our case, the coefficients in question are the coefficients of the parameters A, B, C, and D.

The evaluation of this determinant is extremely laborious and will not he considered in detail. The result is

$$\frac{-(\alpha^2+\beta^2)}{2\alpha\beta}(\sin\alpha a)(\sin\beta b)+(\cos\alpha a)(\cos\beta b)=\cos k(a+b) \qquad (3.19)$$

Equation (3.19) relates the parameter k to the total energy E (through the parameter a) and the potential function V_0 (through the parameter β).

As we mentioned, the more interesting solutions occur for $E < V_0$, which applies lo theelectron bound within the crystal. From Equation (3.7), the parameter β is then an imaginary quantity. We may define

$$\beta = j\gamma \qquad (3.20)$$

where y is areal quantity. Equation (3.19) can he written in terms of γ as

$$\frac{\gamma^2-\alpha^2}{2\alpha\gamma}(\sin\alpha a)(\sinh\gamma b)+(\cos\alpha a)(\cosh\gamma b)=\cos k(a+b) \qquad (3.21)$$

Equation (3.21) does not lend itself to an analytical solution, but must be solved using numerical or graphical techniques to obtain the relation between k, E. and V_0. The solution of Schrodinger's wave equation for a single bound particle resulted in discrete allowed energies. The solution of Equation (3.21) will result in a band of allowed energies.

To obtain an equation that is more susceptible to a graphical solution and thus will illustrate the nature of the results, let the potential barrier width b $\rightarrow$ 0 and the barrier height $V_0 \rightarrow \infty$, but such that the product bV_0 remains finite. Equation (3.21) then reduces to

$$\left(\frac{mV_0ba}{\hbar^2}\right)\frac{\sin\alpha a}{\alpha a}+\cos\alpha a=\cos ka \qquad (3.22)$$

We may define a parameter P' as

$$P'=\frac{mV_0ba}{\hbar^2} \qquad (3.23)$$

Then, finally, we have the relation

$$P'\frac{\sin a o}{\alpha a}+\cos\alpha a=\cos ku \qquad (3.24)$$

Equation (3.24) *again* gives the relation *between* the *parameter* k, total energy E *(through* the parameter a), *and* the potential barrier bV_0. We may note that Equation (3.24) *is* not a *solution of Schrodinger's wave* equation but *gives* the *conditions* for which Schrodinger's wave equation ***will*** *have* a solution. If we assume the crystal is infinitely large, then k in Equation (3.24) can assume a continuum of values and must be real.

3.1.3 The k-Space Diagram

To begin to understand the nature of the solution, initially consider the special case for which $V_0 = 0$. In this case P' = 0, which corresponds to a free particle since there are no potential barriers. From Equation (3.24). we have that

$$\cos aa = \cos ka \tag{3.25}$$

or

$$\alpha = k \tag{3.26}$$

Since the potential is equal to zero, the total energy E is equal to the kinetic energy, so that, from Equation (3.5), Equation (3.26) may be written as

$$\alpha = \sqrt{\frac{2mE}{\hbar^2}} = \sqrt{\frac{2m\left(\frac{1}{2}mv^2\right)}{\hbar^2}} = \frac{p}{\mathrm{h}} = k \tag{3.27}$$

where p is the particle momentum. The constant of the motion parameter k is related to the particle momentum for the free electron. The parameter k is also referred to as a wave number.

We can also relate the energy and momentum as

$$E = \frac{p^2}{2m} = \frac{k^2\hbar^2}{2m} \tag{3.28}$$

Figure 3.7 shows the parabolic relation of Equation (3.28) between the energy E and momentum p for the free particle. Since the momentum and wave number are linearly related, Figure 3.7 is also the *E* versus k curve for the free panicle.

We now want to consider the relation between *E* and k from Equation (3.24) for the particle in the single-crystal lattice. As the parameter P' increases, the particle becomes more tightly bound to the potential well or atom. We may define the left side of Equation (3.24) to be a function f (αa), so that

$$\mathrm{f}(\alpha a) = P'\frac{\sin\alpha a}{\alpha a} + \cos\alpha a \tag{3.29}$$

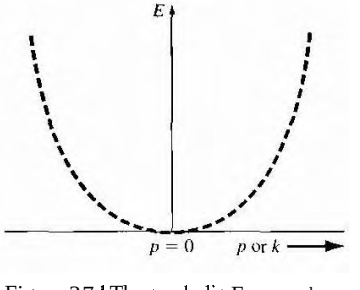


Figure 3.7 The parabolic E versus k curve for the free electron.

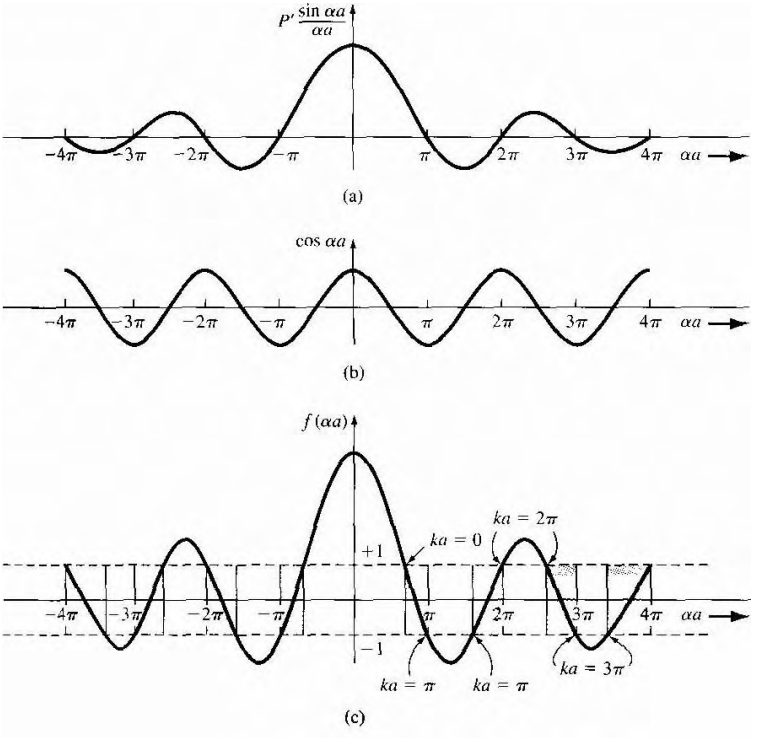


Figure 3.8 | A plot of (a) the first term in Equation (3.29), (b) the second term in Equation (3.29), and (c) the entire $f(\alpha a)$ function. The shaded areas show the allowed values of (αa) corresponding to real values of k.

Figure 3.8a is a plot of the first term of Equation (3.29) versus αa. Figure 3.8b shows a plot of the cos αa term and Figure 3.8c is the sum of the two terms, or $f(\alpha a)$.

Now from Equation (3.24). we also have that

$$f(\alpha a) = \cos ka \tag{3.30}$$

For Equation (3.30) to be valid, the allowed values of the $f(\alpha a)$ function must be bounded between $+1$ and -1. Figure 3.8c shows the allowed values of $f(\alpha a)$ and the allowed values of αa in the shaded areas. Also shown on the figure are the values of ka from the right side of Equation (3.30) which correspond to the allowed values of $f(\alpha a)$.

The parameter a is related to the total energy E of the particle through Equation (3.5), which is $\alpha^2 = 2mE/\hbar^2$. A plot of the energy E of the particle as a function of the wave number k can be generated from Figure 3.8c. Figure 3.9 shows this plot and shows the concept of allowed energy bands for the particle propagating in the

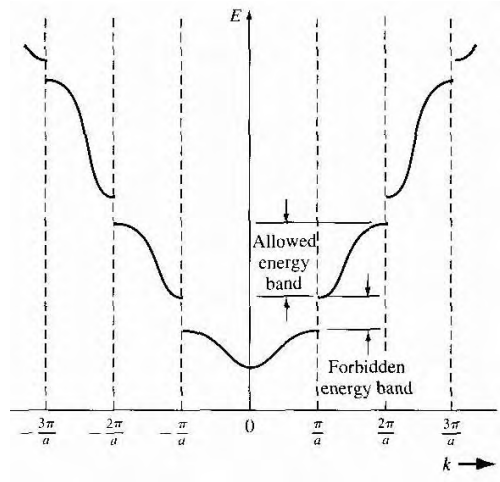


Figure 3.9 | The E versus k diagram generated from Figure 3.8. The allowed energy bands and forbidden energy bandgaps are indicated.

crystal lattice. Since the energy E has discontinuities, we also have the concept of forbidden energies for the particles in the crystal.

EXAMPLE 3.2 | Objective

To determine the lowest allowed energy bandwidth.

Assume that the coefficient $P' = 10$ and that the potential width $a = 5\,\text{Å}$

***8* Solution**

To find the lowest allowed energy handwidth. we need to find the difference in αa values as ka changes from 0 to π (see Figure 3.8c). For $ka = 0$, Equation (3.29) becomes

$$1 = 10\frac{\sin \alpha a}{\alpha a} + \cos \alpha a$$

By trial and error, we find $\alpha a = 2.628$ rad. We see that for $ka = \pi$, $\alpha a = \pi$.

For $\alpha a = \pi$, we have

$$\sqrt{\frac{2mE_2}{\hbar^2}} \cdot a = \pi$$

or

$$E_2 = \frac{\pi^2\hbar^2}{2ma^2} = \frac{\pi^2(1.054 \times 10^{-34})^2}{2(9.11 \times 10^{-31})(5 \times 10^{-10})^2} = 2.407 \times 10^{-19}\,\text{J} = 1.50\,\text{eV}$$

For $\alpha a = 2.628$. we find that $E_1 = 1.68 \times 10^{-19}$ J $= 1.053$ eV. The allowed energy bandwidth is thrn

$$\Delta E = E_2 - E_1 = 1.50 - 1.053 = 0.447\,\text{eV}$$

Comment

We see from Figure 3.8c that. as the energy increases, the widths of the allowed bands increase from this Kronig–Penney model.

TEST YOUR UNDERSTANDING

E3.1 Using the parameters given in Example 3.2, determine the width (in eV) of the forbidden energy band that exists at $ka = \pi$ (see Figure 3.8c). (Ans. 2.79 eV)

Consider again the right side of Equation (3.24), which is the function cos ka. The cosine function is periodic so that

$$\cos ka = \cos(ka + 2n\pi) = \cos(ka - 2n\pi) \tag{3.31}$$

where n is a positive integer. We may consider Figure 3.9 and displace portions of the curve by 2π. Mathematically, Equation (3.24) is still satisfied. Figure 3.10 shows how various segments of the curve can be displaced by the 2π factor. Figure 3.11 shows the case in which the entire E versus k plot is contained within $-\pi/a < k < \pi/a$. This plot is referred to as a reduced k-space diagram, or a reduced-zero representation.

We noted in Equation (3.27) that for a free electron, the particle momentum and the wave number k are related by $p = \hbar k$. Given the similarity between the free

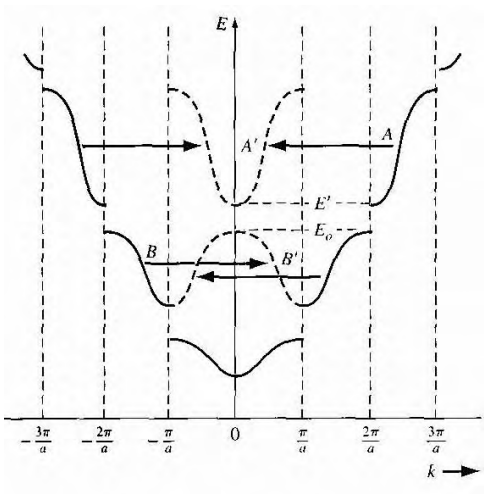


Figure 3.101 The E versus k diagram showing 2π displacements of several sections of allowed energy bands.

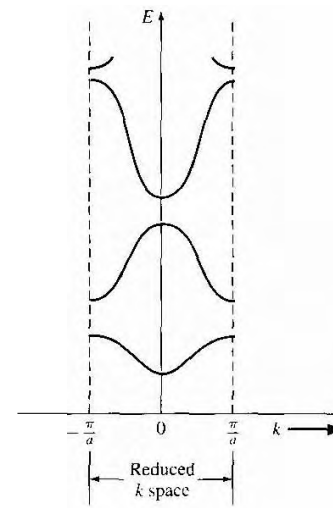


Figure **3.11** | The E versus k diagram in the reduced-zone representation.

electron solution and the results of the single crystal shown in Figure 3.9, the parameter $\hbar k$ in a single crystal is referred to as the *crystal momentum.* This parameter is not the actual momentum of the electron in the crystal, but is a constant of the motion that includes the crystal interaction.

We have been considering the Kronig–Penney model, which is a one-dimensional periodic potential function used to model a single-crystal lattice. The principle result of this analysis, so far, is that electrons in the crystal occupy certain allowed energy bands and are excluded from the forbidden energy bands. For real three-dimensional single-crystal materials, a similar energy-band theory exists. We will obtain additional electron properties from the Kronig–Penney model in the next sections.

3.2 | ELECTRICAL CONDUCTION IN SOLIDS

Again, we are eventually interested in determining the current–voltage characteristics of semiconductor devices. We will need to consider electrical conduction in solids as it relates to the band theory we have just developed. Let us begin by considering the motion of electrons in the various allowed energy hands.

3.2.1 The Energy Band and the Bond Model

In Chapter 1, we discussed the covalent bonding of silicon. Figure 3.12 shows a two-dimensional representation of the covalent bonding in a single-crystal silicon lattice. This figure represents silicon at $T = 0$ K in which each silicon atom is surrounded by eight valence electrons that are in their lowest energy state and are directly involved in the covalent bonding. Figure 3.4b represented the splitting of the discrete silicon energy states into bands of allowed energies as the silicon crystal is formed. At $T = 0$ K, the 4N states in the lower band, the valence band, are tilled with the valence electrons. All of the valence electrons schematically shown in Figure 3.12 are in the valence band. The upper energy band, the conduction band, is completely empty at $T = 0$ K.

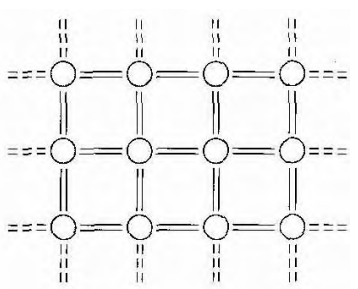


Figure 3.12 | Two-dimensional representation of the covalent bonding in a semiconductor at $T = 0$ K.

As the temperature increases above 0 K, a few valence band electrons may gain enough thermal energy to break the covalent bond and jump into the conduction band. Figure 3.13a shows a two-dimensional representation of this bond-breaking effect and Figure 3.13b, a simple line representation of the energy-band model, shows the same effect.

The semiconductor is neutrally charged. This means that, as the negatively charged electron breaks away from its covalent bonding position, a positively charged "empty state" is created in the original covalent bonding position in the valence band. As the temperature further increases, more covalent bonds are broken, more electrons jump to the conduction hand, and more positive "empty states" are created in the valence band.

We can also relate this hond breaking to the E versus k energy bands. Figure 3.14a shows the E versus k diagram of the conduction and valence bands at

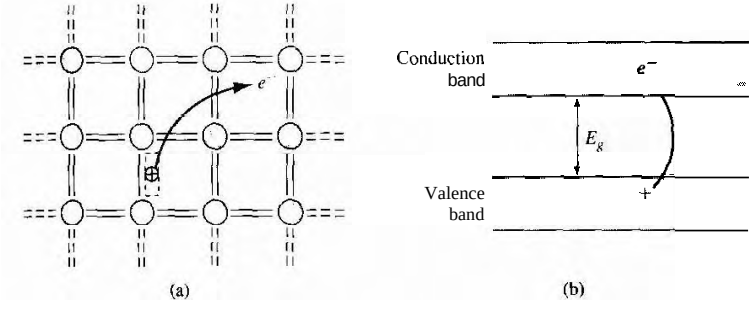


Figure 3.13 (a) Two-dimensional representation of the breaking of a covalent bond. (b) Corresponding line representation **of** the energy band and the generation **of** a negative **and** positive charge with the breaking of a covalent bond.

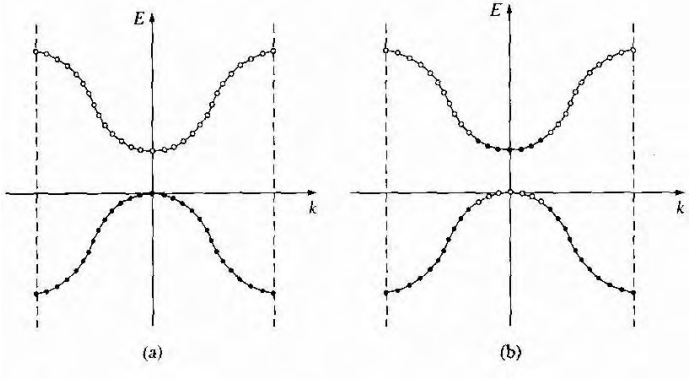


Figure 3.14 The E versus k diagram of the conduction and valence bands of a semiconductor at (a) $T = 0$ K and (b) $T > 0$ K.

$T = 0$ K. The energy states in the valence band are completely full and the states in the conduction band are empty. Figure 3.14b shows these same bands for T > 0 K, in which some electrons have gained enough energy to jump to the conduction band and have left empty states in the valence hand. We are assuming at this point that no external forces are applied so the electron and "empty state" distributions are symmetrical with k.

3.2.2 Drift Current

Current is due to the net flow of charge. If we had a collection of positively charged ions with a volume density N (cm^{-3}) and an average drift velocity v_d (cm/s), then the drift current density would be

$$J = qNv_d \qquad \text{A/cm}^2 \tag{3.32}$$

If, instead of considering the average drift velocity, we considered the individual ion velocities, then we could write the drift current density as

$$J = q\sum_{i=1}^{n} v_i \tag{3.33}$$

where v_i is the velocity of the ith ion. The summation in Equation (3.33) is taken over a unit volume so that the current density J is still in units of A/cm^2.

Since electrons are charged particles, a net drift of electrons in the conduction band will give rise to a current. The electron distribution in the conduction band, as shown in Figure 3.14b, is an even function of k when no external force is applied. Recall that k for a free electron is related to momentum so that, since there are as many electrons with a $+|k|$ value as there are with a $-|k|$ value, the net drift current density due to these electrons is zero. This result is certainly expected since there is no externally applied force.

If aforce is applied to a particle and the particle moves, it must gain energy. This effect is expressed as

$$dE = F\,dx = Fudt \tag{3.34}$$

where F is the applied force, dx is the differential distance the particle moves, v is the velocity, and dE is the increase in energy. If an external force is applied to the electrons in the conduction band, there are empty energy states into which the electrons can move: therefore, because of the external force, electrons can gain energy and a net momentum. The electron distribution in the conduction band may look like that shown in Figure 3.15, which implies that the electrons have gained a net momentum.

We may write the drift current density due to the motion of electrons as

$$J = -e\sum_{i=1}^{n} v_i \tag{3.35}$$

where e is the magnitude of the electronic charge and n is the number of electrons per unit volume in the conduction hand. Again, the summation is taken over a unit

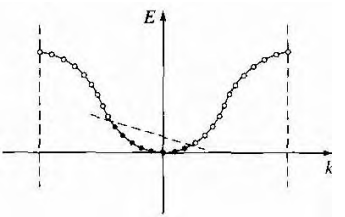


Figure 3.15 | The asymmetric distribution of electrons in the E versus k diagram when an external force is applied.

volume so the current density is A/cm^2. We may note from Equation (3.35) that the current is directly related to the electron velocity; that is, the current is related to how well the electron can move in the crystal.

3.2.3 Electron Effective Mass

The movement of an electron in a lattice will, in general, be different from that of an electron in free space. In addition to an externally applied force, there are internal forces in the crystal due to positively charged ions or protons and negatively charged electrons, which will influence the motion of electrons in the lattice. We can write

$$F_{total} = F_{ext} + F_{int} = ma \tag{3.36}$$

where F_{total}, F_{ext}, and F_{int} are the total force, the externally applied force, and the internal forces, respectively, acting on a particle in a crystal. The parameter a is the acceleration and m is the rest mass of the particle.

Since it is difficult to take into account all of the internal forces, we will write the equation

$$F_{ext} = m^* a \tag{1.37}$$

where the acceleration a is now directly related to the external force. The parameter m^*, called the ***effective mass,*** takes into account the panicle mass and also takes into account the effect of the internal forces.

To use an analogy for the effective mass concept, consider the difference in motion between a glass marble in a container filled with water and in a container filled with oil. In general, the marble will drop through the water at a faster rate than through the oil. The external force in this example is the gravitational force and the internal forces are related to the viscosity of the liquids. Because of the difference in motion of the marble in these two cases, the mass of the marble would appear to be different in water than in oil. (As with any analogy, we must be careful not to be too literal.)

We can also relate the effective mass of an electron in a crystal to the E versus k curves, such as was shown in Figure 3.11. In a semiconductor material, we will be dealing with allowed energy bands that are almost empty of electrons and other energy bands that are almost full of electrons.

To begin, consider the case of a free electron whose E versus k curve was shown in Figure 3.7. Recalling Equation (1.28). the energy and momentum are related by $E = p^2/2m = \hbar^2 k^2/2m$, where m is the mass of the electron. The momentum and wave number k are related by $p = \hbar k$. If we take the derivative of Equation (3.28) with respect to k, we obtain

$$\frac{dE}{dk} = \frac{\hbar^2 k}{m} = \frac{\hbar p}{m} \tag{3.38}$$

Relating momentum to velocity, Equation (3.38) can be written as

$$\frac{1}{\hbar}\frac{dE}{dk} = \frac{p}{m} = v \tag{3.39}$$

where v is the velocity of the particle. The first derivative of E with respect to k is related to the velocity of the panicle.

If we now take the second derivative of E with respect to k, we have

$$\frac{d^2E}{dk^2} = \frac{\hbar^2}{m} \tag{3.40}$$

We may rewrite Equation (3.40) as

$$\boxed{\frac{1}{\hbar^2}\frac{d^2E}{dk^2} = \frac{1}{m}} \tag{3.41}$$

The second derivative of E with respect to k is inversely proportional to the mass of the particle. For the case of a free electron, the mass is a constant (nonrelativistic effect), so the second derivative function is a constant. We may also note from Figure 3.7 that d^2E/dk^2 is a positive quantity, which implies that the mass of the electron is also a positive quantily.

If we apply an electric field to the free electron and use Newton's classical equation of motion, we can write

$$F = ma = -eE \tag{3.42}$$

where a is the acceleration, E is the applied electric field, and e is the magnitude of the electronic charge. Solving for the acceleration, we have

$$a = \frac{-eE}{m} \tag{3.43}$$

The motion of the free electron is in the opposite direction to the applied electric field because of the negative charge.

We may now apply the results to the electron in the bottom of an allowed energy band. Consider the allowed energy band in Figure 3.16a. The energy near the bottom of this energy band may be approximated by a parabola, just as that of a free particle. We may write

$$E - E_c = C_1(k)^2 \tag{3.44}$$

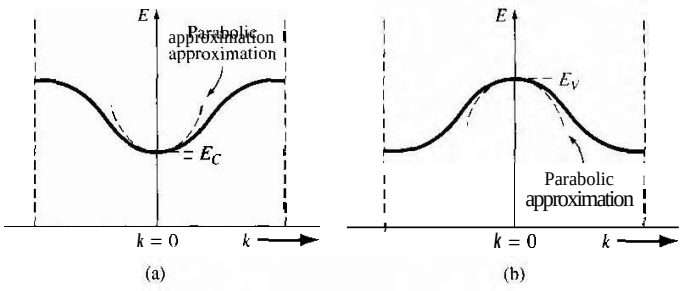


Figure 3.16 | (a) The conduction band in reduced k space, and the parabolic approximation. (b) The valence band in reduced k space, and the parabolic approximation.

Theenergy E_c is the energy at the bottom of the band. Since E > E_c, the parameter C_1 is a positive quantity.

Taking the second derivative of E with respect to k from Equation (3.44), we obtain

$$\frac{d^2E}{dk^2} = 2C_1 \tag{3.45}$$

We may put Equation (3.45) in the form

$$\frac{1}{\hbar^2}\frac{d^2E}{dk^2} = \frac{2C_1}{\hbar^2} \tag{3.46}$$

Comparing Equation (3.46) with Equation (3.41). we may equate $\hbar^2/2C_1$ to the mass of the particle. However, the curvature of the curve in Figure 3.16a will not, in general, be the same as the curvature of the free-particle curve. We may write

$$\frac{1}{\hbar^2}\frac{d^2E}{dk^2} = \frac{2C_1}{\hbar^2} = \frac{1}{m^*} \tag{3.47}$$

wherem* is called the effective mass. Since $C_1 > 0$, we have that $m^* > 0$ also.

The effective mass is a parameter that relates the quantum mechanical results to the classical force equations. In most instances, the electron in the bottom of the conduction band can be thought of as a classical particle whose motion can be modeled by Newtonian mechanics, provided that the internal forces and quantum mechanical properties are taken into account through the effective mass. If we apply an electric field to the electron in the bottom of the allowed energy band, we may write the acceleration as

$$a = \frac{-e\mathrm{E}}{m_n^*} \tag{3.48}$$

where m_n^* is the effective mass of the electron. The effective mass m_n^* of the electron near the bottom of the conduction band is a constant.

3.2.4 Concept of the Hole

In considering the two-dimensional representation of the covalent bonding shown in Figure 3.13a, a positively charged "empty state" was created when a valence electron was elevated into the conduction band. For $T > 0$ K, all valence electrons may gain thermal energy; if a valence electron gains a small amount of thermal energy, it may hop into the empty state. The movement of a valence electron into the empty state is equivalent to the movement of the positively charged empty state itself. Figure 3.17 shows the movement of valence electrons in the crystal alternately filling one empty state and creating a new empty state, a motion equivalent to a positive charge moving in the valence band. The crystal now has a second equally important charge carrier that can give rise to a current. This charge carrier is called a ***hole*** and, as we will see, can also be thought of as a classical particle whose motion can be modeled using Newtonian mechanics.

The drift current density due to electrons in the valence band, such as shown in Figure 3.14b, can be written as

$$J = -e \sum_{i(\text{filled})} v_i \tag{3.49}$$

where the summation extends over all filled states. This summation is inconvenient since it extends over a nearly full valence band and takes into account a very large number of states. We may rewrite Equation (3.49) in the form

$$J = -e \sum_{i(\text{total})} v_i + e \sum_{i(\text{empty})} v_i \tag{3.50}$$

If we consider a band that is totally full, all available states are occupied by electrons. The individual electrons can be thought of as moving with a velocity as given by Equation (3.39):

$$v(E) = \left(\frac{1}{\hbar}\right)\left(\frac{dE}{dk}\right) \tag{3.39}$$

The band is symmetric in k and each state is occupied so that, for every electron with a velocity $|v|$, there is a corresponding electron with a velocity $-|v|$. Since the band is full, the distribution of electrons with respect to k cannot be changed with an externally applied force. The net drift current density generated from a completely full

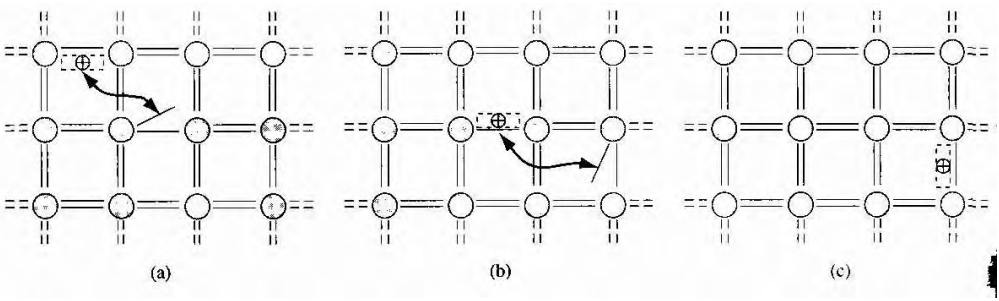


Figure 3.17 | Visualization of the movement of a hole in a semiconductor.

band, then, is zero, or

$$-e \sum_{i(\text{total})} v_i \equiv 0 \tag{3.51}$$

We can now write the drift current density from Equation (3.50) for an almost full band as

$$J = +e \sum_{i(\text{empty})} v_i \tag{3.52}$$

where the v_i in the summation is the

$$v(E) = \left(\frac{1}{\hbar}\right)\left(\frac{dE}{dk}\right)$$

associated with the empty state. Equation (3.52) is entirely equivalent to placing a positively charged particle in the empty states and assuming all other states in the hand are empty, or neutrally charged. This concept is shown in Figure 3.18. Figure 3.18a shows the valence band with the conventional electron-filled states and empty states, while Figure 3.18b shows the new concept of positive charges occupying the original empty states. This concept is consistent with the discussion of the positively charged "empty state" in the valence band as shown in Figure 3.17.

The v_i in the summation of Equation (3.52) is related to how well this positively charged panicle moves in the semiconductor. Now consider an electron near the top of the allowedenergy band shown in Figure 3.16b. The energy near the top of the allowed energy hand may again he approximated by a parabola so that we may write

$$(E - E_v) = -C_2(k)^2 \tag{3.53}$$

The energy E_v is the energy at the top of the energy hand. Since E $<$ E_v for electrons in this band, then the parameter C_2 must be a positive quantity.

Taking the second derivative of E with respect to k from Equation (3.53). we obtain

$$\frac{d^2E}{dk^2} = -2C_2 \tag{3.54}$$

We may rearrange this equation so that

$$\frac{1}{\hbar^2}\frac{d^2E}{dk^2} = \frac{-2C_2}{\hbar^2} \tag{3.55}$$

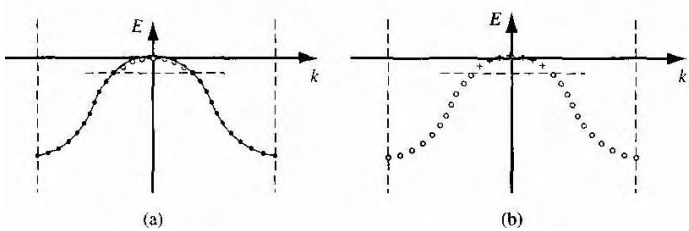


Figure 3.18 f (a) Valence band with conventional electron-filled states and empty states. (b) Concept of positive charges occupying thc original empty states.

Comparing Equation (3.55) with Equation (3.41), we may write

$$\frac{1}{\hbar^2}\frac{d^2E}{dk^2} = \frac{-2C_2}{\hbar^2} = \frac{1}{m^*} \tag{3.56}$$

where m* is again an effective mass. We have argued that C_2 is a positive quantity, which now implies that m^* is a negative quantity. An electron moving near the topo an allowed energy hand behaves as if it has a negative mass.

We must keep in mind that the effective mass parameter is used to relate quantum mechanics and classical mechanics. The attempt to relate these two theories leads to this strange result of a negative effective mass. However, we must recall that solutions to Schrodinger's wave equation also led to results that contradicted classical mechanics. The negative effective mass is another such example.

In discussing the concept of effective mass in the last section, we used an analogy of marbles moving through two liquids. Now consider placing an ice cube in the center of a container filled with water: the ice cube will move upward toward the surface in a direction opposite to the gravitational force. The ice cube appears to have a negative effective mass since its acceleration is opposite to the external force. The effective mass parameter takes into account all internal forces acting on the particle.

If we again consider an electron near the top of an allowed energy band and use Newton's force equation for an applied electric field, we will have

$$F = m^*a = -e\mathrm{E} \tag{3.57}$$

However, m^* is now a negative quantity, so we may write

$$a = \frac{-e\mathrm{E}}{-|m^*|} = \frac{+e\mathrm{E}}{|m^*|} \tag{3.58}$$

An electron moving near the top of an allowed energy band moves in the same direction as the applied electric field.

The net motion of electrons in a nearly full hand can he described by considering just the empty states, provided that a positive electronic charge is associated with each state and that the negative of m* from Equation (3.56) is associated with each state. We now can model this band as having particles with a positive electronic charge and a positive effective mass. The density of these panicles in the valence band is the same as the density of empty electronic energy states. This new panicle is the *hole*. The hole, then, has a positive effective mass denoted by m_p^* and a positive electronic charge, so it will move in the same direction as an applied field.

3.2.5 Metals, Insulators, and Semiconductors

Each crystal has its own energy-band structure. We noted that the splitting of the energy states in silicon, for example, to form the valence and conduction bands, was complex. Complex band splitting occurs in other crystals, leading to large variations in band structures between various solids and to a wide range of electrical characteristics observed in these various materials. We can qualitatively begin to understand

some basic differences in electrical characteristics caused by variations in band structure by considering some simplified energy bands.

There are several possible energy-band conditions to consider. Figure 3.19a shows an allowed energy band that is completely empty of electrons. If an electric field is applied, there are no particles to move, so there will be no current. Figure 3.19b shows another allowed energy band whose energy states are completely full of electrons. We argued in the previous section that a completely full energy band will also not give rise to a current. A material that has energy bands either completely empty or completely full is an insulator. The resistivity of an insulator is very large or, conversely, the conductivity of an insulator is very small. There are essentially no charged particles that can contribute to a drift current. Figure 3.19c shows a simplified energy-band diagram of an insulator. The bandgap energy E_g of an insulator is usually on the order of 3.5 to 6 eV or larger, so that at room temperature, there are essentially no electrons in the conduction band and the valence band remains completely full. There are very few thermally generated electrons and holes in an insulator.

Figure 3.20a shows an energy band with relatively few electrons near the bottom of the band. Now, if an electric field is applied, the electrons can gain energy, move to

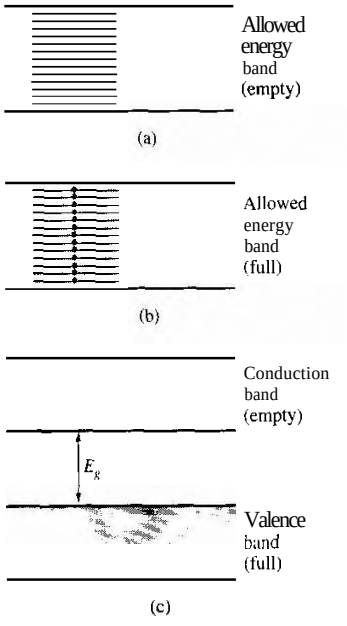


Figure 3.19 | Allowed energy bands showing (a) an empty band, (b) a completely full band, and (c) the bandgap energy between the two allowed bands.

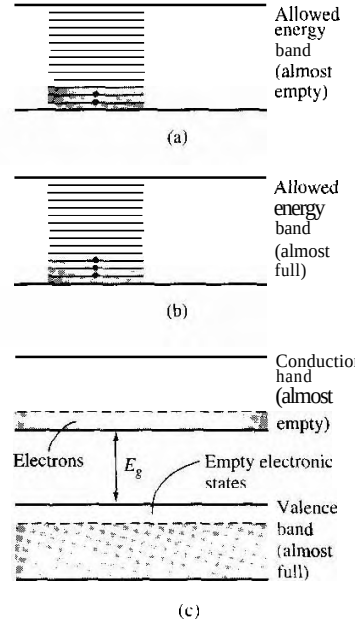


Figure 3.20 | Allowed energy bands showing (a) an almost empty band, (b) an almost full band, and (c) the bandgap energy between the two allowed bands.

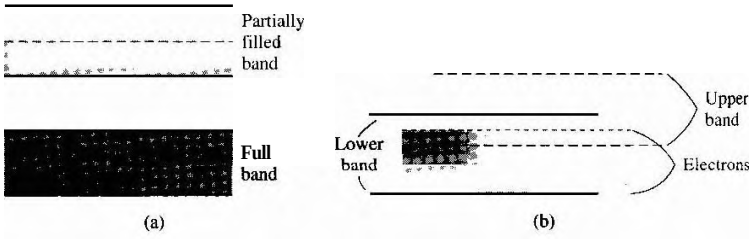


Figure 3.21 | Two possible energy bands of a metal showing (a) a partially filled band and (b) overlapping allowed energy bands.

higher energy states, and move through the crystal. The net flow of charge is a current. Figure 3.20b shows an allowed energy band that is almost full of electrons. which means that we can consider the holes in this band. If an electric field is applied, the holes can move and give rise to a current. Figure 3.20c shows the simplified energy-band diagram for this case. The bandgap energy may be on the order of 1 eV. This energy-band diagram represents a semiconductor for $T > 0$ K. The resistivity of a semiconductor, as we will see in the next chapter, can be controlled and varied over many orders of magnitude.

The characteristics of a metal include a very low resistivity. The energy-band diagram for a metal may be in one of two forms. Figure 3.21a shows the case of a partially full band in which there are many electrons available for conduction, so that the material can exhibit a large electrical conductivity. Figure 3.21b shows another possible energy-band diagram of a metal. The band splitting into allowed and forbidden energy bands is a complex phenomenon and Figure 3.21b shows a case in which the conduction and valence bands overlap at the equilibrium interatomic distance. As in the case shown in Figure 3.21a, there are large numbers of electrons as well as large numbers of empty energy states into which the electrons can move, so this material can also exhibit a very high electrical conductivity.

3.3 | EXTENSION TO THREE DIMENSIONS

The basic concept of allowed and forbidden energy bands and the basic concept of effective mass have been developed in the last sections. In this section, we will extend these concepts to three dimensions and to real crystals. We will qualitatively consider particular characteristics of the three-dimensional crystal in terms of the E versus k plots, bandgap energy, and effective mass. We must emphasize that we will only briefly touch on the basic three-dimensional concepts; therefore, many details will not be considered.

One problem encountered in extending the potential function to a three-dimensional crystal is that the distance between atoms varies as the direction through the crystal changes. Figure 3.22 shows a face-centered cubic structure with the [100] and [110] directions indicated. Electrons traveling in different directions encounter different potential patterns and therefore different k-space boundaries. The E versus k diagrams are in general a function of the k-space direction in a crystal.

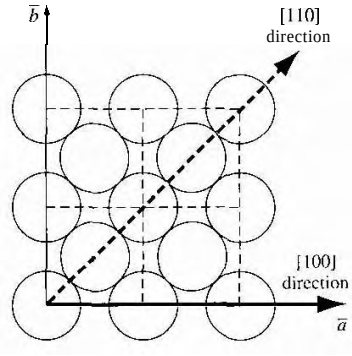


Figure **3.22** | The (100) plane of a face-centered cubic crystal showing the [100] and [110] directions.

3.3.1 The k-Space Diagrams of Si and GaAs

Figure 3.23 shows an E versus k diagram of gallium arsenide and of silicon. These simplified diagrams show the basic properties considered in this text, but do not show many of the details more appropriate for advanced-level courses.

Note that in place of the usual positive and negative k axes, we now show two different crystal directions. The E versus k diagram for the one-dimensional model

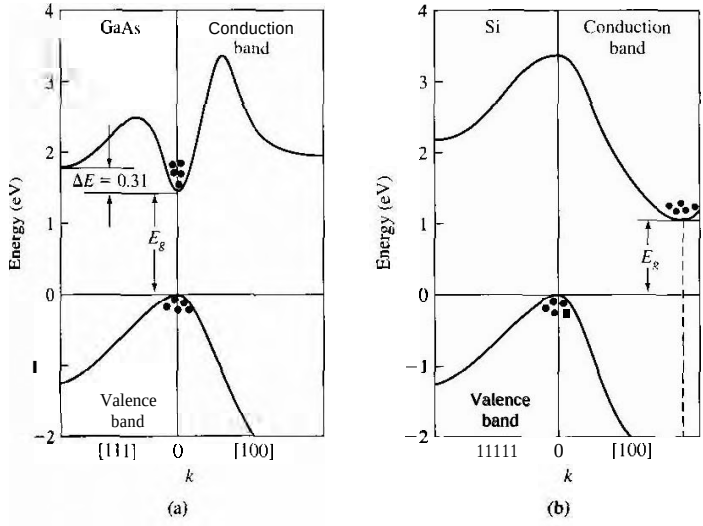


Figure **3.23** | Energy band structures of (a) GaAs and (b) Si *(From Sze [11].)*

was symmetric ink so that no new information is obtained by displaying the negative axis. It is normal practice to plot the [100] direction along the normal $+k$ axis and to plot the [III] portion of the diagram so the $+k$ points to the left. In the case of diamond or zincblende lattices, the maxima in the valence band energy and minima in the conduction band energy occur at k = 0 or along one of these two direction*.

Figure 3.23a shows the E versus k diagram for GaAs. The valence band maximum and the conduction hand minimum both occur at k = 0. The electrons in the conduction band tend to settle at the minimum conduction band energy which is at k = 0. Similarly, holes in the valence band tend to congregate at the uppermost valence band energy. In GaAs, the minimum conduction band energy and maximum valence band energy occur at the same k value. A semiconductor with this property is said to be a *direct* bandgap semiconductor; transitions between the two allowed bands can take place with no change in crystal momentum. This direct nature has significant effect on the optical properties of the material. GaAs and other direct bandgap materials are ideally suited for use in semiconductor lasers and other optical devices.

The E versus k diagram for silicon is shown in Figure 3.23b. The maximum in the valence band energy occurs at k = 0 as before. The minimum in the conduction band energy occurs not at k = 0, hut along the [100] direction. The difference between the minimum conduction band energy and the maximum valence band energy is still defined as the bandgap energy E_g. A semiconductor whose maximum valence band energy and minimum conduction band energy do not occur at the same k value is called an *indirect* bandgap semiconductor. When electrons make a transition between the conduction and valence bands, we must invoke the law of conservation of momentum. A transition in an indirect bandgap material must necessarily include an interaction with the crystal so that crystal momentum is conserved.

Germanium is also an indirect bandgap material, whose valence band maximum occurs at k = 0 and whose conduction band minimum occurs along the [111] direction. GaAs is a direct bandgap semiconductor, but other compound semiconductors, such as GaP and AlAs, have indirect bandgaps.

3.3.2 Additional Effective Mass Concepts

The curvature of the E versus k diagrams near the minimum of the conduction band energy is related to the effective mass of the electron. We may note from Figure 3.23 that the curvature of the conduction band at its minimum value for GaAs is larger than that of silicon, so the effective mass of an electron in the conduction hand of GaAs will be smaller than that in silicon.

For the one-dimensional E versus k diagram, the eftective mass was defined by Equation (3.41) as $1/m^* = 1/\hbar^2 \cdot d^2E/dk^2$. A complication occurs in the effective mass concept in a real crystal. A three-dimensional crystal can be described by t k vectors. The curvature of the E versus kdiagram at the conduction band minimum m not be the same in the three k directions. We will not consider the details of the vario effective mass parameters here. In later sections and chapters, the effective mass parameters used in calculations will be a kind of statistical average that is adequate for most device calculations.

3.4 DENSITY OF STATES FUNCTION

As we have stated. we eventually wish to describe the current-voltage characteristics of semiconductor devices. Since current is due to the flow of charge. an important step in the process is to determine the number of electrons and holes in the semiconductor that will he available for conduction. The number of carriers that can contribute to the conduction process is a function of the number of available energy or quantum states since, by the Pauli exclusion principle, only one electron can occupy a given quantum state. When we discussed the splitting of energy levels into bands of allowed and forbidden energies, we indicated that the band of allowed energies was actually made up of discrete energy levels. We must determine the density of these allowed energy states as a function of energy in order to calculate the electron and hole concentrations.

3.4.1 *Mathematical* Derivation

To determine the density of allowed quantum states as a function of energy, we need to consider an appropriate mathematical model. Electrons are allowed to move relatively freely in the conduction band of a semiconductor, but are confined to the crystal. As a first step, we will consider a free electron confined to a three-dimensional infinite potential well, where the potential well represents the crystal. The potential of the infinite potential well is defined as

$$\begin{aligned} V(x, y, z) &= 0 \qquad \text{for } 0 < x < a \\ &\qquad\qquad\quad 0 < y < a \\ &\qquad\qquad\quad 0 < z < a \\ V(x, y, z) &= \infty \qquad \text{elsewhere} \end{aligned} \tag{3.59}$$

where the crystal is assumed to be a cube with length a. Schrodinger's wave equation in three dimensions can be solved using the separation of variables technique. Extrapolating the results from the one-dimensional infinite potential well, we can show (see Problem 3.21) that

$$\frac{2mE}{\hbar^2} = k^2 = k_x^2 + k_y^2 + k_z^2 = \left(n_x^2 + n_y^2 + n_z^2\right)\left(\frac{\pi^2}{a^2}\right) \tag{3.60}$$

where n_x, n_y, and n_z are positive integers. (Negative values of n_x, n_y, and n_z yield the same wave function, except for the sign, as the positive integer values, resulting in the same probability function and energy, so the negative integers do not represent a different quantum state.)

We can schematically plot the allowed quantum states in k space. Figure 3.24a shows a two-dimensional plot as a function of k_x and k_y. Each point represents an allowed quantum state corresponding to various integral values of n_x and n_y. Positive and negative values of k_x, k_y, or k_z have the same energy and represent the same

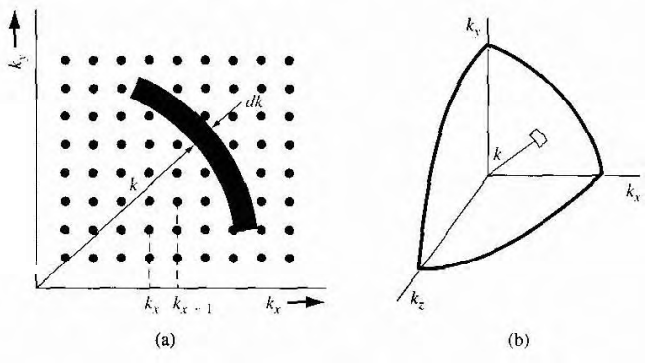


Figure 3.24 I (a)A two-dimensional array of allowed quantum stales in k space. (b) The positive one-eighth of the spherical k space.

energy state. Since negative values of k,, k,, or k_z do not represent additional quantum states, the density of quantum states will be determined by considering only the positive one-eighth of the spherical k space as shown in Figure 3.24b.

The distance between two quantum states in the k_x direction, for example, is given by

$$k_{x+1} - k_x = (n_x + 1)\left(\frac{\pi}{a}\right) - n_x\left(\frac{\pi}{a}\right) = \frac{\pi}{a} \tag{3.61}$$

Generalizing this result to three dimensions, the volume V_k of a single quantum state is

$$V_k = \left(\frac{\pi}{a}\right)^3 \tag{3.62}$$

We can now determine the density of quantum states in k space. **A** differential **volume** in k space is shown in Figure 3.24b and is given hy $4\pi k^2$ dk, so the differential density of quantum states in k space can he written as

$$g_T(k)\,dk = 2\left(\frac{1}{8}\right)\frac{4\pi k^2\,dk}{\left(\frac{\pi}{a}\right)^3} \tag{3.63}$$

The first factor, 2, takes into account the two spin states allowed for each quantum stale; the next factor, $\frac{1}{8}$, takes into account that we are considering only the quantum states for positive values of k_x, k,, and k,. The factor $4\pi k^2\,dk$ is again the differential volume and the factor $(\pi/a)^3$ is the volume of one quantum state. Equation (3.63) may be simplified to

$$g_T(k)\,dk = \frac{\pi k^2\,dk}{\pi^3}\cdot a^3 \tag{3.64}$$

Equation (3.64) gives the density of quantum states as a function of momentum, through the parameter k. We can now determine the density of quantum states as a function of energy E. For a free electron, the parameters E and k are related by

$$k^2 = \frac{2mE}{\hbar^2} \tag{3.65a}$$

or

$$k = \frac{1}{\hbar}\sqrt{2mE} \tag{3.65b}$$

The differential dk is

$$dk = \frac{1}{\hbar}\sqrt{\frac{m}{2E}}\,dE \tag{3.66}$$

Then, substituting the expressions for k^2 and dk into Equation (3.64). the number of energy states between E and $E + dE$ is given by

$$g_T(E)\,dE = \frac{\pi a^3}{\pi^3}\left(\frac{2mE}{\hbar^2}\right)\cdot\frac{1}{\hbar}\sqrt{\frac{m}{2E}}\,dE \tag{3.67}$$

Since $\hbar = h/2\pi$, Equation (3.67) becomes

$$g_T(E)\,dE = \frac{4\pi a^3}{h^3}\cdot(2m)^{3/2}\cdot\sqrt{E}\,dE \tag{3.68}$$

Equation (3.68) gives the total number of quantum states between the energy E and $E + dE$ in the crystal space volume of a^3. If we divide by the volume a^3, then we will obtain the density of quantum states per unit volume of the crystal. Equation (3.68) then becomes

$$g(E) = \frac{4\pi(2m)^{3/2}}{h^3}\sqrt{E} \tag{3.69}$$

The density of quantum states is a function of energy E. As the energy of this free electron becomes small, the number of available quantum states decreases. This density function is really a double density, in that the units are given in terms of states per unit energy per unit volume.

Objective | **EXAMPLE 3.3**

To calculate the density of states per unit volume over a particular energy range.

Consider the density of states for a free electron given by Equation (3.69). Calculate the density of states per unit volume with energies between 0 and 1 eV.

■ **Solution**

The volume density of quantum states, from Equation (3.69), is

$$\blacksquare = \int_0^{1\,\text{eV}} g(E)\,dE = \frac{4\pi(2m)^{3/2}}{h^3} \cdot \int_0^{1\,\text{eV}} \sqrt{E}\,dE$$

or

$$N = \frac{4\pi(2m)^{3/2}}{h^3} \cdot \frac{2}{3} \cdot E^{3/2}$$

The density of states is now

$$N = \frac{4\pi[2(9.11 \times 10^{-31})]^{3/2}}{(6.625 \times 10^{-34})^3} \cdot \frac{2}{3} \cdot (1.6 \times 10^{-19})^{3/2} = 4.5 \times 10^{27}\ \text{m}^{-3}$$

or

$$N = 4.5 \times 10^{21}\ \text{states/cm}^3$$

■ **Comment**

The density of quantum states is typically a large number. An effective density of states in a semiconductor, as we will see in the following sections and in the next chapter, is also a large number, but is usually less than the density of atoms in the semiconductor crystal.

3.4.2 Extension to Semiconductors

In the last section, we derived a general expression for the density of allowed electron quantum states using the model of a free electron with mass m bounded in a three-dimensional infinite potential well. We can extend this same general model to a semiconductor to determine the density of quantum states in the conduction band and the density of quantum states in the valence band. Electrons and holes are confined within the semiconductor crystal so we will again use the basic model of the infinite potential well.

The parabolic relationship between energy and momentum of a free electron was given in Equation (3.28) as $E = p^2/2m = \hbar^2k^2/2m$. Figure 3.16a showed the conduction energy band in the reduced k space. The E versus k curve near $k = 0$ at the bottom of the conduction band can be approximated as a parabola, so we may write

$$E = E_c + \frac{\hbar^2k^2}{2m_n^*} \tag{3.70}$$

where E_c is the bottom edge of the conduction hand and m_n^* is the electron effective mass. Equation (3.70) may be rewritten to give

$$E - E_c = \frac{\hbar^2k^2}{2m_n^*} \tag{3.71}$$

The general form of the E versus k relation for an electron in the bottom of a conduction band is the same as the free electron, except the mass is replaced by the effective mass. We can then think of the electron in the bottom of the conduction band as being a "free" electron with its own particular mass. The right side of Equation (3.71) is of the same form a the right side of Equation (3.28), which was used in the derivation of the density of states function. Because of this similarity, which yields the "free" conduction electron model, we may generalize the free electron results of Equation *(3.69)* and write the density of allowed electronic energy states in the conduction hand as

$$g_c(E) = \frac{4\pi(2m_n^*)^{3/2}}{h^3}\sqrt{E - E_c} \tag{3.72}$$

Equation *(3.72)* is valid for $E \geq E_c$. As the energy of the electron in the conduction band decreases, the number of available quantum states also decreases.

The density of quantum states in the valence band can be obtained by using the same infinite potential well model, since the hole is also confined in the semiconductor crystal and can he treated as a "free" particle. The effective mass of the hole is m_p^*. Figure 3.16b showed the valence energy band in the reduced k space. We may also approximate the E versus k curve near $k = 0$ by a parabola for a "free" hole, so that

$$E = E_v - \frac{\hbar^2 k^2}{2m;} \tag{3.73}$$

Equation *(3.73)* may he rewritten to give

$$E_v - E = \frac{\hbar^2 k^2}{2m_p^*} \tag{3.74}$$

Again, the right side of Equation *(3.74)* is of the same form used in the general derivation of the density of states function. We may then generalize the density of states function from Equation *(3.69)* to apply to the valence band, so that

$$g_v(E) = \frac{4\pi(2m_p^*)^{3/2}}{h^3}\sqrt{E_v - E} \tag{3.75}$$

Equation *(3.75)* is valid for $E \leq E_v$.

We have argued that quantum states do not exist within the forbidden energy band, so $g(E) = 0$ for $E_v < \mathrm{E} < \mathrm{E}_c$. Figure 3.25 shows the plot of the density of quantum states as a function of energy. If the electron and hole effective masses were equal, then the functions $g_c(E)$ and $g_v(E)$ would be symmetrical about the energy midway between E_c and E_v, or the midgap energy. E_{midgap}.

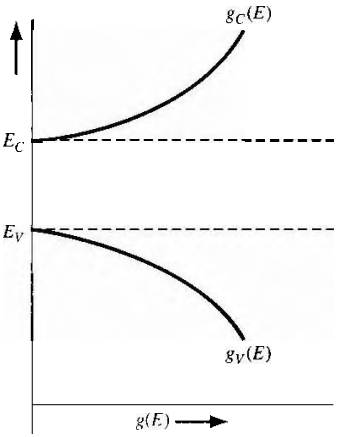


Figure 3.25 | The density of energy states in the conduction band and the density of energy states in the valence band as a function of energy.

TEST YOUR UNDERSTANDING

E3.2 Determine the total number of energy states in silicon between E, and E, $+kT$ at T = 300 K. (Ans. 2.12×10^{19} cm^{-3})

E3.3 Determine the total number of energy states in silicon between E , and E, $-kT$ at T = 300 K. (Ans. 7.92×10^{18} cm^{-3})

3.5 | STATISTICAL MECHANICS

In dealing with large numbers of particles, we are interested only in the statistical behavior of the group as a whole rather than in the behavior of each individual particle. For example, gas within a container will exert an average pressure on the walls of the vessel. The pressure is actually due to the collisions of the individual gas molecules with the walls, but we do not follow each individual molecule as it collides with the wall. Likewise in a crystal, the electrical characteristics will be determined hy the statistical behavior of a large number of electrons.

3.5.1 Statistical Laws

In determining the statistical behavior of particles, we must consider the laws that the particles obey. There are three distribution laws determining the distribution of particles among available energy states.

One distribution law is the Maxwell–Boltzmann probability function. In this case, the panicles are considered to be distinguishable by being numbered, for example, from I to N. with no limit to the number of particles allowed in each energy state. The behavior of gas molecules in a container at fairly low pressure is an example of this distribution.

Asecond distribution law is the Bose–Einstein function. The panicles in this case are indistinguishableand, again, there is no limit to the number of particles permitted in each quantum state. The behavior of photons, or black body radiation, is an example of this law.

The third distribution law is the Fermi–Dirac probability function. In this case, the particles are again indistinguishable, but now only one particle is permitted in each quantum state. Electrons in a crystal obey this law. In each case, the particles are assumed to be noninteracting.

3.5.2 The Fermi–Dirac Probability Function

Figure 3.26 shows the ith energy level with g_i quantum states. A maximum of one particle is allowed in each quantum state by the Pauli exclusion principle. There are g_i ways of choosing where to place the first panicle, $(g_i - 1)$ ways of choosing where to place the second particle, $(g_i - 2)$ ways of choosing where to place the third particle, and so on. Then the total number of ways of arranging N_i particles in the *i*th energy level (where $N_i \leq g_i$) is

$$(g_i)(g_i - 1)\cdots(g_i - (N_i - 1)) = \frac{g_i!}{(g_i - N_i)!} \tag{3.76}$$

This expression includes all permutations of the N_i particles among themselves.

However, since the particles are indistinguishable, the $N_i!$ number of permutations that the particles have among themselves in any given arrangement do not count as separate arrangements. The interchange of any two electrons. for example, does not produce a new arrangement. Therefore, the actual number of independent ways of realizing a distribution of N_i particles in the ith level is

$$W_i = \frac{g_i!}{N_i!(g_i - N_i)!} \tag{3.77}$$

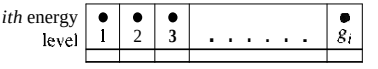


Figure 3.261 The *ith* energy level wrth g_i quantum states.

EXAMPLE 3.4 Objective

To determine the possible number of ways of realizing a particular distribution.

Let $y, = N_i = 10$. Then $(g_i - N_i)! = 1$.

■ Solution

Equation (3.77) becomes

$$\frac{g_i!}{N_i!(g_i - N_i)!} = \frac{10!}{10!} = 1$$

■ Comment

If we have 10 particles to be arranged in 10 quantum states, there is only one possible arrangement. Each quantum state contains one particle.

EXAMPLE 3.5 Objective

To again determine the possible number of ways of realizing a particular distribution.

Let $g_i = 10$ and $N_i = 9$. In this case $g_i - N_i = 1$ so that $(g_i - N_i)! = 1$.

■ Solution

Equation (3.77) becomes

$$\frac{g_i!}{N_i!(g_i - N_i)!} = \frac{10!}{(9!)(1)} = \frac{(10)(9!)}{9!} = 10$$

■ Comment

In this case, if we have 10 quantum states and 9 particles. there is one empty quantum state. There are 10 possible arrangements, or positions, for the one empty state.

Equation (3.77) gives the number of independent ways of realizing a distribution of N, particles in the ith level. The total number of ways of arranging $(N_1, N_2, N_3, \ldots, N_n)$ indistinguishable particles among n energy levels is the product of all distributions, or

$$W = \prod_{i=1}^{n} \frac{g_i!}{N_i!(g_i - N_i)!} \tag{3.78}$$

The parameter W is the total number of ways in which Nelectrons can be arranged in this system, where $N = \sum_{i=1}^{n} N$, is the total number of electrons in the system. We want to find the most probable distribution, which means that we want to find the maximum W. The maximum W is found by varying N_i among the E_i levels, which vanes the distribution, but at the same time, we will keep the total number of particles and total energy constant.

We may write the most probable distribution function as

$$\frac{N(E)}{g(E)} = f_F(E) \simeq \frac{1}{1 + \exp\left(\frac{E - E_F}{kT}\right)} \tag{3.79}$$

where E_F is called the Fermi energy. The number density $N(E)$ is the number of particles per unit volume per unit energy and the function $g(E)$ is the number of quantum states per unit volume per unit energy. The function $f_F(E)$ is called the Fermi–Dirac distribution or probability function and gives the probability that a quantum state at the energy E will be occupied by an electron. Another interpretation of the distribution function is that $f_F(E)$ is the ratio of filled to total quantum states at any energy E.

3.5.3 The Distribution Function and the Fermi Energy

To begin to understand the meaning of the distribution function and the Fermi energy, we can plot the distribution function versus energy. Initially, let T = 0 K and consider the case when $E < E_F$. The exponential term in Equation (1.79) becomes $\exp[(E - E_F)/kT] \rightarrow \exp(-\infty) = 0$. The resulting distribution function is $f_F(E < E_F) = 1$. Again let T = 0 K and consider the case when $E > E_F$. The exponential term in the distribution function becomes $\exp[(E - E_F)/kT] \rightarrow \exp(+\infty) \rightarrow +\infty$. The resulting Fermi–Dirac distribution function now becomes $f_F(E > E_F) = 0$.

The Fermi–Dirac distribution function for T = 0 K is plotted in Figure 3.27. This result shows that, for T = 0 K, the electrons are in their lowest possible energy states. The probability of a quantum state being occupied is unity for $E < E_F$ and the probability of a state being occupied is zero for $E > E_F$. All electrons have energies below the Fermi energy at T = 0 K.

Figure 3.28 shows discrete energy levels of a particular system as well as the number of available quantum states at each energy. If we assume, for this case, that

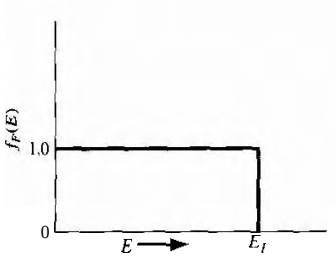


Figure 3.27 | The Fermi probability function versus energy for T = 0 K.

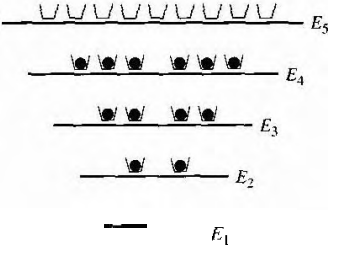


Figure 3.28 | Discrete energy states and quantum states for a particular system at T = 0 K.

the system contains 13 electrons. then Figure 3.28 shows how these electrons are distributed among the various quantum states at T = 0 K. The electrons will he in the lowest possible energy state, so the probability of a quantum state being occupied in energy levels E_1 through E_4 is unity, and the probability of a quantum state being occupied in energy level E_5 is zero. The Fermi energy, for this case, must be above E_4 but less than E_5. The Fermi energy determines the statistical distribution of electrons and does not have to correspond to an allowed energy level.

Now consider a case in which the density of quantum states $g(E)$ is a continuous function of energy as shown in Figure 3.29. If we have N_0 electrons in this system, then the distribution of these electrons among the quantum states at T = 0 K is shown by the dashed line. The electrons are in the lowest possible energy state so that all states below E_F are tilled and all states above E_F are empty. If $g(E)$ and N_0 are known for this particular system, then the Fermi energy E_F can be determined.

Consider the situation when the temperature increases above T = 0 K. Electrons gain a certain amount of thermal energy so that some electrons can jump to higher energy levels, which means that the distribution of electrons among the available energy states will change. Figure 3.30 shows the same discrete energy levels and quantum states as in Figure 3.28. The distribution of electrons among the quantum states has changed from the $T = 0$ K case. Two electrons from the E_4 level have gained enough energy to jump to E_5, and one electron from E_3 has jumped to E_4. As the temperature changes, the distribution of electrons versus energy changes.

The change in the electron distribution among energy levels for T > 0 K can be seen by plotting the Fermi–Dirac distribution function. If we let $E = E_F$ and T > 0K, then Equation (3.79) becomes

$$f_F(E = E_F) = \frac{1}{1 + \exp(0)} = \frac{1}{1 + 1} = \frac{1}{2}$$

The probability of a state being occupied at $E = E_F$ is $\frac{1}{2}$. Figure 3.31 shows the Fermi–Dirac distribution function plotted for several temperatures, assuming the Fermi energy is independent of temperature.

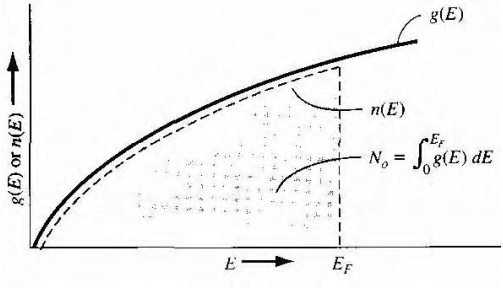


Figure 3.29 | Density of quantum states and electrons in a continuous energy system at T = 0 K.

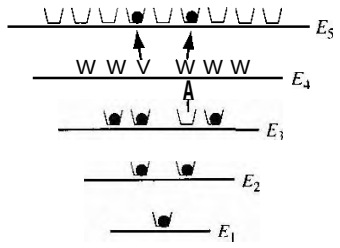


Figure 3.30 | Discrete energy states and quantum states for the same system shown in Figure 3.28 for T > 0 K.

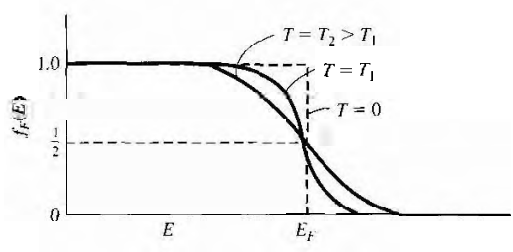


Figure 3.31 | The Fermi probability function versus energy for different temperatures.

We can see that for temperatures above absolute zero, there is a nonzero probability that some energy states above E_F will be occupied by electrons and some energy states below E_F will be empty. This result again means that some electrons have jumped to higher energy levels with increasing thermal energy.

Objective | ***EXAMPLE* 3.6**

To calculate the probability that an energy state above E_F is occupied by an electron.

Let T = 300 K. Determine the probability that an energy level 3kT above the Fermi energy is occupied by an electron.

■ Solution

From Equation (3.79), we can write

$$f_F(E) = \frac{1}{1 + \exp\left(\frac{E - E_F}{kT}\right)} \approx \frac{1}{1 + \exp\left(\frac{3kT}{kT}\right)}$$

which becomes

$$f_F(E) = \frac{1}{1 + 20.09} = 0.0474 = 4.74\%$$

■ Comment

At energies above E_F, the probability of a state being occupied by an electron can become significantly less than unity. or the ratio of electrons to available quantum states can he quite small.

TEST YOUR UNDERSTANDING

E3.4 Assume the Fermi energy level is 0.30 eV below the conduction band energy.
(*a*) Determine the probability of a state being occupied by an electron at E_c.
(*b*) Repeat part (*a*) for an energy state at $E_c + kT$. Assume T = 300 K.
[Ans. (*a*) 9.32×10^{-6}, (*b*) 3.43×10^{-10}]

E3.5 Assume the Fermi energy level is 0.35 eV above the valence band energy. (*a*) Determine the probability of a state being empty of an electron at E,. (*b*) Repeat pan (*a*) for an energy state at $E_V - kT$. Assume $\mathcal{T} = 300\ K$.
[Ans. (*a*) 1.35×10^{-6}, (*b*) 4.98×10^{-1}]

We can see from Figure 3.31 that the probability of an energy above E_F being occupied increases as the temperature increases and the probability of a state below E_F being empty increases as the temperature increases.

EXAMPLE 3.7

Objective

To determine the temperature at which there is a 1 percent probability that an energy state is empty.

Assume that the Fermi energy level for a particular material is 6.25 eV and that the electrons in this material follow the Fermi–Dirac distribution function. Calculate the temperature at which there is a 1 percent probability that a slate 0.30 eV below the Fermi energy level will not contain an electron.

■ Solution

The probability that a state is empty is

$$1 - f_F(E) = 1 - \frac{1}{1 + \exp\left(\dfrac{E - E_F}{kT}\right)}$$

Then

$$0.01 = 1 - \frac{1}{1 + \exp\left(\dfrac{5.95 - 6.25}{kT}\right)}$$

Solving for kT, we find $kT = 0.06529$ eV, so that the temperature is $\mathcal{T} = 756\ K$

■ Comment

The Fermi probability function is a strong function of temperature

TEST YOUR UNDERSTANDING

E3.6 Repeat Exercise E3.4 for $T = 400$ K. [Ans. (*a*) 1.69×10^{-4}, (*b*) 9.20×10^{-5}]
E3.7 Repeat Exercise E3.5 for $T = 400$ K. [Ans. (*a*) 6.53×10^{-5}, (*b*) 1.49×10^{-1}]

We may note that the probability of a state a distance dE above E_F being occupied is the same as the probability of a srate a distance dE below E_F being empty. The function $f_F(E)$ is symmetrical with the function $1 - f_F(E)$ about the Fermi energy, E_F. This symmetry effect is shown in Figure 3.32 and will be used in the next chapter.

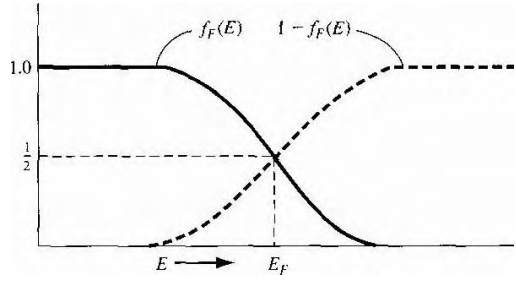


Figure 3.32 The probability of a state being occupied, $f_F(E)$, and the probability of a state being empty, $1 - f_F(E)$

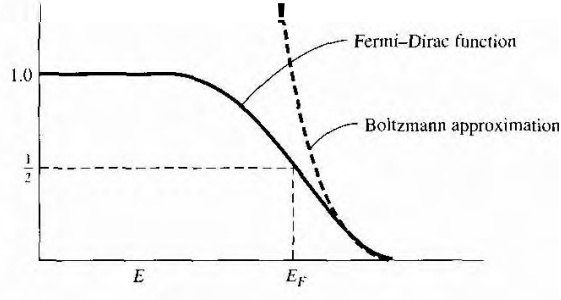


Figure 3.33 The Fermi–Dirac probability function and the Maxwell–Boltrmann approximation.

Consider the case when $E - E_F \gg kT$, where the exponential term in the denominator of Equation (3.79) is much greater than unity. We may neglect the 1 in the denominator, so the Fermi–Dirac distribution function becomes

$$f_F(E) \approx \exp\left[\frac{-(E - E_F)}{kT}\right] \tag{3.80}$$

Equation (3.80) is known as the Maxwell–Boltzmann approximation, or simply the Boltzmann approximation. to the Fermi–Dirac distribution function. Figure 3.33 shows the Fermi–Dirac probability function and the Boltzmann approximation. This figure gives an indication of the range of energies over which the approximation is valid.

EXAMPLE 3.8

Objective

To determine the energy at which the Boltzmann approximation may be considered valid.

Calculate the energy, in terms of kT and E_F, at which the difference between the Boltzmann approximation and the Fermi–Dirac function is 5 percent of the Fermi function.

■ **Solution**

We can write

$$\frac{\exp\left[\dfrac{-(E-E_F)}{kT}\right] - \dfrac{1}{1+\exp\left(\dfrac{E-E_F}{kT}\right)}}{\dfrac{1}{1+\exp\left(\dfrac{E-E_F}{kT}\right)}} = 0.05$$

If we multiply both numerator and denominator by the $1 + \exp(\,)$ function, we have

$$\exp\left[\frac{-(E-E_F)}{kT}\right]\left\{1+\exp\left[\frac{E-E_F}{kT}\right]\right\} - 1 = 0.05$$

which becomes

$$\exp\left[\frac{-(E-E_F)}{kT}\right] = 0.05$$

$$(E-E_F) = kT\,ln\left(\frac{1}{0.05}\right) \approx 3kT$$

■ **Comment**

As seen in this example and in Figure 3.33. the $E - E_F \gg kT$ notation is somewhat misleading. The Maxwell–Boltzmann and Fermi–Dirac functions are within 5 percent of each other when $E - E_F \approx 3kT$.

The actual Boltzmann approximation is valid when $\exp[(E-E_F)/kT] \gg 1$ However, it is still common practice to use the $E - E_F \gg kT$ notation when applying the Boltzmann approximation. We will use this Boltzmann approximation in our discussion of semiconductors in the next chapter.

3.6 SUMMARY

Discrete allowed electron energies split into a band of allowed energies as atoms are brought together to form a crystal.

- The concept of allowed and forbidden energy bands was developed more rigorously by considering quantum mechanics and Schrodinger's wave equation using the Kronig–Penney model representing the potential function of a single crystal material. This result forms the basis of the energy band theory of semiconductors.
- The concept of effective mass was developed. Effective mass relates the motion of a particle in a crystal to an externally applied force and takes into account the effect of the crystal lattice on the motion of the particle.
- Two charged particles exist in a semiconductor An electron is a negatively charged panicle with a positive effective mass existing at the bottom of an allowed energy band. A hole is a positively charged particle with a positive effective mass existing at the top of an allowed energy band.

- The E versus kdiagram of silicon and gallium arsenide were given and the concept of direct and indirect bandgap semiconductors was discussed.
  Energies within an allowed energy band are actually at discrete levels and each contains a finite number of quantum states. The density per unit energy of quantum states was determined by using the three-dimensional infinite potential well as a model.
- In dealing with large numbers of electrons and holes, we must consider the statistical behavior of these particles. The Fermi–Dirac probability function was developed. which gives the probability of a quantum state at an energy E of being occupied by an electron. The Fermi energy was defined.

GLOSSARY OF IMPORTANT TERMS

allowed energy band A band or range of energy levels that an electron in a crystal is allowed to occupy based on quantum mechanics.

density of states function The density of available quantum states as a functiun of energy, given in units of number per unit energy per unit volume.

electron effective mass The parameter that relates the acceleration of an electron in the conduction band of a crystal to an external force: a parameter that takes into account the effect of internal forces in the crystal.

Fermi–Dirae probability function The function describing the statistical distribution of electrons among available energy states and the probability that an allowed energy state is occupied by an electron.

fermi energy In the simplest definition, the energy below which all states are filled with electrons and above which all states are empty at $T = 0$ **K.**

forbidden energy band A hand or range of energy levels that an electron in a crystal is not allowed to occupy based on quantum mechanics.

hole The positively charged "particle" associated with an empty state in the top of the valence band.

hole effective mass The parameter that relates the acceleration of a hole in the valence band of a crystal to an applied external force (**a** positive quantity); a parameter that takes into account the effect of internal forces in **a** crystal.

***k*-space diagram** The plot of electron energy in a crystal versus k, where k is the momentum-related constant of the motion that incorporates the crystal interaction.

Kronig–Penney model The mathematical model of a periodic potential function representing a one-dimensional single-crystill lattice by a series of periodic step functions.

Maxwell–Boltzmann approximation The condition in which the energy is several kT above the Fermi energy or several kT below the Fermi energy so that the Fermi–Dirac probability function can be approximated by **a** simple exponential function.

Pauli exclusion principle The principle which states that no two electrons can occupy the same quantum state.

CHECKPOINT

After studying this chapter, the reader should have the ability to:

- Discuss the concept of allowed and forbidden energy bands in a single crystal both qualitatively and more rigorously from the results of using the Kronig–Penney model

- Discuss the splitting of energy bands in silicon.
  State the definition of effective mass from the E versus k diagram and discuss its meaning in terms of the movement of a particle in a crystal.
  Discuss the concept of a hole.
  Qualitatively, in terms of energy hands, discuss the difference between a metal, insulator, and semiconductor
- Discuss the effective density of states function.
- Understand the meaning of the Fermi–Dirac distribution function and the Fenni energy.

REVIEW QUESTIONS

1. What is the Kronig–Penney model'?
2. State two results of using the Kronig–Penney model with Schrodinger's wave equation.
3. What is effective mass?
4. What is a direct bandgap semiconductor? What is an indirect bandgap semiconductor?
5. What is the meaning of the density of states function'!
6. What was the mathematical model used in deriving the density of states function?
7. In general, what is the relation between density of states and energy?
8. What is the meaning of the Fermi–Dirac probability function?
9. What is the Fermi energy?

PROBLEMS

Section 3.1 Allowed and Forbidden Energy Bands

3.1 Consider Figure 3.4b, which shows the energy-band splitting of silicon. If the equilibrium lattice spacing were to change by a small amount. discuss how you would expect the electrical properties of silicon to change. Determine at what point the material would behave like an insulator or like a metal.

3.2 Show that Equations (3.4) and (3.6) are derived from Schrodinger's wave equation. using the form of solution given by Equation (3.3).

3.3 Show that Equations (3.9) and (3.10) are solutions of the differential equations given by Equations (3.4) and (3.8), respectively.

3.4 Show that Equations (3.12), (3.14), (3.16), and (3.18) result from the boundary conditions in the Kronig–Penney model.



3.5 Plot the function $f(\alpha a) = 9 \sin \alpha a/\alpha a + \text{cosaa}$ for $0 \leq \alpha a \leq 6\pi$. Also, given the function $f(\text{aa}) = \text{cosku}$, indicate the allowed values of αa which will satisfy this equation



3.6 Repeat Problem 3.5 for the function

$$f(\alpha a) = 6 \sin \alpha a/\alpha a + \cos \alpha a = \text{coska}$$

3.7 Using Equation (3.24), show that $dE/dk = 0$ at $\text{k} = n\pi/a$, where $n = 0, 1, 2, \ldots$.

3.8 Using the parameters in Problem 3.5 and letting $a = 5$ A, determine the width (in eV) of the forbidden energy bands that exist at (a) $ka = \pi$. **(b)** $\text{ka} = 2\pi$, (c) $ka = 3\pi$. and (d) $ka = 4\text{n}$. Refer to Figure 3.8c.

3.9 Using the parameters in Problem 3.5 and letting $a = 5$ Å, determine the width (in eV) of the allowed energy bands that exist for (a) $0 < ka < \pi$, (h) $\pi < ka < 2\pi$, (c) $2\pi < ko < 3\pi$, and (d) $3\pi < ka < 4\pi$.

3.10 Repeat Problem 3.8 using the parameters in Problem **3.6.**

3.11 Repeat Problem 3.9 using the parameters in Problem **3.6.**

3.12 The bandgap energy in a semiconductor is usually a slight function of temperature In some cases. the bandgap energy versus temperature can be modeled by

$$E_g = E_g(0) - \frac{\alpha T^2}{(\beta + T)}$$

where $E_g(0)$ is the value of the bandgap energy at T = 0 K. For silicon. the parameter values are $E_g(0) = 1.170$ eV, $\alpha = 4.73 \times 10^{-4}$ eV/K and $\beta = 636$ K. Plot E_g versus Tover the range $0 \leq T \leq 600$ K. In particular, note the value at $T = 300$ K.

Section 3.2 Electrical Conduction in Solids

3.13 Two possible conduction bands are shown in the ***E versus*** k diagram given in Figure 3.34. State which hand will result in the heavier electron effective mass; state why.

3.14 Two possible valence bands are shown in the E versus $\boldsymbol{k}$ diagram given in Figure 3.35. State which band will result in the heavier hole effective mass; state why.

3.15 The E versus $\boldsymbol{k}$ diagram for a panicular allowed energy band is shown in Figure 3.36. Determine (a) the sign of the effective mass and (h) the direction of velocity for a particle at each of the four positions shown.

3.16 Figure 3.37 shows the parabolic E versus $\boldsymbol{k}$ relationship in the conduction band for an electron in two particular semiconductor materials. Determine the effective mass (in units of the free electron mass) of the two electrons.

3.17 Figure 3.38 shows the parabolic E versus $\boldsymbol{k}$ relationship in the valence band for a hole in two particular semiconductor materials. Determine the effective mass (in units of the free electron mass) of the two holes.

3.18 The forbidden energy band of GaAs is 1.42 eV. (a) Determine the minimum frequency of an incident photon that can interact with a valence electron and elevate the electron to the conduction band. (b) What is the corresponding wavelength?

3.19 The E versus k diagrams for a free electron (curve A) and for an electron in a semiconductor (curve B) are shown in Figure 3.39. Sketch (a) dE/dk versus $\boldsymbol{k}$ and

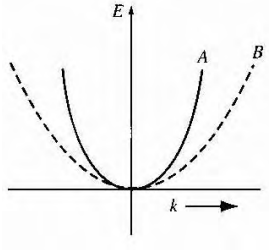


Figure 3.34 | Conduction bands for Problem 3.13.

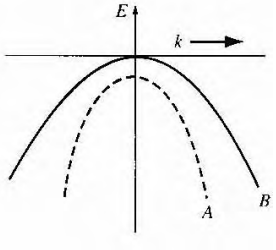


Figure 3.35 | Valence bands for Problem 3.14.

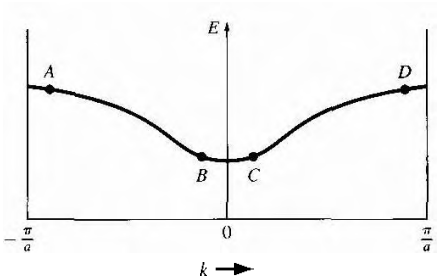


Figure 336 | Figure for Problem 3.15

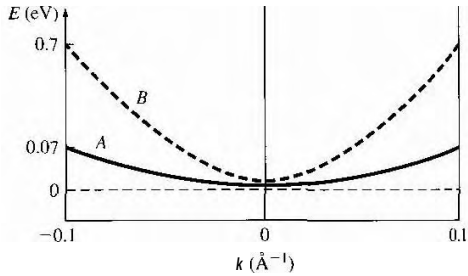


Figure 337 | Figure for Problem 3.16.

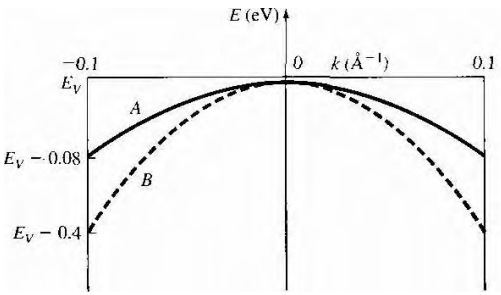


Figure 3.38 | Figure for Problem 3.17.

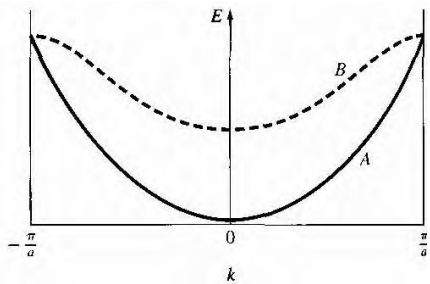


Figure 3.39 | Figure for Problem 3.19.

(h) d^2E/dk^2 versus k for each curve. (c) What conclusion can you make concerning a comparison in effective masses for the two cases?

Section 3.3 Extension to Three Dimensions

3.20 The energy band diagram for silicon is shown in Figure 3.23b. The minimum energy in the conduction band is in the [100] direction. The energy in this one-dimensional direction near the minimum value can be approximated by

$$E = E_0 - E_1 \cos\alpha(k - k_0)$$

where k_0 is the value of k at the minimum energy. Determine the effective mass of the particle at $k = k_0$ in terms of the equation parameters.

Section 3.4 Density of States Function

3.21 Starting with the three-dimensional infinite potential well function given by Equation (3.59) and using the separation of variables technique, derive Equation (3.60).

3.22 Show that Equation (3.69) can be derived from Equation (3.64).

3.23 Determine the total number nf energy states in GaAs between E_c and $E_c + kT$ at $T = 300$ K.

3.24 Determine the total number of energy states in GaAs between E_v and $E_v - kT$ at $T = 300$ K.

3.25 (*a*) Plot the density of states in the conduction band for silicon over the range $E_c \leq E \leq E_c + 0.2$ eV. (*b*) Repeat pan (a) for the density of states in the valence band over the range $E_v - 0.2\,\text{eV} \leq \text{E} \leq E_v$.

3.26 Find the ratio of the effective density of states in the conduction band at $E_c + kT$ to theeffective density of states in the valence band at $E_v - kT$.

Section 3.5 Statistical Mechanics

3.27 Plot the Fermi–Dirac probability function, given by Equation (3.79), over the range $-0.2 \leq (E - E_F) \leq 0.2$ eV for (a) $T = 200$ K, (*b*) $T = 300$ K, and (*c*) T = 400 K.



3.28 Repeat Example 3.4 for the case when $g_c = 10$ and $N_t = 8$.

3.29 **(a)** If $E_F = E_t$, find the probability of a state being occupied at $E = E_c + kT$. (*b*) If $E_F = E_v$, find the probability of a state being empty at $E = E_v - kT$.

3.30 Determine the probability that an energy level is occupied by an electron if the state is above the Fermi level by (*a*) kT, (*b*) $5kT$, and (*c*) $10kT$.

3.31 Determine the probability that an energy level is empty of an electron if the state is below the Fermi level by *(a)* kT, (b) $5kT$, and *(c)* $10kT$.

3.32 The Fermi energy in silicon is 0.25 eV below the conduction band energy E_c. (*a*) Plot the probability of a state being occupied by an electron over the range $E_c \leq \text{E} \leq E_c + 2kT$. Assume $T = 300$ K. (*b*) Repeat part (*a*) for T = 400 K.

333 Four electrons exist in a one-dimensional infinite potential well of width $a = 10$ Å. Assuming the free electron mass, what is the Fermi energy at $T = 0$ K.

334 **(a)** Five electrons exist in a three-dimensional infinite potential well with all three widths equal to $a = 10$ Å. Assuming the free electron mass. what is the Fermi energy at T = 0 K. *(b)* Repeat pan (*a*) for 13 electrons.

3.35 Show that the probability of an energy state being occupied A E above the Fermi energy is the same as the probability of a state being empty A E below the Fermi level.

3.36 **(a)** Determine for what energy above E_F (in terms of kT) the Fermi–Dirac probability function is within I percent of the Boltzmann approximation. (*b*) Give the value of the probability function at this energy.

3.37 The Fermi energy level for a particular material at $T = 300$ K is 6.25 eV. The electrons in this material follow the Fermi–Dirac distribution function. (*a*) Find the probability of an energy level at 6.50 eV being occupied by an electron. (*b*) Repeat pan (a) if the temperature is increased to $T = 950$ K. (Assume that E_F is a constant.) (*c*) Calculate the temperature at which there is a 1 percent probability that a state 0.30 eV below the Fermi level will be empty of an electron.

3.38 The Fermi energy for copper at $T = 300$ K is 7.0 eV. The electrons in copper follow the Fermi–Dirac distribution function. (a) Find the probability of an energy level at 7.15 eV being occupied by an electron. (*b*) Repeat part (*a*) for $T = 1000$ K. (Assume that E_F is a constant.) (c) Repeat part (*a*) for E = 6.85 eV and $T = 300$ K. (*d*) Determine the probability of the energy state at $E = E_F$ being occupied at $T = 300$ K and at T = 1000 K.

3.39 Consider the energy levels shown in Figure 3.40. Let $T = 300$ K. (*a*) If $E_1 - E_F = 0.30$ eV, determine the probability that an energy state at $E = E_1$ is occupied by an electron and the probability that an energy state at $E = E_2$ is empty. (*b*) Repeat pan (a) if $E_F - E_2 = 0.40$ eV.

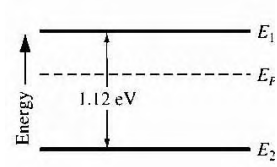


Figure 3.40 | Energy levels for Problem 3.39.

3.40 Repeat problem 3.39 for the case when $E_1 - E_2 = 1.42$ eV.

3.41 Determine the derivative with respect to energy of the Fermi–Dirac distribution function. Plot the derivative with respect to energy for (*a*) $T = 0$ K, (*b*) T = 300 K, and (*c*) $T = 500$ K.

3.42 Assume the Fermi energy level is exactly in the center of the bandgap energy of a semiconductor at $T = 300$ K. (*a*) Calculate the probability that an energy state in the bottom of the conduction band is occupied by an electron for Si, Ge. and GaAs. (*b*) Calculate the probability that an energy state in the top of the valence hand is empty for Si, Ge, and GaAs.

3.43 Calculate the temperature at which there is a 10-▾▾ robability that an energy state 0.55 eV above the Fermi energy level is occupied by an electron.

3.44 Calculate the energy range (in eV) between $f_t(E) = 0.95$ and $f_t(E) = 0.05$ for $E_F = 7.0$ eV and for (*a*) $T = 300$ K and (b) $T = 500$ K.

READING LIST

1. Kano, K. *Semiconductor* ***Devices.*** Upper Saddle River. NJ: Prentice Hall, 1998.
2. Kittel, C. ***Introduction to*** *Solid State Physics,* 7th ed. Berlin: Springer-Verlag, 1993.
3. McKelvey, J. P ***Solid*** *State Physics* ***for*** *Engineering and Materials* ***Science.*** Malabar, FL.: Krieger, 1993.
4. Pierret, R. F. ***Semiconducror*** *Device Fundamentals.* Reading, ***MA:*** Addison-Wesley. 1996.

*5. Shockley, W. ***Electrons and Holes in*** *Semiconductors.* New York: D. Van Nostrand, 1950.

6. Shur, M. ***Introduction to*** *Electronic* ***Devices.*** New York: John Wiley and Sons, 1996.

*7. Shur, M. ***Physics*** *of Semiconductor* ***Devices.*** Englewood Cliffs. NJ: Prentice Hall, 1990.

8. Singh, J. *Semiconductor Devices: An Introduction.* New York: McGraw-Hill, 1994.
9. Singh, J. *Semiconductor Devices: Basic* ***Principles.*** New York: John Wiley and Sons, 2001.
10. Streetman, B. G., and S. Banerjee. *Solid State Electronic* ***Devices,*** 5th ed. Upper Saddle River, NJ: Prentice-Hall, 2000.
11. Sze, S. M. ***Semiconductor*** *Devices:* ***Physics*** *and* ***Technology,*** 2nd ed. New York: John Wiley and Sons, 2001

*12. Wang, S. ***Fundamentals of Semiconductor Theory*** *and Device Physics.* Englewood Cliffs, ***NJ:*** Prentice Hall, 1988.

CHAPTER 4

The Semiconductor in Equilibrium

PREVIEW

So far, we have been considering a general crystal and applying to it the concepts of quantum mechanics in order to determine a few of the characteristics ofelectrons in a single-crystal lattice. In this chapter, we will apply these concepts specifically to a semiconductor material. In particular, we will use the density of quantum states in the conduction band and the density of quantum states in the valence band along with the Fermi–Dirac probability function to determine the concentration of electrons and holes in the conduction and valence bands, respectively. We will also apply the concept of the Fermi energy to the semiconductor material.

This chapter deals with the semiconductor in equilibrium. Equilibrium, or thermal equilibrium, implies that no external forces such as voltages, electric fields. magnetic fields, or temperature gradients are acting on the semiconductor. All properties of the semiconductor will be independent of time in this case. Equilibrium is our starting point for developing the physics of the semiconductor. We will then be able to determine the characteristics that result when deviations from equilibrium occur, such as when a voltage is applied to a semiconductor device.

We will initially consider the properties of an intrinsic semiconductor, that is, a pure crystal with no impurity atoms or defects. We will see that the electrical properties of a semiconductor can be altered in desirable ways by adding controlled amounts of specific impurity atoms. called *dopant atoms,* to the crystal. Depending upon the type of dopant atom added, the dominant charge carrier in the semiconductor will be either electrons in the conduction band or holes in the valence band. Adding dopant atoms changes the distribution of electrons among the available energy states, so the Fermi energy becomes a function of the type and concentration of impurity atoms.

Finally, as part of this discussion, we will attempt to add more insight into the significance of the Fermi energy. ■

4.1 | CHARGE CARRIERS IN SEMICONDUCTORS

Current is the rate at which charge flows. In a semiconductor. two types of charge carrier, the electron and the hole, can contribute to a current. Since the current in a semiconductor is determined largely by the number of electrons in the conduction band and the number of holes in the valence hand, an important characteristic of the semiconductor is the density of these charge carriers. The density of electrons and holes is related to the density of states function and the Fermi distribution function, both of which we have considered. Aqualitative discussion of these relationships will be followed by a more rigorous mathematical derivation of the thermal-equilibrium concentration of electrons and holes.

4.1.1 Equilibrium Distribution of Electrons and Holes

The distribution (with respect to energy) of electrons in the conduction band is given by the density of allowed quantum states times the probability that a state is occupied by an electron. This statement is written in equation form as

$$n(E) = g_c(E) f_F(E) \tag{4.1}$$

where $f_F(E)$ is the Fermi–Dirac probability function and $g_c(E)$ is the density of quantum states in the conduction band. The total electron concentration per unit volume in the conduction band is then found by integrating Equation (4.1) over the entire conduction-band energy.

Similarly, the distribution (with respect to energy) of holes in the valence bend is the density of allowed quantum states in the valence hand multiplied by the probability that a state is nor occupied by an electron. We may express this as

$$p(E) = g_v(E)[1 - f_F(E)] \tag{4.2}$$

The total hole concentration per unit volume is found by integrating this function over the entire valcncc-band energy.

To find the thermal-equilibrium electron and hole concentrations, we need to determine the position of the Fermi energy E_F with respect to the bottom of the conduction-band energy E_c and the top of the valence-band energy E_v. To address this question, we will initially consider an intrinsic semiconductor. An ideal intrinsic semiconductor is a pure semiconductor with no impurity atoms and no lattice defects in the crystal (e.g., pure silicon). We have argued in the previous chapter that, for an intrinsic semiconductor at $T = 0$ K, all energy states in the valence band are filled with electrons and all energy states in the conduction band are empty of electrons. The Fermi energy must, therefore, be somewhere between E_c and E,.. (The Fermi energy does not need to correspond to an allowed energy.)

As the temperature begins to increase above 0 K, the valence electrons will gain thermal energy. A few electrons in the valence band may gain sufficient energy to jump to the conduction band. As an electron jumps from the valence band to the conduction band, an empty state, or hole, is created in the valence band. In an intrinsic semiconductor, then, electrons and holes are created in pairs by the thermal energy so

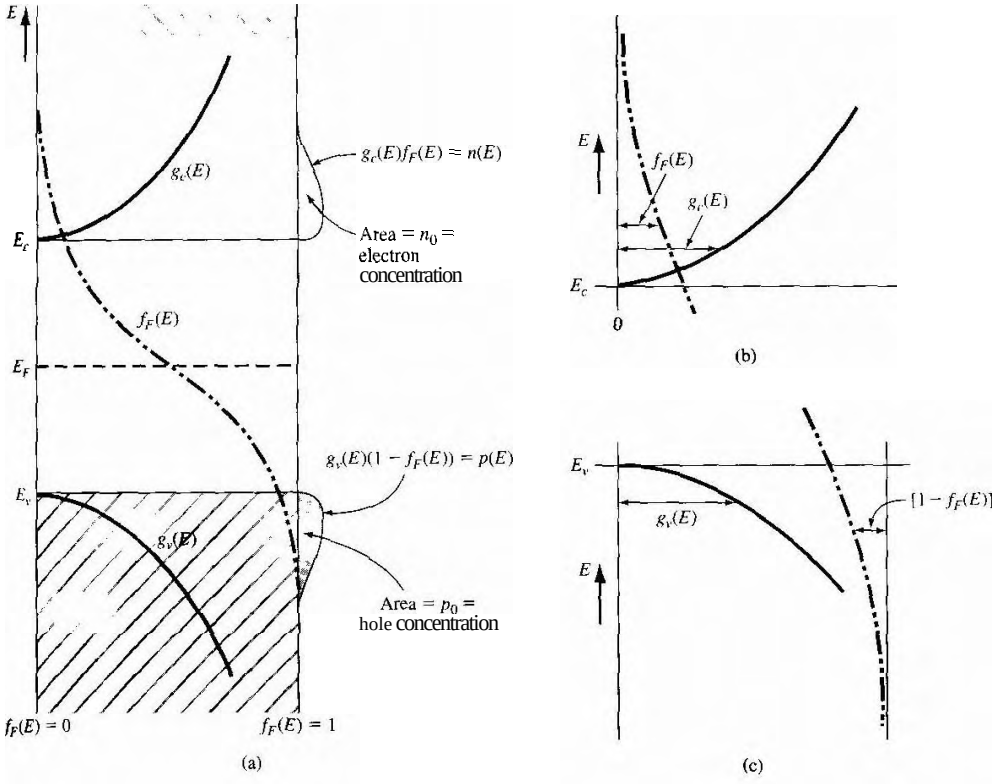


Figure 4.1 | (a) Density of states functions, Fermi–Dirac probability function, and areas representing electron and hole concentrations for the case when E_F is near the midgap energy; (b) expanded view near the conduction band energy; and (c) expanded view near the valence band energy.

that the number of electrons in the conduction band is equal to the number of holes in the valence band.

Figure 4.1a shows a plot of the density of states function in the conduction band $g_c(E)$, the density of states function in the valence band $g_v(E)$, and the Fermi–Dirac probability function for T > 0 K when E_F is approximately halfway between E_c and E_v. If we assume, for the moment, that the electron and hole effective masses are *equal,* then $g_c(E)$ and $g_v(E)$ are symmetrical functions about the midgap energy (the energy midway between E_c and E_v). We noted previously that the function $f_F(E)$ for $E > E_F$ is symmetrical to the function $1 - f_F(E)$ for $E < E_F$ about the energy $E = E_F$. This also means that the function $f_F(E)$ for $E = E_F + dE$ is equal to the function $1 - f_F(E)$ for $E = E_F - dE$.

Figure 4.1b is an expanded view of the plot in Figure 4.1a showing $f_F(E)$ and $g_c(E)$ above the conduction band energy E,. The product of $g_c(E)$ and $f_F(E)$ is the distribution of electrons $n(E)$ in the conduction band given by Equation (4.1). This product is plotted in Figure 4.1a. Figure 4.1c is an expanded view of the plot in Figure 4.1a showing $[1 - f_F(E)]$ and $g_v(E)$ below the valence band energy E_v. The product of $g_v(E)$ and $[1 - f_F(E)]$ is the distribution of holes $p(E)$ in the valence band given by Equation (4.2). This product is also plotted in Figure 4.1a. The area under these curves are then the total density of electrons in the conduction band and the total density of holes in the valence band. From this we see that if $g_c(E)$ and $g_v(E)$ are symmetrical, the Fermi energy must be at the midgap energy in order to obtain equal electron and hole concentrations. If the effective masses of the electron and hole are not exactly equal, then the effective density of states functions $g_c(E)$ and $g_v(E)$ will not be exactly symmetrical about the midgap energy. The Fermi level for the intrinsic semiconductor will then shift slightly from the midgap energy in order to obtain equal electron and hole concentrations.

4.1.2 The n_0 and p_0 Equations

We have argued that the Fermi energy for an intrinsic semiconductor is near midgap. In deriving the equations for the thermal-equilibrium concentration of electrons n_0 and the thermal-equilibrium concentration of holes p_0, we will not be quite so restrictive. We will see later that, in particular situations, the Fermi energy can deviate from this midgap euergy. We will assume initially, however, that the Fermi level remains within the bandgap energy.

The equation for the thermal-equilibrium concentration of electrons may be found by integrating Equation (4.1) over the conduction band energy, or

$$n_0 = \int g_c(E) f_F(E)\, dE \tag{4.3}$$

The lower limit of integration is E_c and the upper limit of integration should be the top of the allowed conduction band energy. However, since the Fermi probability function rapidly approaches zero with increasing energy as indicated in Figure 4.1a we can take the upper limit of integration to be infinity.

We are assuming that the Fermi energy is within the forbidden-energy bandgap. For electrons in the conduction hand, we have E > E,. If $(E_c - E_F) \gg kT$, then $(E - E_F) \gg kT$, so that the Fermi probability function reduces to the Boltzmann approximation,[1] which is

$$f_F(E) = \frac{1}{1 + \exp\dfrac{(E - E_F)}{kT}} \approx \exp\frac{[-(E - E_F)]}{kT} \tag{4.4}$$

[1] The Maxwell–Boltzmann and Fermi–Dirac distribution functions are within 5 percent of each other when $E - E_F \approx 3kT$ (see Figure 3.33). The $\gg$ notation is then somewhat misleading to indicate when the Boltzmann approximation is valid, although it is commonly used.

Applying the Boltzmann approximation to Equation (4.3), the thermal-equilibrium density of electrons in the conduction band is found from

$$n_0 = \int_{E_c}^{\infty} \frac{4\pi(2m_n^*)^{3/2}}{h^3} \sqrt{E - E_c} \exp\left[\frac{-(E - E_F)}{kT}\right] dE \tag{4.5}$$

The integral of Equation *(4.5)* may be solved more easily by making a change of variable. If we let

$$\eta = \frac{E - E_c}{kT} \tag{4.6}$$

then Equation *(4.5)* becomes

$$n_0 = \frac{4\pi(2m_n^* kT)^{3/2}}{h^3} \exp\left[\frac{-(E_c - E_F)}{kT}\right] \int_0^{\infty} \eta^{1/2} \exp(-\eta)\, d\eta \tag{4.7}$$

The integral is the gamma function, with a value of

$$\int_0^{\infty} \eta^{1/2} \exp(-\eta)\, d\eta = \frac{1}{2}\sqrt{\pi} \tag{4.8}$$

Then Equation *(4.7)* becomes

$$n_0 = 2\left(\frac{2\pi m_n^* kT}{h^2}\right)^{3/2} \exp\left[\frac{-(E_c - E_F)}{kT}\right] \tag{4.9}$$

We may define a parameter N_c as

$$N_c = 2\left(\frac{2\pi m_n^* kT}{h^2}\right)^{3/2} \tag{4.10}$$

so that the thermal-equilibrium electron concentration in the conduction band can be written as

$$\boxed{n_0 = N_c \exp\left[\frac{-(E_c - E_F)}{kT}\right]} \tag{4.11}$$

The parameter N_c is called the *effective density of states function in the conduction band.* If we were to assume that $m_n^* = m_0$, then the value of the effective density of states function at T = 300 K is $N_c = 2.5 \times 10^{19}$ cm^{-3}, which is the order of magnitude of N_c for most semiconductors. If the effective mass of the electron is larger or smaller than m_0, then the value of the effective density of states function changes accordingly, but is still of the same order of magnitude.

EXAMPLE 4.1 Objective

Calculate the probability that a state in the conduction band is occupied by an electron and calculate the thermal equilibrium electron concentration in silicon at $T = 100$ K.

Assume the Fermi energy is 0.25 eV below the conduction band. The value of N_c for silicon at T = 100 K is $N_c = 2.8 \times 10^{19}$ cm^{-3}.

■ **Solution**

The probability that an energy state at $E = E_c$ is occupied by an electron is given by

$$f_F(E_c) = \frac{1}{1 + \exp\left(\frac{E_c - E_F}{kT}\right)} \approx \exp\left[\frac{-(E_c - E_F)}{kT}\right]$$

or

$$f_F(E_c) = \exp\left(\frac{-0.25}{0.0259}\right) = 6.43 \times 10^{-5}$$

The electron concentration is given by

$$n_0 = N_c \exp\left[\frac{-(E_c -}{kT}\right] = (2.8 \times 10^{19}) \exp\left(\frac{-0.25}{0.0259}\right)$$

or

$$n_0 = 1.8 \times 10^{15}\ \text{cm}^{-3}$$

■ **Comment**

The probability of a state being occupied can be quite small, but the fact that there are a large number of states means that the electron concentration is a reasonable value.

The thermal-equilibrium concentration of holes in the valence band is found by integrating Equation (4.2) over the valence band energy, or

$$p_0 = \int g_v(E)[1 - f_F(E)]\, dE \tag{4.12}$$

We may note that

$$1 - f_F(E) = \frac{1}{1 + \exp\left(\frac{E_F - E}{kT}\right)} \tag{4.13a}$$

For energy states in the valence band, $E < E_v$. If $(E_F - E_v) \gg kT$ (the Fermi function is still assumed to be within the bandgap), then we have a slightly different form of the Boltzmann approximation. Equation (4.13a) may be written as

$$1 - f_F(E) = \frac{1}{1 + \exp\left(\frac{E_F - E}{kT}\right)} \approx \exp\left[\frac{-(E_F - E)}{kT}\right] \tag{4.13b}$$

Applying the Boltzmann approximation of Equation (4.13b) to Equation (4.12), we find the thermal-equilibrium concentration of holes in the valence band is

$$p_0 = \int_{-\infty}^{E_v} \frac{4\pi(2m_p^*)^{3/2}}{h^3} \sqrt{E_v - E} \exp\left[\frac{-(E_F - E)}{kT}\right] dE \tag{4.14}$$

where the lower limit of integration is taken as minus infinity instead of the bottom of the valence band. The exponential term decays fast enough so that this approximation is valid.

Equation (4.14) may be solved more easily by again making a change of variable. If we let

$$\eta' = \frac{E_v - E}{kT} \tag{4.15}$$

then Equation (4.14) becomes

$$p_0 = \frac{-4\pi(2m_p^* kT)^{3/2}}{h^3} \exp\left[\frac{-(E_F - E_v)}{kT}\right] \int_{+\infty}^{0} (\eta')^{1/2} \exp(-\eta')\, d\eta' \tag{4.16}$$

where the negative sign comes from the differential $dE = -kT\,d\eta'$. Note that the lower limit of η' becomes $+\infty$ when $E = -\infty$. If we change the order of integration, we introduce another minus sign. From Equation (4.8). Equation (4.16) becomes

$$p_0 = 2\left(\frac{2\pi m_p^* kT}{h^2}\right)^{3/2} \exp\left[\frac{-(E_F - E_v)}{kT}\right] \tag{4.17}$$

We may define a parameter N_v as

$$N_v = 2\left(\frac{2\pi m_p^* kT}{h^2}\right)^{3/2} \tag{4.18}$$

which is called the *effective density of states function in the valence band*. The thermal-equilibrium concentration of holes in the valence band may now be written as

$$\boxed{p_0 = N_v \exp\left[\frac{-(E_F - E_v)}{kT}\right]} \tag{4.19}$$

The magnitude of N_v is also on the order of 10^{19} cm^{-3} at T = 300 K for most semiconductors.

Objective | **EXAMPLE 4.2**

Calculate the thermal equilibrium hole concentration in silicon at T = 400 K.

Assume that the Fermi energy is 0.27 eV above the valence band energy. The value of N_v for silicon at T = 300 K is $N_v = 1.04 \times 10^{19}$ cm^{-3}.

Solution

The parameter values at T = 400 K are found as:

$$N_v = (1.04 \times 10^{19})\left(\frac{400}{300}\right)^{3/2} = 1.60 \times 10^{19}\,\text{cm}^{-3}$$

$$kT = (0.0259)\left(\frac{400}{300}\right) = 0.03453\ \text{eV}$$

The hole concentration is then

$$p_0 = N_v \exp\left[\frac{-(E_F - E_v)}{kT}\right] = (1.60 \times 10^{19}) \exp\left(\frac{-0.27}{0.03453}\right)$$

or

$$p_0 = 6.43 \times 10^{15}\ \text{cm}^{-3}$$

■ **Comment**

The parameter values at any temperature can easily he found hy using the 300 K values and the temperature dependence.

The effective density of states functions, N_c and N_v, are constant for a given semiconductor material at a fixed temperature. Table 4.1 gives the values of the density of states function and of the effective masses for silicon, gallium arsenide, and germanium. Note that the value of N, for gallium arsenide is smaller than the typical $10^{19}\ \text{cm}^{-3}$ value. This difference is due to the small electron effective mass in gallium arsenide.

The thermal equilibrium concentrations of electrons in the conduction band and of holes in the valence band are directly related to the effective density of states constants and to the Fermi energy level,

TEST YOUR UNDERSTANDING

E4.1 Calculate the thermal equilibrium electron and hole concentration in silicon at T = 300 K for the case when the Fermi energy level is 0.22 eV below the conduction hand energy E_c. The value of E_g is given in Appendix B.4.
(Ans. $n_0 = 5.73 \times 10^{15}\ \text{cm}^{-3}$, $p_0 = 8.43 \times 10^{3}\ \text{cm}^{-3}$)

E4.2 Determine the thermal equilibrium electron and hole concentration in GaAs at T = 300 K for the case when the Fermi energy level is 0.30 eV above the valence band energy E_v. The value of E_g is given in Appendix B.4.
(Ans. $n_0 = 0.0779\ \text{cm}^{-3}$, $p_0 = 6.53 \times 10^{13}\ \text{cm}^{-3}$)

4.1.3 The Intrinsic Carrier Concentration

For an intrinsic semiconductor, the concentration of electrons in the conduction band is equal to the concentration of holes in the valence band. We may denote n_i and p_i

Table 4.1 | Effectivedensity of states function and effective mass values

	$N_c\ (\text{cm}^{-3})$	$N_v\ (\text{cm}^{-3})$	m_n^*/m_0	m_p^*/m_0
Silicon	2.8×10^{19}	1.04×10^{19}	1.08	0.56
Gallium arsenide	4.7×10^{17}	7.0×10^{18}	0.067	0.48
Germanium	1.04×10^{19}	6.0×10^{18}	0.55	0.37

as the electron and hole concentrations, respectively, in the intrinsic semiconductor. These parameters are usually referred to as the intrinsic electron concentration and intrinsic hole concentration. However, $n_i = p_i$, so normally we simply use the parameter n_i as the intrinsic carrier concentration, which refers to either the intrinsic electron or hole concentration.

The Fermi energy level for the intrinsic semiconductor is called the intrinsic Fermi energy, or $E_F = E_{Fi}$. If we apply Equations (4.11) and (4.19) to the intrinsic semiconductor, then we can write

$$n_0 = n_i = N_c \exp\left[\frac{-(E_c - E_{Fi})}{kT}\right] \tag{4.20}$$

$$p_0 = p_i = n_i = N_v \exp\left[\frac{-(E_{Fi} - E_v)}{kT}\right] \tag{4.21}$$

If we take the product of Equation.; (4.20) and (4.21). we obtain

$$n_i^2 = N_c N_v \exp\left[\frac{-(E_c - E_{Fi})}{kT}\right] \cdot \exp\left[\frac{-(E_{Fi} - E_v)}{kT}\right] \tag{4.22}$$

$$\boxed{n_i^2 = N_c N_v \exp\left[\frac{-(E_c - E_v)}{kT}\right] = N_c N_v \exp\left[\frac{-E_g}{kT}\right]} \tag{4.23}$$

where E_g is the bandgap energy. For a given semiconductor material at a constant temperature, the value of n, is a constant, and independent of the Fermi energy.

The intrinsic carrier concentration for silicon at T = 300 K may be calculated by using the effective density of states function values from Table 4.1. The value of n_i calculated from Equation (4.23) for $E_g = 1.12$ eV is $n_i = 6.95 \times 10^9$ cm^{-3}. The commonly accepted value' of n_i for silicon at T = 300 K is approximately 1.5×10^{10} cm^{-3}. This discrepancy may arise from several sources. First, the values of the effective masses are determined at a low temperature where the cyclotron resonance experiments are performed. Since the effective mass is an experimentally determined parameter, and since the effective mass is a measure of how well a particle moves in a crystal, this parameter may be a slight function of temperature. Next, the density of states function for a semiconductor was obtained by generalizing the model of an electron in a three-dimensional infinite potential well. This theoretical function may also not agree exactly with experiment. However, the difference between the theoretical value and the experimental value of n, is approximately a factor

[2]Various references may list slightly different values of the intrinsic silicon concentration at room temperature. In general, they are all between 1×10^{10} and 1.5×10^{10} cm^{-3}. This difference is, in most carer, not significant.

Table 4.2 | Commonly accepted values of n_i at $T = 300$ K

Silicon	$n_i = 1.5 \times 10^{10}$ cm^{-3}
Gallium arsenide	$n_i = 1.8 \times 10^{6}$ cm^{-3}
Germanium	$n_i = 2.4 \times 10^{13}$ cm^{-3}

of 2, which, in many cases, is not significant. Table 4.2 lists the commonly accepted values of n_i for silicon, gallium arsenide, and germanium at $T = 300$ K.

The intrinsic carrier concentration is a very strong function ot temperature.

EXAMPLE 4.3

Objective

To calculate the intrinsic carrier concentration in gallium arsenide at T = 300 K and at T = 450 K.

The values of N_c and N, at 300 K for gallium arsenide are 4.7×10^{17} cm^{-3} and 7.0×10^{18} cm^{-3}, respectively. Both N_c and N, vary as $T^{3/2}$. Assume the bandgap energy of gallium arsenide is 1.42 eV and does not vary with temperature over this range. The value of kT at 450 K is

$$kT = (0.0259)\left(\frac{450}{300}\right) = 0.03885 \text{ eV}$$

Solution

Using Equation (4.23). we find for T = 300 K

$$n_i^2 = (4.7 \times 10^{17})(7.0 \times 10^{18}) \exp\left(\frac{-1.42}{0.0259}\right) = 5.09 \times 10^{12}$$

so that

$$n_i = 2.26 \times 10^{6} \text{ cm}^{-3}$$

At $T = 450$ K, we find

$$n_i^2 = (4.7 \times 10^{17})(7.0 \times 10^{18})\left(\frac{450}{300}\right)^3 \exp\left(\frac{-1.42}{0.03885}\right) = 1.48 \times 10^{21}$$

so that

$$n_i = 3.85 \times 10^{10} \text{ cm}^{-3}$$

Comment

We may note from this example that the intrinsic carrier concentration increased by over 4 orders of magnitude as the temperature increased by 150°C.

Figure 4.2 is a plot of n_i from Equation (4.23) for silicon, gallium arsenide, and germanium as a function of temperature. As seen in the figure, the value of n_i fa these semiconductors may easily vary over several orders of magnitude as the temperature changes over a reasonable range.

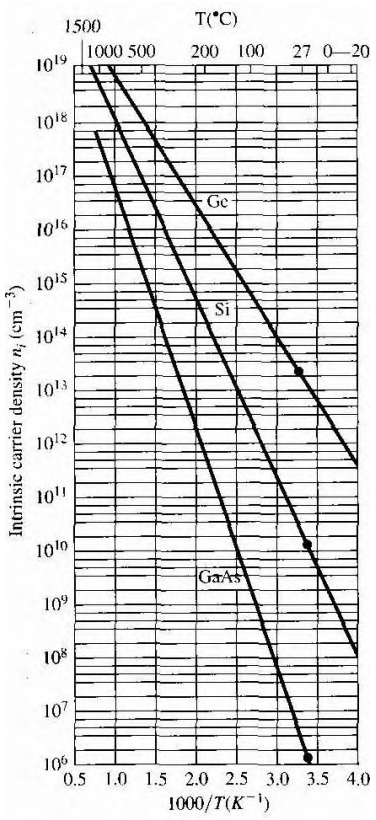


Figure 4.2 | The intrinsic carrier concentration of Ce, Si, and GaAs as a function of temperature.
(From Sze [13].)

TEST YOUR UNDERSTANDING

E4.3 Find the intrinsic carrier concentration in silicon at (*a*) $T = 200$ K and (*b*) $T = 400$ K. [Ans. (*a*) 7.68×10^{4} cm^{-3}, (*b*) 2.38×10^{12} cm^{-3}]

E4.4 Repeat **E4.3** for GaAs. [Ans. (*a*) 1.38 cm^{-3}, (*b*) 3.28×10^{9} cm^{-3}]

E4.5 Repeat **E4.3** for Ge. [Ans. (*a*) 2.16×10^{10} cm^{-3}, (*b*) 8.6×10^{14} cm^{-3}]

4.1.4 The Intrinsic Fermi-Level Position

We have qualitatively argued that the Fermi energy level is located near the center of the forbidden bandgap for the intrinsic semiconductor. We can specifically calculate

the intrinsic Fermi-level position. Since the electron and hole concentrations are equal, setting Equations (4.20) and (4.21) equal to each other, we have

$$N_c \exp\left[\frac{-(E_c - E_{Fi})}{kT}\right] = N_v \exp\left[\frac{-(E_{Fi} - E_v)}{kT}\right] \tag{4.24}$$

If we take the natural log of both sides of this equation and solve for E_{Fi}, we obtain

$$E_{Fi} = \frac{1}{2}(E_c + E_v) + \frac{1}{2} kT \ln\left(\frac{N_v}{N_c}\right) \tag{4.25}$$

From the delinitions for N_c and N, given by Equations (4.10) and (4.18), respectively, Equation (4.25) may be written as

$$E_{Fi} = \frac{1}{2}(E_c + E_v) + \frac{3}{4} kT \ln\left(\frac{m_p^*}{m_n^*}\right) \tag{4.26a}$$

The first term, $\frac{1}{2}(E_c + E_v)$, is the energy exactly midway between E, and E,. or the midgap energy. We can define

$$\frac{1}{2}(E_c + E_v) = E_{\text{midgap}}$$

so that

$$\boxed{E_{Fi} - E_{\text{midgap}} = \frac{3}{4} kT \ln\left(\frac{m_p^*}{m_n^*}\right)} \tag{4.26b}$$

If the electron and hole effective masses are equal so that $m_p^* = m_n^*$, then the intrinsic Fermi level is exactly in the center of the bandgap. If $m_p^* > m_n^*$, the intrinsic Fermi level is slightly above the center, and if $m_p^* < m_n^*$, it is slightly below the center of the bandgap. The density of states function is directly related to the carrier effective mass: thus a larger effective mass means a larger density of states function. The intrinsic Fermi level must shift away from the band with the larger density of states in order to maintain equal numbers of electrons and holes.

EXAMPLE 4.4

Objective

To calculate the position of the intrinsic Fermi level with respect to the center of the bandgap in silicon at T = *300 K*.

The density of states effective carrier masses in silicon are $m_n^* = 1.08m_0$ and $m_p^* = 0.56m_0$.

■ **Solution**

The intrinsic Fermi level with respect to the center of the bandgap is

$$E_{Fi} - E_{\text{midgap}} = \frac{3}{4} \text{kT} \ln\left(\frac{m_p^*}{m_n^*}\right) = \frac{3}{4}(0.0259) \ln\left(\frac{0.56}{1.08}\right)$$

or

$$E_{Fi} - E_{midgap} = -0.0128 \text{ eV} = -12.8 \text{ meV}$$

Comment

The intrinsic Fermi level in silicon is 12.8 meV below the midgap energy. If we compare 12.8meV to 560 meV, which is one-half of the bandgap energy of silicon, we can, in many applications, simply approximate the intrinsic Fermi level to be in the center of the bandgap.

TEST YOUR UNDERSTANDING

E4.6 Determine the position of the intrinsic Fermi level with respect to the center of the bandgap in GaAs at T = 300 K. (Ans. −38.2 meV)

4.2 | DOPANT ATOMS AND ENERGY LEVELS

The intrinsic semiconductor may be an interesting material, but the real power of semiconductors is realized by adding *small*, controlled amounts of specific dopant, or impurity, atoms. This doping process, described briefly in Chapter I, can greatly alter the electrical characteristics of the semiconductor. The doped semiconductor, called an ***extrinsic*** material, is the primary reason we can fabricate the various semiconductor devices that we will consider in later chapters.

4.2.1 Qualitative Description

In Chapter **3,** we discussed the covalent bonding of silicon and considered the simple two-dimensional representation of the single-crystal silicon lattice as shown in Figure 4.3. Now consider adding a group V element, such as phosphorus, as a substitutional impurity. The group V element has five valence electrons. Four of these will contribute to the covalent bonding with the silicon atoms, leaving the fifth more loosely hound to the phosphorus atom. This effect is schematically shown in Figure 4.4. We refer to the fifth valence electron as a donor electron.

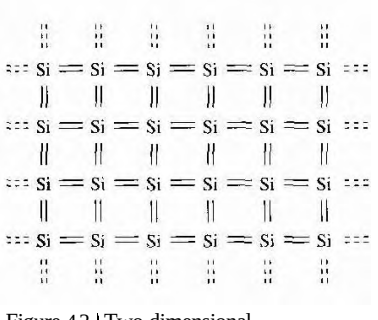


Figure 4.3 | Two-dimensional representation of the intrinsic silicon lattice.

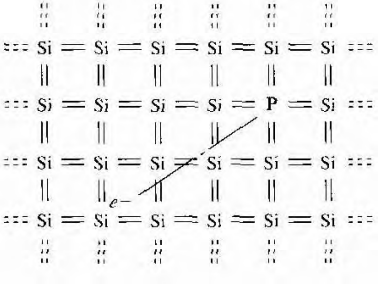


Figure 4.4 | Two-dirnsnsional representation of the silicon lattice doped with a phosphorus atom.

The phosphorus atom without the donor electron is positively charged. At very low temperatures, the donor electron is bound to the phosphorus atom. However. by intuition, it should seem clear that the energy required to elevate the donor electron into the conduction band is considerably less than that for the electrons involved in the covalent bonding. Figure 4.5 shows the energy-hand diagram that we would expect. The energy level, E_d, is the energy state of the donor electron.

If a small amount of energy, such as thermal energy. is added to the donor electron, it can be elevated into the conduction band, leaving behind a positively charged phosphorus ion. The electron in the conduction band can now move through the crystal generating a current, while the positively charged ion is fixed in the crystal. This type of impurity atom donates an electron to the conduction band and so is called a ***donor*** *impurity atom.* The donor impurity atoms add electrons to the conduction band without creating holes in the valence band. The resulting material is referred to as an *n-type* semiconductor (*n* for the negatively charged electron).

Now consider adding a group III element, such as boron, as a substitutional impurity to silicon. The group III element has three valence electrons, which are all taken up in the covalent bonding. As shown in Figure 4.6a, one covalent bonding position appears to he empty. If an electron were to occupy this "empty" position. its

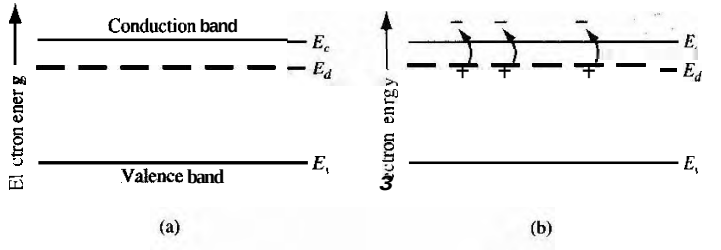


Figure 4.5 | The energy-hand diagram showing (a) the discrete donor energy state and (b) the effect of a donor state being ionized.

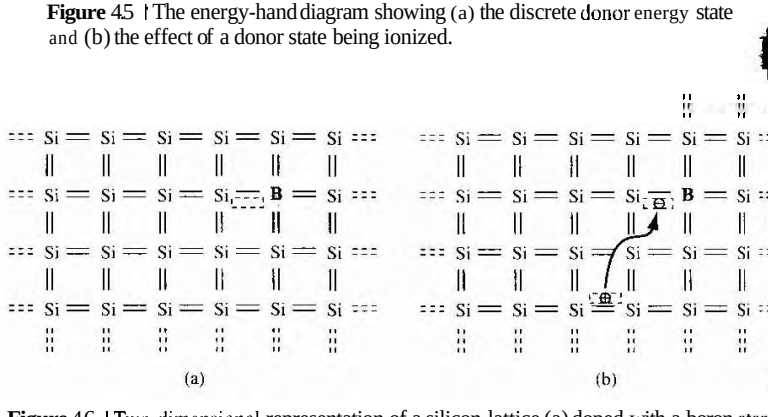


Figure 4.6 | Two-dimensional representation of a silicon lattice (a) doped with a boron atom and (b) showing the ionization of the boron atom resulting in a hole.

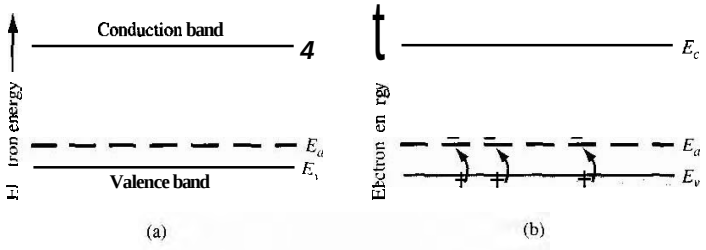


Figure 4.7 | The energy-band diagram showing (a) the discrete acceptor energy state and (b) the effect of an acceptor state being ionized.

energy would have to be greater than that of the valence electrons, since the net charge state of the boron atom would now be negative. However, the electron occupying this "empty" position does not have sufficient energy to he in the conduction band, so its energy is far smaller than the conduction-band energy. Figure 4.6b shows how valence electrons may gain a small amount of thermal energy and move about in the crystal. The "empty" position associated with the boron atom becomes occupied, and other valence electron positions become vacated. These other vacated electron positions can he thought of as holes in the semiconductor material.

Figure 4.7 shows the expected energy state of the "empty" position and also the formation of a hole in the valence hand. The hole can move through the crystal generating a current, while the negatively charged boron atom is fixed in the crystal. The group III atom accepts an electron from the valence band and so is referred to as an *acceptor impurity atom.* The acceptor atom can generate holes in the valence hand without generating electrons in the conduction band. This type of semiconductor material is referred to as a *p-type* material (p for the positively charged hole).

The pure single-crystal semiconductor material is called an intrinsic material. Adding controlled amounts of dopant atoms, either donors or acceptors, creates a material called an *extrinsic semiconductor.* An extrinsic semiconductor will have either a preponderance of electrons (n type) or a preponderance of holes (p type).

42.2 Ionization Energy

We can calculate the approximate distance of the donor electron from the donor impurity ion, and also the approximate energy required to elevate the donor electron into the conduction band. This energy is referred to as the ionization energy. We will use the Bohr model of the atom for these calculations. The justification for using this model is that the most probable distance of an electron from the nucleus in a hydrogen atom, determined from quantum mechanics, is the same as the Bohr radius. The energy levels in the hydrogen atom determined from quantum mechanics are also the same as obtained from the Bohr theory.

In the case of the donor impurity atom, we may visualize the donor electron orbiting the donor ion, which is embedded in the semiconductor material. We will need to use the permittivity of the semiconductor material in the calculations rather than

the permittivity of free space as is used in the case of the hydrogen atom. We will al use the effective mass of the electron in the calculations.

The analysis begins by setting the coulomb force of attraction between the ek tron and ion equal to the centripetal force of the orbiting electron. This condition w give a steady orbit. We have

$$\frac{\mathbf{e}^2}{4\pi\epsilon r_n^2} = \frac{m^* v^2}{r_n}$$

where v is the magnitude of the velocity and r_n is the radius of the orbit. If we assu the angular momentum is also quantized, then we can write

$$m^* r_n v = n\hbar \qquad (4.2$$

where n is a positive integer. Solving for v from Equation (4.28), substituting in Equation (4.27), and solving for the radius, we obtain

$$r_n = \frac{n^2\hbar^2 4\pi\epsilon}{m^* e^2}$$

The assumption of the angular momentum being quantized leads to the radius al being quantized.

The Bohr radius is defined as

$$a_0 = \frac{4\pi\epsilon_0\hbar^2}{m_0 e^2} = 0.53\ \text{Å}$$

We can normalize the radius of the donor orbital to that of the Bohr radius, which giv

$$\frac{r_n}{a_0} = n^2\epsilon_r\left(\frac{m_0}{m^*}\right)$$

where ϵ_r is the relative dielectric constant of the semiconductor material, m_0 is I rest mass of an electron, and m^* is the conductivity effective mass of the electron the semiconductor.

If we consider the lowest energy state in which n = 1, and if we consider silic in which $\epsilon_r = 11.7$ and the conductivity effective mass is $m^*/m_0 = 0.26$. then' have that

$$\frac{r_1}{a_0} = 45$$

or $r_1 = 23.9$ ~This radius corresponds to approximately four lattice constants silicon. Recall that one unit cell in silicon effectively contains eight atoms, so the dius of the orbiting donor electron encompasses many silicon atoms. The donor el tron is not tightly bound to the donor atom.

The total energy of the orbiting electron is given by

$$E = T + V \qquad (4.3$$

where T is the kinetic energy and V is the potential energy of the electron. The kinetic energy is

$$T = \frac{1}{2}m^*v^2 \tag{4.34}$$

Using the velocity v from Equation (4.28) and the radius r_n from Equation (4.29), the kinetic energy becomes

$$T = \frac{m^*e^4}{2(n\hbar)^2(4\pi\epsilon)^2} \tag{4.35}$$

The potential energy is

$$V = \frac{-e^2}{4\pi\epsilon r_n} = \frac{-m^*e^4}{(n\hbar)^2(4\pi\epsilon)^2} \tag{4.36}$$

The total energy is the sum of the kinetic and potential energies, so that

$$E = T + V = \frac{-m^*e^4}{2(n\hbar)^2(4\pi\epsilon)^2} \tag{4.37}$$

For the hydrogen atom, $m^* = m_0$ and $\epsilon = \epsilon_0$. The ionization energy of the hydrogen atom in the lowest energy state is then E = −13.6 eV. If we consider silicon, the ionization energy is $E = -25.8$ meV, much less than the bandgap energy of silicon. This energy is the approximate ionization energy of the donor atom, or the energy required to elevate the donor electron into the conduction band.

For ordinary donor impurities such as phosphorus or arsenic in silicon or germanium, this hydrogenic model works quite well and gives some indication of the magnitudes of the ionization energies involved. Table 4.3 lists the actual experimentally measured ionization energies for a few impurities in silicon and germanium. Germanium and silicon have different relative dielectric constants and effective masses; thus we expect the ionization energies to differ.

4.2.3 Group III–V Semiconductors

In the previous sections, we have been discussing the donor and acceptor impurities in a group IV semiconductor, such as silicon. The situation in the group III–V

Table 4.3 Impurity ionization energies in silicon and germanium

	Ionization energy (eV)	
Impurity	**Si**	***Ge***
Donors		
Phosphorus	0.045	0.012
Arsenic	0.05	0.0127
Acceptors		
Boron	0.045	0.0104
Aluminum	0.06	0.0102

Table 4.4 | Impurity ionization energies in gallium arsenide

Impurity	Ionization energy (eV)
Donors	
Selenium	0.0059
Tellurium	0.0058
Silicon	0.0058
Germanium	0.0061
Acceptors	
Beryllium	0.028
Zinc	0.0307
Cadmium	0.0347
Silicon	0.0345
Germanium	0.0404

compound semiconductors, such as gallium arsenide, is more complicated. Group II elements, such as beryllium, zinc, and cadmium, can enter the lattice as substitutional impurities, replacing the group III gallium element to become acceptor impurities. Similarly, group VI elements, such as selenium and tellurium, can enter the lattice substitutionally, replacing the group V arsenic element to become donor impurities. The corresponding ionization energies for these impurities are smaller than for the impurities in silicon. The ionization energies for the donors in gallium arsenide are also smaller than the ionization energies for the acceptors, because of the smaller effective mass of the electron compared to that of the hole.

Group IV elements, such as silicon and germanium, can also be impurity atoms in gallium arsenide. If a silicon atom replaces a gallium atom, the silicon impurity will act as a donor. but if the silicon atom replaces an arsenic atom. then the silicon impurity will act as an acceptor. The same is true for germanium as an impurity atom. Such impurities are called ***amphoreric.*** Experimentally in gallium arsenide, it is found that germanium is predominantly an acceptor and silicon is predominantly a donor. Table 4.4 lists the ionization energies for the various impurity atoms in gallium arsenide.

TEST YOUR UNDERSTANDING

E4.7 Calculate the radius (normalized to a Bohr radius) of a donor electron in its lowest energy state in GaAs. (Ans. 19.5)

4.3 | THE EXTRINSIC SEMICONDUCTOR

We defined an intrinsic semiconductor as a material with no impurity atoms present in the crystal. An ***extrinsic*** *semiconductor* is defined as a semiconductor in which controlled amounts of specific dopant or impurity atoms have been added so that the thermal-equilibrium electron and hole concentrations are different from the intrinsic

canier concentration. One type of carrier will predominate in an extrinsic semiconductor.

4.3.1 Equilibrium Distribution of Electrons and Holes

Adding donor or acceptor impurity atoms to a semiconductor will change the distribution of electrons and holes in the material. Since the Fermi energy is related to the distribution function, the Fermi energy will change as dopant atoms are added. If the Fermi energy changes from near the midgap value, the density of electrons in the conduction band and the density of holes in the valence band will change. These effects are shown in Figures 4.8 and 4.9. Figure 4.8 shows the case for $E_F > E_{Fi}$ and Figure 4.9 shows the case for $E_F < E_{Fi}$. When $E_F > E_{Fi}$, the electron concentration is larger than the hole concentration, and when $E_F < E_{Fi}$, the hole concentration

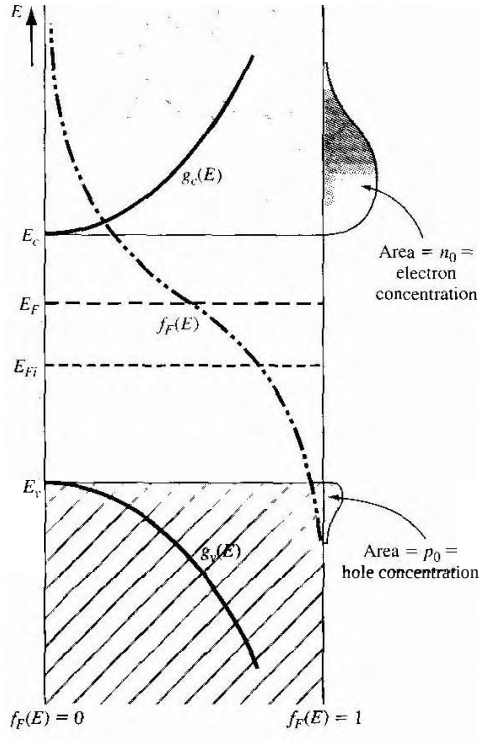


Figure 4.81 Density of states functions. Fermi–Dirac probability function, and areas representing electron and hole concentrations for the case when E_F is **above** the intrinsic Fermi energy.

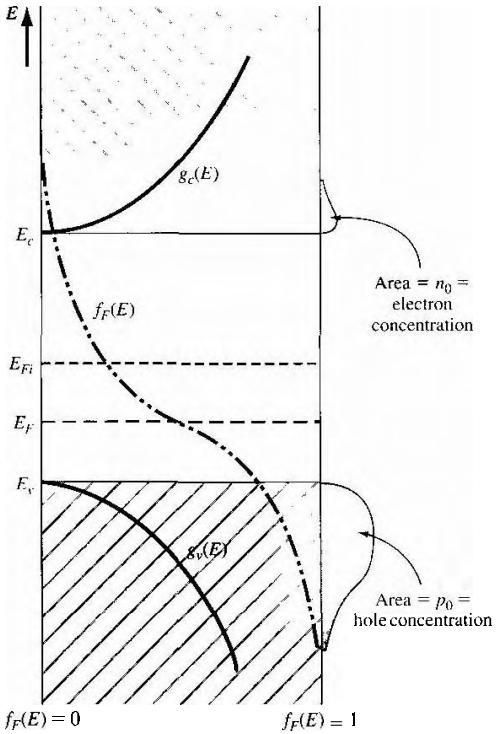


Figure 4.9 | Density of states functions, Fermi–Dirac probability function, and areas representing electron and hole concentrations for the case when E_I is below the intrinsic Fermi energy.

is larger than the electron concentration. When the density of electrons is greater than the density of holes, the semiconductor is n type; donor impurity atoms have been added. When the density of holes is greater than the density of electrons, the semiconductor is p type; acceptor impurity atoms have been added. The Fermi energy level in a semiconductor changes as the electron and hole concentrations change and, again, the Fermi energy changes as donor or acceptor impurities are added. The change in the Fermi level as a function of impurity concentrations will be considered in Section 4.6.

The expressions previously derived for the thermal-equilibrium concentrations of electrons and holes, given by Equations (4.11) and (4.19) are general equations for n_0 and p_0 in terms of the Fermi energy. These equations are again given as

$$n_0 = N_c \exp\left[\frac{-(E_c - E_F)}{kT}\right]$$

and

$$p_0 = N_v \exp\left[\frac{-(E_F - E_v)}{kT}\right]$$

As we just discussed, the Fermi energy may vary through the bandgap energy, which will then change the values of n_0 and p_0.

Objective

EXAMPLE 4.5

To calculate the thermal equilibrium concentrations of electrons and holes for a given Fermi energy.

Consider silicon at T = 300 K so that $N_c = 2.8 \times 10^{19}$ cm^{-3} and $N_v = 1.04 \times 10^{19}$ cm^{-3}. Assume that the Fermi energy is 0.25 eV below the conduction band. If we assume that the bandgap energy of silicon is 1.12 eV, then the Fermi energy will be 0.87 eV above the valence band.

■ Solution

Using Equation (4.11), we have

$$n_0 = (2.8 \times 10^{19}) \exp\left(\frac{-0.25}{0.0259}\right) = 1.8 \times 10^{15} \text{ cm}^{-3}$$

From Equation (4.19), we can write

$$p_0 = (1.04 \times 10^{19}) \exp\left(\frac{-0.87}{0.0259}\right) = 2.7 \times 10^{4} \text{ cm}^{-3}$$

■ Comment

The change in the Fermi level is actually a function of the donor or acceptor impurity concentrations that are added to the semiconductor. However, this example shows that electron and hole concentrations change by orders of magnitude from the intrinsic carrier concentration as the Fermi energy changes by a few tenths of an electron-volt.

In this example, since $n_0 > p_0$, the semiconductor is n type. In an n-type semiconductor, electrons are referred to as the majority carrier and holes as the minority carrier. By comparing the relative values of n_0 and p_0 in the example, it is easy to see how this designation came about. Similarly, in a p-type semiconductor where $p_0 > n_0$, holes are the majority carrier and electrons are the minority carrier.

We may derive another form of the equations for the thermal-equilibrium concentrations of electrons and holes. If we add and subtract an intrinsic Fermi energy in the exponent of Equation (4.11), we can write

$$n_0 = N_c \exp\left[\frac{-(E_c - E_{Fi}) + (E_F - E_{Fi})}{kT}\right] \tag{4.38a}$$

or

$$n_0 = N_c \exp\left[\frac{-(E_c - E_{Fi})}{kT}\right] \exp\left[\frac{(E_F - E_{Fi})}{kT}\right] \tag{4.38b}$$

The intrinsic carrier concentration is given by Equation (4.20) as

$$n_i = N_c \exp\left[\frac{-(E_c - E_{Fi})}{kT}\right]$$

so that the thermal-equilibrium electron concentration can be written as

$$n_0 = n_i \exp\left[\frac{E_F - E_{Fi}}{kT}\right] \tag{4.39}$$

Similarly, if we add and subtract an intrinsic Fermi energy in the exponent of Equation (4.19), we will obtain

$$p_0 = n_i \exp\left[\frac{-(E_F - E_{Fi})}{kT}\right] \tag{4.40}$$

As we will see, the Fermi level changes when donors and acceptors are added, hut Equations (4.39) and (4.40) show that, as the Fermi level changes from the intrinsic Fermi level, n_0 and p_0 change from then; value. If $E_F > E_{Fi}$, then we will have $n_0 > n$, and $p_0 < n_i$. One characteristic of an n-type semiconductor is that $E_F > E_{Fi}$ so that $n_0 > p_0$. Similarly, in a p-type semiconductor, $E_F < E_{Fi}$ so that $p_0 > n_i$ and $no < n_i$; thus $p_0 > no$.

We can see the functional dependence of n_0 and p_0 with E_F in Figures 4.8 and 4.9. As E_F moves above or below E_{Fi}, the overlapping probability function with the density of states functions in the conduction band and valence band changes. As E_F moves above E_{Fi}, the probability function in the conduction band increases, while the probability, $1 - f_F(E)$, of an empty state (hole) in the valence band decreases. As E_F moves below E_{Fi}, the opposite occurs.

4.3.2 The $n_0 p_0$ Product

We may take the product of the general expressions for n_0 and p_0 as given in Equations (4.11) and (4.19). respectively. The result is

$$n_0 p_0 = N_c N_v \exp\left[\frac{-(E_c - E_F)}{kT}\right] \exp\left[\frac{-(E_F - E_v)}{kT}\right] \tag{4.41}$$

which may be written as

$$n_0 p_0 = N_c N_v \exp\left[\frac{-E_g}{kT}\right] \tag{4.42}$$

As Equation (4.42) was derived for a general value of Fermi energy, the values of n_0 and p_0 are not necessarily equal. However, Equation (4.42) is exactly the same as Equation (4.23), which we derived for the case of an intrinsic semiconductor. We

then have that, for the semiconductor in thermal equilibrium,

$$\boxed{n_0 p_0 = n_i^2} \tag{4.43}$$

Equation (4.43) states that the product of n_0 and p_0 is always a constant for a given semiconductor material at a given temperature. Although this equation seems very simple, it is one of the fundamental principles of semiconductors in thermal equilibrium. The significance of this relation will become more apparent in the chapters that follow. It is important to keep in mind that Equation (4.43) was derived using the Boltrmann approximation. If the Boltzmann approximation is not valid, then likewise, Equation (4.43) is not valid.

An extrinsic semiconductor in thermal equilibrium does not, strictly speaking, contain an intrinsic carrier concentration, although some thermally generated carriers are present. The intrinsic electron and hole carrier concentrations are modified by the donor or acceptor impurities. However, we may think of the intrinsic concentration n_i in Equation (4.41) simply as a parameter of the semiconductor material.

*4.3.3 The Fermi–Dirac Integral

In the derivation of the Equations (4.11) and (4.19) for the thermal equilibrium electron and hole concentrations, we assumed that the Boltzmann approximation was valid. If the Boltrmann approximation does not hold. the thermal equilibrium electron concentration is written from Equation (4.3) as

$$n_0 = \frac{4\pi}{h^3}(2m_n^*)^{3/2}\int_{E_c}^{\infty}\frac{(E-E_c)^{1/2}\,dE}{1+\exp\left(\dfrac{E-E_F}{kT}\right)} \tag{4.44}$$

If we again make a change of variable and let

$$\eta = \frac{E-E_c}{kT} \tag{4.45a}$$

and also define

$$\eta_F = \frac{E_F-E_c}{kT} \tag{4.45b}$$

then we can rewrite Equation (4.44) as

$$n_0 = 4\pi\left(\frac{2m_n^*kT}{h^2}\right)^{3/2}\int_0^{\infty}\frac{\eta^{1/2}\,d\eta}{1+\exp(\eta-\eta_F)} \tag{4.46}$$

The integral is defined as

$$F_{1/2}(\eta_F) = \int_0^{\infty}\frac{\eta^{1/2}\,d\eta}{1+\exp(\eta-\eta_F)} \tag{4.47}$$

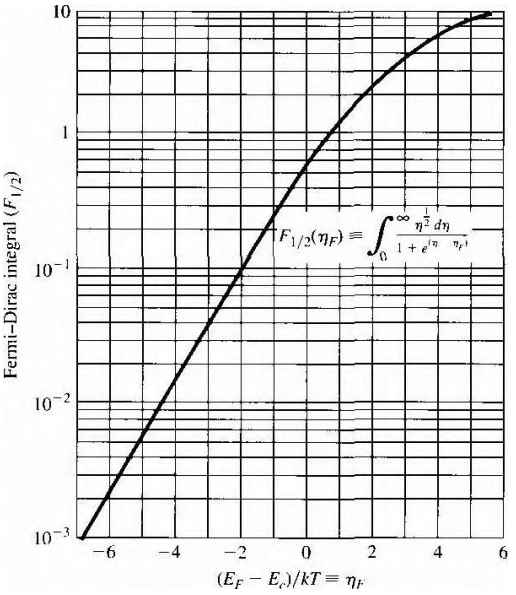


Figure 4.10 | The Fermi–Dirac integral $F_{1/2}$ as a function of the Fermi energy.
(From Sze [13].)

This function, called the Fermi–Dirac integral, is a tabulated function of the variabl η_F. Figure 4.10 is a plot of the Fermi–Dirac integral. Note that if $\eta_F > 0$. the $E_F > E_c$; thus the Fermi energy is actually in the conduction band.

EXAMPLE 4.6 | Objective

To calculate the electron concentration using the Fermi–Dirac integral.

Let $\eta_F = 2$ so that the Fermi energy is above the conduction hand by approximatel 52 meV at T = 300 K.

■ Solution

Equation (4.46)can he written as

$$n_0 = \frac{2}{\sqrt{\pi}} N_c F_{1/2}(\eta_F)$$

For silicon at 300 K, $N_c = 2.8 \times 10^{19}$ cm $^{-3}$ and. from Figure 4.10, the Fermi–Dirac integr has a value of $F_{1/2}(2) = 2.3$. Then

$$n_0 = \frac{2}{\sqrt{\pi}}(2.8 \times 10^{19})(2.3) = 7.27 \times 10^{19}\ \text{cm}^{-3}$$

Comment

Note that if we had used Equation (4.11). the thermal equilibrium value of n_0 would be $n_0 = 2.08 \times 10^{20}\ cm^{-3}$, which is incorrect since the Boltzmann approximation is not valid for this case.

We may use the same general method to calculate the thermal equilibrium con centration of holes. We obtain

$$p_0 = 4\pi \left(\frac{2m_p^* kT}{h^2}\right)^{3/2} \int_0^\infty \frac{(\eta')^{1/2}\, d\eta'}{1 + \exp(\eta' - \eta_F')} \tag{4.48}$$

where

$$\eta' = \frac{E_v - E}{kT} \tag{4.49a}$$

$$\eta_F' = \frac{E_v - E_F}{kT} \tag{4.49b}$$

The integral in Equation (4.48) is the same Fermi–Dirac integral defined hy Equation (4.47), although the variables have slightly different definitions. We may note that if $\eta_F' > 0$, then the Fermi level is in the valence hand.

TEST YOUR UNDERSTANDING

E4.8 Calculate the thermal equilibrium electron concentration in silicon for the case when $E_F = E$, and $T = 300$ K. (Ans. $1.9 \times 10^{16}\ cm^{-3}$)

4.3.4 Degenerate and Nondegenerate Semiconductors

In our discussion of adding dopant atoms to a semiconductor, we have implicitly assumed that the concentration of dopant atoms added is small when compared to the density of host or semiconductor atoms. The small number of impurity atoms are spread far enough apart so that there is no interaction between donor electrons, for example, in an n-type material. We have assumed that the impurities introduce discrete, noninteracting donor energy states in the n-type semiconductor and discrete. noninteracting acceptor states in the p-type semiconductor. These types of semiconductors are referred to as nondegenerate semiconductors.

If the impurity concentration increases, the distance between the impurity atoms decreases and a point will he reached when donor electrons, for example, will begin to interact with each other When this occurs, the single discrete donor energy will split into a band of energies. As the donor concentration further increases, the band of donor states widens and may overlap the bottom of the conduction band. This overlap occurs when the donor concentration becomes comparable with the effective density of states. When the concentration of electrons in the conduction band exceeds

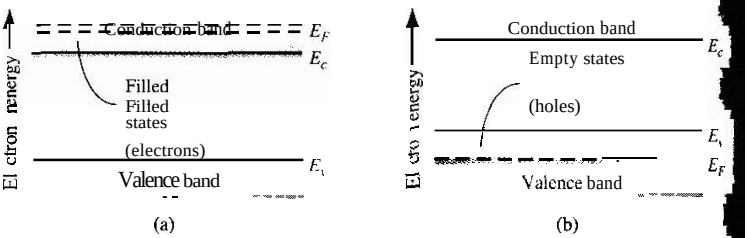


Figure 4.11 | Simplified energy-band diagrams for degenerately doped (a) n-type and (b) p-type semiconductors.

the density of states N_c, the Fermi energy lies within the conduction band. This type of semiconductor is called a degenerate n-type semiconductor.

In a similar way, as the acceptor doping concentration increases in a p-type semiconductor, the discrete acceptor energy states will split into a band of energies and may overlap the top of the valence band. The Fermi energy will lie in the valence band when the concentration of holes exceeds the density of states N_v. This type of semiconductor is called a degenerate p-type semiconductor.

Schematic models of the energy-band diagrams for a degenerate n-type and degenerate p-type semiconductor are shown in Figure 4.11. The energy states below E_F are mostly filled with electrons and the energy states above E_F are mostly empty. In the degenerate n-type semiconductor, the states between E_F and E_c are mostly filled with electrons; thus, the electron concentration in the conduction band is very large. Similarly, in the degenerate p-type semiconductor, the energy states between E_v and E_F are mostly empty; thus, the hole concentration in the valence band is very large.

4.4 | STATISTICS OF DONORS AND ACCEPTORS

In the last chapter, we discussed the Fermi–Dirac distribution function, which gives the probability that a particular energy state will be occupied by an electron. We need to reconsider this function and apply the probability statistics to the donor and acceptor energy states.

4.4.1 Probability Function

One postulate used in the derivation of the Fermi–Dirac probability function was the Pauli exclusion principle, which states that only one particle is permitted in each quantum state. The Pauli exclusion principle also applies to the donor and acceptor states.

Suppose we have N_i electrons and g_i quantum states, where the subscript i indicates the ith energy level. There are g_i ways of choosing where to put the first particle. Each donor level has two possible spin orientations for the donor electron; thus each donor level has two quantum states. The insertion of an electron into one quantum state, however, precludes putting an electron into the second quantum state. By

adding one electron, the vacancy requirement of the atom is satisfied, and the addition of a second electron in the donor level is not possible. The distribution function of donor electrons in the donor energy states is then slightly different than the Fermi–Dirac function.

The probability function of electrons occupying the donor state is

$$n_d = \frac{N_d}{1 + \frac{1}{2}\exp\left(\frac{E_d - E_F}{kT}\right)} \tag{4.50}$$

where n_d is the density of electrons occupying the donor level and E_d is the energy of the donor level. The factor $\frac{1}{2}$ in this equation is a direct result of the spin factorjust mentioned. The $\frac{1}{2}$ factor is sometimes written as $1/g$, where g is called a degeneracy factor.

Equation (4.50) can also be written in the form

$$n_d = N_d - N_d^+ \tag{4.51}$$

where N_d^+ is the concentration of ionized donors. In many applications, we will be interested more in the concentration of ionized donors than in the concentration of electrons remaining in the donor states.

If we do the same type of analysis for acceptor atoms, we obtain the expression

$$p_a = \frac{N_a}{1 + \frac{1}{g}\exp\left(\frac{E_F - E_a}{kT}\right)} = N_a - N_a^- \tag{4.52}$$

where N_a is the concentration of acceptor atoms. E_a is the acceptor energy level, p, is the concentration of holes in the acceptor states, and N_a^- is the concentration of ionized acceptors. A hole in an acceptor state corresponds to an acceptor atom that is neutrally charged and still has an "empty" bonding position as we discussed in Section 4.2.1. The parameter g is, again, a degeneracy factor. The ground state degeneracy factor g is normally taken as four for the acceptor level in silicon and gallium arsenide because of the detailed band structure.

4.4.2 Complete Ionization and Freeze-Out

The probability function for electrons in the donor energy state was just given by Equation (4.50). If we assume that $(E_d - E_F) \gg kT$, then

$$n_d \approx \frac{N_d}{\frac{1}{2}\exp\left(\frac{E_d - E_F}{kT}\right)} = 2N_d \exp\left[\frac{-(E_d - E_F)}{kT}\right] \tag{4.53}$$

If $(E_d - E_F) \gg kT$, then the Boltzmann approximation is also valid for the electrons in the conduction band so that. from Equation (4.11),

$$n_0 = N_c \exp\left[\frac{-(E_c - E_F)}{kT}\right]$$

We can determine the relative number of electrons in the donor state compa with the total number of electrons; therefore we can consider the ratio of electrons the donor state to the total number of electrons in the conduction band plus dot state. Using the expressions of Equations (4.53) and (4.11), we write

$$\frac{n_d}{n_d + n_0} = \frac{2N_d \exp\left[\dfrac{-(E_d - E_F)}{kT}\right]}{2N_d \exp\left[\dfrac{-(E_d - E_F)}{kT}\right] + N_c \exp\left[\dfrac{-(E_c - E_F)}{kT}\right]} \quad (4.5$$

The Fermi energy cancels out of this expression. Dividing by the numerator term, obtain

$$\frac{n_d}{n_d + n_0} = \frac{1}{1 + \dfrac{N_c}{2N_d} \exp\left[\dfrac{-(E_c - E_d)}{kT}\right]} \quad (4.5$$

The factor $(E_c - E_d)$ is just the ionization energy of the donor electrons.

EXAMPLE 4.7

Objective

To determine the fraction of total electmns still in the donor states at T = 300 K.

Consider phosphorus doping in silicon, for T = 300 K, at a concentration of N_d 10^{16} cm^{-3}.

Solution

Using Equation (4.55), we find

$$\frac{n_d}{n_0 + n_d} = \frac{1}{1 + \dfrac{2.8 \times 10^{19}}{2(10^{16})} \exp\left(\dfrac{-0.045}{0.0259}\right)} = 0.0041 = 0.41\%$$

Comment

This example shows that there are very few electmns in the donor state compared with th conduction band. Essentially all of the electrons from the donor states are in the conductio band and. since only about 0.4 percent of the donor states contain electrons, the donor state are said to be completely ionized.

At room temperature, then, the donor states are essentially completely ionize and, fora typical doping of 10^{16} cm^{-3}, almost all donor impurity atoms have donate an electron to the conduction band.

At room temperature, there is also essentially *complete ionization* of the accep tor atoms. This means that each acceptor atom has accepted an electron from the va lence band so that p_a is zero. At typical acceptor doping concentrations, a hole is cre ated in the valence hand for each acceptor atom. This ionization effect and the creation of electrons and holes in the conduction band and valence band, respec tively, are shown in Figure 4.12.

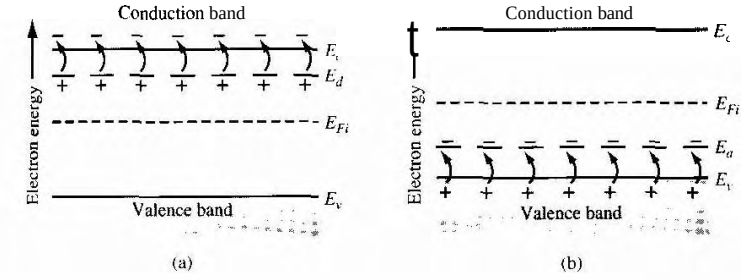


Figure 4.12 | Energy-hand diagrams showing complete ionization of (a) donor states and (b) acceptor states.

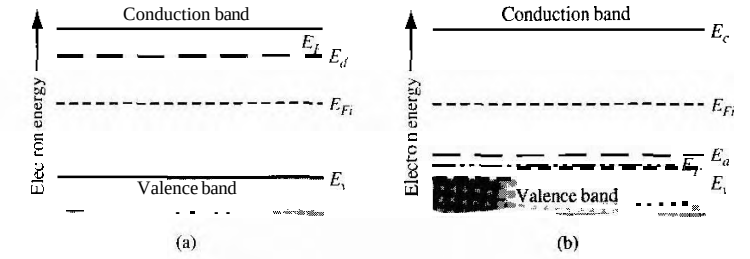


Figure 4.13 | Energy-band diagram at $T = 0$ K for (a) n-type and (b) p-type semiconductors.

The opposite of complete ionization occurs at $T = 0\ K$. At absolute zero degrees, all electrons are in their lowest possible energy state; that is, for an n-type semiconductor, each donor state must contain an electron, therefore $n_d = N_d$ or $N_d^+ = 0$. We must have, then, from Equation (4.50) that $\exp[(E_d - E_F)/kT] = 0$. Since T = 0 K, this will occur for $\exp(-\infty) = 0$, which means that $E_F > E_d$. The Fermi energy level must he above the donor energy level at absolute zero. In the case of a p-type semiconductor at absolute zero temperature, the impurity atoms will not contain any electrons, so that the Fermi energy level must be below the acceptor energy state. The distribution of electrons among the various energy states, and hence the Fermi energy, is a function of temperature.

Adetailed analysis, not given in this text, shows that at $T = 0\ K$, the Fermi energy is halfway between E, and E_d for the n-type material and halfway between E_a and E_v for the p-type material. Figure 4.13 shows these effects. No electrons from the donor state are thermally elevated into the conduction band; this effect is called freeze-our. Similarly, when no electrons from the valance band are elevated into the acceptor states, the effect is also called freeze-out.

Between $T = 0$ K. freeze-out, and T = 300 K, complete ionization, we have partial ionization of donor or acceptor atoms.

EXAMPLE 4.8 Objective

To determine the temperature at which 90 percent of acceptor atoms are ionized.

Consider p-type silicon doped with boron at a concentration of $N_a = 10^{17}$ cm^{-3}.

■ Solution

Find the ratio of holes in the acceptor state to the total number of holes in the valence band pl acceptor state. Taking into account the Boltzmann approximation and assuming the degener acy factor is $g = 4$, we write

$$\frac{p_a}{p_0 + p_a} = \frac{1}{1 + \frac{N_v}{4N_a} \cdot \exp\left[\frac{-(E_a - E_v)}{kT}\right]}$$

For 90 percent ionization,

$$\frac{p_a}{p_0 + p_a} = 0.10 = \frac{1}{1 + \frac{(1.04 \times 10^{19})\left(\frac{T}{300}\right)^{3/2}}{4(10^{16})} \cdot \exp\left[\frac{-0.045}{0.0259\left(\frac{T}{300}\right)}\right]}$$

Using trial and error, wc find that T = 193 K.

■ Comment

This example shows that at approximately 100°C below room temperature. we still have YO percent of the acceptor atoms ionized; in other words. 90 percent of the acceptor atoms have "donated" a hole to the valence band.

TEST YOUR UNDERSTANDING

E4.9 Determine the fraction of total holes still in the acceptor states in silicon at T = 300 K for a boron impurity concentration of $N_a = 10^{17}$ cm^{-3}. (6Ll'0 'suA)

E4.10 Consider silicon with a phosphorus impurity concentration of $N_d = 5 \times 10^{15}$ cm^{-3}. Plot the percent of ionized impurity atoms versus temperature over the range $100 \leq T \leq 400$ K.

4.5 | CHARGE NEUTRALITY

In thermal equilibrium, the semiconductor crystal is electrically neutral. The elec trons are distributed among the various energy states. creating negative and positive charges, but the net charge density is zero. This charge-neutrality condition is used to determine the thermal-equilibrium electron and hole concentrations as a function of

the impurity doping concentration. We will define a compensated semiconductor and then determine the electron and hole concentrations as a function of the donor and acceptor concentrations.

4.5.1 compensated Semiconductors

A compensated semiconductor is one that contains both donor and acceptor impurity atoms in the same region. A compensated semiconductor can he formed, for example, by diffusing acceptor impurities into an n-type material, or by diffusing donor impurities into a p-type material. An n-type compensated semiconductor occurs when $N_d > N_a$, and a p-type compensated semiconductor occurs when $N_a > N_d$. If $N_a = N_d$, we have a completely compensated semiconductor that has. as we will show, the characteristics of an intrinsic material. Compensated semiconductors are created quite naturally during device fabrication as we will see later.

4.5.2 Equilibrium Electron and Hole Concentrations

Figure 4.14 shows the energy-hand diagram of a semiconductor when both donor and acceptor impurity atoms are added to the same region to form a compensated

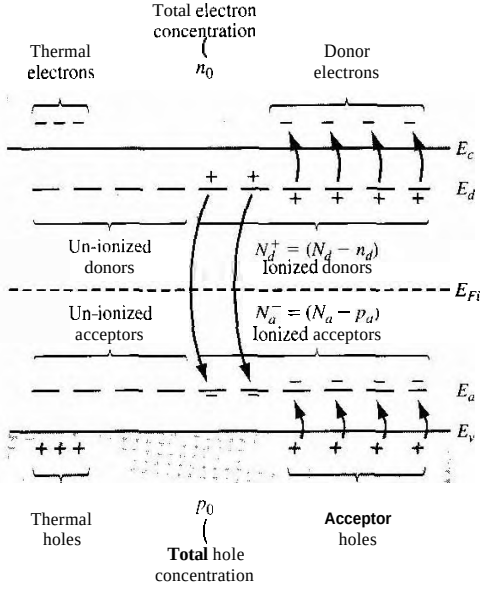


Figure 4.14 **I** Energy-band diagram of a campenrated semiconductor showing ionized and un-ionized donors and acceptors.

semiconductor. The figure shows how the electrons and holes can be distributed among the various states.

The charge neutrality condition is expressed by equating the density of negative charges to the density of positive charges. We then have

$$n_0 + N_a^- = p_0 + N_d^+ \tag{4.56}$$

or

$$n_0 + (N_a - p_a) = p_0 + (N_d - n_d) \tag{4.57}$$

where n_0 and p_0 are the thermal-equilibrium concentrations of electrons and holes the conduction band and valence band, respectively. The parameter n_d is the conce tration of electrons in the donor energy states, so $N_d^+ = N_d - n_d$ is the concentrati of positively charged donor states. Similarly, p, is the concentration of holes in t acceptor states, so $N_a^- = N_{,,} - p$, is the concentration of negatively charged accep tor states. We have expressions for n_0, p_0, n_d, and p_a in terms of the Fermi energy and temperature.

If we assume complete ionization, n_d and p_a are both zero, and Equation (4.57) becomes

$$n_0 + N_a = p_0 + N_d \tag{4.58}$$

If we express p_0 as n_i^2/n_0, then Equation (4.58) can be written as

$$n_0 + N_a = \frac{n_i^2}{n_0} + N_d \tag{4.59a}$$

which in turn can be written as

$$n_0^2 - (N_d - N_a)n_0 - n_i^2 = 0 \tag{4.59b}$$

The electron concentration n_0 can be determined using the quadratic formula, or

$$\boxed{n_0 = \frac{(N_d - N_a)}{2} + \sqrt{\left(\frac{N_d - N_a}{2}\right)^2 + n_i^2}} \tag{4.60}$$

The positive sign in the quadratic formula must be used, since, in the limit of an intrinsic semiconductor when $N_a = N_d = 0$. the electron concentration must he a positive quantity, or $n_0 = n_i$.

Equation (4.60) is used to calculate the electron concentration in an n-type semiconductor, or when $N_d > N_a$. Although Equation (4.60) was derived for a compensated semiconductor, the equation is also valid for $N_a = 0$.

EXAMPLE 4.9 Objective

To determine the thermal equilibrium electron and hole concentrations fur a given doping concentration.

Consider an n-type silicon semiconductor at $T = 300$ K in which $N_d = 10^{16}$ cm ^{3} and $N_a = 0$. The intrinsic carrier concentration is assumed to be $n_i = 1.5 \times 10^{10}$ cm^{-3}

Solution

From Equation (4.60), the majority carrier electron concentration is

$$n_0 = \frac{10^{16}}{2} + \sqrt{\left(\frac{10^{16}}{2}\right)^2 + (1.5 \times 10^{10})^2} \approx 10^{16} \text{ cm}^{-3}$$

The minority carrier hole concentration is found as

$$p_0 = \frac{n_i^2}{n_0} = \frac{(1.5 \times 10^{10})^2}{1 \times 10^{16}} = 2.25 \times 10^4 \text{ cm}^{-3}$$

Comment

In this example, $N_d \gg n_i$, so that the thermal-equilibrium majority carrier electron concentration is essentially equal to the donor impurity concentration. The thermal-equilibrium majority and minority carrier concentrations can differ by many orders of magnitude.

We have argued in our discussion and we may note from the results of Example 4.9 that the concentration of electrons in the conduction band increases above the intrinsic carrier concentration as we add donor impurity atoms. At the same time, the minority carrier hole concentration decreases below the intrinsic carrier concentration as we add donor atoms. We must keep in mind that as we add donor impurity atoms and the corresponding donor electrons, there is a redistribution of electrons among available energy states. Figure 4.15 shows a schematic of this physical redistribution. A few of the donor electrons will fall into the empty states in the valence

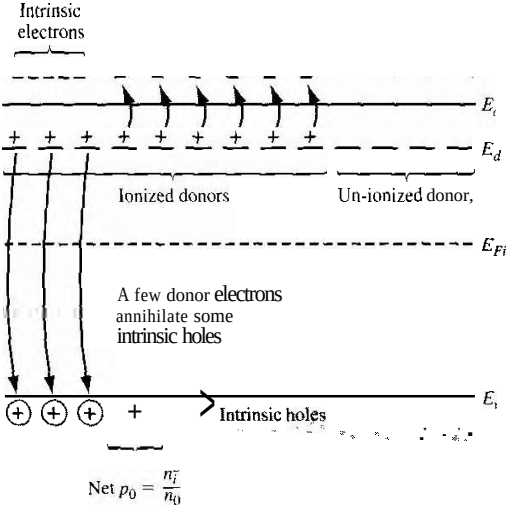


Figure 4.15 | Energy-band diagram showing the redistribution of electrons when donors are added.

band and, in doing so, will annihilate some of the intrinsic holes. The minority carrier hole concentration will therefore decrease as we have seen in Example 4.9. At the same time, because of this redistribution, the net electron concentration in the conduction band is *not* simply equal to the donor concentration plus the intrinsic electron concentration.

EXAMPLE 4.10

Objective

To calculate the thermal-equilibrium electron and hole concentrations in a germanium sample for a given doping density.

Consider a germanium sample at T = 300 K in which $N_d = 5 \times 10^{13}$ cm^{-3} and $N_a =$ Assume that $n_i = 2.4 \times 10^{13}$ cm^{-3}.

■ Solution

Again, from Equation (4.60). the majority carrier electron concentration is

$$n_0 = \frac{5 \times 10^{13}}{2} + \sqrt{\left(\frac{5 \times 10^{13}}{2}\right)^2 + (2.4 \times 10^{13})^2} = 5.97 \times 10^{13} \text{ cm}^{-3}$$

The minority carrier hole concentration is

$$p_0 = \frac{n_i^2}{n_0} = \frac{(2.4 \times 10^{13})^2}{5.97 \times 10^{13}} = 9.65 \times 10^{12} \text{ cm}^{-3}$$

Comment

If the donor impurity concentration is not too different in magnitude from the intrinsic carrier concentratiun. then the thermal-equilibrium majority carrier electron concentration is influenced by the intrinsic concentration.

We have seen that the intrinsic carrier concentration n_i is a very strong function of temperature. As the temperature increases, additional electron-hole pairs are thermally generated so that the n_i^2 term in Equation (4.60) may begin to dominate. The semiconductor will eventually lose its extrinsic characteristics. Figure 4.16 shows the electron concentration versus temperature in silicon doped with 5×10^{14} donors per cm^3. As the temperature increases, we can see where the intrinsic concentration begins to dominate. Also shown is the partial ionization, or the onset of freeze-out, at the low temperature.

If we reconsider Equation (4.58) and express n_0 as n_i^2/p_0, then we have

$$\frac{n_i^2}{p_0} + N_a = p_0 + N_d \tag{4.61a}$$

which we can write as

$$p_0^2 - (N_a - N_d)p_0 - n_i^2 = 0 \tag{4.61b}$$

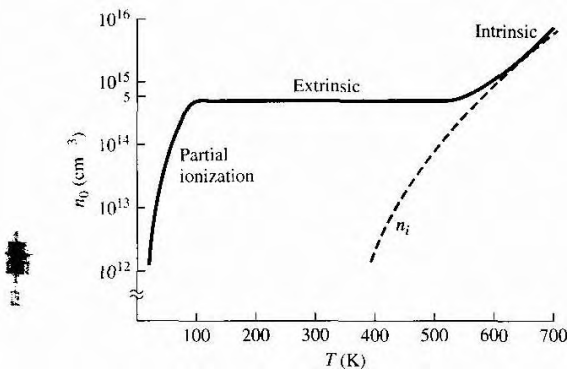


Figure 4.16 | Electron concentration versus temperature showing the ***three*** regions: partial ionization, extrinsic, and intrinsic.

Using the quadratic formula, the hole concentration is given by

$$p_0 = \frac{N_a - N_d}{2} + \sqrt{\left(\frac{N_a - N_d}{2}\right)^2 + n_i^2} \tag{4.62}$$

where the positive sign, again, must he used. Equation (4.62) is used to calculate the thermal-equilibrium majority carrier hole concentration in a p-type semiconductor, or when $N_a > N_d$. This equation also applies for $N_d = 0$.

Objective | EXAMPLE 4.11

To calculate the thermal-equilibrium electron and hole concentrations in a compensated p-type semiconductor.

Consider a silicon semiconductor at T = 300 K in which $N_a = 10^{16}$ cm^{-3} and $N_d = 3 \times 10^{15}$ cm^{-3}. Assume $n_i = 1.5 \times 10^{10}$ cm^{-3}.

Solution

Since $N_a > N_d$, the compensated semiconductor is p-type and the thermal-equilibrium majority carrier hole concentration is given by Equation (4.62) as

$$p_0 = \frac{10^{16} - 3 \times 10^{15}}{2} + \sqrt{\left(\frac{10^{16} - 3 \times 10^{15}}{2}\right)^2 + (1.5 \times 10^{10})^2}$$

so that

$$p_0 \approx 7 \times 10^{15} \text{ cm}^{-3}$$

The minority carrier electron concentration is

$$n_0 = \frac{n_i^2}{p_0} = \frac{(1.5 \times 10^{10})^2}{7 \times 10^{15}} = 3.21 \times 10^4 \text{ cm}^{-3}$$

■ Comment

If we assume complete ionization and if $(N_a - N_d) \gg n_i$, then the majority carrier hole concentration is, to a very good approximation, just the difference between the acceptor and donor concentrations.

We may note that, for a *compensated p-type* semiconductor, the minority carrier electron concentration is determined from

$$n_0 = \frac{n_i^2}{p_0} = \frac{n_i^2}{(N_a - N_d)}$$

DESIGN EXAMPLE 4.12



Objective

To determine the required impurity doping concentration in a semiconductor material.

A silicon device with n-type material is to be operated at T = 550 K. At this temperature the intrinsic carrier concentration must contribute no more than 5 percent of the total electron concentration. Determine the minimum donor concentration required to meet this specification.

■ Solution

At T = 550 K, the intrinsic carrier concentration is found from Equation (4.23) as

$$n_i^2 = N_c N_v \exp\left(\frac{-E_g}{kT}\right) = (2.8 \times 10^{19})(1.04 \times 10^{19})\left(\frac{550}{300}\right)^3 \exp\left[\frac{-1.12}{0.0259}\left(\frac{300}{550}\right)\right]$$

or

$$n_i^2 = 1.02 \times 10^{29}$$

so that

$$n_i = 3.20 \times 10^{14} \text{ cm}^{-3}$$

For the intrinsic carrier concentration to contribute no more than 5 percent of the total electron concentration, we set $n_0 = 1.05 N_d$.

From Equation (4.60), we have

$$n_0 = \frac{N_d}{2} + \sqrt{\left(\frac{N_d}{2}\right)^2 + n_i^2}$$

or

$$1.05 N_d = \frac{N_d}{2} + \sqrt{\left(\frac{N_d}{2}\right)^2 + (3.20 \times 10^{14})^2}$$

which yields

$$N_d = 1.39 \times 10^{15} \text{ cm}^{-3}$$

Comment

If the temperature remains less than $T = 550$ K, then the intrinsic carrier concentration will contribute less than 5 percent of the total electron concentration for this donor impurity concentration.

Equations (4.60) and (4.62) are used to calculate the majority carrier electron concentration in an n-type semiconductor and majority carrier hole concentration in a p-type semiconductor, respectively. The minority carrier hole concentration in an n-type semiconductor could, theoretically, be calculated from Equation (4.62). However, we would be subtracting two numbers on the order of 10^{16} cm^{-3}, for example, to obtain a number on the order of 10^4 cm^{-3}, which from a practical point of view is not possible. The minority carrier concentrations are calculated from $n_0 p_0 = n_i^2$ once the majority carrier concentration has been determined.

TEST YOUR UNDERSTANDING

E4.11 Consider a compensated GaAs semiconductor at T = 300 K doped at $N_d = 5 \times 10^{15}$ cm^{-3} and $N_a = 2 \times 10^{16}$ cm^{-3}. Calculate the thermal equilibrium electron and hole concentrations. ($n_0 = 1.5 \times 10^{15}$ cm^{-3}, $p_0 = 2.17 \times 10^{15}$ cm^{-3})

E4.12 Silicon is doped at $N_d = 10^{15}$ cm^{-3} and $N_a = 0$. *(a)* Plot the concentration of electrons versus temperature over the range $300 \leq T \leq 600$ K. *(b)* Calculate the temperature at which the electron concentration is equal to 1.1×10^{15} cm^{-3}. ($T \approx 552$ K)

4.6 POSITION OF FERMI ENERGY LEVEL

We discussed qualitatively in Section 4.3.1 how the electron and hole concentrations change as the Fermi energy level moves through the bandgap energy. Then, in Section 4.5, we calculated the electron and hole concentrations as a function of donor and acceptor impurity concentrations. We can now determine the position of the Fermi energy level as a function of the doping concentrations and as a function of temperature. The relevance of the Fermi energy level will be further discussed after the mathematical derivations.

4.6.1 Mathematical Derivation

The position of the Fermi energy level within the bandgap can be determined by using the equations already developed for the thermal-equilibrium electron and hole concentrations. If we assume the Boltzmann approximation to be valid, then from Equation (4.11) we have $n_0 = N_c \exp[-(E_c - E_F)/kT]$. We can solve for $E_c - E_F$

from this equation and obtain

$$E_c - E_F = kT \ln\left(\frac{N_c}{n_0}\right) \tag{4.6}$$

where n_0 is given by Equation (4.60). If we consider an n-type semiconductor in which $N_d \gg n_i$, then $n_0 \approx N_d$, so that

$$E_c - E_F = kT \ln\left(\frac{N_c}{N_d}\right) \tag{4.64}$$

The distance between the bottom of the conduction band and the Fermi energy is a logarithmic function of the donor concentration. As the donor concentration increases, the Fermi level moves closer to the conduction band. Conversely, if the Fermi level moves closer to the conduction band, then the electron concentration in the conduction band is increasing. We may note that if we have a compensated semiconductor, then the N_d term in Equation (4.64) is simply replaced by $N_d - N_a$, or the net effective donor concentration.

DESIGN EXAMPLE 4.13



Objective

To determine the required donor impurity concentration to obtain a specified Fermi energy.

Silicon at T = 300 K contains an acceptor impurity concentration of $N_a = 10^{16}$ cm^{-3}. Determine the concentration of donor impurity atoms that must be added so that the silicon is n type and the Fermi energy is 0.20 eV below the conduction band edge.

■ Solution

From Equation (4.64). we have

$$E_c - E_F = kT \ln\left(\frac{N_c}{N_d - N_a}\right)$$

which can be rewritten as

$$N_d - \mathrm{N}, = N_c \exp\left[\frac{-(E_c - E_F)}{kT}\right]$$

Then

$$N_d - \mathrm{N}, = 2.8 \times 10^{19} \exp\left[\frac{-0.20}{0.0259}\right] = 1.24 \times 10^{16}\ \mathrm{cm}^{-3}$$

or

$$N_d = 1.24 \times 10^{16} + N_a = 2.24 \times 10^{16}\ \mathrm{cm}^{-3}$$

■ Comment

A compensated semiconductor can be fabricated to provide a specific Fermi energy level.

We may develop a slightly different expression for the position of the Fermi level. We had from Equation (4.39) that $n_0 = n_i \exp[(E_F - E_{Fi})/kT]$. We can solve for $E_F - E_{Fi}$ as

$$E_F - E_{Fi} = kT \ln\left(\frac{n_0}{n_i}\right) \tag{4.65}$$

Equation (4.65) can be used specifically for an n-type semiconductor, where n_0 is given by Equation (4.60), to find the difference between the Fermi level and the intrinsic Fermi level as a function of the donor concentration. We may note that, if the net effective donor concentration is zero, that is, $N_d - N_a = 0$, then $n_0 = n$, and $E_F = E_{Fi}$. A completely compensated semiconductor has the characteristics of an intrinsic material in terms of carrier concentration and Fenni level position.

We can derive the same types of equations for a p-type semiconductor. From Equation (4.19), we have $p_0 = N_v \exp[-(E_F - E_v)/kT]$, so that

$$E_F - E_v = kT \ln\left(\frac{N_v}{p_0}\right) \tag{4.66}$$

If we assume that $N_a \gg n_i$, then Equation (4.66) can be written as

$$E_F - E_v = kT \ln\left(\frac{N_v}{N_a}\right) \tag{4.67}$$

The distance between the Fermi level and the top of the valence-band energy for a p-type semiconductor is a logarithmic function of the acceptor concentration: as the acceptor concentration increases, the Fermi level moves closer to the valence band. Equation (4.67) still assumes that the Boltzmann approximation is valid. Again. if we have a compensated p-type semiconductor, then the N_a term in Equation (4.67) is replaced by $N_a - N_d$, or the net effective acceptor concentration.

We can also derive an expression for the relationship between the Fermi level and the intrinsic Fermi level in terms of the hole concentration. We have from Equation (4.40) that $p_0 = n_i \exp[-(E_F - E_{Fi})/kT]$, which yields

$$E_{Fi} - E_F = kT \ln\left(\frac{p_0}{n_i}\right) \tag{4.68}$$

Equation (4.68) can be used to find the difference between the intrinsic Fermi level and the Fermi energy in terms of the acceptor concentration. The hole concentration p_0 in Equation (4.68) is given by Equation (4.62).

We may again note from Equation (4.65) that, for an n-type semiconductor, $n_0 > n_i$ and $E_F > E_{Fi}$. The Fermi level for an n-type semiconductor is above E_{Fi}. For a p-type semiconductor, $p_0 > n_i$, and from Equation (4.68) we see that

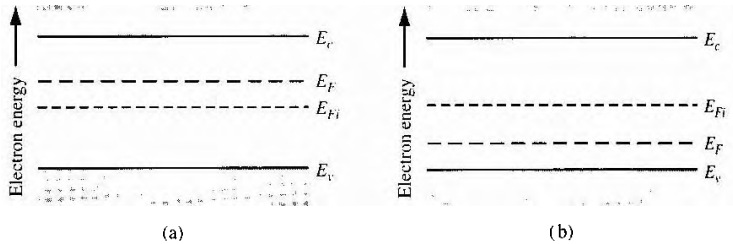


Figure 4.17 | Position of Fermi level for an (a) n-type ($N_d > N_a$) and (b) p-type ($N_d > N_a$) semiconductor.

$E_{Fi} > E_F$. The Fenni level for a p-type semiconductor is below E_{Fi}. These result are shown in Figure 4.17.

4.6.2 Variation of E_F with Doping Concentration and Temperature

We may plot the position of the Fermi energy level as a function of the doping concentration. Figure 4.18 shows the Fermi energy level as a function of donor concentration (n type) and as a function of acceptor concentration (p type) for silicon at T = 300 K. As the doping levels increase, the Fermi energy level moves closer to the conduction band for the n-type material and closer to the valence band for the p-type material. Keep in mind that the equations for the Fermi energy level that we have derived assume that the Boltzmann approximation is valid.

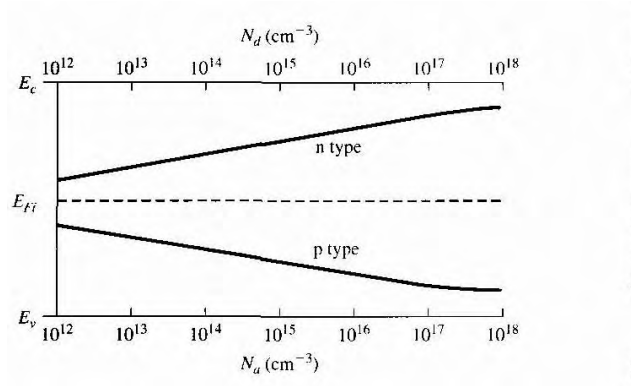


Figure 4.18 | Position of Fermi level as a function of donor concentration (n type) and acceptor concentration (p type).

Objective

EXAMPLE 4.14

To determine the Fermi-level position and the maximum doping at which the Boltzmann approximation is still valid.

Consider p-type silicon, at T = 300 K, doped with boron. We may assume that the limit of the Boltzmann approximation occurs when $E_F - E_a = 3kT$. (See Section 4.1.2.)

■ Solution

From Table 4.3, we find the ionization energy is $E_a - E_v = 0.045$ eV for boron in silicon. If we assume that $E_{Fi} \approx E_{midgap}$, then from Equation (4.68), the position of the Fermi level at the maximum doping is given by

$$E_{Fi} - E_F = \frac{E_g}{2} - (E_a - E_v) - (E_F - E_a) = kT \ln\left(\frac{N_a}{n_i}\right)$$

$$0.56 - 0.045 - 3(0.0259) = 0.437 = (0.0259) \ln\left(\frac{N_a}{n_i}\right)$$

We can then solve for the doping as

$$N_a = n_i \exp\left(\frac{0.437}{0.0259}\right) = 3.2 \times 10^{17} \text{ cm}^{-3}$$

■ Comment

If the acceptor (or donor) concentration in silicon is greater than approximately 3×10^{17} cm^{-3}, then the Boltzmann approximation of the distribution function becomes less valid and the equations for the Fermi-level position are no longer quite as accurate.

TEST YOUR UNDERSTANDING

E4.13 Determine the position of the Fermi level with respect to the valence band energy in p type GaAs at T = 300 K. The doping concentrations are $N_a = 5 \times 10^{16}$ cm^{-3} and $N_d = 4 \times 10^{15}$ cm^{-3}. (Ans. $E_F - E_v = 0.130$ eV)

E4.14 Calculate the position of the Fermi energy level in n-type silicon at $T = 300$ K with respect to the intrinsic Fermi energy level. The doping concentrations are $N_d = 2 \times 10^{17}$ cm^{-3} and $N_a = 3 \times 10^{16}$ cm^{-3}. (Ans. $E_F - E_{Fi} = 0.421$ eV)

The intrinsic carrier concentration n_i, in Equations (4.65) and (4.68), is a strong function of temperature, so that E_F is a function of temperature also. Figure 4.19 shows the variation of the Fermi energy level in silicon with temperature for several donor and acceptor concentrations. As the temperature increases, n_i increases, and E_F moves closer to the intrinsic Fermi level. At high temperature, the semiconductor material begins to lose its extrinsic characteristics and begins to behave more like an intrinsic semiconductor. At the very low temperature, freeze-out occurs; the Boltzmann approximation is no longer valid and the equations we derived for the

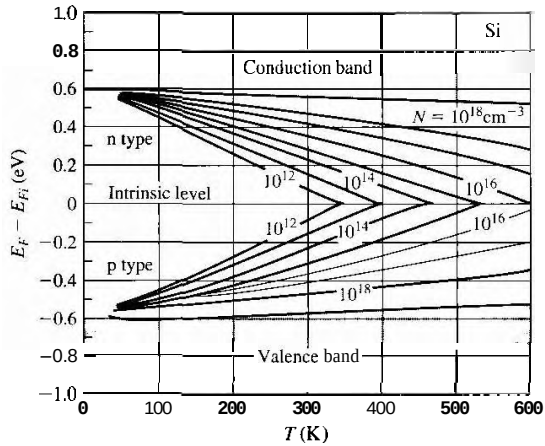


Figure 4.19 | Position of Fermi level as a function of temperature for various doping concentrations. (From *Sze [13].*)

Fermi-level position no longer apply. At the low temperature where freeze-out occurs, the Fermi level goes above E_d for the n-type material and below $E_$, for the p-type material. At absolute zero degrees, all energy states below E_F are full and all energy states above E_F are empty.

4.6.3 Relevance of the Fermi Energy

We have been calculating the position of the Fermi energy level as a function of doping concentrations and temperature. This analysis may seem somewhat arbitrary and fictitious. However, these relations do become significant later in our discussion of pn junctions and the other semiconductor devices we consider. An important point is that, in thermal equilibrium, the Fermi energy level is a constant throughout a system. We will not prove this statement, but we can intuitively see its validity by considering the following example.

Suppose we have a particular material, A, whose electrons are distributed in the energy states of an allowed band as shown in Figure 4.20a. Most of the energy states below E_{FA} contain electrons and most of the energy states above E_{FA} are empty of electrons. Consider another material, B, whose electrons are distributed in the energy states of an allowed band as shown in Figure 4.20b. The energy states below E_{FB} are mostly full and the energy states above E_{FB} are mostly empty. If these two materials are brought into intimate contact, the electrons in the entire system will tend to seek the lowest possible energy. Electrons from material A will flow into the lower energy states of material B, as indicated in Figure 4.20c, until thermal equilibrium is reached. Thermal equilibrium occurs when the distribution of electrons, as

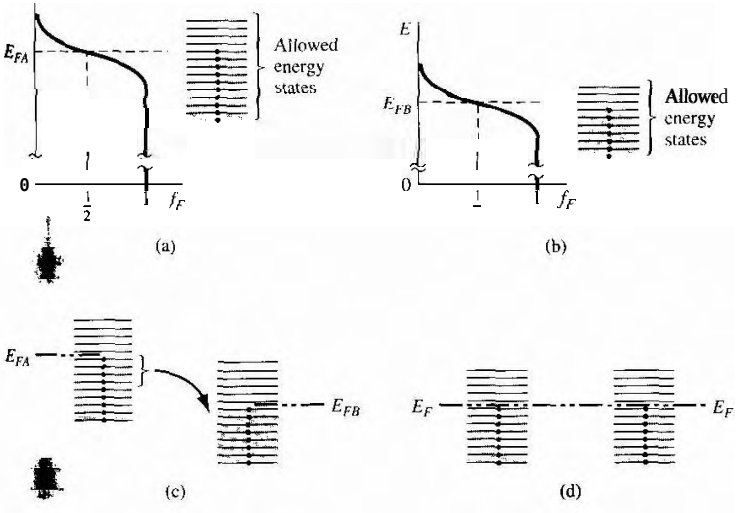


Figure 4.20 | The Fermi energy of (a) material A in thermal equilibrium, (b) material B in thermal equilibrium, (c) materials A and B at the instant they are placed in contact, md (d) materials A and B in contact at thermal equilibrium.

a function of energy, is the same in the two materials. This equilibrium state occurs when the Fermi energy is the same in the two materials as shown in Figure 4.20d. The Fermi energy, important in the physics of the semiconductor, also provides a good pictorial representation of the characteristics of the semiconductor materials and devices.

4.7 | SUMMARY

- The concentration of electrons in the conduction band is the integral over the conduction band energy of the product of the density of states function in the conduction hand and the Fermi–Dirac probahility function.
- The concentration of holes in the valence band is the integral over the valence bend energy of the product of the density of states function in the valence band and the probability of a state being empty, which is $[1 - f_F(E)]$.
- Using the Maxwell–Boltzmann approximation, the thermal equilibrium concentration of electrons in the conduction band is given by

$$n_0 = \mathrm{N}_c \exp\left[\frac{-(E_c - E_F)}{kT}\right]$$

where N_c is the effective density of states in the conduction band

- Using the Maxwell–Boltzmann approximation, the thermal equilibrium concentration of holes in the valence band is given by

$$p_0 = N_v \exp\left[\frac{-(E_F - E_v)}{kT}\right]$$

where N_v is the effective density of states in the valence hand
The intrinsic carrier concentration is found from

$$n_i^2 = N_c N_v \exp\left[\frac{-E_g}{kT}\right]$$

- The concept of doping the semiconductor with donor (group V elements) impurities and acceptor (group III elements) impurities to form n-type and p-type extrinsic semiconductors was discussed.
The fundamental relationship of $n_0 p_0 = n_i^2$ was derived.
Using the concepts of complete ionization and charge neutrality, equations for the electron and hole concentrations as a function of impurity doping concentrations were derived.
The position of the Fermi energy level as a function of impurity doping concentrations was derived.
The relevance of the Fermi energy was discussed. The Fermi energy is a constant throughout a semiconductor that is in thermal equilibrium.

GLOSSARY OF IMPORTANT TERMS

acceptor atoms Impurity atoms added to a semiconductor to create a p-type material

charge carrier The electron and/or hole that moves inside the semiconductor and gives rise to electrical currents.

compensated semiconductor A semiconductor that contains both donors and acceptors in the same semiconductor region.

complete ionization The condition when all donor atoms are positively charged by giving up their donor electrons and all acceptor atoms arc negatively charged by accepting electrons.

degenerate semiconductor A semiconductor whose electron concentration or hole concentration is greater than the effective density of states, so that the Fermi level is in the conduction band (n type) or in the valence band (p type).

donor atoms Impurity atoms added to a semiconductor to create an n-type material.

effective density of states The parameter N_c. which results from integrating the density of quantum states $g_c(E)$ times the Fermi function $f_F(E)$ over the conduotion-band energy, and the parameter N_v. which results from integrating the density of quantum states $g_v(E)$ times $[1 - f_F(E)]$ over the valence-band energy.

extrinsic semiconductor A semiconductor in which controlled amounts of donors and/or acceptors have been added so that the electron and hole concentrations change from the intrinsic carrier concentration and a preponderance of either electrons (n type) or holes (p type) is created.

freeze-out The condition that occurs in a semiconductor when the temperature is lowered and the donors and acceptors become neutrally charged. The electron and hole concentrations become very small.

intrinsic carrier concentration n_i The electron concentration in the conduction hand and the hole concentration in the valence band (equal values) in an intrinsic semiconductor.

intrinsic Fermi level E_{Fi} The position of the Fermi level in an intrinsic semiconductor.

intrinsic semiconductor A pure semiconductor material with no impurity atoms and no lattice defects in the crystal.

nondegenerate **semiconductor** A semiconductor in which a relatively small number of donors and/or acceptors have been added so that discrete, nonintcracting donor states and/or discrete, noninteracting acceptor states are introduced.

CHECKPOINT

After studying this chapter. the reader should have the ability to:

- Derive the equations for the thermal equilibrium concentrations of electrons and holes in terms of the Fermi energy.
- Derive the equation for the intrinsic carrier concentration.
- State the value of the intrinsic carrier concentration for silicon at T = 300 K.
- Derive the expression for rhe intrinsic Fcrmi level.
- Describe the effect of adding donor and acceptor impurity atoms to a semiconductor.
- Understand the concept of complete ionization.
- Understand the derivation of the fundamental relationship $n_0 p_0 = n_i^2$
- Describe the meaning of degenerate and nondegenerate semiconductors.
- Discuss the concept of charge neutrality.
- Derive the equations for n_0 and p_0 in terms of impurity doping concentrations.
- Discuss the variation of the Fermi energy with doping concentration and temperature.

REVIEW QUESTIONS

1. Write the equation for $n(E)$ as a function of the density of states and the Fermi probability function. Repeat for the function $p(E)$.
2. In deriving the equation for n_0 in terms of the Fermi function. the upper limit of the integral should be the energy at the top of the conduction hand. Justify using infinity instead.
3. Assuming the Boltzmann approximation applies, write the equations for n_0 and p_0 in terms of the Fcrmi energy.
4. What is the value of the intrinsic carrier concentration in silicon at $T = 300$ K?
5. Under what condition would the intrinsic Fermi level be at the midgap energy?
6. What is a donor impurity'! What is an acceptor impurity?
7. What is meant by complete ionization? What is meant by freeze-out?
8. What is the product of n_0 and p_0 equal to?
9. Write the equation for charge neutrality for the condition of complete ionization.
10. Sketch a graph of n_0 versus temperature for an n-type material.
11. Sketch graphs of the Fenni energy versus donor impurity concentration and versus temperature.

PROBLEMS

Section 4.1 Charge Carriers in Semiconductors

4.1 Calculate the intrinsic carrier concentration, n_i, at T = 200, 400, and 600 K for (*a*) silicon, (*b*) germanium, and (*c*) gallium arsenide.

4.2 The intrinsic carrier concentration in silicon is to be no greater than $n_i = 1 \times 10^{12}$ cm^{-3}. Assume $E_g = 1.12$ eV. Determine the maximum temperature allowedfc the silicon.



4.3 Plot the intrinsic carrier concentration, n_i, for a temperature range of $200 \leq T \leq 600$ K for (*a*)silicon, (b) germanium, and (*c*) gallium arsenide. (Use a log scale for n_i.)

4.4 In a particular semiconductor material, the effective density of states functions are given hy $N_c = N_{c0}(T)^{3/2}$ and $N_v = N_{v0}(T)^{3/2}$ where N_{c0} and N_{v0} are consrants independent of temperature. The experimentally determined intrinsic carrier concentrations as a function of temperature are given in Table 4.5. Determine the product $N_{c0}N_{v0}$ and the bandgap energy E_g. (Assume E_g is independent of temperature.)

4.5 (*a*) The magnitude of the product $g_C(E)f_F(E)$ in the conduction band is a function energy as shown in Figure 4.1. Assume the Boltrmann approximation is valid. Determine the energy with respect to E_c at which the maximum occurs. (*b*) Repeat pan (*a*) for the magnitude of the product $g_V(E)$ [I $- f_F(E)$] in the valence band.

4.6 Assume the Boltzmann approximation in a semiconductor is valid. Determine the ratio of $n(E) = g_C(E)f_F(E)$ at $E = E_c + 4kT$ to that at $E = E_c + kT/2$.

4.7 Assume that E, $- E_F = 0.20$ eV in silicon. Plot $n(E) = g_C(E)f_F(E)$ over the range $E_c \leq E \leq E_c + 0.10$ eV for (*a*) $T = 200$ K and (*b*) $T = 400$ K.

4.8 Two semiconductor materials have exactly the same properties except that material has a bandgap energy of 1.0 eV and material B has a bandgap energy of 1.2 eV. Determine the ratio of n_i of material A to that of material B for $T = 300$ K.



4.9 (*a*) Consider silicon at $T = 300$ K. Plot the thermal equilibrium electron concentration n_0 (on a log scale) over the energy range $0.2 \leq E_c - E_F \leq 0.4$ eV. (*b*) Repeat part (*a*) lor the hole concentration over the range $0.2 \leq E_F - E_v \leq 0.4$ eV.

4.10 Given the effective messes of electrons and holes in silicon. germanium, and gallium arsenide, calculate the position of the intrinsic Fermi energy level with respect to the center of the bandgap for each semiconductor at $T = 300$ K.

4.11 (a)The carrier effective masses in a semiconductor are $m_n^* = 0.62m_0$ and $m_p^* = 1.4m$ Determine the oosition of the intrinsic Fermi level with resnect to the center of the bandgap at T = 300 K. ***(b)***Repeat pan (*a*) if $m_n^* = 1.10m_0$ and $m_p^* = 0.25m_0$.

Table 4.5 | Intrinsic concentration as a function of temperature

T (K)	n_i (cm^{-3})
200	1.82×10^2
300	5.83×10^7
400	3.74×10^{10}
500	1.95×10^{12}

4.12 Calculate E_{Fi} with respect to the center of the bandgap in silicon for $T = 200$. 400, and 600 K.

4.13 Plot the intrinsic Fermi energy E_{Fi} with respect to the center of the bandgap in silicon for $200 \leq T \leq 600$ K.



4.14 If the density of states function in the conduction band of a particular semiconductor is a constant equal to K, derive the expression for the thermal-equilibrium concentration of electrons in the conduction hand, assuming Fermi–Dirac statistics and assuming the Boltzmann approximation is valid.

4.15 Repeat Problem 4.14 if the density of states function is given by $g_c(E) = C_1(E - E_c)$ for $E \geq E_c$ where C_1 is aconstant.

Section 4.2 Dopant Atoms and Energy Levels

4.16 Calculate the ionization energy and radius of the donor electron in germanium using the Bohr theory. (Use the density of states effective mass as a first approximation.)

4.17 Repeat Problem 4.16 for gallium arsenide.

Section 4.3 The Extrinsic Semiconductor

4.18 Theelectron concentration in silicon at $T = 300$ K is $n_0 = 5 \times 10^4$ cm^{-3}. (*a*) Determine p_0. Is this n- or p-type matenal? (*b*) Determine the position of the Fermi level with respect to the intrinsic Fermi level.

4.19 Determine the values of n_0 and p_0 for silicon at $T = 300$ K if the Fermi energy is 0.22 eV above the valence band energy.

420 (*a*) If $E_c - E_f = 0.25$ eV in gallium arsenide at $T = 400$ K. calculate the values of n_0 and p_0. (b) Assuming the value of n_0 from part (a) remains constant, determine $E_c - E_F$ and p_0 at T = 300 K.

4.21 The value of p_0 in silicon at T = 300 K is 10^{15} cm^{-3}. Determine (a) $E_c - E_F$ and (*b*) n_0.

4.22 (*a*) Consider silicon at $T = 300$ K. Determine p_0 if $E_{Fi} - E_{F} = 0.35$ eV. (*b*) Assuming that p_0 from part (*a*) remains constant, determine the value of $E_{Fi} - E_F$ when $T = 400$ K. (*c*) Find the value of n_0 in both parts (**a**) and (b).

4.23 Repeat problem 4.22 for GaAs.

***4.24** Assume that $E_F = E_i$ at $T = 300$ K in silicon. Determine p_0.

***4.25** Consider silicon at T = 300 K, which has $n_0 = 5 \times 10^{19}$ cm^{-3}. Determine $E_i - E_F$.

Section 4.4 Statistics of Donors and Acceptors

***4.26** The electron and hole concentrations as a function of energy in the conduction hand and valence band peak at a particular energy as shown in Figure 4.8. Consider silicon and assume $E_c - E_F = 0.20$ eV. Determine the energy, relative to the band edges, at which the concentrations peak.

***4.27** For the Boltzmann approximation to he valid for a semiconductor, the Fermi level must be at least 3kT below the donor level in an n-type material and at least 3kT above the acceptor level in a p-type material. If T = 300 K, determine the maximum electron concentration in an n-type semiconductor and the maximum hole concentration

in a p-type semiconductor for the Boltzmann approximation to be valid in (*a*) silicon and (b) gallium arsenide.



4.28 Plot the ratio of un-ionized donor atoms to the total electron concentration versus temperature for silicon over the range $50 \leq T \leq 200$ K.

Section 4.5 Charge Neutrality

4.29 Consider a germanium semiconductor at T = 300 K. Calculate the thermal equilibrium concentrations of n_0 and p_0 for *(a)* $N_a = 10^{13}$ c m ^{3}. $N_d = 0$, and *(b)* $N_d = 5 \times 10^{15}$ cm^{-3}, $N_a = 0$.

***4.30** The Fermi level in n-type silicon at $T = 300$ K is 245 meV below the conduction band and 200 meV below the donor level. Determine the probability of finding an electron (*a*) in the donor level and (*b*) in a state in the conduction band kT above the conduction band edge.

4.31 Determine the equilibrium electron and hole concentrations in silicon for the following conditions:

(a) $T = 300$ K, $N_d = 2 \times 10^{15}$ cm^{-3}, $N_a = 0$

(b) $T = 300$ K. $N_d = 0$, N,, $= 10^{16}$ cm^{-3}

(*c*) T = 300 K, $N_d = N_a = 10^{15}$ cm^{-3}

(*d*) $T = 400$ K. $N_d = 0$. N,, $= 10^{14}$ cm^{-3}

(*e*) 7' = 500 K. $N_d = 10^{14}$ cm^{-3}. $N_a = 0$

4.32 Repeat problem 4.31 for GaAs.

4.33 Assume that silicon. germanium, and gallium arsenide each have dopant concentrations of $N_d = 1 \times 10^{13}$ cm^{-3} and N, $= 2.5 \times 10^{13}$ cm^{-3} at T = 300 K. For eachof the three materials (*a*) Is this material n type or p type'?(*b*) Calculate n_0 and p_0.

4.34 A sample of silicon at T = 450 K is doped with boron at a concentration of 1.5×10^{15} cm ^{3} and with arsenic at a concentration of 8×10^{14} cm^{-3}. (*a*) Is the material n or p type? (*b*) Determine the electron and hole concentrations. (*c*) Calculate the total ionized impurity concentration.

4.35 The thermal equilibrium hole concentration in silicon at $T = 300$ K is $p_0 = 2 \times 10^5$ cm^{-3}. Determine the thermal equilibrium electron concentration. Is the material n type or p type?

4.36 In a sample of GaAs at $T = 200$ K, we have experimentally determined that $n_0 = 5p$ and that $N_a = 0$. Calculate n_0, p_0, and N_d.



4.37 Consider a sample of silicon doped at $N_d = 0$ and $N_a = 10^{14}$ cm^{-3}. Plot the majority carrier concentration versus temperature over thr range $200 \leq T \leq 500$ K.

4.38 The temperature of a sample of silicon is T = 300 K and the acceptor doping conceo tration is N,, = 0. Plat the minority carrier concentratiun (on a log–log plot) versus N_d over the range $10^{15} \leq N_d \leq 10^{18}$ cm^{-3}.



4.39 Repeat problem **4.38** for GaAs.



4.40 A particular semiconductor material is doped at $N_d = 2 \times 10^{13}$ cm^{-3}. $N_a = 0$, and the intrinsic carrier concentration is $n_i = 2 \times 10^{13}$ cm^{-3}. Assume complete ionization Determine the thermal equilibrium majority and minority carrier concenuations.

4.41 *(a)* Silicon at $T = 300$ K is uniformly doped with arsenic atoms at a concentrationof 2×10^{16} cm^{-3} and boron atoms at a concentration of 1×10^{16} cm ^{3}. Determine the thermal equilibrium concentrations of majority and minority carriers. (b) Repeat

part (a) if the impurity concentrations are 2×10^{15} cm^{-3} phosphorus atoms and 3×10^{16} cm^{-3} boron atoms.

4.42 In silicon at T = 300 K. we have experimentally found that $n_0 = 4.5 \times 10^{}$ cm^{-3} and $N_d = 5 \times 10^{}$ cm^{-3}. *(a)* Is the material n type or p type? *(b)* Determine the majority and minority carrier concentrations. *(c)* What types and concentrations of impurity atoms exist in the material?

Section 4.6 Position of Fermi Energy Level

4.43 Consider germanium with an acceptor concentration of $N_a = 10^{15}$ cm^{-3} and a donor concentration of $N_d = 0$. Consider temperatures of T = 200,400. and 600 K. Calculate the position of the Fermi energy with respect to the intrinsic Fermi level at these temperatures.

4.44 Consider germanium at T = 300 K with donor concentrations of $N_d = 10^{14}$, 10^{16}, and 10^{18} cm^{-3}. Let $N_a = 0$. Calculate the position of the Fermi energy level with respect to the intrinsic Fermi level for these doping concentrations.

4.45 A GaAs device is doped with a donor concentration of 3×10^{15} cm^{-3}. For the device lo operate properly. the intrinsic carrier concentration must remain less than 5 percent of the total electron concentration. What is the maximum temperature that the device may operate?

4.46 Consider germanium with an acceptor concentration of $N_a = 10^{15}$ cm^{-3} and a donor concentration of $N_d = 0$. Plot the position of the Fermi energy with respect to the intrinsic Fermi level as a function of temperature over the range $200 \leq T \leq 600$ K.

4.47 Consider silicon at T = 300 K with $N_a = 0$. Plot the position of the Fermi energy level with respect to the intrinsic Fermi level as a function of the donor doping concentration over the range $10^{14} \leq N_d \leq 10^{18}$ cm^{-3}

4.48 For a particular semiconductor, $E_g = 1.50$ eV, $m_p^* = 10m_n^*$, $T = 300$ K, and $n_i = 1 \times 10^5$ cm^{-3}. *(a)* Determine the position of the intrinsic Fermi energy level with respect to the center of the bandgap. *(b)* Impurity atoms are added so that the Fermi energy level is 0.45 eV below the center of the bandgap. *(i)* Are acceptor or donor atoms added? *(ii)* What is the concentration of impurity atoms added?

4.49 Silicon at T = 300 K contains acceptor atoms at a concentration of $N_a = 5 \times 10^{15}$ cm^{-3}. Donor atoms are added forming an n-type compensated semiconductor such that the Fermi level is 0.215 eV below the conduction band edge. What concentration of donor atoms are added?

4.50 Silicon at $T = 300$ K is doped with acceptor atoms at a concentration of $N_a = 7 \times 10^{15}$ cm^{-3}. (a) Determine $E_F - E_v$. *(b)* Calculate the concentralinn of additional acceptor atoms that must be added to move the Fermi level a distance kT closer to the valence-band edge.

4.51 (a) Determine the position of the Fermi level with respect to the intrinsic Fermi level in silicon at T = 300 K that is doped with phosphorus atoms at a concentration of 10^{15} cm^{-3}. *(b)* Repeat part *(a)* if the silicon is doped with boron atoms at a concentration of 10^{15} cm^{-3}. *(c)* Calculate the electron concentration in the silicon for parts *(a)* and *(b)*.

4.52 Gallium arsenide at T = 300 K contains acceptor impurity atoms at a density of 10^{15} cm^{-3}. Additional impurity atoms are to be added so that the Fermi level is 0.45 eV below the intrinsic level. Determine the concentration and type (donor or acceptor) of impurity atoms to be added.

4.53 Determine the Fermi energy level with respect to the intrinsic Fermi level for each condition given in Problem 4.31

4.54 Find the Fermi energy level with respect to the valence band energy for the conditio given in Problem 4.32.

4.55 Calculate the position of the Fermi energy level with respect to the intrinsic Fermi fo the conditions given in Problem 4.42.

Summary and Review

4.56 A special semiconductor material is to he "designed." The semiconductor is to be n-type and doped with 1×10^{15} cm^{-3} donor atoms. Assume complete ionization and assume $N_a = 0$. The effective density of states functions are given by $N_c = N_v = 1.5 \times 10^{19}$ cm^{-3} and are independent of temperature. A particular semiconductor device fabricated with this material requires the electron concentration to he no greater than 1.01×10^{15} cm^{-3} at T = 400 K. What is the minimum value of the bandgap energy'!

4.57 Silicon atoms, at a concentration of 10^{10} c m ^{3}, are added to gallium arsenide. Assum that the silicon atoms act as fully ionized dopant atoms and that 5 percent of the c n- centration added replace gallium atoms and 95 percent replace arsenic atoms. Let T = 300 K. *(a)* Determine the donor and acceptor concmtrations. *(h)* Calculate the electron and hole concentrations and the position of the Fcrmi level with respect to $E_{,,}$.

4.58 Defects in a semiconductor material introduce allowed energy states within the forbidden bandgap. Assume that a particular defect in silicon introduces two discrete lev els: a donor level 0.25 eV above the top of the valence band, and an acceptor level 0.65 eV above the top of the valence band. The charge state of each defect is a function of the position of the Fermi level. *(a)* Sketch the charge density of each defect as the Fermi level moves from $E_,$ to $E_{,}$. Which defect level dominates in heavily doped n-type material'? In heavily doped p-type material? *(h)* Determine the electron and hole concentrations and the location of the Fermi level in (*i*) an n-type sample doped at $N_d = 10^{17}$ cm^{-3} and *(ii)* in a p-type sample doped at $N_a = 10^{17}$ cm^{-3}. (*c*) Determine the Fermi level position if no dopant atoms are added. Is the material n-type, p-type, or intrinsic?

READING LIST

***1.** Hess, K. *Advanced Theory of Semiconductor Devices.* Englewood Cliffs, NJ: Prentice Hall, 1988.

2. Kano, K. *Semiconductor Devices.* Upper Saddle River, NJ: Prentice Hall, 1998.

***3.** Li, S. S. *Semiconductor Physical Electronics.* New York: Plenum Press. 1993.

4. McKelvey, J. P. *Solid State Physics for Engineering and Materials Science.* Malabar, FL.: Krieger Publishing, 1993.

5. Navon, D. H. *Semiconductor Microdevices and Materials.* New York: Halt, Rinehart &Winston. 1986.

6. Pierret, R. F. *Semiconductor Device Fundamentals.* Reading. *MA:* Addison-Wesley, 1996.

7. Shur, M. *Introduction to Electronic Devices.* New York: John Wiley and Sons, 1996.

*8. Shur, M. *Physics of Semiconductor Devices.* Englewood Cliffs, NJ: Prentice Hall. 1990.

9. Singh, J. *Semiconductor Devices: An Introduction.* New York: McGraw-Hill, 1994.

10. Singh, J. *Semiconductor Devices: Basic Principles.* New York: John Wiley and Sons, 2001.

*11. Smith, R. A. *Semiconductors.* 2nd ed. New York; Cambridge University Press, 1978.

12. Streetman, B. G., and S. Banerjee. *Solid State Electronic Devices.* 5th ed. Upper Saddle River, NJ: Prentice Hall, 2000.

13. Sze, S. M. *Physics of Semiconductor Devices.* 2nd ed. New York: Wiley, 1981.

*14. Wang, S. *Fundamentals of Semiconductor Theory and Device Physics.* Englewood Cliffs, NJ: Prentice Hall, 1989.

*15. Wolfe, C. M., N. Holonyak, Jr, and G. E. Stillman. *Physical Properties of Semiconductors.* Englewood Cliffs, NJ: Prentice Hall. 1989.

16. Yang, E. S. *Microelectronic Devices.* New York: McGraw-Hill, 1988.

CHAPTER 5

Carrier Transport Phenomena

PREVIEW

In the previous chapter, we considered the semiconductor in equilibrium and determined electron and hole concentrations in the conduction and valence bands, respectively. A knowledge of the densities of these charged particles is important toward an understanding of the electrical properties of a semiconductor material. The net flow of the electrons and holes in a semiconductor will generate currents. The process by which these charged particles move is called ***transport.*** In this chapter we will consider the two basic transport mechanisms in a semiconductor crystal: drift—the movement of charge due to electric fields, and diffusion—the flow of charge due to density gradients. We should mention, in passing, that temperature gradients in a semiconductor can also lead to carrier movement. However, as the semiconductor device size becomes smaller, this effect can usually be ignored. The carrier transport phenomena are the foundation for finally determining the current-voltage characteristics of semiconductor devices. We will implicitly assume in this chapter that, though there will be a net flow of electrons and holes due to the transport processes, thermal equilihrium will not be substantially disturbed. Nonequilibrium processes will be considered in the next chapter. ■

5.1 | CARRIER DRIFT

An electric field applied to a semiconductor will produce a force on electrons and holes so that they will experience a net acceleration and net movement, provided there are available energy states in the conduction and valence bands. This net movement of charge due to an electric field is called *drift.* The net drift of charge gives rise to a *drift* ***current.***

5.1.1 Drift Current Density

If we have a positive volume charge density ρ moving at an average drift velocity v_d, the drift current density is given by

$$J_{drf} = \rho v_d \tag{5.1}$$

where J is in units of C/cm^2-s or amps/cm^2. If the volume charge density is due to positively charged holes, then

$$J_{p|drf} = (ep)v_{dp} \tag{5.2}$$

where $J_{p|drf}$ is the drift current density due to holes and v_{dp} is the average drift velocity of the holes.

The equation of motion of a positively charged hole in the presence of an electric field is

$$F = m_p^* a = e\mathrm{E} \tag{5.3}$$

where e is the magnitude of the electronic charge, a is the acceleration, E is the electric field, and m_p^* is the effective mass of the hole. If the electric field is constant, then we expect the velocity to increase linearly with time. However, charged particles in a semiconductor are involved in collisions with ionized impurity atoms and with thermally vibrating lattice atoms. These collisions, or scattering events, alter the velocity characteristics of the particle.

As the hole accelerates in a crystal due to the electric field, the velocity increases. When the charged particle collides with an atom in the crystal, for example, the panicle loses most, or all, of its energy. The particle will again begin to accelerate and gain energy until it is again involved in a scattering process. This continues over and over again. Throughout this process, the particle will gain an average drift velocity which, for low electric fields, is directly proportional to the electric field. We may then write

$$v_{dp} = \mu_p \mathrm{E} \tag{5.4}$$

where μ_p is the proportionality factor and is called the *hole mobility*. The mobility is an important parameter of the semiconductor since it describes how well a particle will move due to an electric tield. The unit of mobility is usually expressed in terms of cm^2/V-s.

By combining Equations *(5.2)* and (5.4). we may write the drift current density due to holes as

$$J_{p|drf} = (ep)v_{dp} = e\mu_p p\mathrm{E} \tag{5.5}$$

The drift current due to holes is in the same direction as the applied electric field.

The same discussion of drift applies to electrons. We may write

$$J_{n|drf} = \rho v_{dn} = (-en)v_{dn} \tag{5.6}$$

where $J_{n|drf}$ is the drift current density due to electrons and v_{dn} is the average drift velocity of electrons. The net charge density of electrons is negative.

Table 5.1 | Typical mobility values at $T = 300$ K and luw doping concentrations

	μ_n (cm²/V-s)	μ_p (cm²/V-s)
Silicon	1350	480
Gallium arsenide	8500	400
Germanium	3900	1900

The average drift velocity of an electron is also propostional to the electric fieh for small fields. However, since the electron is negatively charged, the net motiona the electron is opposite to the electric field direction. We can then write

$$\upsilon_{dn} = -\mu_n \mathrm{E} \tag{5}$$

where μ_n is the *electron mobility* and is a positive quantity. Equation (5.6) may not be written as

$$J_{n|drf} = (-en)(-\mu_n \mathrm{E}) = e\mu_n n\mathrm{E} \tag{5.8}$$

The conventional drift current due to electrons is also in the same direction as th applied electric field even though the electron movement is in the opposite direction

Electron and hole mobilities are functions of temperature and doping concentrations, as we will see in the next section. Table 5.1 shows some typical mobility values at T = 300 K for low doping concentrations.

Since both electrons and holes contribute to the drift current, the total *drift current density* is the sum of the individual electron and hole drift current densities, so we may write

$$\boxed{J_{drf} = e(\mu_n n + \mu_p p)\mathrm{E}} \tag{5.}$$

EXAMPLE 5.1 Objective

To calculate the drift current density in a semiconductor for a given electric field.

Consider a gallium arsenide sample at $\mathbf{7} = 300$ K with doping concentrations of $N_a =$ and $N_d = 10^{16}$ cm^{-3}. Assume complete ionization and assume electron and hole mobilitie given in Table 5.1. Calculate the drift current density if the appliedelecuic field is E = 10 V/cm

■ **Solution**

Since $N_d > N_a$, the semiconductor is *n* type and the majority carrier electron concentration from Chapter 4 is given by

$$n = \frac{N_d - N_a}{2} + \sqrt{\left(\frac{N_d - N_a}{2}\right)^2 + n_i^2} \approx 10^{16} \text{ cm}^{-3}$$

The minority carrier hole concentration is

$$p = \frac{n_i^2}{n} = \frac{(1.8 \times 10^6)^2}{10^{16}} = 3.24 \times 10^{-4} \text{ cm}^{-3}$$

For this extrinsic n-type semiconductor, the drift current density is

$$J_{drf} = e(\mu_n n + \mu_p p)\mathrm{E} \approx e\mu_n N_d \mathrm{E}$$

Then

$$J_{drf} = (1.6 \times 10^{-19})(8500)(10^{16})(10) = 136 \text{ A/cm}^2$$

Comment

Significant drift current densities can be obtained in a semiconductor applying relatively small electric fields. We may note from this example that the drift current will usually he due primarily to the majority carrier in an extrinsic semiconductor.

TEST YOUR UNDERSTANDING

E5.1 Consider a sample of silicon at T = 300 K doped at an impurity concentration of $N_d = 10^{15}$ cm^{-3} and $N_a = 10^{14}$ cm^{-3}. Assume electron and hole mobilities given in Table5.1. Calculate the drift current density if the applied electric field is E = 35 V/cm. (Ans. 6.80 A/cm^2)

E5.2 Adrift current density of $J_{drf} = 120$ A/cm^2 is required in a particular semiconductor device using p-type silicon with an applied electric field of E = 20 V/cm. Determine the required impurity doping concentration to achieve this specification. Assume electron and hole mobilities given in Table 5.1. (Ans. $N_d = 7.81 \times 10^{16}$ cm^{-3})

5.1.2 Mobility Effects

In the last section, we defined mobility, which relates the average drift velocity of a carrier to the electric field. Electron and hole mobilities are important semiconductor parameters in the characterization of carrier drift, as seen in Equation (5.9).

Equation (5.3)related the acceleration of a hole to a force such as an electric field. We may write this equation as

$$F = m_p^* \frac{dv}{dt} = e\mathrm{E} \tag{5.10}$$

where v is the velocity of the particle due to the electric field and does not include the random thermal velocity. If we assume that the effective mass and electric field are constants, then we may integrate Equation (5.10) and obtain

$$v = \frac{e\mathrm{E}t}{m_p^*} \tag{5.11}$$

where we have assumed the initial drift velocity to be zero.

Figure 5.1a shows a schematic model of the random thermal velocity and motion of a hole in a semiconductor with zero electric field. There is a mean time between collisions which may be denoted by τ_{mn}. If a small electric field (E-field) is

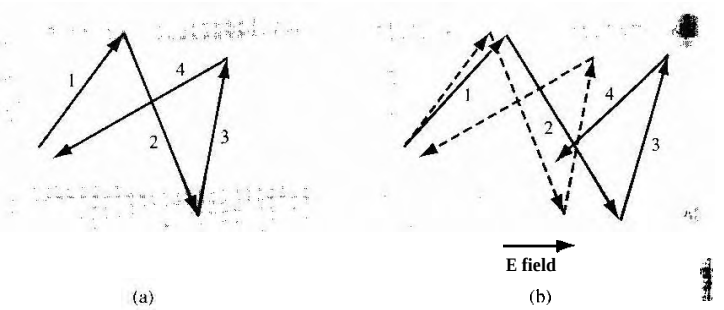


Figure 5.1 | Typical random behavior of a hole in a semiconductor (a) without an electric field and (b) with an electric field.

applied as indicated in Figure 5.1b, there will be a net drift of the hole in the directiw of the E-field, and the net drift velocity will be a small perturbation on the random thermal velocity, so the time between collisions will not be altered appreciably. If we use the mean time between collisions τ_{cp} in place of the time t in Equation (5.11), then the mean peak velocity just prior to a collision or scattering event is

$$v_{d|\text{peak}} = \left(\frac{e\tau_{cp}}{m_p^*}\right)\text{E} \tag{5.12a}$$

The average drift velocity is one half the **peak** value so that we can write

$$\langle v_d \rangle = \frac{1}{2}\left(\frac{e\tau_{cp}}{m_p^*}\right)\text{E} \tag{5.12b}$$

However, the collision process is not as simple as this model, but is statistical in nature. In a more accurate model including the effect of a statistical distribution, the factor $\frac{1}{2}$ in Equation (5.12b) does not appear. The hole mobility is then given by

$$\mu_p = \frac{v_{dp}}{\text{E}} = \frac{e\tau_{cp}}{m_p^*}$$

The same analysis applies to electrons; thus we can write the electron mobility as

$$\mu_n = \frac{e\tau_{cn}}{m_n^*} \tag{5.14}$$

where τ_{cn} is the mean time between collisions for an electron.

There are two collision or scattering mechanisms that dominate in a semiconductor and affect the carrier mobility: phonon or lattice scattering, and ionized impurity scattering.

The atoms in a semiconductor crystal have a certain amount of thermal energy at temperatures above absolute zero that causes the atoms to randomly vibrate about their lattice position within the crystal. The lattice vibrations cause a disruption in the

perfect periodic potential function. A perfect periodic potential in a solid allows electrons to move unimpeded, or with no scattering, through the crystal. But the thermal vibrations cause a disruption of the potential function, resulting in an interaction between the electrons or holes and the vibrating lattice atoms. This *lattice scattering* is also referred to as *phonon scattering.*

Since lattice scattering is related to the thermal motion of atoms, the rate at which the scattering occurs is a function of temperature. If we denote μ_L as the mobility that would be observed if only lattice scattering existed, then the scattering theory states that to first order

$$\mu_L \propto T^{-3/2} \tag{5.15}$$

Mobility that is due to lattice scattering increases as the temperature decreases. Intuitively, we expect the lattice vibrations to decrease as the temperature decreases, which implies that the probability of a scattering event also decreases, thus increasing mobility.

Figure 5.2 shows the temperature dependence of electron and hole mobilities in silicon. In lightly doped semiconductors, lattice scattering dominates and the carrier mobility decreases with temperature as we have discussed. The temperature dependence of mobility is proportional to T^{-n}. The inserts in the figure show that the parameter n is not equal to $\frac{3}{2}$ as the first-order scattering theory predicted. However, mobility does increase as the temperature decreases.

The second interaction mechanism affecting carrier mobility is called *ionized impurity scattering.* We have seen that impurity atoms are added to the semiconductor to control or alter its characteristics. These impurities are ionized at room temperature so that a coulomb interaction exists between the electrons or holes and the ionized impurities. This coulomb interaction produces scattering or collisions and also alters the velocity characteristics of the charge canier. If we denote μ_I as the mobility that would be observed if only ionized impurity scattering existed, then to first order we have

$$\mu_I \propto \frac{T^{+3/2}}{N_I} \tag{5.16}$$

where $N_I = N_d^+ + N_a^-$ is the total ionized impurity concentration in the semiconductor. If temperature increases, the random thermal velocity of a carrier increases, reducing the time the carrier spends in the vicinity of the ionized impurity center. The less timespent in the vicinity of a coulomb force, the smaller the scattering effect and the larger the expected value of μ_I. If the number of ionized impurity centers increases, then the probability of a carrier encountering an ionized impurity center increases, implying a smaller value of μ_I.

Figure 5.3 is a plot of electron and hole mobilities in germanium, silicon, and gallium arsenide at T = 300 K as a function of impurity concentration. More accurately, these curves are of mobility versus ionized impurity concentration N_I. As the impurity concentration increases, the number of impurity scattering centers increases, thus reducing mobility.

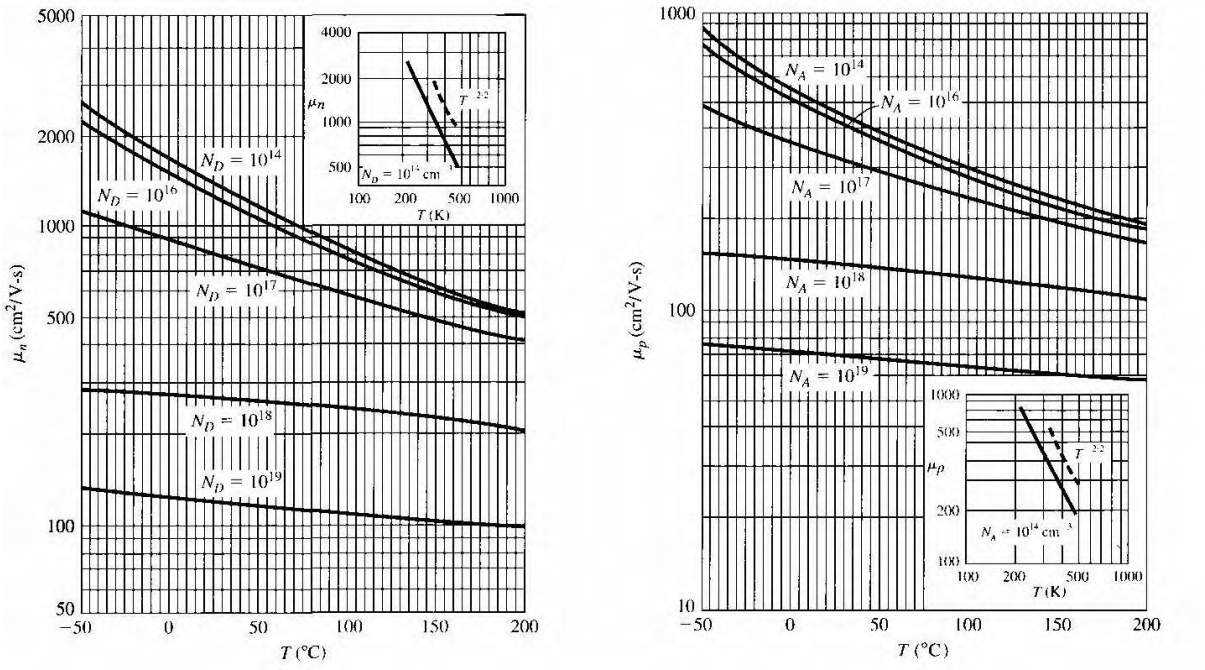


Figure 5.2 | (a) Electron and (b) hole mobilities in silicon versus temperature for various doping concentrations. Inserts show temperature d[illegible] for "almost" [illegible]

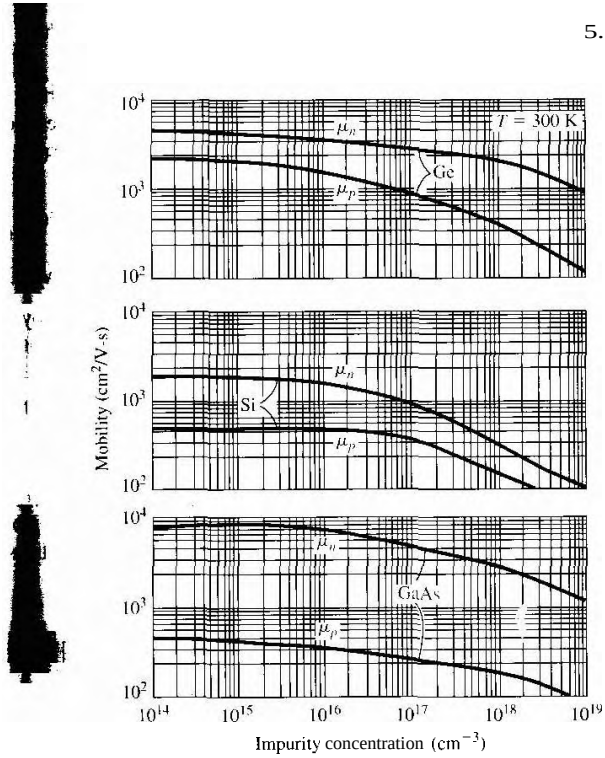


Figure 5.3 | Electron and hale mobilities versus impurity concentrations for germanium, silicon. and gallium arsenide at T = 300K.
(From Sze [12].)

TEST YOUR UNDERSTANDING

E5.3 ***(a)*** Using Figure 5.2, find the electron mobility for ***(i)*** $N_d = 10^{\cdot\cdot}\ cm^{-3}$, T = *150°C* and ***(ii)*** $N_d = 10^{16}\ cm^{-3}$, T = OC. (*b*) Find the hole mobilities for (*i*) $N_a = 10^{16}\ cm^{-3}$, T = 50°C; and (*ii*) $N_a = 10^{17}\ cm^{-3}$, T = 150°C.
[Ans. (*a*) (*i*) 500 cm^2/V-s, (*ii*) ~1500 cm^2/V-s; (*b*) (*i*) ~380 cm^2/V-s, (*ii*) ~200 cm^2/V-s]

E5.4 Using Figure 5.3, determine the electron and hole mobilities in ***(a)*** silicon for $N_d = 10^{15}\ cm^{-3}$, $N_a = 0$; (*b*) silicon for $N_d = 10^{17}\ cm^{-3}$, $N_a = 5 \times 10^{16}\ cm^{-3}$; (c) silicon for $N_d = 10^{16}\ cm^{-3}$, $N_a = 10^{17}\ cm^{-3}$; and (*d*) GaAs for $N_d = N_a = 10^{17}\ cm^{-3}$. [Ans. (*a*) $\mu_n \approx 1350$, $\mu_p = 480$; (*b*) $\mu_n \approx 700$, $\mu_p \approx 300$; (*c*) $\mu_n \approx 800$, $\mu_p \approx 310$; (*d*) $\mu_n \approx 4500$, $\mu_p \approx 220\ cm^2/V$-s]

If τ_L is the mean time between collisions due to lattice scattering, then dt/τ_L is the probability of a lattice scattering event occurring in a differential time dt. Likewise, if τ_I is the mean time between collisions due to ionized impurity scattering,

then dt/τ_I is the probability of an ionized impurity scattering event occurring in th differential time **dt.** If these two scattering processes are independent, then the tota probability of a scattering event occurring in the differential time **dt** is the sum ofth individual events, or

$$\frac{dt}{\tau} = \frac{dt}{\tau_I} + \frac{dt}{\tau_L} \tag{5.1}$$

where r is the mean time between any scattering event.

Comparing Equation (5.17) with the definitions of mobility given by Equation (5.13) or (5.14), we can write

$$\boxed{\frac{1}{\mu} = \frac{1}{\mu_I} + \frac{1}{\mu_L}}$$

where μ_I is the mobility due to the ionized impurity scattering process and μ_L is th mobility due to the lattice scattering process. The parameter μ is the net mobility With two or more independent scattering mechanisms, the inverse mobilities add which means that the net mobility decreases.

5.1.3 Conductivity

The drift current density, given by Equation (5.9), may be written as

$$J_{drf} = e(\mu_n n + \mu_p p)\mathrm{E} = \sigma \mathrm{E} \tag{5.19}$$

where σ is the **conductivity** of the semiconductor material. The conductivity is given in units of $(\Omega\text{-cm})^{-1}$ and is a function of the electron and hole concentrations and mobilities. We have just seen that the mobilities are functions of impurity concentration, conductivity, then is a somewhat complicated function of impurity concentration.

The reciprocal of conductivity is **resistivity,** which is denoted by ρ and is given in units of ohm-cm. We can write the formula for resistivity as

$$\boxed{\rho = \frac{1}{\sigma} = \frac{1}{e(\mu_n n + \mu_p p)}} \tag{5.20}$$

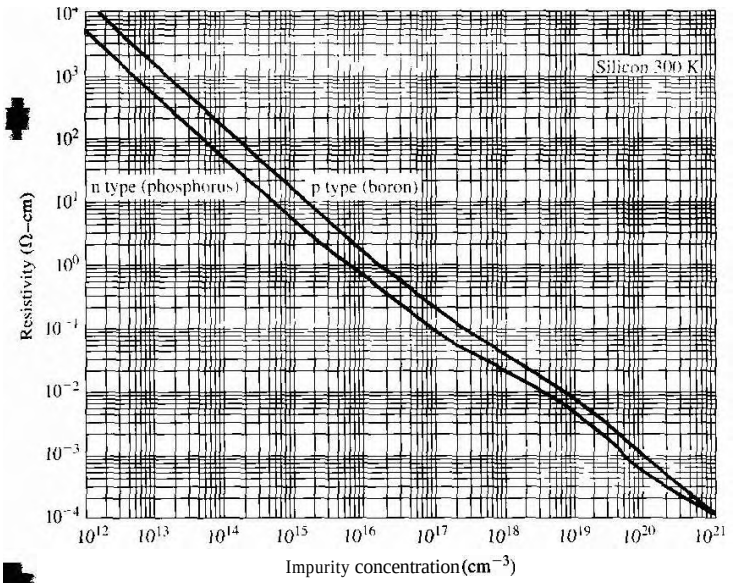
Figure 5.4 is a plot of resistivity as a function of impurity concentration in silicon germanium, gallium arsenide, and gallium phosphide at T = 300 K. Obviously, th curves are not linear functions of N_d or N_a because of mobility effects.

If we have a bar of semiconductor material as shown in Figure 5.5 with a voltage applied that produces a current I, then we can write

$$J = \frac{I}{\mathrm{A}} \tag{5.21a}$$

and

$$\mathrm{E} = \frac{V}{L} \tag{5.21b}$$



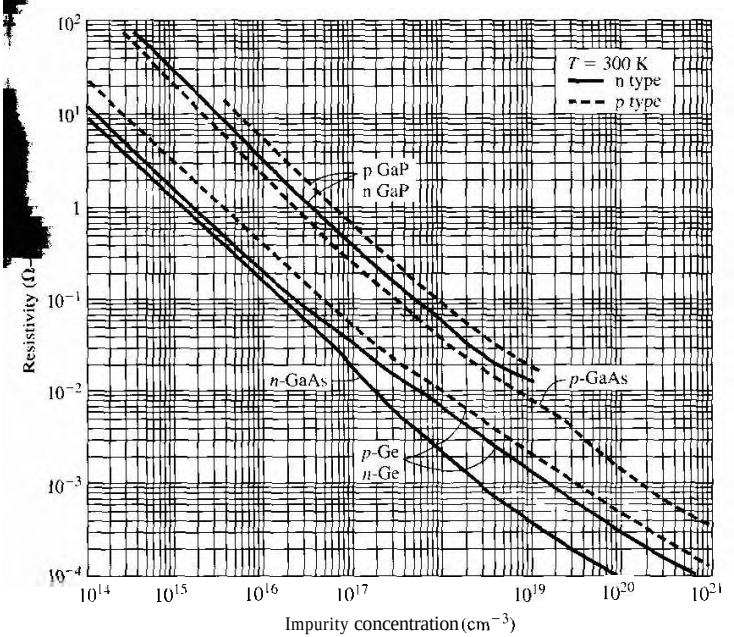


Figure 5.4 | Resistivity versus impurity concentration at T = 300 K in (a) silicon and (b) germanium, gallium arsenide, and gallium phosphide.
(From Sze [12].)

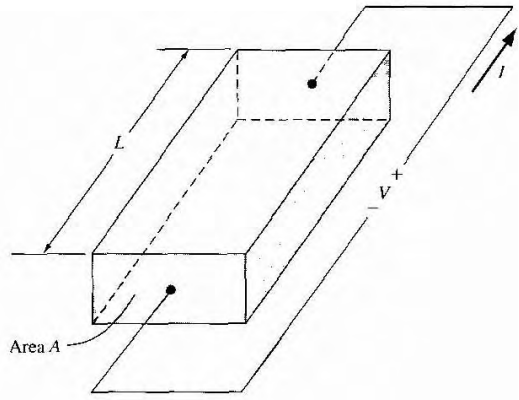


Figure 5.5 | Bar of semiconductor material as a resistor.

We can now rewrite Equation (5.19) as

$$\frac{I}{A} = \sigma\left(\frac{V}{L}\right) \tag{5.22}$$

or

$$V = \left(\frac{L}{\sigma A}\right) I = \left(\frac{\rho L}{A}\right) I = IR \tag{5.22}$$

Equation (5.22b) is Ohm's law for a semiconductor. The resistance is a function of resistivity, or conductivity, as well as the geometry of the semiconductor.

If we consider. for example, a p-type semiconductor with an acceptor doping $N_a (N_d = 0)$ in which $N_a \gg n_i$, and if we assume that the electron and hole mobilities are of the same order of magnitude, then the conductivity becomes

$$\sigma = e(\mu_n n + \mu_p p) \approx e\mu_p p \tag{5.23}$$

If we also assume complete ionization, then Equation (5.23) becomes

$$\sigma \approx e\mu_p N_a \approx \frac{1}{\rho} \tag{5.24}$$

The conductivity and resistivity of an extrinsic semiconductor are a function primarily of the majority carrier parameters.

We may plot the carrier concentration and conductivity of a semiconductor as a function of temperature for a particular doping concentration. Figure 5.6 shows the electron concentration and conductivity of silicon as a function of inverse temperature for the case when $N_d = 10^{15}$ cm^{-3}. In the midtemperature range, or extrinsic range, as shown, we have complete ionization—the electron concentration remains essentially constant. However, the mobility is a function of temperature so the conductivity

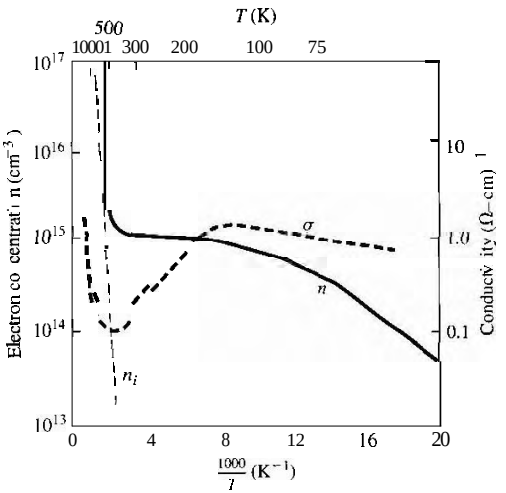


Figure 5.6 | Electron concentration and conductivity versus inverse temperature for silicon.
(After Sze [12].)

varies with temperature in this range. At higher temperatures, the intrinsic carrier concentration increases and begins to dominate the electron concentration as well as the conductivity. In the lower temperature range, freere-out begins to occur; the electron concentration and conductivity decrease with decreasing temperature.

Objective | EXAMPLE **5.2**

To determine the doping concentration and majority carrier mobility given the type and conductivity of a cnmpensated semiconductor

Consider compensated n-type silicon at T = 300 K, with a conductivity of $\sigma = 16\,(\Omega\text{-cm})^{-1}$ and an acceptor doping concentration of 10^{17} cm^{-3}. Determine the donor concentration and the electron mobility.

■ Solution

For n-type silicon at $T = 300$ K, we can assume complete ionization; therefore the conductivity, assuming $N_d - N_a \gg$ n, , is given hy

$$\sigma \approx e\mu_n n = e\mu_n (N_d - N_a)$$

We have that

$$16 = (1.6 \times 10^{-19})\mu_n (N_d - 10^{17})$$

Since mobility is a function of the ionized impurity cnncentration. we can use Figure 5.3 along with trial and error to determine μ_n and N_d. For example, if we choose $N_d = 2 \times 10^{17}$, then

$N_i = N_d^+ + N_a^- = 3 \times 10^{17}$ so that $\mu_n \approx 510$ cm^2/V-s which gives $\sigma = 8.16$ $(\Omega\text{-cm})^{-1}$. If we choose $N_d = 5 \times 10^{17}$, then $N_I = \text{h} \times 10^{17}$ so that $\mu_n \approx 325$ cm^2/V-s, which gives $\sigma = 20.8\ (\Omega\text{-cm})^{-1}$. The doping is bounded between these two values. Funhcr trial and error yields

$$N_d \approx 3.5 \times 10^{17}\ \text{cm}^{-3}$$

and

$$\mu_n \approx 400\ \text{cm}^2/\text{V-s}$$

which gives

$$\sigma \approx 16\ (\Omega\text{-cm})^{-1}$$

■ Comment

We can see from this example that, in high-conductivity semicunductor material, mobility is a strong function of carrier concentration.

DESIGN EXAMPLE 5.3



Objective

To design a semiconductor resistor with a specified resistance to handle a given current density.

A silicon semiconductor at T = 300 K is initially doped with donors at a concentration of $N_d = 5 \times 10^{15}$ cm^{-3}. Acceptors are to be added to form a compensated p-type material. The resistor is to have a resistance of 10 kΩ and handle a current density of 50 A/cm^2 when 5 V is applied.

Solution

For 5 V applied to a 10-kΩ resistor, the total current is

$$I = \frac{V}{R} = \frac{5}{10} = 0.5\ \text{mA}$$

If the current density is limited to 50 A/cm^2, then the cross-sectional area is

$$A = \frac{I}{J} = \frac{0.5 \times 10^{-3}}{50} = 10^{-5}\ \text{cm}^2$$

If we, somewhat arbitrarily at this point, limit the electric field to E = 100 V/cm, then the length of the resistor is

$$L = \frac{V}{\text{E}} = \frac{5}{100} = 5 \times 10^{-2}\ \text{cm}$$

From Equation (5.22b), the conductivity of the semiconductor is

$$\sigma = \frac{L}{RA} = \frac{5 \times 10^{-2}}{(10^4)(10^{-5})} = 0.50\ (\Omega\text{-cm})^{-1}$$

The conductivity of a compensated p-type semiconductor is

$$\sigma \approx e\mu_p p = e\mu_p (N_a - N_d)$$

where the mobility is a function of the total ionized impurity concentration $N_a + N_d$.

Using trial and error, if $N_a = 1.25 \times 10^{16}$ cm^{-3}, then $N_a + N_d = 1.75 \times 10^{16}$ cm^{-3}, and the hole mobility, from Figure 5.3, is approximately $\mu_p = 410$ cm^2/V-s. The conductivity is then

$$\sigma = e\mu_p(N_a - N_d) = (1.6 \times 10^{-19})(410)(1.25 \times 10^{16} - 5 \times 10^{15}) = 0.492$$

which is very close to the value we need

■ Comment

Since the mobility is related to the total ionized impurity concentration, the determination of the impurity concentration to achieve a particular conductivity is not straightforward.

TEST YOUR UNDERSTANDING

E5.5 Silicon at T = 300 K is doped with impurity concentrations of $N_d = 5 \times 10^{16}$ cm^{-3} and $N_a = 2 \times 10^{16}$ cm^{-3}. (a) What are the electron and hole mobilities? (*b*) Determine the conductivity and resistivity of the material. [Ans. (*a*) $\mu_n = 1000$ cm^2/V-s, $\mu_p = 350$ cm^2/V-s; (*b*) $\sigma = 4.8$ $(\Omega\text{-cm})^{-1}$, $\rho = 0.208$ Ω-cm]

E5.6 For a particular silicon semiconductor device at T = 300 K, the required material is n type with a resistivity of 0.10 Ω-cm. (*a*) Determine the required impurity doping concentration and (b) the resulting electron mobility. [Ans. (*a*) From Figure 5.4, $N_d \approx 9 \times 10^{16}$ cm^{-3}, (*b*) $\mu_n \approx 695$ cm^2/V-s]

E5.7 A bar of p-type silicon, such as shown in Figure 5.5, has a cross-sectional area of $A = 10^{-6}$ cm^2 and a length of L = 1.2×10^{-3} cm. For an applied voltage of 5 V, a current of 2 mA is required. What is the required (*a*) resistance, (*b*) resistivity of the silicon, and (*c*) impurity doping concentration? [Ans. (*a*) 2.5 kΩ, (*b*) 2.08 Ω-cm, (*c*) $N_a \approx 7 \times 10^{15}$ cm^{-3}]

For an intrinsic material, the conductivity can be written as

$$\sigma_i = e(\mu_n + \mu_p)n_i \tag{5.25}$$

The concentrations of electrons and holes are equal in an intrinsic semiconductor, so the intrinsic conductivity includes both the electron and hole mohility. Since, in general, the electron and hole mobilities are not equal, the intrinsic conductivity is not the minimum value possible at a given temperature.

5.1.4 Velocity Saturation

So far in our discussion of drift velocity, we have assumed that mobility is not a function of electric field, meaning that the drift velocity will increase linearly with applied electric field. The total velocity of a particle is the sum of the random thermal velocity and drift velocity. At T = 300 K, the average random thermal energy is given by

$$\tfrac{1}{2}mv_{th}^2 = \tfrac{3}{2}kT = \tfrac{3}{2}(0.0259) = 0.03885 \text{ eV} \tag{5.26}$$

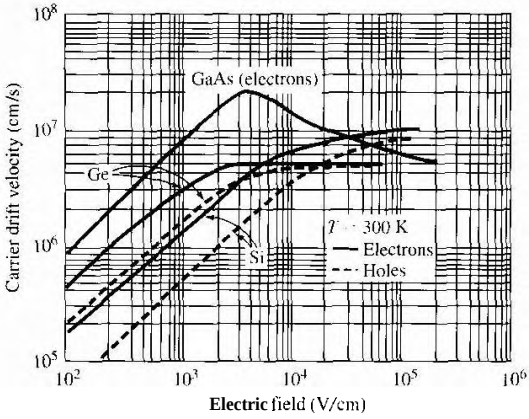


Figure 5.7 | Carrier drift velocity versus electric field for high-purity silicon. germanium, and gallium arsenide. *(From Sze [12].)*

This energy translates into a mean thermal velocity of approximately 10^7 cm/s for an electron in silicon. If we assume an electron mobility of $\mu_n = 1350$ cm^2/V-s in low-doped silicon, a drift velocity of 10^5 cm/s, or 1 percent of the thermal velocity, is achieved if the applied electric field is approximately 75 V/cm. This applied electric field does not appreciably alter the energy of the electron.

Figure 5.7 is a plot of average drift velocity as a function of applied electric field for electrons and holes in silicon, gallium arsenide, and germanium. At low electric fields, where there is a linear variation of velocity with electric field, the slope of the drift velocity versus electric field curve is the mobility. The behavior of the drift velocity of carriers at high electric fields deviates substantially from the linear relationship observed at low fields. The drift velocity of electrons in silicon, for example. saturates at approximately 10^7 cm/s at an electric field of approximately 30 kV/cm. If the drift velocity of a charge carrier saturates, then the drift current density also saturates and becomes independent of the applied electric field.

The drift velocity versus electric field characteristic of gallium arsenide is more complicated than for silicon or germanium. At low fields, the slope of the drift velocity versus E-field is constant and is the low-field electron mobility, which is approximately 8500 cm^2/V-s for gallium arsenide. The low-field electron mobility in gallium arsenide is much larger than in silicon. As the field increases, the electron drift velocity in gallium arsenide reaches a peak and then decreases. A differential mobility is the slope of the v_d versus E curve at a particular point on the curve and the negative slope of the drift velocity versus electric field represents a negative differential mobility. The negative differential mobility produces a negative differential resistance; this characteristic is used in the design of oscillators.

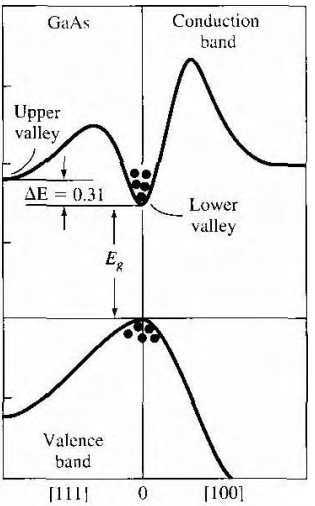


Figure 5.8 | Energy-hand structure for gallium arsenide showing the upper valley and lower valley in the conduction band. *(From Sze [13].)*

The negative differential mobility can be understood by considering the E versus k diagram for gallium arsenide, which is shown again in Figure 5.8. The density of states effective mass of the electron in the lower valley is $m_n^* = 0.067m_0$. The small effective mass leads to a large mobility. As the E-field increases. the energy of the electron increases and the electron can be scattered into the upper valley, where the density of states effective mass is $0.55m_0$. The larger effective mass in the upper valley yields a smaller mobility. This intervalley transfer mechanism results in a decreasing average drift velocity of electrons with electric field, or the negative differential mobility characteristic.

5.2 | CARRIER DIFFUSION

There is a second mechanism, in addition to drift, that can induce a current in a semiconductor. We may consider a classic physics example in which a container, as shown in Figure 5.9, is divided into two parts by a membrane. The left side contains gas molecules at a particular temperature and the right side is initially empty. The gas molecules are in continual random thermal motion so that, when the membrane is broken, the gas molecules How into the right side of the container. *Diffusion* is the process whereby panicles flow from a region of high concentration toward a region of low

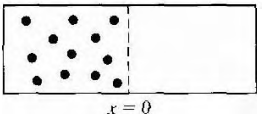


Figure 5.9 | Container divided by a membrane with gas molecules on one side.

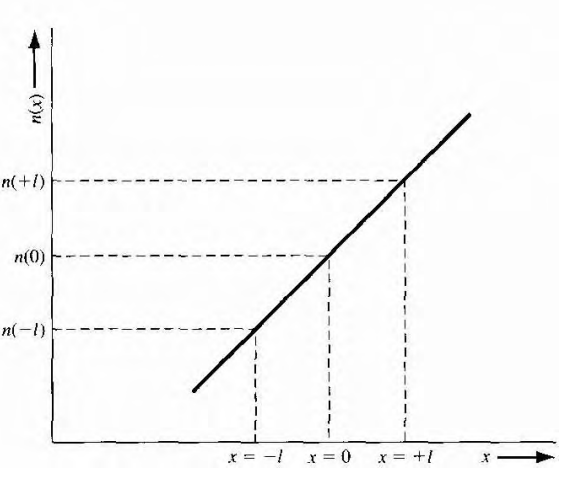


Figure 5.10 | Electron concentration versus distance.

concentration. If the gas molecules were electrically charged, the net flow of cha would result in a *diffusion current.*

5.2.1 Diffusion Current Density

To begin to understand the diffusion process in a semiconductor, we will consider simplified analysis. Assume that an electron concentration varies in one dimensio shown in Figure 5.10. The temperature is assumed to be uniform so that the averag thermal velocity of electrons is independent of x. To calculate the current, we will de termine the net Row of electrons per unit time per unit area crossing the planed $x = 0$. If the distance l shown in Figure 5.10 is the mean-free path of an electron, tha is, the average distance an electron travels between collisions ($l = v_{th}\tau_{cn}$), then o the average, electrons moving to the right at $x = -l$ and electrons moving to the le at $x = +l$ will cross the $x = 0$ plane. One half of the electrons at $x = -l$ will be trav eling to the right at any instant of time and one half of the electrons at $x = +l$ will b traveling to the left at any given time. The net rate of electron flow, $F_{,,}$, in the +

direction at $x = 0$ is given by

$$F_n = \tfrac{1}{2}n(-l)\upsilon_{th} - \tfrac{1}{2}n(+l)\upsilon_{th} = \tfrac{1}{2}\upsilon_{th}[n(-l) - n(+l)] \tag{5.27}$$

If we expand the electron concentration in a Taylor series about $x = 0$ keeping only the first two terms, then we can write Equation (5.27) as

$$F_n = \frac{1}{2}\upsilon_{th}\left\{\left[n(0) - l\frac{dn}{dx}\right] - \left[n(0) + l\frac{dn}{dx}\right]\right\} \tag{5.28}$$

which becomes

$$F_n = -\upsilon_{th} l \frac{\mathrm{dn}}{dx} \tag{5.29}$$

Each electron has a charge $(-e)$, so the current is

$$J = -eF_n = +e\upsilon_{th} l \frac{dn}{dx} \tag{5.30}$$

The current described by Equation *(5.30)* is the electron diffusion current and is proportional to the spatial derivative, or density gradient, of the electron concentration.

The diffusion of electrons from a region of high concentration to a region of low concentration produces a flux of electrons flowing in the negative x direction for this example. Since electrons have a negative charge, the conventional current direction is in the positive ***x*** direction. Figure 5.11a shows these one-dimensional flux and current directions. We may write the electron diffusion current density for this one-dimensional case. in the form

$$\boxed{J_{nx|dif} = eD_n \frac{dn}{dx}} \tag{5.31}$$

where D_n is called the ***electron*** *diffusion coefficient,* has units of cm^2/s, and is a positive quantity. If the electron density gradient becomes negative, the electron diffusion current density will be in the negative ***x*** direction.

Figure 5.11b shows an example of a hole concentration as a function of distance in a semiconductor. The diffusion of holes, from a region of high concentration to a region of low concentration, produces a flux of holes in the negative x direction. Since holes are positively charged particles, the conventional diffusion current density is also in the negative ***x*** direction. The hole diffusion current density is proportional to the hole density gradient and to the electronic charge, so we may write

$$\boxed{J_{px|dif} = -eD_p \frac{dp}{dx}}$$

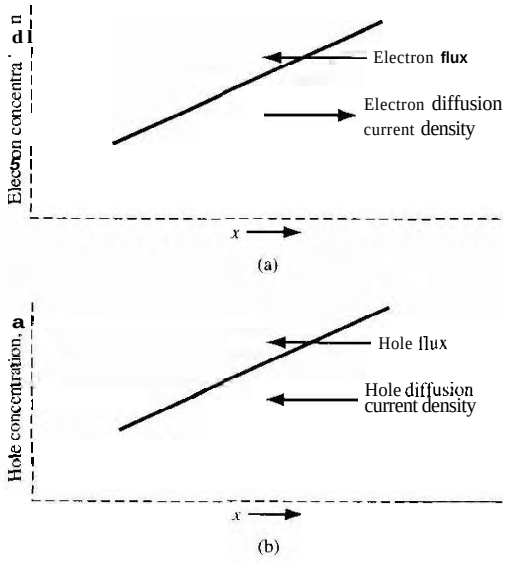


Figure 5.11 | (a) Diffusion of electrons due to a density gradient. (b) Diffusion of holes due to a density gradient.

for the one-dimensional case. The parameter D_p is called the ***hole*** *diffusion coefficient*, has units of cm^2/s, and is a positive quantity. If the hole density gradient comes negative, the hole diffusion current density will be in the positive x directi

EXAMPLE 5.4

Objective

To calculate the diffusion current density given a density gradient.

Assume that. in an n-type gallium arsenide semiconductor at T = 300 K, the electr concentration varies linearly from 1 x 10^{18} to 7 x 10^{17} cm^{-3} over a distance of 0.10 cm Ca culate the diffusion current density if the electron diffusion coefficient is $D_n = 225$ cm^2/s

■ Solution

The diffusion current density is given by

$$J_{n|dif} = eD_n \frac{dn}{dx} \approx eD_n \frac{\Delta n}{\Delta x}$$

$$= (1.6 \times 10^{-19})(225)\left(\frac{1 \times 10^{18} - 7 \times 10^{17}}{0.10}\right) = 108 \text{ A/cm}^2$$

■ Comment

A significant diffusion current density can be generated in a semiconductor material with o a modest density gradient.

TEST YOUR UNDERSTANDING

E5.8 The electron concentration in silicon is given by $n(x) = 10^{15}\, e^{-(x/L_n)}\ \text{cm}^{-3}$ $(x \geq 0)$ where $L_n = 10^{-4}$ cm. The electron diffusion coefficient is $D_n = 25\ \text{cm}^2/\text{s}$. Determine the electron diffusion current density at (a) $x = 0$, (b) $x = 10^{-4}$ cm, and (c) $x \to \infty$.

[Ans. (a) $-40\ \text{A/cm}^2$, (b) $-14.7\ \text{A/cm}^2$, (c) 0]

E5.9 The hole concentration in silicon varies linearly from $x = 0$ to $x = 0.01$ cm. The hole diffusion coefficient is $D_p = 10\ \text{cm}^2/\text{s}$, the hole diffusion current density is $20\ \text{A/cm}^2$, and the hole concentration at $x = 0$ is $p = 4 \times 10^{17}\ \text{cm}^{-3}$. What is the value of the hole concentration at $x = 0.01$ cm?

(Ans. $10^{15}\ \text{cm}^{-3}$)

E5.10 The hole concentration in silicon is given by $p(x) = 2 \times 10^{15} e^{-(x/L_p)}\ \text{cm}^{-3}$ $(x \geq 0)$. The hole diffusion coefficient is $D_p = 10\ \text{cm}^2/\text{s}$. The value of the diffusion current density at $x = 0$ is $J_{\text{dif}} = +6.4\ \text{A/cm}^2$. What is the value of L_p?

(Ans. $L_p = 5 \times 10^{-4}$ cm)

5.2.2 Total Current Density

We now have four possible independent current mechanisms in a semiconductor. These components are electron drift and diffusion currents and hole drift and diffusion currents. The total current density is the sum of these four components, or, for the one-dimensional case,

$$J = en\mu_n \mathcal{E}_x + ep\mu_p \mathcal{E}_x + eD_n \frac{dn}{dx} - eD_p \frac{dp}{dx} \tag{5.33}$$

This equation may be generalized to three dimensions as

$$J = en\mu_n \mathcal{E} + ep\mu_p \mathcal{E} + eD_n \nabla n - eD_p \nabla p \tag{5.34}$$

The electron mobility gives an indication of how well an electron moves in a semiconductor as a result of the force of an electric field. The electron diffusion coefficient gives an indication of how well an electron moves in a semiconductor as a result of a density gradient. The electron mobility and diffusion coefficient are not independent parameters. Similarly, the hole mobility and diffusion coefficient are not independent parameters. The relationship between mobility and the diffusion coefficient will be developed in the next section.

The expression for the total current in a semiconductor contains four terms. Fortunately in most situations, we will only need to consider one term at any one time at a particular point in a semiconductor.

5.3 | GRADED IMPURITY DISTRIBUTION

In most cases so far, we have assumed that the semiconductor is uniformly doped. In many semiconductor devices, however, there may be regions that are nonuniformly doped. We will investigate how a nonuniformly doped semiconductor reaches thermal

equilihrium and, from this analysis, we will derive the Einstein relation, which re
mobility and the diffusion coefficient.

5.3.1 Induced Electric Field

Consider a semiconductor that is nonunifnrmly doped with donor impurity atoms. I
the semiconductor is in thermal equilibrium, the Fermi energy level is co
through the crystal so the energy-band diagram may qualitatively look like th
shown in Figure 5.12. The doping concentration decreases as x increases in this case
There will be a diffusion of majority carrier electrons from the region of high con
centration to the region of low concentration, which is in the $+x$ direction. The flo
of negative electrons leaves behind positively charged donor ions. The separation o
positive and negative charge induces an electric field that is in a direction to oppos
the diffusion process. When equilibrium is reached, the mobile carrier concentratio
is not exactly equal to the fixed impurity concentration and the induced electric fiel
prevents any further separation of charge. In most cases of interest, the space charg
induced by this diffusion process is a small fraction of the impurity concentration
thus the mobile carrier concentration is not too different from the impurity dopa
density.

The electric potential ϕ is related to electron potential energy by the char
$(-e)$, so we can write

$$\phi = +\frac{1}{e}(E_F - E_{Fi}) \tag{5}$$

The electric field for the one-dimensional situation is defined as

$$\mathrm{E}_x = -\frac{d\phi}{dx} = \frac{1}{e}\frac{dE_{Fi}}{dx} \tag{5}$$

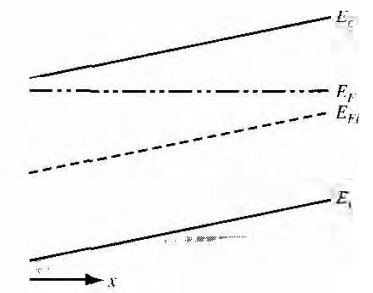


Figure 5.12 | Energy-band diagram for a semiconductor in thermal equilibrium with a nonuniform **donor** impurity concentration

If the intrinsic Fermi level changes as a function of distance through a semiconductor in thermal equilibrium, an electric field exists in the semiconductor.

If we assume a quasi-neutrality condition in which the electron concentration is almost equal to the donor impurity concentration, then we can still write

$$n_0 = n_i \exp\left[\frac{E_F - E_{Fi}}{kT}\right] \approx N_d(x) \tag{5.37}$$

Solving for $E_F - E_{Fi}$, we obtain

$$E_F - E_{Fi} = kT \ln\left(\frac{N_d(x)}{n_i}\right) \tag{5.38}$$

The Fermi level is constant for thermal equilibrium so when we take the derivative with respect to x we obtain

$$-\frac{dE_{Fi}}{dx} = \frac{kT}{N_d(x)}\frac{dN_d(x)}{dx} \tag{5.39}$$

Theelectric field can then be written, combining Equations (5.39) and (5.36), as

$$\mathrm{E}_x = -\left(\frac{kT}{e}\right)\frac{1}{N_d(x)}\frac{dN_d(x)}{dx} \tag{5.40}$$

Since we have an electric field, there will be a potential difference through the semiconductor due to the nonuniform doping.

Objective | **EXAMPLE 5.5**

To determine the induced electric field in a semiconductor in thermal equilibrium, given a linear variation in doping concentration.

Assume that the donor concentration in an n-type semiconductorat T = 300 K is given by

$$N_d(x) = 10^{16} - 10^{19}x \qquad (\mathrm{cm}^{-3})$$

where x is given in cm and ranges between $0 \le x \le 1\ \mu\mathrm{m}$

■ Solution

Taking the derivative of the donor concentration, we have

$$\frac{dN_d(x)}{dx} = -10^{19} \qquad (\mathrm{cm}^{-4})$$

Theelectric field is given hy Equation (5.40), so we have

$$\mathrm{E}_x = \frac{-(0.0259)(-10^{19})}{(10^{16} - 10^{19}x)}$$

Atx = 0, for example, wc find

$$\mathrm{E}_x = 25.9\ \mathrm{V/cm}$$

■ **Comment**

We may recall from our previous discussion of drift current that fairly small electric fields c produce significant drift current densities, so that an induced electric field from nonuni o doping can significantly influence semiconductor device characteristics.

5.3.2 The Einstein Relation

If we consider the nonuniformly doped semiconductor represented by the ener band diagram shown in Figure 5.12 and assume there are no electrical connections that the semiconductor is in thermal equilibrium, then the individual electron a hole currents must be zero. We can write

$$J_n = 0 = en\mu_n \mathrm{E}_x + eD_n \frac{dn}{dx} \tag{5.4}$$

If we assume quasi-neutrality so that $n \approx N_d(x)$, then we can rewrite Equ tion (5.41) as

$$J_n = 0 = e\mu_n N_d(x)\mathrm{E}_x + eD_n \frac{dN_d(x)}{dx} \tag{5.4}$$

Substituting the expression for the electric field from Equation (5.40) into Equ tion (5.42). we obtain

$$0 = -e\mu_n N_d(x)\left(\frac{kT}{e}\right)\frac{1}{N_d(x)}\frac{dN_d(x)}{dx} + eD_n \frac{dN_d(x)}{dx} \tag{5.}$$

Equation (5.43) is valid for the condition

$$\frac{D_n}{\mu_n} = \frac{kT}{e} \tag{5.4}$$

The hole current must also be zero in the semiconductor. From this conditi we can show that

$$\frac{D_p}{\mu_p} = \frac{kT}{e} \tag{5.4}$$

Combining Equations (5.44a) and (5.44b) gives

$$\boxed{\frac{D_n}{\mu_n} = \frac{D_p}{\mu_p} = \frac{kT}{e}} \tag{5.}$$

The diffusion coefficient and mobility are not independent parameters. This relat between the mobility and diffusion coefficient, given by Equation (5.45), is known the *Einstein relation.*

Table 5.2 | Typical mobility and diffusion coefficient values at T = 300 K ($\mu = cm^2/V\text{-}s$ and $D = cm^2/s$)

	μ_m	D_n	a	D_p
Silicon	1350	35	480	12.4
Gallium arsenide	8500	220	400	10.4
Germanium	3900	101	1900	49.2

Objective

EXAMPLE 5.6

To determine the diffusion coefficient given the carrier mobility. Assume that the mobility of a particular carrier is 1000 $cm^2/V\text{-}s$ at $T = 300$ K.

■ **Solution**

Using the Einstein relation. we have that

$$D = \left(\frac{kT}{e}\right)\mu = (0.0259)(1000) = 25.9\ cm^2/s$$

■ **Comment**

Although this example is fairly simple and straightforward. it is important to keep in mind the relative orders of magnitude of the mobility and diffusion coefficient. The diffurioncoefficient is approximately 40 times smaller than the mobility at room temperature.

Table 5.2 shows the diffusion coefficient values at $T = 300$ K corresponding to the mobilities listed in Table 5.1 tor silicon. gallium arsenide. and germanium.

The relation between the mobility and diffusion coefficient given by Equation (5.45) contains temperature. It is important to keep in mind that the major temperature effects are a result of lattice scattering and ionized impurity scattering processes, as discussed in Section 5.1.2. As the mobilities are strong functions of temperature because of the scattering processes, the diffusion coefficients are also strong functions of temperature. The specific temperature dependence given in Equation (5.45) is a small fraction of the real temperature characteristic.

*5.4 | THE HALL EFFECT

The Hall effect is a consequence of the forces that are exerted on moving charges by electric and magnetic fields. The Hall effect is used to distinguish whether a semiconductor is n type or p type[1] and to measure the majority carrier concentration and tnajority carrier mobility. The Hall effect device, as discussed in this section, is used to experimentally measure semiconductor parameters. However, it is also used extensively in engineering applications as a magnetic probe and in other circuit applications.

[1]We will assume an extrinsic semiconductor material in which the majority carrier concentration is much larger than the minority carrier concentration.

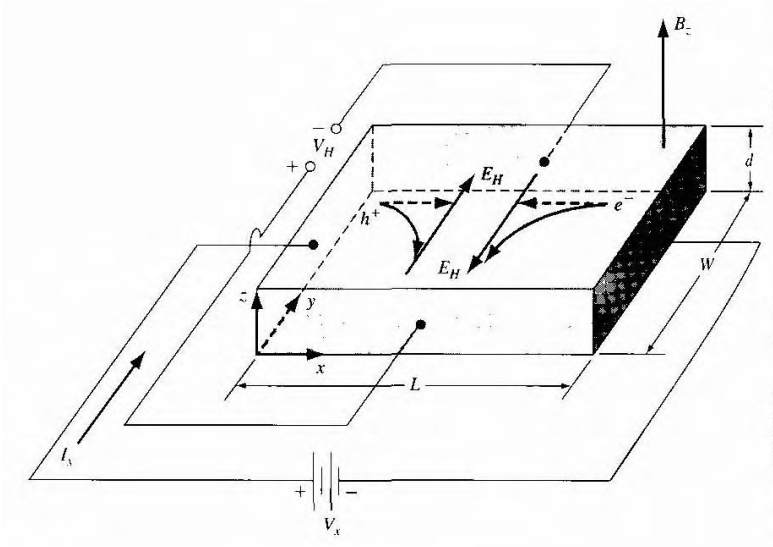


Figure 5.13 | *Geometry for* measuring the Hall *effect*

The force on a particle having a charge q and moving in a magnetic field is given by

$$F = q\upsilon \times B \tag{5.46}$$

where the cross product is taken between velocity and magnetic field so that the force vector is perpendicular to both the velocity and magnetic field.

Figure 5.13 illustrates the Hall effect. A semiconductor with a current I_x placed in a magnetic field perpendicular to the current. In this case, the magnetic field is in the z direction. Electrons and holes flowing in the semiconductor will experience a force as indicated in the figure. The force on both electrons and holes is in the $(-y)$ direction. In a p-type semiconductor ($p_0 > n_0$), there will be a buildup of positive charge on the $y = 0$ surface of the semiconductor and. in an n-type semiconductor ($n_0 > p_0$), there will be a buildup of negative charge on the $y = 0$ surface. This net charge induces an electric field in the y-direction as shown in the figure. In steady state, the magnetic field force will be exactly balanced by the induced electric field force. This balance may be written as

$$F = q[\mathrm{E} + \upsilon \times B] = 0 \tag{5.47a}$$

which becomes

$$q\mathrm{E}_y = q\upsilon_x B_z \tag{5.47b}$$

The induced electric field in the y-direction is called the *Hall field.* The Hall field produces a voltage across the semiconductor which is called the *Hall voltage.* We can write

$$V_H = +\mathrm{E}_H \mathrm{W} \tag{5.48}$$

The hole rnobility is then given by

$$\mu_p = \frac{I_x L}{epV_x Wd} \quad (5$$

Similarly for an n-type semiconductor, the low-field electron mobility is determi from

$$\mu_n = \frac{I_x L}{enV_x Wd}$$

EXAMPLE 5.7

Objective

To determine the majority carrier concentration and mobility. given Hall effect parameter

Consider the geometry shown in Figure 5.13. Let L = 10^{-1} cm, W = 10^{-2} cm, d = 10^{-3} cm. Also assume that I, = 1.0 mA, V_x = 12.5 V, B_z = 500 gauss = 5 x 10^{-2} t and V_H = −6.25 mV.

■ **Solution**

A negative Hall voltage for this geometry implies that we have an n-type semicondu Using Equation (5.54). we can calculate the electron concentration as

$$n = \frac{-(10^{-3})(5 \times 10^{-2})}{(1.6 \times 10^{-19})(10^{-5})(-6.25 \times 10^{-3})} = 5 \times 10^{21}\ \text{m}^{-3} = 5 \times 10^{15}\ \text{cm}^{-3}$$

The electron mobility is then determined from Equation (5.58) as

$$\mu_n = \frac{(10^{-3})(10^{-3})}{(1.6 \times 10^{-19})(5 \times 10^{21})(12.5)(10^{-4})(10^{-5})} = 0.10\ \text{m}^2/\text{V-s}$$

or

$$\mu_n = 1000\ \text{cm}^2/\text{V-s}$$

■ **Comment**

It is important to note that the MKS units must be used consistently in the Hall effect equat to yield correct results.

5.5 | SUMMARY

- The two basic transport mechanisms are drift, due to an applied electric field. and diffusion. due to a density gradient.
  Carriers reach an average drift velocity in the presence of an applied electric field, d to scattering events. Two scattering proccsscs within a semiconductor are lattice scattering and impurity scattering.
- The average drift velocity is a linear function of the applied electric field for small values of electric field, hut the drift velocity reaches a saturation limit that is on the order of 10^7 cm/s at high electric fields.

- Carrier mobility is the ratio of the average drift velocity and applied elechic field. The electron and hole mobilities are functions of temperature and of the ionized impurity concentration.
- The drift current density is the product of conductivity and electric field (a form of Ohm's law). Conductivity is a function of the carrier concentrations and mobilities. Resistivity is the inverse of conductivity
- The diffusion current density is proportional to the carrier diffusion coefficient and the cartier density gradient.
- The diffusion coefficient and mohilitv are related through the Einstein relation.
- The Hall effect is a consequence of a charged carrier moving in the presence of perpendicular electric and magnetic fields. The charged carrier is deflected. inducing a Hall voltage. The polarity of the Hall voltage is a function of the semiconductor conductivity type. The majority carrier concentration and mobility can he determined from the Hall voltage.

GLOSSARY OF IMPORTANT TERMS

conductivity **A** material parameter related to carrier drift; quantitatively, the ratio of drift current density to electric field.

diffusion The process whereby particles flow from a region of high concentration to a region of low concentration.

diffusion coefficient The parameter relating particle flux to the particle density gradient.

diffusion **current** The current that results from the diffusion of charged particle?.

drift The process wherehy charged particles move while under the influence of an electric field.

drift current The current that results from the drift of charged particles.

drift velocity The average velocity of charged particles in the presence of an electric field.

Einstein **relation** The relation between the mobility and the diffusion coefficient.

Hall voltage The voltage induced across a semiconductor in a Hall effect measurement.

ionized impurity scattering The interaction between a charged carrier and an ionized impurity center.

latticescattering The interaction between a charged carrier and a thermally vibrating lattice atom.

mobility The parameter relating carrier drift velocity and electric field.

resistivity The reciprocal of conductivity; a material parameter that is a measure of the resistance to current.

velocity saturation The saturation of carrier drift velocity with increasing electric field.

CHECKPOINT

After studying this chapter, the reader should have the ability to:

- Discuss carrier drift current density.
- Explain why carriers reach an average drift velocity in the presence of an applied electric field.
- Discuss the mechanisms of lattice scattering and impurity scattering.

- Define mobility and discuss the temperature and ionized impurity concentration dependence on mobility.
  Define conductivity and resistivity.
  Discuss velocity saturation.
  Discuss carrier diffusion current density.
- State the Einstein relation.
- Describe the Hall effect.

REVIEW QUESTIONS

1. Write the equation for the total drift current density.
2. Define carrier mobility. What is the unit of mobility'?
3. Explain the temperature dependence of mobility. Why is the carrier mobility a functi of the ionized impurity concentrations?
4. Define conductivity. Define resistivity. What are the units of conductivity and resistiv
5. Sketch the drift velocity of electrons in silicon versus electric field. Repeat for GaAs
6. Write the equations for the diffusion current densities of electrons and holes.
7. What is the Einstein relation?
8. Describe the Hall effect.
9. Explain why the polarity of the Hall voltage changes depending on the conductivity (n type or p type) of the semiconductor.

PROBLEMS

(Note; Use the semiconductor parameters given in Appendix B if the parameters are specifically given in a problem.)

Section 5.1 Carrier Drift

5.1 Consider a homogeneous gallium arsenide semiconductor at T = 300 K with $N_d = 10^{16}$ cm^{-3} and $N_a = 0$. *(a)* Calculate the thermal-equilibrium values of electron a hole concentrations. *(b)* For an applied E-field of 10 V/cm, calculate the drift cum density. *(c)* Repeat parts *(a)* and *(b)* if $N_d = 0$ and $N_a = 10^{16}$ cm^{-3}.

5.2 A silicon crystal having a cross-sectional area of 0.001 cm^2 and a length of 10^{-3} c connected at its ends to a 10-V battery. At T = 300 K, we wanr a current of 100 m in the silicon. Calculate: *(a)* the required resistance R. *(b)* the required conductivit *(c)* the density of donor atoms to be added to achieve this conductivity. and id) th concentration of acceptor atoms to be added to form a compensated p-type materi with the conductivity given from part *(b)* if the initial concentration of donor atom $N_d = 10^{15}$ cm^{-3}.

53 *(a)* A silicon semiconductor is in the shape of a rectangular har with a cross-sectio area of 100 μm^2, a length of 0.1 cm, and is doped with 5×10^{16} cm^{-3} arsenic ator The temperature is T = 300 K. Determine the current if 5 V is applied across the length. *(b)* Repeat part *(a)* if the length is reduced to 0.01 cm. *(c)* Calculate the average drift velocity of electrons in parts *(a)* and *(b)*.

5.4 *(a)* A GaAs semiconductor resistor is doped with acceptor impurities at a concentr tion of $N_a = 10^{17}$ cm^{-3}. The cross-sectional area is 85 μm^2. The current in the

resistor is to be I = 20 mA with 10 V applied. Determine the required length of the device. (b) Repeat part *(n)* for silicon.

5.5 (a) Three volts is applied across a 1-cm-long semiconductor bar. The average electron drift velocity is 10^4 cm/s. Find the electron mobility. (*b*) If the electron mobility in part (a) were 800 cm^2/V-s, what is the average electron drift velocity?

5.6 Use the velocity-field relations for silicon and gallium arsenide shown in Figure 5.7 to determine the transit time of electrons through a 1-μm distance in these materials for an electric field of (a) 1 kV/cm **and** (*b*) 50 kV/cm.

5.7 A perfectly compensated semiconductor is one in which the donor and acceptor impurity concentrations are exactly equal. Assuming complete ionization, determine the conductivity of silicon at T = 300 Kin which the impurity concentrations are (a) $N_a = N_d = 10^{14}$ cm^{-3} and (*b*) $N_a = N_d = 10^{18}$ cm^{-3}.

5.8 (*a*) In a p-type gallium arsenide semiconductor, the conductivity is $\sigma = 5$ $(\Omega\text{-cm})^{-1}$ at T = 300 K. Calculate the thermal-equilibrium values of the electron and hole concentrations. (b) Repeat part (*a*) for n-type silicon if the resistivity is $\rho = 8$ Ω-cm.

5.9 In a particular semiconductor material, $\mu_n = 1000$ cm^2/V-s, $\mu_p = 600$ cm^2/V-s, and $N_C = N_V = 10^{19}$ cm^{-3}. These parameters arc independent of temperature. The measured conductivity of the intrinsic material is $\sigma = 10^{-6}$ $(\Omega\text{-cm})^{-1}$ at $T = 300$ K. Find the conductivity at $T = 500$ K.

5.10 (*a*) Calculate the resistivity at T = 300 K of intrinsic (*i*) silicon, (*ii*) germanium. and (iii) gallium arsenide. (*b*) If rectangular semiconductor bars are fabricated using the materials in part (a),determine the resistance of each bar if its cross-sectional area is 85 μm^2 and length is 200 μm.

5.11 **An** n-type silicon sample has a resistivity of 5 Ω-cm at T = 300 K. (a) What is the donor impurity concentration? (*b*) What is the expected resistivity at (*i*) T = 200 K and (ii) T = 400 K.

5.12 Consider silicon doped at impurity concentrations of $N_d = 2 \times 10^{16}$ cm^{-3} and $N_a = 0$. **An** empirical expression relating electron drift velocity to electric field is given by

$$v_d = \frac{\mu_{n0}E}{\sqrt{1 + \left(\frac{\mu_{n0}E}{v_{sat}}\right)^2}}$$

where $\mu_{n0} = 1350$ cm^2/V-s, $v_{sat} = 1.8 \times 10^7$ cm/s, and E is given in V/cm. Plot electron drift current density (magnitude) versus electric field (log–log scale) over the range $0 \leq E \leq 10^6$ Vlcm.

5.13 Consider silicon at $T = 300$ K. Assume the electron mobility is $\mu_n = 1350$ cm^2/V-s. The kinetic energy of an electron in the conduction band is $(1/2)m_n^* v_d^2$, where m_n^* is the effective mass and v_d is the drift velocity Determine the kinetic energy of an electron in the conductian hand if the applied electric field is *(a)* 10 V/cm and (*b*) 1 kV/cm.

5.14 Consider a semiconductor that is uniformly doped with $N_d = 10^{14}$ cm^{-3} and $N_a = 0$, with an applied electric field of E = 100 V/cm. Assume that $\mu_n = 1000$ cm^2N-s and $\mu_p = 0$. Also assume the following parameters:

$$N_c = 2 \times 10^{19}(T/300)^{3/2} \text{ cm}^{-3}$$
$$N_v = 1 \times 10^{19}(T/300)^{3/2} \text{ cm}^{-3}$$
$$E_g = 1.10 \text{ eV}$$

(a) Calculate the electric-current density at $T = 300$ K. (b) At what temperature will this current increase by 5 percent? (Assume the mobilities are independent of temperature.)

5.15 A semiconductor material has electron and hole mobilities μ_n and μ_p, respectively. When the conductivity is considered as a function of the hole concentration p_0, (a) show that the minimum value of conductivity, σ_{min}, can be written as

$$\sigma_{min} = \frac{2\sigma_i(\mu_n\mu_p)^{1/2}}{(\mu_n + \mu_p)}$$

where σ_i is the intrinsic conductivity, and (b) show that the corresponding hole concentration is $p_0 = n_i(\mu_n/\mu_p)^{1/2}$.

5.16 A particular intrinsic semiconductor has a resistivity of 50 Ω-cm at $T = 300$ K and 5 Ω-cm at $T = 330$ K. Neglecting the change in mobility with temperature, determine the bandgap energy of the semiconductor

5.17 Three scattering mechanisms are present in a particular semiconductor material. If only the first scattering mechanism were present, the mobility would be $\mu_1 = 2000$ cm^2/V-s, if only the second mechanism were present, the mobility would be $\mu_2 = 1500$ cm^2/V-s, and if only the third mechanism were present, the mobility would be $\mu_3 = 500$ cm^2/V-s. What is the net mobility?

5.18 Assume that the mahility ofelectrons in silicon at T = 300 K is $\mu_n = 1300$ cm^2/V Also assume that the mobility is limited by lattice scattering and varies as $T^{-3/2}$. Determine the electron mobility at (a) $T = 200$ K and (b) $T = 400$ K.

5.19 Two scattering mechanisms exist in a semiconductor. If only the first mechanism w present, the mobility would he 250 cm^2/V-s. If only the second mechanism wer p sent, the mobility would be 500 cm^2/V-s. Determine the mobility when both scatte mechanisms exist at the same time.

5.20 The effective density of states functions in silicon can be written in the form

$$N_c = 2.8 \times 10^{19}\left(\frac{T}{300}\right)^{3/2} \qquad N_v = 1.04 \times 10^{19}\left(\frac{T}{300}\right)^{3/2}$$

Assume the mobilities are given by

$$\mu_n = 1350\left(\frac{T}{300}\right)^{-3/2} \qquad \mu_p = 480\left(\frac{T}{300}\right)^{-3/2}$$

Assume the bandgap energy is $E_g = 1.2$ eV and independent of temperature Plot the intrinsic conductivity as a function of T over the range 200 ≤ T ≤ 600 K

5.21 (a) Assume that the electron mobility in an n-type semiconductor is given by

$$\mu_n = \frac{1350}{\left(1 + \frac{N_d}{5 \times 10^{16}}\right)^{1/2}} \text{ cm}^2/\text{V-s}$$

where N_d is the donor concentration in cm^{-3}. Assuming complete ionization. plot the conductivity as a function of N_d over the range $10^{15} \le N_d \le 10^{18}$ cm^{-3}. (b) Compare the results of part (a) to that if the mobility were assumed to he a constant equal to

1350 cm^2/V-s. (*c*) If an electric field of $E = 10$ V/cm is applied to the semiconductor, plot the electron drift current density of pans (*a*) and (*b*)

Section 5.2 Carrier Diffusion

5.22 Consider a sample of silicon at T = 300 K. Assume that the electron concentration varies linearly with distance, as shown in Figure 5.14. The diffusion current density is found to he $J_n = 0.19$ A/cm^2. If the electron diffusion coefficient is $D_n = 25$ cm^2/s, determine the electron concentration at $x = 0$.

5.23 The electron concentration in silicon decreases linearly from 10^{16} cm^{-3} to 10^{15} cm^{-3} over a distance of 0.10 cm. The cross-sectional area of the sample is 0.05 cm'. The electron diffusion coefficient is 25 cm^2/s. Calculate the electron diffusion current.

5.24 The electron concentration in a sample of n-type silicon varies linearly from 10^{17} cm^{-3} at x = 0 to 6×10^{16} cm^{-3} at $x = 4$ μm. There is no applied electric field. The electron current density is experimentally measured to bc -400 A/cm^2. What is the electron diffusion coefficient?

5.25 The holeconcentration in p type GaAs is given by $p = 10^{16}(1 - x/L)$ cm^{-3} for $0 \le x \le$ L where $L = 10$ μm. The hole diffusion coeflicient is 10 cm^2/s. Calculate the hole diffusion current density at (*a*) x = 0, (*b*) x = 5 μm, and (*c*) $x = 10$ μm.

5.26 The hole concentralion is given by $p = 10^{15} \exp(-x/L_p)$ cm^{-3} for $x \ge 0$ and the electron concentration is given by $5 \times 10^{14} \exp(+x/L_n)$ cm^{-3} for $x \le 0$. The values of L_n and L_n are 5×10^{-4} cm and 10^{-3} cm, respectively. The hole and electron diffusion coefficients are 10 cm^2/s and 25 cm^2/s, respectively. The total current density is defined as the sum of the hole diffusion current density at $x = 0$ and the electron diffusion current density at $x = 0$. Calculate the total current density.

5.27 The hole concentration in germanium at T = 300 K varies as

$$p(x) = 10^{15} \exp\left(\frac{-x}{22.5}\right) \text{cm}^{-3}$$

where x is measured in μm. If the hole diffusion coefficient is $D_p = 48$ cm^2/s, determine the hole diffusion current density as a function of x.

5.28 The electron concentration in silicon at T = 300 K is given by

$$n(x) = 10^{16} \exp\left(\frac{-x}{18}\right) \text{cm}^{-3}$$

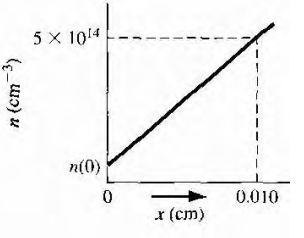


Figure 5.14 | Figure for Problem 5.22.

where x is measured in μm and is limited to $0 \leq x \leq 25\ \mu$m. The electron diffusion coefficient is $D_n = 25$ cm'ls and the electron mobility is $\mu_n = 960$ cm^2/V-s. The total electron current density through the semiconductor is constant and equal to $J_n = -40$ Alcm^2. The electron current has both diffusion and drift current components. Determine the electric field as a function of x which must exist in the semiconductor.

5.29 The total current in a semiconductor is constant and is composed of electron drift current and hole diffusion current. The electron concentration is constant and is equal to 10^{16} cm^{-3}. The hole concentration is given by

$$p(x) = 10^{15} \exp\left(\frac{-x}{L}\right) \text{cm}^{-3} \qquad (x \geq 0)$$

where L = 12 μm. The hole diffusion coefficient is $D_p = 12$ cm^2/s and the electron mobility is $\mu_n = 1000$ cm^2/V-s. The total current density is $J = 4.8$ A/cm^2. Calculate ***(a)*** the hole diffusion current density versus x, (b) the electron current density versus x, and ***(r)*** the electric field versus x.

***5.30** A constant electric field, E = 12 V/cm, exists in the $+x$ direction of an n-type gallium arsenide semiconductor for $0 \leq x \leq 50\ \mu$m. The total current density is a constant and is $J = 100$ A/cm^2. At $x = 0$, the drift and diffusion currents are equal. Let T = 300 K and $\mu_n = 8000$ cm^2/V-s. (a) Determine the expression for the electron concentration $n(x)$. (b) Calculate the electron concentration at $x = 0$ and at $x = 50\ \mu$m. (c) Calculate the drift and diffusion current densities at $x = 50\ \mu$m.

5.31** In n-type silicon, the Fermi energy level varies linearly with distance over a short range. At $x = 0$, $E_F - E_{Fi} = 0.4$ eV and, at $x = 10^{-3}$ cm, $E_f - E_{Fi} = 0.15$ eV. ***(a) Write the expression for the electron concentration over the distance. (b) If the electron diffusion coefficient is $D_n = 25$ cm^2/s, calculate the electron diffusion current density at (i) $x = 0$ and (ii) $x = 5 \times 10^{-4}$ cm.

***5.32** (a) The electron concentration in a semiconductor is given by $n = 10^{16}(1 - x/L)$ cm^{-3} for $0 \leq x \leq$ L. where L = 10 μm. The electron mobility and diffusion coefficient are $\mu_n = 1000$ cm^2/V-s and $D_n = 25.9$ cm^2/s. An electric field is applied such that the total electron current density is a constant over the given range of x and is $J_n = -80$ A/cm^2. Determine the required electric field versus distance function. (b) Repeat part (a) if $J_n = -20$ A/cm^2.

Section 5.3 Graded Impurity Distribution

5.33 Consider a semiconductor in thermal equilibrium (no current). Assume that the donor concenvation varies exponentially as

$$N_d(x) = N_{d0} \exp(-ax)$$

over the range $0 \leq x \leq 1/\alpha$ where N_{d0} is a constant. ***(a)*** Calculate the electric field as a function of x for $0 \leq x \leq 1/\alpha$. (b) Calculate the potential difference between $x = 0$ and $x = 1/\alpha$.

5.34 Using the data in Example 5.5, calculate the potential difference between $x = 0$ and $x = 1\ \mu$m.

5.35 Determine a doping profile in a semiconductor at T = 300 K that will induce an electric field of 1 kV/cm over a length of 0.2 μm.

***5.36** In GaAs, the donor impurity concentration varies as $N_{d0}\exp(-x/L)$ for $0 \le x \le L$, where $L = 0.1\ \mu m$ and $N_{d0} = 5 \times 10^{16}\ cm^{-3}$. Assume $\mu_n = 6000\ cm^2/V\text{-s}$ and $T = 300$ K. (o) Derive the expression for the electron diffusion current density versus distance over the given range of x. (b) Determine the induced electric field that generates a drift current density that compensates the diffusion current density.

5.37 ***(a)*** Consider the electron mobility in silicon for $N_d = 10^{17}\ cm^{-3}$ from Figure 5.2a. Calculate and plot the electron diffusion coefficient versus temperature over the range $-50 \le T \le 200°C$. (b) Repeat pan (a) if the electron diffusion coefficient is given by $D_n = (0.0259)\mu_n$ for all temperatures. What conclusion can be made about the temperature dependence of the diffusion coefficient?

5.38 (a) Assume that the mobility of a carrier at $T = 300$ K is $\mu = 925\ cm^2/V\text{-s}$. Calculate the carrier diffusion coefficient. ***(b)*** Assume that the diffusion coefficient of a carrier at $T = 300$ K is $D = 28.3\ cm^2/s$. Calculate the carrier mobility.

Section 5.4 The Hall Effect

(*Note:* Refer to Figure 5.13 fur the geometry of the Hall effect.)

5.39 A sample of silicon is doped with 10^{16} baron atoms per cm'. The Hall sample has the same geometrical dimensions given in Example 5.7. The current is $I_x = 1$ mA with $B_z = 350$ gauss $= 3.5 \times 10^{-2}$ tesla. Determine (a) the Hall voltage and ***(b)*** the Hall held.

5.40 Germanium is doped with 5×10^{15} donor atotns per cm^3 at $T = 300$ K. The dimensions of the Hall device are $d = 5 \times 10^{-3}$ cm, $W = 2 \times 10^{-2}$ cm, and $L = 10^{-1}$ cm. The current is $I_x = 250\ \mu A$, the applied voltage is $V_x = 100$ mV, and the magnetic flux density is $B_z = 500$ gauss $= 5 \times 10^{-2}$ tesla. Calculate: (a) the Hall voltage, ***(b)*** the Hall field. and (c) the carrier mobility

5.41 Asilicon Hall device at $T = 300$ K has the following geometry: $d = 10^{-3}$ cm, $W = 10^{-2}$ cm, and $L = 10^{-1}$ cm. The following parameters are measured: $I_x = 0.75$ mA, $V_x = 15$ V, $V_H = +5.8$ mV, and $B_z = 1000$ gauss $= 10^{-1}$ tesla. Determine (a) the conductivity type, (b) the majority carrier concentration, and ***(c)*** the majority carrier mobility.

5.42 Consider silicon at $T = 300$ K. A Hall effect device is fabricated with the following geometry: $d = 5 \times 10^{-3}$ cm, $W = 5 \times 10^{-2}$ cm, and $L = 0.50$ cm. The electrical parameters measured are: $I_x = 0.50$ mA, $V_x = 1.25$ V, and $B_z = 650$ gauss $= 6.5 \times 10^{-2}$ tesla. The Hall field is $E_H = -16.5$ mV/cm. Determine (a) the Hall voltage, (b) the conductivity type, (c) the majority carrier concentretion, and (d) the majority carrier mobility.

5.43 Consider a gallium arsenide sample at $T = 300$ K. A Hall effect device has been fabricated with the following geometry: $d = 0.01$ cm, $W = 0.05$ cm. and $L = 0.5$ cm The electrical parameters are: $I_x = 2.5$ mA, $V_x = 2.2$ V, and $B_z = 2.5 \times 10^{-2}$ tesla. The Hall voltage is $V_H = -4.5$ mV. Find: (a) the conductivity type, ***(h)*** the majority carrier concentration, (c) the mobility, and (d) the resistivity.

Summary and Review

5.44 An n-type silicon semiconductor resistor is to be designed so that it carries a current of 5 mA with an applied voltage of 5 V. (a) If $N_d = 3 \times 10^{14}\ cm^{-3}$ and $N_a = 0$, design a resistor lo meet the required specifications. (b) If $N_d = 3 \times 10^{16}\ cm^{-3}$ and

$N_a = 2.5 \times 10^{16}$ cm^{-3}, redesign the resistor. (*c*) Discuss the relative lengths of the two designs compared to the doping concentration. Is there a linear relationship'!

5.45 In fabricating a Hall effect device, the two points at which the Hall voltage is measured may not be lined up exactly perpendicular to the current ***I,*** (see Figure 5.13). Discuss the effect this misalignment will have on the Hall voltage. Show that a valid Hall voltage can be obtained from two measurements: first with the magnetic field in the $+z$ direction, and then in the $-z$ direction.

5.46 Another technique for determining the conductivity type of a semiconductor is called the hot probe method. It consists of two probes and an ammeter that indicates the direction of current. One probe is heated and the other is at room temperature. No voltage is applied, but a current will exist when the probes touch the semiconductor Explain the operation of this hot probe technique and sketch a diagram indicating the direction of current for p- and n-type semiconductor samples.

READING LIST

1. Bube, R. H. *Electrons in Solids: An Introductory Survey.* 3rd ed. San Diego, CA: Academic Press, 1992.
2. Kano, K. ***Semiconductor Devices.*** Upper Saddle River, NJ: Prentice Hall, 1998.
*3. Lundstrom, M. *Fundamentals of Carrier* ***Transport.*** Vol. *X* of ***Modular Series on Solid State Devices.*** Reading, MA: Addison-Wesley, 1990.
4. Muller, R. S., and T. I. Kamins. ***Device*** *Electronics for Integrated* ***Circuits.*** 2nd ed. New York: Wiley, 1986.
5. Navon, D. H. ***Semiconductor*** *Microdevices and Materials.* New ***York:*** Holt, Rinehart & Winston, 1986.
6. Pierret, R. F. *Semiconductor* ***Device Fundamentals.*** Reading, MA: Addison-Wesley Publishing Co., 1996.
7. Shur, M. ***Introduction to Electronic Devices.*** New York: John Wiley and Sons, 1996.
*8. Shur, M. ***Physics of Semiconducror Devices.*** Englewood Cliffs, NJ: Prentice Hall, 1990.
9. Singh, J. ***Semiconductor Devices: An Introduction.*** New York: McGraw-Hill, 1994.
10. Singh, J. ***Semiconductor Devices: Basic Principles.*** New York: John Wiley and Sons. 2001.
11. Streetman, B. G., and S. Banerjee. *Solid State Electronic* ***Devices.*** 5th ed. Upper Saddle River, NJ: Prentice Hall, 2000.
12. Sre, S. M. ***Physics of Semiconducror*** *Devices.* 2nd ed. New York: John Wiley and Sons. 1981.
13. Sze, S. M. ***Semiconducror Devices:*** *Physics* ***and Technology*** 2nd ed. New York: John Wiley and Sons, 2001.
*14. van der Ziel, A. *Solid State* ***Physical*** *Electronics.* 2nd ed. Englewood Cliffs, NJ: Prentice Hall, 1968.
15. Wang, S. *Fundamentals* of ***Semiconductor Theory*** *and* ***Device Physics.*** Englewood Cliffs, NJ: Prentice Hall, 1989.
16. Yang, E. S. *Microelectronic* ***Devices.*** New York: McGraw-Hill, 1988.

CHAPTER 6

Nonequilibrium Excess Carriers in Semiconductors

PREVIEW

Our discussion of the physics of semiconductors in Chapter 4 was based on thermal equilibrium. When a voltage is applied or a current exists in a semiconductor device, the semiconductor is operating under nonequilibrium conditions. In our discussion of current transport in Chapter 5, we did not address nonequilibrium conditions but implicitly assumed that equilibrium was not significantly disturbed. Excess electrons in the conduction band and excess holes in the valence band may exist in addition to the thermal-equilibrium concentrations if an external excitation is applied to the semiconductor. In this chapter, we will discuss the behavior of nonequilibrium electron and hole concentrations as functions of time and space coordinates.

Excess electrons and excess holes do not move independently of each other. They diffuse, drift, and recombine with the same effective diffusion coefficient, drift mobility, and lifetime. This phenomenon is called ambipolar transport. We will develop the ambipolar transport equation which describes the behavior of the excess electrons and holes. The behavior of excess carriers is fundamental to the operation of semiconductor devices. Several examples of the generation of excess carriers will be explored to illustrate the characteristics of the ambipolar transport phenomenon.

The Fermi energy was previously defined for a semiconductor in thermal equilibrium. The creation of excess electrons and holes means that the semiconductor is no longer in thermal equilibrium. We can define two new parameters that apply to the nonequilibrium semiconductor: the quasi-Fermi energy for electrons and the quasi-Fermi energy for holes.

Semiconductor devices are generally fabricated at or near a surface. We will study the effect of these surfaces on the characteristics of excess electrons and holes. These effects can significantly influence the semiconductor device properties. ■

6.1 | CARRIER GENERATION AND RECOMBINATION

In this chapter, we discuss carrier generation and recombination, which we can define as follows: ***generation*** is the process whereby electrons and holes are created, and *recombination* is the process whereby electron? and holes are annihilated.

Any deviation from thermal equilibrium will tend to change the electron and hole concentrations in a semiconductor. A sudden increase in temperature, for example, will increase the rate at which electrons and holes are thermally generated so that their concentrations will change with time until new equilibriuin values are reached. An external excitation, such as light (a flux of photons), can also generate electrons and holes, creating a nonequilibrium condition. To understand the generation and recombination processes, we will first consider direct band-to-band generation and recombination, and then, later, the effect of allowed electronic energy states within the bandgap, referred to as traps or recombination centers.

6.1.1 The Semiconductor in Equilibrium

We have determined the thermal-equilibrium concentration of electrons and holes in the conduction and valence bands, respectively. In thermal equilibrium, these concentrations are independent of time. However, electrons are continually being thermally excited from the valence band into the conduction band by the random nature of the thermal process. At the same time, electrons moving randomly through the crystal in the conduction band may come in close proximity to holes and "fall" into the empty states in the valence band. This recombination process annihilates both the electron and hole. Since the net carrier concentrations are independent of time in thermal equilibrium, the rate at which electrons and holes are generated and the rate at which they recombine must be equal. The generation and recombination processes are schematically shown in Figure 6.1.

Let G_{n0} and G_{p0} be the thermal-generation rates of electrons and holes, respectively, given in units of #/cm^3-s. For the direct band-to-band generation. the electrons and holes are created in pairs, so we must have that

$$G_{n0} = G_{p0} \tag{6.1}$$

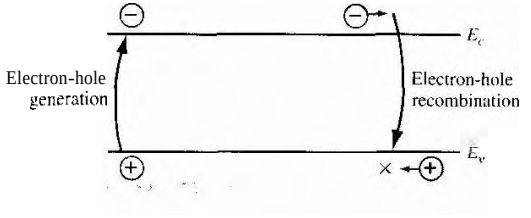


Figure 6.1 | Electrun-hole generation and recombination

Let R_{n0} and R_{p0} be the recomhination rates of electrons and holes, respectively, for a semiconductor in thermal equilibrium, again given in units of #/cm^3-s. In direct band-to-hand recombination, electrons and holes recombine in pairs, so that

$$R_{n0} = R_{p0} \tag{6.2}$$

In thermal equilibrium, the concentrations of electrons and holes are independent of time; therefore, the generation and recomhination rates are equal, so we have

$$G_{n0} = G_{p0} = R_{n0} = R_{p0} \tag{6.3}$$

6.1.2 Excess Carrier Generation and Recombination

Additional notation is introduced in this chapter. Table 6.1 lists some of the more pertinent symbols used throughout the chapter. Other symbols will be defined as we advance through the chapter.

Electrons in the valence hand may be excited into the conduction hand when, for example, high-energy photons are incident on a semiconductor. When this happens, not only is an electron created in the conduction band, but a hole is created in the valence band; thus an electron-hole pair is generated. The additional electrons and holes created are called ***excess electrons*** and ***excess holes.***

The excess electrons and holes are generated by an external force at a particular rate. Let g'_n be the generation rate of excess electrons and g'_p be that of excess holes. These generation rates also have units of #/cm^3-s. For the direct hand-to-hand generation, the excess electrons and holes are also created in pairs, so we must have

$$g'_n = g'_p \tag{6.4}$$

When excess electrons and holes are created, the concentration of electrons in the conduction band and of holes in the valence band increase above their thermal-equilibrium value. We may write

$$n = n_0 + \delta n \tag{6.5a}$$

and

$$p = p_0 + \delta p \tag{6.5b}$$

Table **6.1** | Relevant notation used in Chapter 6

Symbol	Definition
n_0, p_0	Thermal equilibrium electron and hole concentrations (independent of time and also usually position).
n, p	Total electron and hole concentrations (may be functions of time and/or position).
$\delta n = n - n_0$ $\delta p = p - p_0$	Excess electron and hole concentrations (may be functions of time and/or position).
g'_n, g'_p	Excess electron and hole generation rates.
R'_n, R'_p	Excess electron and hole recombination rates.
τ_{n0}, τ_{p0}	Excess minority carrier electron and hole lifetimes.

where n_0 and p_0 are the thermal-equilibrium concentrations, and δn and δp are the excess electron and hole concentrations. Figure 6.2 shows the excess electron-hole generation process and the resulting carrier concentrations. The external force has perturbed the equilibrium condition so that the semiconductor is no longer in thermal equilibrium. We may note from Equations (6.5a) and (6.5b) that, in a nonequilibrium condition, $np \neq n_0 p_0 = n_i^2$.

A steady-state generation of excess electrons and holes will not cause a continual buildup of the carrier concentrations. As in the case of thermal equilibrium, an electron in the conduction band may "fall down" into the valence band, leading to the process of excess electron-hole recombination. Figure 6.3 shows this process. The recombination rate for excess electrons is denoted by R'_n and for excess holes by R'_p. Both parameters have units of #/cm^3-s. The excess electrons and holes recombine in pairs, so the recombination rates must be equal. We can then write

$$R'_n = R'_p \tag{6.6}$$

In the direct band-to-band recombination that we are considering, the recombination occurs spontaneously: thus, the probability of an electron and hole recombining is constant with time. The rate at which electrons recombine must be proportional

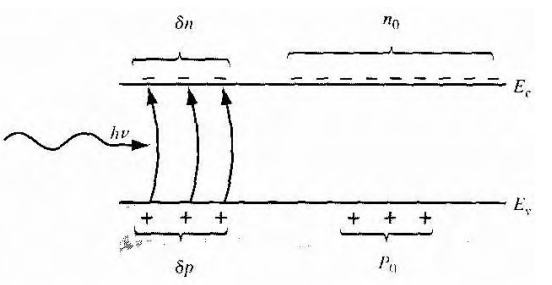


Figure 6.2 | Creation of excess electron and hole densities by photons.

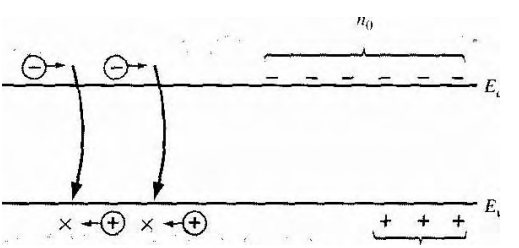


Figure 6.3 | Recombination of excess carriers reestablishing thermal equilibrium.

to the electron concentration and must also be proportional to the hole concentration. If there are no electrons or holes, there can be no recombination.

The net rate of change in the electron concentration can be written as

$$\frac{dn(t)}{dt} = \alpha_r\left[n_i^2 - n(t)p(t)\right] \tag{6.7}$$

where

$$n(t) = n_0 + \delta n(t) \tag{6.8a}$$

and

$$p(t) = p_0 + \delta p(t) \tag{6.8b}$$

The first term, $\alpha_r n_i^2$, in Equation *(6.7)* is the thermal-equilibrium generation rate. Since excess electrons and holes are created and recombine in pairs, we have that $\delta n(t) = \delta p(t)$. (Excess electron and hole concentrations are equal so we can simply use the phrase excess carriers to mean either.) The thermal-equilibrium parameters, n_0 and p_0, being independent of time, Equation *(6.7)* becomes

$$\begin{aligned}\frac{d(\delta n(t))}{dt} &= \alpha_r\left[n_i^2 - (n_0 + \delta n(t))(p_0 + \delta p(t))\right] \\ &= -\alpha_r \delta n(t)[(n_0 + p_0) + \delta n(t)]\end{aligned} \tag{6.9}$$

Equation *(6.9)* can easily be solved if we impose the condition of *low-level injection*. Low-level injection puts limits on the magnitude of the excess carrier concentration compared with the thermal equilibrium carrier concentrations. In an extrinsic n-type material, we generally have $n_0 \gg p_0$ and, in an extrinsic p-type material, we generally have $p_0 \gg n_0$. Low-level injection means that the excess carrier concentration is much less than the thermal equilibrium majority carrier concentration. Conversely. high-level injection occurs when the excess carrier concentration becomes comparable to or greater than the thermal equilibrium majority carrier concentrations.

If we consider a p-type material ($p_0 \gg n_0$) under low-level injection ($\delta n(t) \ll p_0$), then Equation *(6.9)* becomes

$$\frac{d(\delta n(t))}{dt} = -\alpha_r p_0 \delta n(t) \tag{6.10}$$

The solution to the equation is an exponential decay from the initial excess concentration, or

$$\delta n(t) = \delta n(0) e^{-\alpha_r p_0 t} = \delta n(0) e^{-t/\tau_{n0}} \tag{6.11}$$

where $\tau_{n0} = (\alpha_r p_0)^{-1}$ and is a constant for the low-level injection. Equation (6.11) describes the decay of excess minority carrier electrons so that τ_{n0} is often referred to as the ***excess minority carrier lifetime.***[1]

[1] In Chapter 5 we defined τ as a mean rime between collisions. We definer here as the mean time before a recombination event occurs. The two parameters are not related.

The recombination rate—which is defined as a positive quantity--of excess minority carrier electrons can be written, using Equation (6.10). as

$$R'_n = \frac{-d(\delta n(t))}{dt} = +\alpha_r p_0 \delta n(t) = \frac{\delta n(t)}{\tau_{n0}} \tag{6.12}$$

For the direct hand-to-hand recombination, the excess majority carrier holes recombine at the same rate, so that for the p-type material

$$\boxed{R'_n = R'_p = \frac{\delta n(t)}{\tau_{n0}}} \tag{6.13}$$

In the case of an n-type material ($n_0 \gg p_0$) under low-level injection ($\delta n(t) \ll n_0$), the decay of minority carrier holes occurs with a time constant $\tau_{p0} = (\alpha_r n_0)^{-1}$, where τ_{p0} is also referred to as the excess minority carrier lifetime. The recombination rate of the majority carrier electrons will be the same as that of the minority carrier holes, so we have

$$\boxed{R'_n = R'_p = \frac{\delta n(t)}{\tau_{p0}}} \tag{6.14}$$

The generation rates of excess carriers are not functions of electron or hole concentrations. In general, the generation and recombination rates may be functions of the space coordinates and time.

TEST YOUR UNDERSTANDING

E6.1 Excess electrons have been generated in a semiconductor to aconcentrationof $\delta n(0) = 10^{15}$ cm^{-3}. The excess carrier lifetime is $\tau_{n0} = 10^{-6}$ s. The forcing function generating the excess carriers turns off at $t = 0$ so the semicdnductor is allowed to return to an equilibrium condition for $t > 0$. Calculate the excess electron concentration for (a) t = 0, (*b*) t = 1 μs, and (*c*) t = 4 μs.

[(a) 10^{15} cm^{-3}, (*b*) 1.83×10^{14} cm^{-3}, (*c*) 3.68×10^{13} cm^{-3}]

E6.2 Using the parameters given in E6.1, calculate the recombination rate of the excess electrons for (a) t = 0, ***(b)t*** = 1 μs, and (c) t = 4 μs.

[(a) 10^{21} cm^{-3} s^{-1}, (*b*) 3.68×10^{20} cm^{-3} s^{-1}, (*c*) 1.83×10^{19} cm^{-3} s^{-1}]

6.2 | CHARACTERISTICS OF EXCESS CARRIERS

The generation and recombination rates of excess carriers are important parameters, but how the excess carriers behave with time and in space in the presence of electric fields and density gradients is of equal importance. As mentioned in the preview section, the excess electrons and holes do not move independently of each other, but they diffuse and drift with the same effective diffusion coefficient and with the same

effective mobility. This phenomenon is called amhipolar transport. The question that must be answered is what is the effective diffusion coefficient and what is the effective mobility that characterizes the behavior of these excess carriers? To answer these questions, we must develop the continuity equations for the carriers and then develop the ambipolar transport equations.

The final results show that. for an extrinsic semiconductor under low injection (this concept will be defined in the analysis). the effective diffusion coefficient and mobility parameters are those of the minority carrier. This result is thoroughly developed in the following derivations. As will he seen in the following chapters, the behavior of the excess carriers has a profound impact on the characteristics of semiconductor devices.

6.2.1 Continuity Equations

The continuity equations for electrons and holes are developed in this section. Figure 6.4 shows a differential volume element in which a one-dimensional holepanicle flux is entering the differential element at x and is leaving the element at $x + dx$. The parameter F_{px}^{+} is the hole-particle flux, or flow, and has units of number of holes/cm^2-s. For the x component of the particle current density shown, we may write

$$F_{px}^{+}(x + dx) = F_{px}^{+}(x) + \frac{\partial F_{px}^{+}}{\partial x} \cdot dx \tag{6.15}$$

This equation is a Taylor expansion of $F_{px}^{+}(x + dx)$, where the differential length d x is small, so that only the first two terms in the expansion are significant. The net increase in the number of holes per unit time within the differential volume element due to the x-component of hole flux is given by

$$\frac{\partial p}{\partial t}\, dx\, dy\, dz = [F_{px}^{+}(x) - F_{px}^{+}(x + dx)]\, dy\, dz = -\frac{\partial F_{px}^{+}}{\partial x}\, dx\, dy\, dz \tag{6.16}$$

If $F_{px}^{+}(x) > F_{px}^{+}(x + dx)$, for example, there will be a net increase in the number of holes in the differential volume element with time. If we generalize to a three-dimensional hole flux, then the right side of Equation (6.16) may be written as

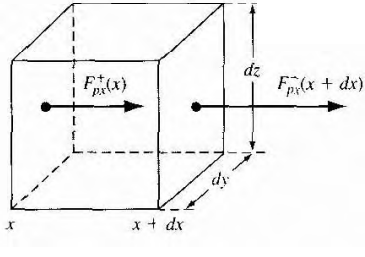


Figure 6.4 | Differential volume showing x component of the hole-particle flux.

$-\nabla \cdot F_p^+ \, dx\, dy\, dz$, where $V \cdot F_p^+$ is the divergence of the flux vector. We will limit ourselves to a one-dimensional analysis.

The generation rate and recombination rate of holes will also affect the hole concentration in the differential volume. The net increase in the number of holes per unit time in the differential volume element is then given by

$$\frac{\partial p}{\partial t} dx\, dy\, dz = -\frac{\partial F_p^+}{\partial x} dx\, dy\, dz + g_p\, dx\, dy\, dz - \frac{p}{\tau_{pt}} dx\, dy\, dz \tag{6.17}$$

where p is the density of holes. The first term on the right side of Equation (6.17) is the increase in the number of holes per unit time due to the hole flux, the second term is the increase in the number of holes per unit time due to the generation of holes, and the last term is the decrease in the number of holes per unit time due to the recombination of holes. The recombination rate for holes is given by p/τ_{pt} where τ_{pt} includes the thermal equilibrium carrier lifetime and the excess carrier lifetime.

If we divide both sides of Equation (6.17) by the differential volume $dx\, dy\, dz$, the net increase in the hole concentration per unit time is

$$\frac{\partial p}{\partial t} = -\frac{\partial F_p^+}{\partial x} + g_p - \frac{p}{\tau_{pt}} \tag{6.18}$$

Equation (6.18) is known as the continuity equation for holes.

Similarly, the one-dimensional continuity equation for electrons is given by

$$\frac{\partial n}{\partial t} = -\frac{\partial F_n^-}{\partial x} + g_n - \frac{n}{\tau_{nt}} \tag{6.19}$$

where F_n^- is the electron-particle flow, or flux, also given in units of number of electrons/cm^2-s.

6.2.2 Time-Dependent Diffusion Equations

In Chapter 5, we derived the hole and electron current densities, which are given, in one dimension, by

$$J_p = e\mu_p p\mathrm{E} - eD_p \frac{\partial p}{\partial x} \tag{6.20}$$

and

$$J_n = e\mu_n n\mathrm{E} + eD_n \frac{\partial n}{\partial x} \tag{6.21}$$

If we divide the hole current density by $(+e)$ and the electron current density by $(-e)$, we obtain each particle flux. These equations become

$$\frac{J_p}{(+e)} = F_p^+ = \mu_p p\mathrm{E} - D_p \frac{\partial p}{\partial x} \tag{6.22}$$

and

$$\frac{J_n}{(-e)} = F_n^- = -\mu_n n\mathrm{E} - D_n \frac{\partial n}{\partial x} \tag{6.23}$$

Taking the divergence of Equations (6.22) and (6.23), and substituting back into the continuity equations of (6.18) and (6.19), we obtain

$$\frac{\partial p}{\partial t} = -\mu_p \frac{\partial (p\mathrm{E})}{\partial x} + D_p \frac{\partial^2 p}{\partial x^2} + g_p - \frac{p}{\tau_{pt}} \tag{6.24}$$

$$\frac{\partial n}{\partial t} = +\mu_n \frac{\partial (n\mathrm{E})}{\partial x} + D_n \frac{\partial^2 n}{\partial x^2} + g_n - \frac{n}{\tau_{nt}} \tag{6.25}$$

Keeping in mind that we are limiting ourselves to a one-dimensional analysis, we can expand the derivative of the product as

$$\frac{\partial (p\mathrm{E})}{\partial x} = \mathrm{E}\frac{\partial p}{\partial x} + p\frac{\partial \mathrm{E}}{\partial x} \tag{6.26}$$

In a more generalized three-dimensional analysis, Equation (6.26) would have to be replaced by a vector identity. Equations (6.24) and (6.25) can be written in the form

$$D_p \frac{\partial^2 p}{\partial x^2} - \mu_p \left(\mathrm{E}\frac{\partial p}{\partial x} + p\frac{\partial \mathrm{E}}{\partial x} \right) + g_p - \frac{p}{\tau_{pt}} = \frac{\partial p}{\partial t} \tag{6.27}$$

and

$$D_n \frac{\partial^2 n}{\partial x^2} + \mu_n \left(\mathrm{E}\frac{\partial n}{\partial x} + n\frac{\partial \mathrm{E}}{\partial x} \right) + g_n - \frac{n}{\tau_{nt}} = \frac{\partial n}{\partial t} \tag{6.28}$$

Equations (6.27) and (6.28) are the time-dependent diffusion equations for holes and electrons, respectively. Since both the hole concentrationp and the electron concentration n contain the excess concentrations, Equations (6.27) and (6.28) describe the space and time behavior of the excess carriers.

The hole and electron concentrations are functions of both the thermal equilibrium and the excess values are given in Equations (6.5a) and (6.5b). The thermal-equilibrium concentrations, n_0 and p_0, are not functions of time. For the special case of a homogeneous semiconductor, n_0 and p_0 are also independent of the space coordinates. Equations (6.27) and (6.28) may then be written in the form

$$D_p \frac{\partial^2 (\delta p)}{\partial x^2} - \mu_p \left(\mathrm{E}\frac{\partial (\delta p)}{\partial x} + p\frac{\partial \mathrm{E}}{\partial x} \right) + g_p - \frac{p}{\tau_{pt}} = \frac{\partial (\delta p)}{\partial t} \tag{6.29}$$

$$D_n \frac{\partial^2 (\delta n)}{\partial x^2} + \mu_n \left(\mathrm{E}\frac{\partial (\delta n)}{\partial x} + n\frac{\partial \mathrm{E}}{\partial x} \right) + g_n - \frac{n}{\tau_{nt}} = \frac{\partial (\delta n)}{\partial t} \tag{6.30}$$

Note that the Equations (6.29) and (6.30) contain terms involving the total concentrations, p and n, and terms involving only the excess concentrations, δp and δn.

6.3 | AMBIPOLAR TRANSPORT

Originally, we assumed that the electric field in the current Equations (6.20) and (6.21) was an applied electric field. This electric field term appears in the time-dependent diffusion equations given by Equations (6.29) and (6.30). If a pulse of

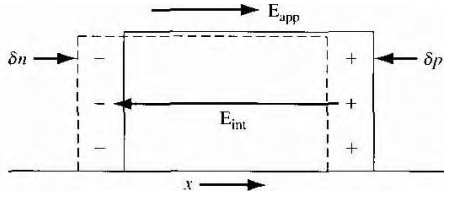


Figure 6.5 | The creation ot an internal electric field as excess electrons and holes rend to separate.

excess electrons and a pulse of excess holes are created at a particular point in a semiconductor with an applied electric field, the excess holes and electrons *will tend* to drift in opposite directions. However, because the electrons and holes are charged particles, any separation will induce an internal electric field between the two sets of particles. This internal electric field will create a force attracting the electrons and holes back toward each other. This effect is shown in Figure 6.5. The electric field term in Equations (6.29) and (6.30) is then composed of the externally applied field plus the induced internal field. This E-field may be written as

$$E = E_{app} + E_{int} \tag{6.31}$$

where E_{app} is the applied electric field and E_{int} is the induced internal electric field.

Since the internal E-field creates a force attracting the electrons and holes. this E-field will hold the pulses of excess electrons and excess holes together. The negatively charged electrons and positively charged holes then will drift or diffuse together with a single effective mobility or diffusion coefficient. This phenomenon is called *ambipolar diffusion* or *ambipolar transport.*

6.3.1 Derivation of the Ambipolar Transport Equation

The time-dependent diffusion Equations (6.29) and (6.30) describe the behavior of the excess carriers. However, a third equation is required to relate the excess electron and hole concentrations to the internal electric field. This relation is Poisson's equation, which may be written as

$$\nabla \cdot E_{int} = \frac{e(\delta p - \delta n)}{\epsilon_s} = \frac{\partial E_{int}}{\partial x} \tag{6.32}$$

where ϵ_s is the permittivity of the semiconductor material.

To make the solution of Equations (6.29), (6.30). and (6.32) more tractable, we need to make some approximations. We can show that only a relatively small internal electric field is sufficient to keep the excess electrons and holes drifting and diffusing together. Hence, we can assume that

$$|E_{int}| \ll |E_{app}| \tag{6.33}$$

However. the V . E_{int} term may not be negligible. We will impose the condition of charge neutrality: We will assume that the excess electron concentration is just balanced by an equal excess hole concentration at any point in space and time. If this assumption were exactly true, there would be no induced internal electric field to keep the two sets of panicles together. However, only a very small difference in the excess electron concentration and excess hole concentration will set up an internal E-field rufficient to keep the particles diffusing and drifting together. We can show that a I percent difference in δp and *Sn.* for example, will result in non-negligible values of the V . $E = V$. E_{int} term in Equations *(6.29)* and *(6.30).*

We can combine Equations *(6.29)* and *(6.30)* to eliminate the V . E term. Considering Equations (6.1) and (6.4), we can define

$$g_n = g_p \equiv g \tag{6.34}$$

and considering Equations *(6.2)* and (6.6), we can define

$$R_n = \frac{n}{\tau_{nt}} = R_p = \frac{p}{\tau_{pt}} \equiv R \tag{6.35}$$

The lifetimes in Equation *(6.35)* include the thermal-equilibrium carrier lifetimes and theexcess-carrier lifetimes. If we impose thecharge neutrality condition, then $\delta n \approx Sp$. We will denote both the excess electron and excess hole concentrations in Equations *(6.29)* and *(6.30)* by δn. We may then rewrite Equations *(6.29)* and *(6.30)* as

$$D_p \frac{\partial^2(\delta n)}{\partial x^2} - \mu_p \left(E \frac{\partial(\delta n)}{\partial x} + p \frac{\partial E}{\partial x} \right) + g - R = \frac{\partial(\delta n)}{\partial t} \tag{6.36}$$

$$D_n \frac{\partial^2(\delta n)}{\partial x^2} + \mu_n \left(E \frac{\partial(\delta n)}{\partial x} + n \frac{\partial E}{\partial x} \right) + g - R = \frac{\partial(\delta n)}{\partial t} \tag{6.37}$$

If we multiply Equation *(6.36)* by $\mu_n n$, multiply Equation *(6.37)* by $\mu_p p$, and add the two equations, the V . E = $\partial E/\partial x$ term will be eliminated. The result of this addition gives

$$(\mu_n n D_p + \mu_p p D_n) \frac{\partial^2(\delta n)}{\partial x^2} + (\mu_n \mu_p)(p - n) E \frac{\partial(\delta n)}{\partial x}$$

$$+ (\mu_n n + \mu_p p)(g - R) = (\mu_n n + \mu_p p) \frac{\partial(\delta n)}{\partial t} \tag{6.38}$$

If we divide Equation *(6.38)* by the term $(\mu_n n + \mu_p p)$, this equation becomes

$$\boxed{D' \frac{\partial^2(\delta n)}{\partial x^2} + \mu' E \frac{\partial(\delta n)}{\partial x} + g - R = \frac{\partial(\delta n)}{\partial t}} \tag{6.39}$$

where

$$D' = \frac{\mu_n n D_p + \mu_p p D_n}{\mu_n n + \mu_p p} \tag{6.40}$$

and

$$\mu' = \frac{\mu_n \mu_p (p - n)}{\mu_n n + \mu_p p} \tag{6.41}$$

Equation (6.39) is called the *ambipolar* transport *equation* and describes the behavior of the excess electrons and holes in time and space. The parameter D' is called the *ambipolar diffusion* coefficient and μ' is called the *ambipolar mobility.*

The Einstein relation relates the mobility and diffusion coefficient by

$$\frac{\mu_n}{D_n} = \frac{\mu_p}{D_p} = \frac{e}{kT} \tag{6.42}$$

Using these relations, the ambipolar diffusion coefficient may be written in the form

$$D' = \frac{D_n D_p (n + p)}{D_n n + D_p p} \tag{6.43}$$

The ambipolar diffusion coefficient, D', and the ambipolar mobility, μ', are functions of the electron and hole concentrations, n and p, respectively. Since both n and p contain the excess-carrier concentration δn, the coefficient in the ambipolar transport equation are not constants. The ambipolar transport equation, given by Equation (6.39), then, is a nonlinear differential equation.

6.3.2 Limits of Extrinsic Doping and Low Injection

The ambipolar transport equation may be simplified and linearized by considering an extrinsic semiconductor and by considering low-level injectinn. The ambipolar diffusion coefficient, from Equation (6.43), may be written as

$$D' = \frac{D_n D_p [(n_0 + \delta n) + (p_0 + \delta n)]}{D_n (n_0 + \delta n) + D_p (p_0 + \delta n)} \tag{6.44}$$

where n_0 and p_0 are the thermal-equilibrium electron and hole concentrations, respectively. and Sn is the excess carrier concentration. If we consider a p-type semiconductor, we can assume that $p_0 \gg n_0$. The condition of low-levelinjection, or just low injection, means that the excess carrier concentration is much smaller than the thermal-equilibrium majority carrier concentration. For the p-type semiconductor, then, low injection implies that $\delta n \ll p_0$. Assuming that $n_0 \ll p_0$ and $\delta n \ll p_0$, and assuming that D_n and D_p are on the same order of magnitude. the ambipolar diffusion coefficient from Equation (6.44) reduces to

$$D' = D_n \tag{6.45}$$

If we apply the conditions of an extrinsic p-type semiconductor and low injection to the ambipolar mobility, Equation (6.41) reduces to

It is important to *note that for* an *extrinsic p-type semiconductor* under low injection, the ambipolar *diffusion coefficient* and the *ambipolar* mobility *coefficient reduce* to *the* minorig-currier *electron parameter values, which* are constants. The ambipolar transport equation reduces to a linear differential equation with constant coefficients.

If we now consider an extrinsic n-type semiconductor under low injection, we may assume that $p_0 \ll n_0$ and $\delta n \ll n_0$. The ambipolar diffusion coefficient from Equation *(6.43)* reduces to

$$D' = D_p \tag{6.47}$$

and the ambipolar mobility from Equation *(6.41)* reduces to

$$\mu = -\mu_p \tag{6.48}$$

The ambipolar parameter5 *again* reduce to *the* minorit?-currier *values,* which are constants. Note that, for the n-type semiconductor, the amhipolar mobility is a negative value. The ambipolar mobility term is associated with carrier drift; therefore. the sign of the drift term depends on the charge of the particle. The equivalent amhipolar particle is negatively charged, as one can see by comparing Equations (6.30) and *(6.39).* If the amhipolar mobility reduces to that of a positively charged hole, a negative sign is introduced as shown in Equation *(6.48).*

The remaining terms we need to consider in the ambipolar transport equation are the generation rate and the recombination rate. Recall that the electron and hole recombination rates are equal and were given by Equation (6.35) as $R_n = R_p = n/\tau_{nt} = p/\tau_{pt} \equiv$ R. where τ_{nt} and τ_{pt} are the mean electron and hole lifetimes, respectively. If we consider the inverse lifetime functions, then $1/\tau_{nt}$ is the probability per unit time that an electron will encounter a hole and recombine. Likewise, $1/\tau_{pt}$ is the probability per unit time that a hole will encounter an electron and recombine. If we again consider an extrinsic p-type semiconductor under low injection, the concentration of majority carrier holes will be essentially constant, even when excess carriers are present. Then, the probability per unit time of a minority carrier electron encountering a majority carrier hole will be essentially constant. Hence $\tau_{nt} \equiv \tau_n$, the minority carrier electron lifetime, will remain a constant for the extrinsic p-type semiconductor under low injection.

Similarly, if we consider an extrinsic n-type semiconductor under low injection, the minority carrier hole lifetime, $\tau_{pt} \equiv$ r,,, will remain constant. Even under the condition of low injection, the minority carrier hole concentration may increase by several orders of magnitude. The probability per unit time of a majority carrier electron encountering a hole may change drastically. The majority carrier lifetime, then, may change substantially when excess carriers are present.

Consider, again, the generation and recombination terms in the ambipolar transport equation. For electrons we may write

$$g - R = g_n - R_{,,} = (G_{n0} + g_n') - (R_{n0} + R_n') \tag{6.49}$$

where G_{n0} and g_n' are the thermal-equilibrium electron and excess electron generation rates, respectively. The terms R_{n0} and R_n' are the thermal-equilibrium electron

and excess electron recombination rates, respectively. For thermal equilibrium, we have that

$$G_{n0} = R_{n0} \tag{6.50}$$

so Equation (6.49) reduces to

$$g - R = g_n' - R_n' = g_n' - \frac{\delta n}{\tau_n} \tag{6.51}$$

where τ_n is the excess minority carrier electron lifetime.

For the case of holes, we may write

$$g - R = g_p - R_p = (G_{p0} + g_p') - (R_{p0} + R_p') \tag{6.52}$$

where G_{p0} and g_p' are the thermal-equilibrium hole and excess hole generation rates, respectively. The terms R_{p0} and R_p' are the thermal-equilibrium hole and excess hole recombination rates, respectively. Again, for thermal equilibrium, we have that

$$G_{p0} = R_{p0} \tag{6.53}$$

so that Equation *(6.52)* reduces to

$$g - R = g_p' - R_p' = g_p' - \frac{\delta p}{\tau_p} \tag{6.54}$$

where τ_p is the excess minority carrier hole lifetime.

The generation rate for excess electrons must equal the generation rate for excess holes. We may then define a generation rate for excess carriers as g', so that $g_n' = g_p' \equiv$ g'. We also determined that the minority carrier lifetime is essentially a constant for low injection. Then the term g – R in the ambipolar transport equation may he written in terms of the minority-carrier parameters.

The ambipolar transport equation, given by Equation (6.39), for a p-type semiconductor under low injection then becomes

$$\boxed{D_n \frac{\partial^2(\delta n)}{\partial x^2} + \mu_n \mathrm{E} \frac{\partial(\delta n)}{\partial x} + g' - \frac{\delta n}{\tau_{n0}} = \frac{\partial(\delta n)}{\partial t}} \tag{6.55}$$

The parameter δn is the excess minority carrier electron concentration, the parameter τ_{n0} is the minority carrier lifetime under low injection, and the other parameters are the usual minority carrier electron parameters.

Similarly, for an extrinsic n-type semiconductor under low injection, the ambipolar transport equation becomes

$$\boxed{D_p \frac{\partial^2(\delta p)}{\partial x^2} - \mu_p \mathrm{E} \frac{\partial(\delta p)}{\partial x} + g' - \frac{\delta p}{\tau_{p0}} = \frac{\partial(\delta p)}{\partial t}} \tag{6.56}$$

The parameter δp is the excess minority carrier hole concentration, the parameter τ_{p0} is the minority carrier hole lifetime under low injection, and the other parameters are the usual minority carrier hole parameters.

Table 6.2 | Common ambipolar transport equation simplifications

Specification	Effect
Steady slate	$\frac{\partial(\delta n)}{\partial t} = 0, \quad \frac{\partial(\delta p)}{\partial t} = 0$
Uniform distribution of excess carriers (uniform generation rate)	$D_n \frac{\partial^2(\delta n)}{\partial x^2} = 0, \quad D_p \frac{\partial^2(\delta n)}{\partial x^2} = 0$
Zero electric field	$E\frac{\partial(\delta n)}{\partial x} = 0, \quad E\frac{\partial(\delta p)}{\partial x} = 0$
No excess carrier generation	$g' = 0$
No excess carrier recombination (infinite lifetime)	$\frac{\delta n}{\tau_{n0}} = 0, \quad \frac{\delta p}{\tau_{p0}} = 0$

It is extremely important to note that the transport and recombination parameters in Equations (6.55) and (6.56) are those of the minority carrier. *Equations (6.55) and (6.56) describe the drift, diffusion, and recombination of excess minority carriers as a function of spatial coordinates and as a function of time.* Recall that we had imposed the condition of charge neutrality; the excess minority carrier concentration is equal to the excess majority carrier concentration. The excess majority carriers, then, diffuse and drift with the excess minority carriers; thus, the behavior of the excess majority carrier is determined by the minority carrier parameters. This ambipolar phenomenon is extremely important in semiconductor physics, and is the basis for describing the characteristics and behavior of semiconductor devices.

6.3.3 Applications of the Ambipolar Transport Equation

We will solve the ambipolar transport equation for several problems. These examples will help illustrate the behavior of excess carriers in a semiconductor material, and the results will be used later in the discussion of the pn junction and the other semiconductor devices.

The following examples use several common simplifications in the solution of the ambipolar transport equation. Table 6.2 summarizes these simplifications and their effects.

EXAMPLE 6.1

Objective

To determine the time behavior of excess carriers as a semiconductor returns to thermal equilibrium.

Consider an infinitely large, homogeneous n-type semiconductor with zero applied electric field. Assume that at time $t = 0$, a uniform concentration of excess carriers exists in the crystal, but assume that $g' = 0$ for $t > 0$. If we assume that the concentration of excess carriers is much smaller than the thermal-equilibrium electron concentration, then the low-injection condition applies. Calculate the excess carrier concentration as a function of time for $t \geq 0$.

■ Solution

For the n-type semiconductor, we need to consider the ambipolar transport equation for the minority carrier holes, which was given by Equation (6.56). The equation is

$$D_p \frac{\partial^2(\delta p)}{\partial x^2} - \mu_p E \frac{\partial(\delta p)}{\partial x} + g' - \frac{\delta p}{\tau_{p0}} = \frac{\partial(\delta p)}{\partial t}$$

We are assuming a uniform concentration of excess holes so that $\partial^2(\delta p)/\partial x^2 = \partial(\delta p)/\partial x = 0$. For $t > 0$, we are also assuming that $g' = 0$. Equation (6.56) reduces to

$$\frac{d(\delta p)}{dt} = -\frac{\delta p}{\tau_{p0}} \tag{6.57}$$

Since there is no spatial variation, the total time derivative may be used. At low injection, the minority carrier hole lifetime, τ_{p0}, is a constant. The solution to Equation *(6.57)* is

$$\boxed{\delta p(t) = \delta p(0) e^{-t/\tau_{p0}}} \tag{6.58}$$

where $\delta p(0)$ is the uniform concentration of excess carriers that exists at time $t = 0$. The concentration of excess holes decays exponentially with time, with a time constant equal to the minority carrier hole lifetime.

From the charge-neutrality condition, we have that $\delta n = \delta p$, so the excess electron concentration is given by

$$\delta n(t) = \delta p(0) e^{-t/\tau_{p0}} \tag{6.59}$$

■ Numerical Calculation

Consider n-type gallium arsenide doped at $N_d = 10^{16}$ cm^{-3}. Assume that 10^{14} electron-hole pairs per cm^3 have been created at $t = 0$, and assume the minority carrier hole lifetime is $\tau_{p0} = 10$ ns.

We may note that $\delta p(0) \ll n_0$, so low injection applies. Then from Equation (6.58) we can write

$$\delta p(t) = 10^{14} e^{-t/10^{-8}} \text{ cm}^{-3}$$

The excess hole and excess electron concentrations will decay to $1/e$ of their initial value in 10 ns.

■ Comment

The excess electrons and holes recombine at the rate determined by the excess minority carrier hole lifetime in the n-type semiconductor.

EXAMPLE 6.2 Objective

To determine the time dependence of excess carriers in reaching a steady-state condition.

Again consider an infinitely large, homogeneous n-type semiconductor with a zero applied electric field. Assume that, for $t < 0$, the semiconductor is in thermal equilibrium and

that, for $t \geq 0$, a uniform generation rate exists in the crystal. Calculate the excess carrier concentration as a function of time assuming the condition of low injection.

Solution

The condition of a uniform generation rate and a homogeneous semiconductor again implies that $\partial^2(\delta p)/\partial x^2 = \partial(\delta p)/\partial x = 0$ in Equation *(6.56)*. The equation, for this case, reduces to

$$g' - \frac{\delta p}{\tau_{p0}} = \frac{d(\delta p)}{dt} \tag{6.60}$$

The solution to this differential equation is

$$\boxed{\delta p(t) = g'\tau_{p0}(1 - e^{-t/\tau_{p0}})} \tag{6.61}$$

■ Numerical Calculation

Considern-type silicon at T = 300 K doped at $N_d = 2 \times 10^{16}$ cm^{-3}. Assume that $\tau_{p0} = 10^{-7}$ s and g' = 5×10^{21} cm^{-3} s^{-1}. From Equation *(6.61)* we can write

$$\delta p(t) = (5 \times 10^{21})(10^{-7})\left[1 - e^{-t/10^{-7}}\right] = 5 \times 10^{14}\left[1 - e^{-t/10^{-7}}\right] \text{ cm}^{-3}$$

■ Comment

We may note that fort $\rightarrow \infty$. we will create a steady-state excess hole and electron concentration of 5×10^{14} cm^{-3}. We may note that $\delta p \ll n_0$, so low injection is valid.

The excess minority carrier hole concentration increases with time with the same time constant τ_{p0}, which is the excess minority carrier lifetime. The excess carrier concentration reaches a steady-state value as time goes to infinity, even though a steady-state generation of excess electrons and holes exists. This steady-state effect can he seen from Equation (6.60) by setting $d(\delta p)/dt = 0$. The remaining tenns simply state that, in steady state, the generation rate is equal to the recotnbination rate.

TEST YOUR UNDERSTANDING

E6.3 Silicon at T = 300 K has been doped with boron atoms to a concentration of $N_a = 5 \times 10^{16}$ cm^{-3}. Excess carriers have been generated in the uniformly doped material to a concentration of 10^{15} cm^{-3}. The minority carrier lifetime is 5 μs. (o) What carrier type is the minority carrier? *(b)* Assuming g' = E = 0 for $t > 0$, determine the minority carrier concentration for $t > 0$.

[(a) electrons (b) $10^{15}e^{-t/5\times10^{-6}}$ cm^{-3}]

E6.4 Consider silicon with the same parameters as given in **E6.3.** The material is in thermal equilibrium fort < 0. At $t = 0$, a source generating excess carriers is turned on, producing a generation rate of g' = 10^{20} cm^{-3}-s^{-1}. ***(a)*** What carrier type is the minority carrier? *(b)* Determine the minority carrier concentration fort > 0. ***(c)*** What is the minority carrier concentration as $t \rightarrow \infty$?

[(a) electrons (b) $5 \times 10^{14}[1 - e^{-t/5\times10^{-6}}]$ cm^{-3}, (c) 5×10^{14} cm^{-3}]

EXAMPLE 6.3 Objective

To determine the steady-state spatial dependence of the excess carrier concentration.

Consider a p-type semiconductor that is homogeneous and infinite in extent. Assume a zero applied electric field. For a one-dimensional crystal, assume that excess carriers are being generated at $x = 0$ only, as indicated in Figure 6.6. The excess carriers being generated at $x = 0$ will begin diffusing in both the $+x$ and $-x$ directions. Calculate the steady-state excess carrier concentration as a function of x.

Solution

The ambipolar transport equation lor excess minority carrier electrons was given by Equation (6.55), and is written as

$$D_n \frac{\partial^2(\delta n)}{\partial x^2} + \mu_n E \frac{\partial(\delta n)}{\partial x} + g' - \frac{\delta n}{\tau_{n0}} = \frac{\partial(\delta n)}{\partial t}$$

From our assumptions. we have $E = 0, g' = 0$ for $x \neq 0$, and $\partial(\delta n)/\partial t = 0$ fur steady state. Assuming a one-dimensional crystal, Equation (6.55) reduces to

$$D_n \frac{d^2(\delta n)}{dx^2} - \frac{\delta n}{\tau_{n0}} = 0 \tag{6.62}$$

Dividing by the diffusion coefficient, Equation (6.62) may he written as

$$\frac{d^2(\delta n)}{dx^2} - \frac{\delta n}{D_n \tau_{n0}} = \frac{d^2(\delta n)}{dx^2} - \frac{\delta n}{L_n^2} = 0 \tag{6.63}$$

where we have defined $L_n^2 = D_n \tau_{n0}$. The parameter L_n has the unit of length and is called the minority carrier electron diffusion length. The general solution to Equation (6.63) is

$$\delta n(x) = Ae^{-x/L_n} + Be^{x/L_n} \tag{6.64}$$

As the minority carrier electrons diffuse away from $x = 0$. they will recombine with the majority carrier holes. The minority carrier electron concentration will then decay toward zero at both $x = +\infty$ and $x = -\infty$. These boundary conditions mean that $B \equiv 0$ for $x > 0$ and $A \equiv 0$ for $x < 0$. The solution to Equation (6.63) may then he written as

$$\boxed{\delta n(x) = \delta n(0) e^{-x/L_n} \qquad x \geq 0} \tag{6.65a}$$

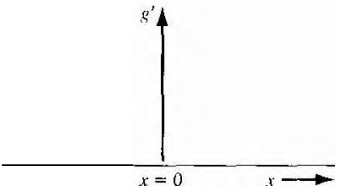


Figure 6.6 | Steady-state generation rate at $x = 0$.

$$\delta n(x) = \delta n(0) e^{+x/L_n} \qquad x \leq 0 \tag{6.65b}$$

where $\delta n(0)$ is the value of the excess electron concentration at x = 0. The steady-state excess electron concentration decays exponentially with distance away from the source at $x = 0$.

■ Numerical Calculation

Consider p-type silicon at T = 300 K doped at $N_a = 5 \times 10^{16}$ cm^{-3}. Assume that $\tau_{n0} = 5 \times 10^{-7}$ s, $D_n = 25$ cm^2/s, and $\delta n(0) = 10^{15}$ cm^{-3}.

The minority carrier diffusion length is

$$L_n = \sqrt{D_n \tau_{n0}} = \sqrt{(25)(5 \times 10^{-7})} = 35.4\ \mu\text{m}$$

Then for $x \geq 0$, we have

$$\delta n(x) = 10^{15} e^{-x/35.4 \times 10^{-4}}\ \text{cm}^{-3}$$

■ Comment

We may note that the steady-state excess concentration decays to $1/e$ of its value at x = 35.4 μm.

As before, we will assume charge neutrality; thus, the steady-state excess majority carrier hole concentration also decays exponentially with distance with the same characteristic minority carrier electron diffusion length L_n. Figure 6.7 is a plot of the total electron and hole concentrations as a function of distance. We are assuming low injection, that is, $\delta n(0) \ll p_0$ in the p-type semiconductor. The total concentration of majority carrier holes barely changes. However, we may have $\delta n(0) \gg n_0$ and still satisfy the low-injection condition. The minority carrier concentration may change by many orders of magnitude.

TEST YOUR UNDERSTANDING

E6.5 Excess electrons and holes are generated at the end of a silicon bar ($x = 0$). The silicon is doped with phosphorus atoms to a concentration of $N_d = 10^{17}$ cm^{-3}. The minority carrier lifetime is 1 μs, the electron diffusion coefficient is $D_n = 25$ cm^2/s, and the hole diffusion coefficient is $D_p = 10$ cm^2/s. If $\delta n(0) = \delta p(0) = 10^{15}$ cm^{-3}, determine the steady-state electron and hole concentrations in the silicon for $x > 0$. (Ans. $\delta n = \delta p = 10^{15} e^{-x/3.16 \times 10^{-3}}$ cm^{-3}, where x is in cm)

E6.6 Using the parameters given in E6.5, calculate the electron and hole diffusion current densities at x = 10 μm. (Ans. $J_p = +0.369$ A/cm^2, $J_n = -0.369$ A/cm^2)

The three previous examples, which applied the amhipolar transport equation to specific situations, assumed either a homogeneous or a steady-state condition; only the time variation or the spatial variation was considered. Now consider an example in which both the time and spatial dependence are considered in the same problem.

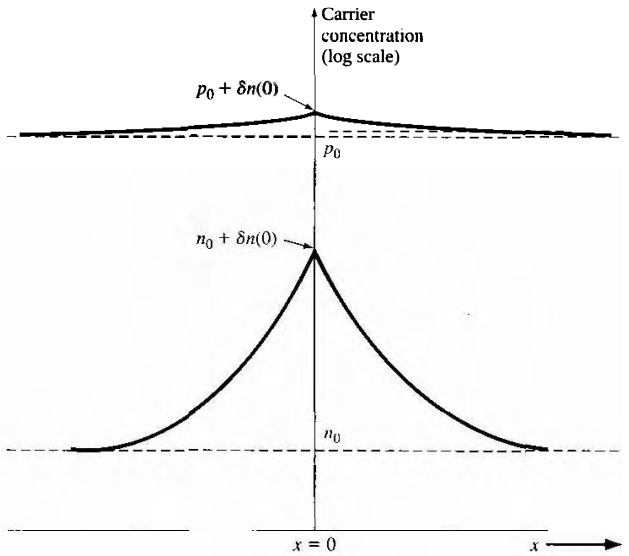


Figure 6.7 l Steady-state electron and hole concentrations for the case when excess electrons and holes are generated at $x = 0$.

EXAMPLE 6.4 Objective

To determine both the time dependence and spatial dependence of the excess carrier concentration.

Assume that a finite number of electron–hole pairs is generated instantaneously at time $t = 0$ and at $x = 0$, but assume $g' = 0$ for $t > 0$. Assume we have an n-type semiconductor with a constant applied electric field equal to E_0, which is applied in the $+x$ direction. Calculate the excess carrier concentration as a function of x and t.

Solution

The one-dimensional ambipolar transport equation for the minority carrier holes can be written from Equation (6.56) as

$$D_p \frac{\partial^2(\delta p)}{\partial x^2} - \mu_p E_0 \frac{\partial(\delta p)}{\partial x} - \frac{\delta p}{\tau_{p0}} = \frac{\partial(\delta p)}{\partial t} \tag{6.66}$$

The solution to this partial differential equation is of the form

$$\delta p(x, t) = p'(x, t) e^{-t/\tau_{p0}} \tag{6.67}$$

By substituting Equation (6.67) into Equation (6.66), we are left with the partial differential equation

$$D_p \frac{\partial^2 p'(x,t)}{\partial x^2} - \mu_p E_0 \frac{\partial p'(x,t)}{\partial x} = \frac{\partial p'(x,t)}{\partial t} \tag{6.68}$$

Equation (6.68) is normally salved using Laplace transform techniques. The solution, without going through the mathematical details, is

$$p'(x,t) = \frac{1}{(4\pi D_p t)^{1/2}} \exp\left[\frac{-(x-\mu_p E_0 t)^2}{4D_p t}\right] \tag{6.69}$$

The total solution. from Equations (6.67) and (6.69). for the excess minority carrier hole concentration is

$$\delta p(x,t) = \frac{e^{-t/\tau_{p0}}}{(4\pi D_p t)^{1/2}} \exp\left[\frac{-(x-\mu_p E_0 t)^2}{4D_p t}\right] \tag{6.70}$$

■ Comment

We could show that Equation (6.70) is a solution hy direct substitution back into the partial differential equation. Equation (6.66).

Equation (6.70) can be plotted as a function of distance x, for various times. Figure 6.8 shows such a plot for the case when the applied electric field is zero. For $t > 0$, the excess minority carrier holes diffuse in both the $+x$ and $-x$ directions. During this time, the excess majority carrier electrons, which were generated, diffuse at exactly the same rate as the holes. **As** time proceeds, the excess holes recombine

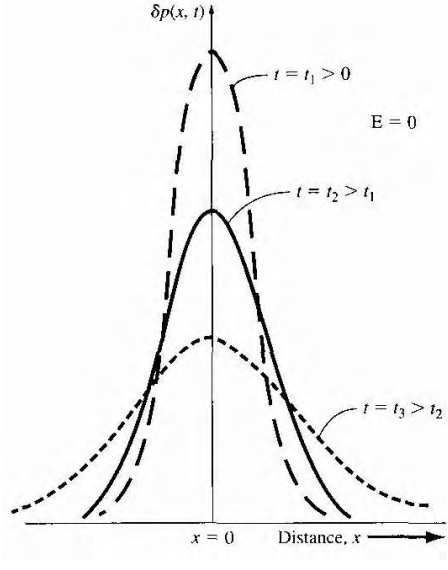


Figure 6.8 | Excess-hole concentration versus distance at various times for zero applied electric field.

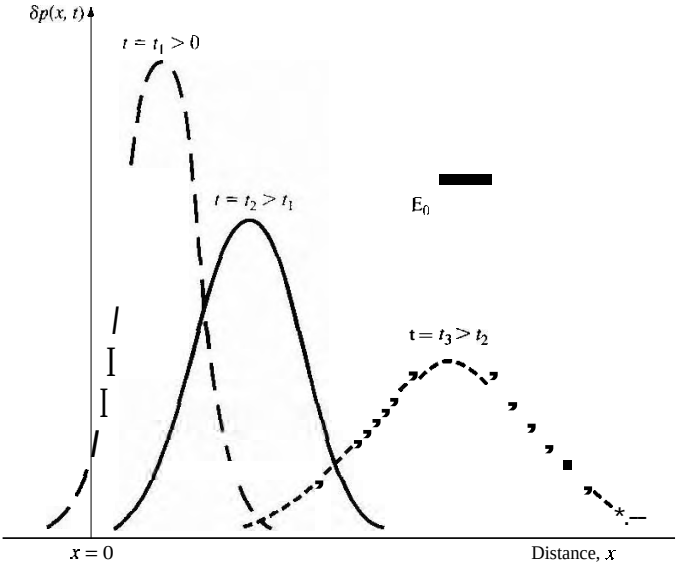


Figure **6.9** | Excess-hole concentration versus distance at various times for a constant applied electric field.

with the excess electrons so that ar $t = \infty$ the excess hole concentration is zero. In this particular example, both diffusion and recombination processes are occurring at the same time.

Figure 6.9 shows a plot of Equation (6.70) as a function of distance x at various times for the case when the applied electric field is not zero. In this case, the pulse of excess minority carrier holes is drifting in the $+x$ direction, which is the direction of the electric field. We still have the same diffusion and recombination processes as we had before. An important point to consider is that, with charge neutrality, $\delta n = \text{Spat}$ any instant of time and at any point in space. The excess-electron concentration is equal to the excess-hole concentration. In this case, then, the excess-electron pulse is moving in the same direction as the applied electric field even though the electrons have a negative charge. In the ambipolar transport process, the excess carriers are characterized by the minority carrier parameters. In this example, the excess carriers behave according to the minority carrier hole parameters, which include D_p, μ_p, and τ_{p0}. The excess majority carrier electrons are being pulled along by the excess minority canier holes.

TEST YOUR UNDERSTANDING

E6.7 As a good approximation, the peak value of a normalized excess canierconcentration, given by Equation (6.70), occurs at $x = \mu_p E_0 t$. Assume the following parameters:

$\tau_{p0} = 5\ \mu s$, $D_p = 10\ cm^2/s$, $\mu_p = 386\ cm^2/V\text{-}s$, and $E_0 = 10\ V/cm$. Calculate the peak value at times of (a) $t = 1\ \mu s$, (b) $t = 5\ \mu s$, (c) $t = 15\ \mu s$, and (d) $t = 25\ \mu s$. What are the corresponding values of x for parts (a) to (d)? [Ans. (a) 73.0, $x = 38.6\ \mu m$; (b) 14.7, $x = 193\ \mu m$; (c) 1.15, $x = 579\ \mu m$; (d) 0.120, $x = 965\ \mu m$]

E6.8 The excess carrier concentration, given by Equation (6.70), is to be calculated at distances of one diffusion length away from the peak value. Using the parameters given in E6.7, calculate the values of δp for (a) $t = 1\ \mu s$ at (i) 1.093×10^{-2} cm and (ii) $x = -3.21 \times 10^{-3}$ cm; (b) $t = 5\ \mu s$ at (i) $x = 2.64 \times 10^{-2}$ cm and (ii) $x = 1.22 \times 10^{-2}$ cm; (c) $t = 15\ \mu s$ at (i) $x = 6.50 \times 10^{-2}$ cm and (ii) $x = 5.08 \times 10^{-2}$ cm.
[Ans. (a) (i) 20.9, (ii) 20.9; (b) (i) 11.4, (ii) 11.4; (c) (i) 1.05, (ii) 1.05]

E6.9 Using the parameters given in E6.7, (a) plot $\delta p(x, t)$ from Equation (6.70) versus x for (i) $t = 1\ \mu s$, (ii) $t = 5\ \mu s$, and (iii) $t = 15\ \mu s$, and (b) plot $\delta p(x, t)$ versus time for (i) $x = 10^{-2}$ cm, (ii) $x = 3 \times 10^{-2}$ cm, and (iii) $x = 6 \times 10^{-2}$ cm.

6.3.4 Dielectric Relaxation Time Constant

We have assumed in the previous analysis that a quasi-neutrality conditions exists—that is, the concentration of excess holes is balanced by an equal concentration of excess electrons. Suppose that we have a situation as shown in Figure 6.10, in which a uniform concentration of holes δp is suddenly injected into a portion of the surface of a semiconductor. We will instantly have a concentration of excess holes and a net positive charge density that is not balanced by a concentration of excess electrons. How is charge neutrality achieved and how fast?

There are three defining equations to he considered. Poisson's equation is

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon} \tag{6.71}$$

The current equation, Ohm's law, is

$$J = \sigma \mathbf{E} \tag{6.72}$$

The continuity equation, neglecting the effects of generation and recombination, is

$$\nabla \cdot J = -\frac{\partial \rho}{\partial t} \tag{6.73}$$

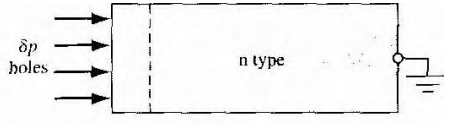


Figure 6.10 | The injection of a concentration of holes into a small region at the surface of an n-type semiconductor.

The parameter ρ is the net charge density and the initial value is given by $e(\delta p)$. We will assume that δp is uniform over a short distance at the surface. The parameter ϵ is the permittivity of the semiconductor.

Taking the divergence of Ohm's law and using Poisson's equation, we find

$$\nabla \cdot J = \sigma \nabla \cdot E = \frac{\sigma \rho}{\epsilon} \tag{6.74}$$

Substituting Equation (6.74) into the continuity equation, we have

$$\frac{\sigma \rho}{\epsilon} = -\frac{\partial \rho}{at} = -\frac{d\rho}{dr} \tag{6.75}$$

Since Equation (6.75) is a function of time only, we can write the equation as a total derivative. Equation (6.75) can be rearranged as

$$\frac{d\rho}{dt} + \left(\frac{\sigma}{\epsilon}\right)\rho = 0 \tag{6.76}$$

Equation (6.76) is a first-order differential equation whose solution is

$$\rho(t) = \rho(0)e^{-(t/\tau_d)} \tag{6.77}$$

where

$$\tau_d = \frac{\epsilon}{\sigma} \tag{6.78}$$

and is called the dielectric relaxation time constant.

EXAMPLE 6.5 Objective

Calculate the dielectric relaxation time constant for a particular semiconductor.

Assume an n-type semiconductor with a donor impurity concentration of $N_d = 10^{16}$ cm^{-3}.

■ **Solution**

The conductivity is found as

$$\sigma \approx e\mu_n N_d = (1.6 \times 10^{-19})(1200)(10^{16}) = 1.92\ (\Omega\text{-cm})^{-1}$$

where the value of mobility is the approximate value found from Figure 5.3. The permittivity of silicon is

$$\epsilon = \epsilon_r \epsilon_0 = (11.7)(8.85 \times 10^{-14})\ \text{F/cm}$$

The dielectric relaxation time constant is then

$$\tau_d = \frac{\epsilon}{\sigma} = \frac{(11.7)(8.85 \times 10^{-14})}{1.92} = 5.39 \times 10^{-13}\ \text{s}$$

■ Comment

Equation (6.77) then predicts that in approximately four time constants, or in approximately 2 ps, the net charge density is essentially zero; that is, quasi-neutrality has been achieved. Since the continuity equation, Equation (6.73), does not contain any generation or recombination terms, the initial positive charge is then neutralized by pulling in electrons from the bulk n-type material to create excess electrons. This process occurs very quickly compared to the normal excess carrier lifetimes of approximately 0.1 μs. The condition of quasi-charge-neutrality is then justified.

*6.3.5 Haynes–Shockley Experiment

We have derived the mathematics describing the behavior of excess carriers in a semiconductor. The Haynes–Shockley experiment was one of the first experiments to actually measure excess-carrier behavior.

Figure 6.11 shows the basic experimental arrangement. The voltage source V_1 establishes an applied electric field E_0 in the $+x$ direction in the n-type semiconductor sample. Excess carriers are effectively injected into the semiconductor at contact A. Contact B is a rectifying contact that is under reverse bias by the voltage source V_2. The contact B will collect a fraction of the excess carriers as they drift through the semiconductor. The collected carriers will generate an output voltage, V_0.

This experiment corresponds to the problem we discussed in Example 6.4. Figure 6.12 shows the excess-carrier concentrations at contacts A and B for two conditions. Figure 6.12a shows the idealized excess-carrier pulse at contact A at time $t = 0$. For a given electric field E_{01}, the excess carriers will drift along the semiconductor producing an output voltage as a function of time given in Figure 6.12b. The peak of the pulse will arrive at contact B at time t_0. If the applied electric field is reduced to a value E_{02}, $E_{02} < E_{01}$, the output voltage response at contact B will look approximately as shown in Figure 6.12c. For the smaller electric field, the drift velocity of the pulse of excess carriers is smaller, and so it will take a longer time for the

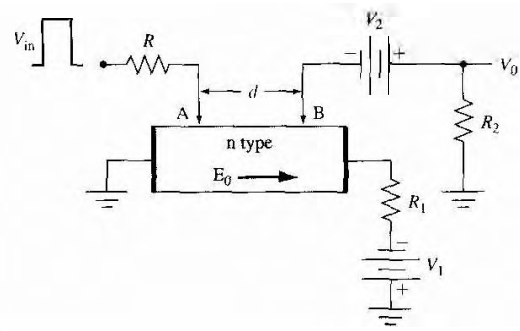


Figure 6.11 | The basic Haynes–Shockley experimental arrangement.

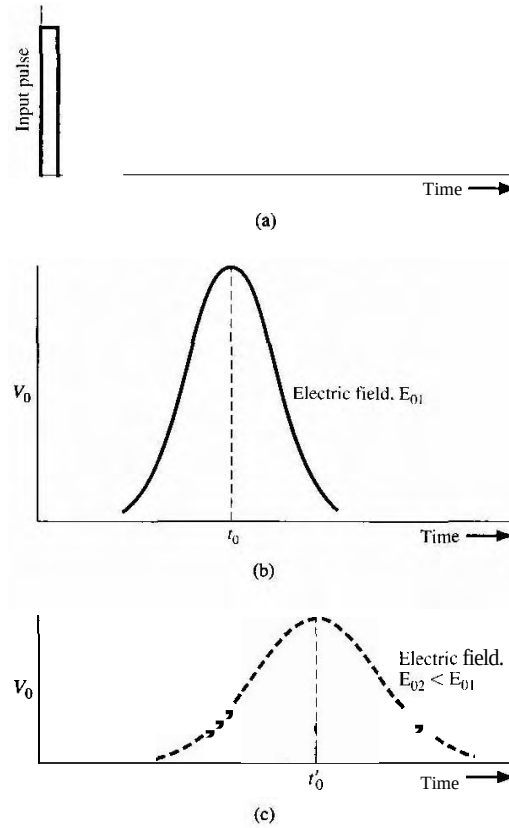


Figure 6.12 | (a) The idealized excess-carrier pulse at terminal **A** at $t = 0$. (b) The excess-carrier pulse versus time at terminal B for a given applied electric field. (c) The excess-carrier pulse versus time at terminal B for a smaller applied electric field.

pulse to reach the contact B. During this longer time period, there is more diffusion and more recombination. The excess-carrier pulse shapes shown in Figures 6.12b and 6.12c are different for the two electric field conditions.

The minority carrier mobility, lifetime, and diffusion coefficient can be determined from this single experiment. As a good first approximation, the peak of the minority carrier pulse will arrive at contact B when the exponent involving distance and time in Equation (6.70) is zero, or

In this case $x = d$, where d is the distance between contacts A and B, and $t = t_0$, where t_0 is the time at which the peak of the pulse reaches contact B. The mobility may be calculated as

$$\mu_p = \frac{d}{E_0 t_0} \tag{6.79b}$$

Figure 6.13 again shows the output response as a function of time. At times t_1 and t_2, the magnitude of the excess concentration is e^{-1} of its peak value. If the time difference between t_1 and t_2 is not too large, $e^{-t/\tau_{p0}}$ and $(4\pi D_p t)^{1/2}$ do not change appreciably during this time; then the equation

$$(d - \mu_p E_0 t)^2 = 4 D_p t \tag{6.80}$$

is satisfied at both $t = t_1$ and $t = t_2$. If we set $t = t_1$ and $t = t_2$ in Equation (6.80) and add the two resulting equations, we may show that the diffusion coefficient is given by

$$D_p = \frac{(\mu_p E_0)^2 (\Delta t)^2}{16 t_0} \tag{6.81}$$

where

$$\Delta t = t_2 - t_1 \tag{6.82}$$

The area S under the curve shown in Figure 6.13 is proportional to the number of excess holes that have not recombined with majority carrier electrons. We may write

$$S = K \exp\left(\frac{-t_0}{\tau_{p0}}\right) = K \exp\left(\frac{-d}{\mu_p E_0 \tau_{p0}}\right) \tag{6.83}$$

where K is a constant. By varying the electric field, the area under the curve will change. A plot of $\ln(S)$ as a function of $(d/\mu_p E_0)$ will yield a straight line whose slope is $(1/\tau_{p0})$, so the minority carrier lifetime can also be determined from this experiment.

The Haynes–Shockley experiment is elegant in the sense that the three basic processes of drift, diffusion, and recombination are all observed in a single experiment.

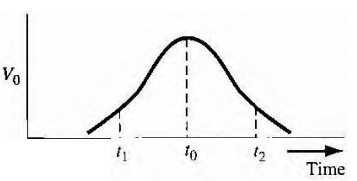


Figure 6.13 | The output excess-carrier pulse versus time to determine the diffusion coefficient.

The determination of mobility is straightforward and can yield accurate values. The determination of the diffusion coefficient and lifetime is more complicated and may lead to some inaccuracies.

6.4 | QUASI-FERMI ENERGY LEVELS

The thermal-equilibrium electron and hole concentrations are functions of the Fermi energy level. We can write

$$= n_i \exp\left(\frac{E_F - E_{Fi}}{kT}\right) \tag{6.84a}$$

and

$$p_0 = n_i \exp\left(\frac{E_{Fi} - E_F}{kT}\right) \tag{6.84b}$$

where E_F and E_{Fi} are the Fermi energy and intrinsic Fermi energy, respectively, and n_i is the intrinsic carrier concentration. Figure 6.14a shows the energy-band diagram for an n-type semiconductor in which $E_F > E_{Fi}$. For this case, we may note from Equations (6.84a) and (6.84b) that $n_0 > n_i$ and $p_0 < n_i$, as we would expect. Similarly, Figure 6.14b shows the energy-band diagram for a p-type semiconductor in which $E_F < E_{Fi}$. Again we may note from Equations (6.84a) and (6.84b) that $n_0 < n_i$ and $p_0 > n_i$, as we would expect for the p-type material. These results are for thermal equilibrium.

If excess carriers are created in a semiconductor, we are no longer in thermal equilibrium and the Fermi energy is strictly no longer defined. However, we may define a quasi-Fermi level for electrons and a quasi-Fermi level for holes that apply for nonequilibrium. If δn and δp are the excess electron and hole concentrations, respectively, we may write:

$$\boxed{n_0 + \delta n = n_i \exp\left(\frac{E_{Fn} - E_{Fi}}{kT}\right)} \tag{6.85a}$$

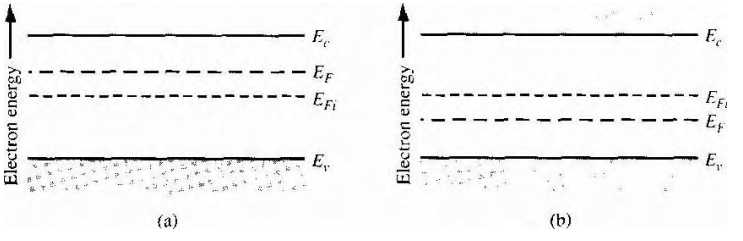


Figure 6.14 | Thermal-equilibrium energy-band diagrams for (*a*) n-type semiconductor and (*b*) p-type semiconductor.

$$p_0 + \delta p = n_i \exp\left(\frac{E_{Fi} - E_{Fp}}{kT}\right) \tag{6.85b}$$

where E_{Fn} and E_{Fp} are the quasi-Fermi energy levels for electrons and holes, respectively. The total electron concentration and the total hole concentration are functions of the quasi-Fermi levels.

Objective

To calculate the quasi-Fermi energy levels.

EXAMPLE 6.6

Consider an n-type semiconductor at T = 300 K with carrier concentrations of $n_0 = 10^{15}\,cm^{-3}$, $n_i = 10^{10}\,cm^{-3}$, and $p_0 = 10^5\,cm^{-3}$. In nonequilibrium, assume that the excess carrier concentrations are $\delta n = \delta p = 10^{13}\,cm^{-3}$.

■ Solution

The Fermi level for thermal equilibrium can be determined from Equation (6.84a). We have

$$E_F - E_{Fi} = kT \ln\left(\frac{n_0}{n_i}\right) = 0.2982\ eV$$

We can use Equation (6.85a) to determine the quasi-Fermi level for electrons in nonequilibrium. We can write

$$E_{Fn} - E_{Fi} = kT \ln\left(\frac{n_0 + \delta n}{n_i}\right) = 0.2984\ eV$$

Equation (6.85b) can be used to calculate the quasi-Fermi level for holes in nonequilibnum. We can write

$$E_{Fi} - E_{Fp} = kT \ln\left(\frac{p_0 + \delta p}{n_i}\right) = 0.179\ eV$$

Comment

We may note that the quasi-Fermi level for electrons is above E_{Fi} while the quasi-Fermi level for holes is below E_{Fi}

Figure 6.15a shows the energy-band diagram with the Fermi energy level corresponding to thermal equilibrium. Figure 6.15b now shows the energy-band diagram under the nonequilibrium condition. Since the majority carrier electron concentration does not change significantly for this low-injection condition, the quasi-Fermi level for electrons is not much different from the thermal-equilibrium Fermi level. The quasi-Fermi energy level for the minority carrier holes is significantly different from the Fermi level and illustrates the fact that we have deviated from thermal equilibrium significantly. Since the electron concentration has increased, the quad-Fermi level for electrons has moved slightly closer to the conduction band. The hole

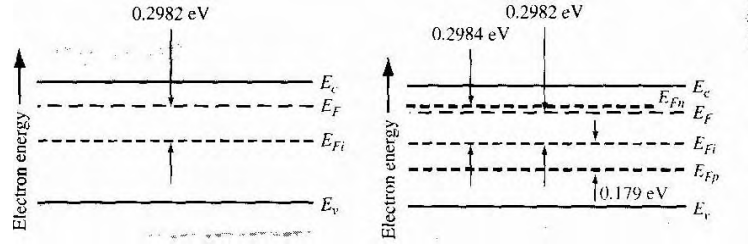


Figure 6.15 | (a) Thermal-equilibrium energy-band diagram for $N_d = 10^{15}$ cm^{-3} and $n_i = 10^{10}$ cm^{-3}. (b) Quasi-Fermi levels for electrons and holes if 10^{13} cm^{-3} excess carriers are present.

concentration has increased significantly so that the quasi-Fermi level for holes has moved much closer to the valence band. We will consider the quasi-Fermi energy levels again when we discuss forward-biased pn junctions.

TEST YOUR UNDERSTANDING

E6.10 Silicon at $T = 300$ K is doped at impurity concentrations of $N_d = 10^{16}$ cm^{-3} and $N_a = 0$. Excess carriers are generated such that the steady-state values are $\delta n = \delta p = 5 \times 10^{14}$ cm^{-3}. *(a)* Calculate the thermal equilibrium Fermi level with respect to E_{Fi}. *(b)* Determine E_{Fn} and E_{Fp} with respect to E_{Fi}.

[Ans. (a) $E_F - E_{Fi} = 0.3473$ eV; (b) $E_{Fn} - E_{Fi} = 0.3486$ eV, $E_{Fp} - E_{Fi} = -0.2697$ eV]

E6.11 Impurity concentrations of $N_d = 10^{15}$ cm^{-3} and $N_a = 6 \times 10^{15}$ cm^{-3} are added to silicon at T = 300 K. Excess carriers are generated in the material such that the steady-state concentrations are $\delta n = \delta p = 2 \times 10^{14}$ cm^{-3}. *(a)* Find the thermal equilibrium Fermi level with respect to E_{Fi}. (b) Calculate E_{Fn} and E_{Fp} with respect to E_{Fi}.

[Ans. (a) $E_F - E_{Fi} = -0.3294$ eV; (b) $E_{Fn} - E_{Fi} = -0.2460$ eV, $E_{Fp} - E_{Fi} = -0.3304$ eV]

*6.5 | EXCESS-CARRIER LIFETIME

The rate at which excess electrons and holes recombine is an important characteristic of the semiconductor and influences many of the device characteristics, as we will see in later chapters. We considered recombination briefly at the beginning of this chapter and argued that the recombination rate is inversely proportional to the mean carrier lifetime. We have assumed up to this point that the mean carrier lifetime is simply a parameter of the semiconductor material.

We have been considering an ideal semiconductor in which electronic energy states do not exist within the forbidden-energy bandgap. This ideal effect is present in a perfect single-crystal material with an ideal periodic-potential function. In a real

semiconductor material, defects occur within the crystal and disrupt the perfect periodic-potential function. If the density of these defects is not too great, the defects will create discrete electronic energy states within the forbidden-energy band. These allowed energy states may be the dominant effect in determining the mean carrier lifetime. The mean carrier lifetime may be determined from the Shockley–Read–Hall theory of recombination.

6.5.1 Shockley–Read–Hall Theory of Recombination

An allowed energy state, also called a trap, within the forbidden bandgap may act as a *recombination center,* capturing both electrons and holes with almost equal probability. This equal probability of capture means that the capture cross sections for electrons and holes are approximately equal. The Shockley–Read–Hall theory of recombination assumes that a single recombination center, or trap, exists at an energy E_t within the bandgap. There are four basic processes, shown in Figure 6.16, that may occur at this single trap. We will assume that the trap is an acceptor-type trap; that is, it is negatively charged when it contains an electron and is neutral when it does not contain an electron.

The four basic processes are as follows:

Process 1: The capture of an electron from the conduction band by an initially neutral empty trap.

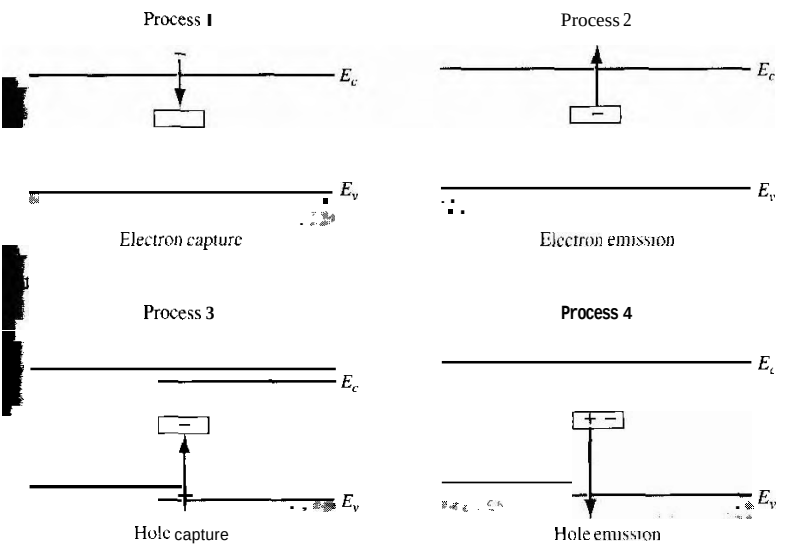


Figure 6.16 | The four basic trapping and emission processes for the case of an acceptor-type trap.

Process 2: The inverse of process 1—the emission of an electron that is initially occupying a trap level back into the conduction band.

Process 3: The capture of a hole from the valence band by a trap containing an electron. (Or we may consider the process to be the emission of an electron from the trap into the valence band.)

Process 4: The inverse of process 3—the emission of a hole from a neutral trap into the valence band. (Or we may consider this process to be the capture of an electron from the valence band.)

In process 1, the rate at which electrons from the conduction band are captured by the traps is proportional to the density of electrons in the conduction band and proportional to the density of empty trap states. We can then write the electron capture rate as

$$R_{cn} = C_n N_t (1 - f_F(E_t)) n \tag{6.86}$$

where

R_{cn} = capture rate (#/cm^3-s)
C_n = constant proportional to electron-capture cross section
N_t = total concentration of trapping centers
n = electron concentration in the conduction band
$f_F(E_t)$ = Fermi function at the trap energy

The Fermi function at the trap energy is given by

$$f_F(E_t) = \frac{1}{1 + \exp\left[\dfrac{E_t - E_F}{kT}\right]} \tag{6.87}$$

which is the probability that a trap will contain an electron. The function $(1 - f_F(E_t))$ is then the probability that the trap is empty. In Equation (6.87), we have assumed that the degeneracy factor is one, which is the usual approximation made in this analysis. However, if a degeneracy factor is included, it will eventually be absorbed in other constants later in the analysis.

For process 2, the rate at which electrons are emitted from filled traps back into the conduction band is proportional to the number of filled traps, so that

$$R_{en} = E_n N_t f_F(E_t) \tag{6.88}$$

where

R_{en} = emission rate (#/cm^3-s)
E_n = constant
$f_F(E_t)$ = probability that the trap is occupied

In thermal equilibrium, the rate of electron capture from the conduction band and the rate of electron emission back into the conduction band must be equal. Then

$$R_{en} = R_{cn} \tag{6.89}$$

so that

$$E_n N_t f_{F0}(E_t) = C_n N_t (1 - f_{F0}(E_t)) n_0 \tag{6.90}$$

where f_{F0} denotes the thermal-equilibrium Fermi function. Note that, in thermal equilibrium, the value of the electron concentration in the capture rate term is the equilibrium value n_0. Using the Boltrmann approximation for the Fermi function, we can find E_n in terms of C_n as

$$E_n = n' C_n \tag{6.91}$$

where n' is defined as

$$n' = N_c \exp\left[\frac{-(E_c - E_t)}{kT}\right] \tag{6.92}$$

The parameter n' is equivalent to an electron concentration that would exist in the conduction band if the trap energy E, coincided with the Fermi energy E_F.

In nonequilibrium, excess electrons exist, so that the net rate at which electrons are captured from the conduction band is given by

$$R_n = R_{cn} - R_{en} \tag{6.93}$$

which is just the difference between the capture rate and the emission rate. Combining Equations (6.86) and (6.88) with (6.93) gives

$$R_n = [C_n N_t (1 - f_F(E_t)) n] - [E_n N_t f_F(E_t)] \tag{6.94}$$

We may note that, in this equation, the electron concentration n is the total concentration, which includes the excess electron concentration. The remaining constants and terms in Equation (6.94) are the same as defined previously and the Fermi energy in the Fermi probability function needs to be replaced by the quasi-Fermi energy for electrons. The constants E_n and C_n are related by Equation (6.91). so the net recombination rate can be written as

$$R_n = C_n N_t [n(1 - f_F(E_t)) - n' f_F(E_t)] \tag{6.95}$$

If we consider processes 3 and 4 in the recombination theory, the net rate at which holes are captured from the valence band is given by

$$R_p = C_p N_t [p f_F(E_t) - p'(1 - f_F(E_t))] \tag{6.96}$$

where C_p is a constant proportional to the hole capture rate, and p' is given by

$$p' = N_v \exp\left[\frac{-(E_t - E_v)}{kT}\right] \tag{6.97}$$

In a semiconductor in which the trap density is not too large, the excess electron and hole concentrations are equal and the recombination rates of electrons and holes are equal. If we set Equation (6.95) equal to Equation (6.96) and solve for the Fermi function, we obtain

$$f_F(E_t) = \frac{C_n n + C_p p'}{C_n(n + n') + C_p(p + p')} \tag{6.98}$$

We may note that $n'p' = n_i^2$. Then, substituting Equation (6.98) back into either Equation (6.95) or (6.96) gives

$$R_{,,} = R_p = \frac{C_n C_p N_t (np - n_i^2)}{C_n(n + n') + C_p(p + p')} \equiv R \tag{6.99}$$

Equation (6.99) is the recombination rate of electrons and holes due to the recombination center at E = E,. If we consider thermal equilibrium, then $np = n_0 p_0 = n_i^2$ so that $R_{,,} = R_p = 0$. Equation (6.99). then, is the recombination rate of excess electrons and holes.

Since R in Equation (6.99) is the recombination rate of the excess carriers, we may write

$$R = \frac{\delta n}{\tau} \tag{6.100}$$

where δn is the excess-carrier concentration and τ is the lifetime of the excess carriers.

6.5.2 Limits of Extrinsic Doping and Low Injection

We simplified the ambipolar transport equation, Equation (6.39), from a nonlinear differential equation to a linear differential equation by applying limits of extrinsic doping and low injection. We may apply these same limits to the recombination rate equation.

Consider an n-type semiconductor under low injection. Then

$$\text{no} \gg p_0, \quad n_0 \gg \delta p, \quad \text{no} \gg n', \quad n_0 \gg p'$$

where δp is the excess minority carrier hole concentration. The assumptions of $n_0 \gg n'$ and $n_0 \gg p'$ imply that the trap level energy is near midgap so that n' and p' are not too different from the intrinsic carrier concentration. With these assumptions, Equation (6.99) reduces to

$$R = C_p N_t \delta p \tag{6.101}$$

The recombination rate of excess carriers in the n-type semiconductor is a function of the parameter $C_{,,}$ which is related to the minority carrier hole capture cross section. The recombination rate, then, is a function of the minority carrier parameter in the same way that the ambipolar transport parameters reduced to their minority carrier values.

The recombination rate is related to the mean carrier lifetime. Comparing Equations (6.100) and (6.101), we may write

$$R = \frac{\delta n}{\tau} = C_p N_t \delta p \equiv \frac{\delta p}{\tau_{p0}} \tag{6.102}$$

where

$$\tau_{p0} = \frac{1}{C_p N_t} \tag{6.103}$$

and where τ_{p0} is defined as the excess minority carrier hole lifetime. If the trap concentration increases, the probability of excess carrier recombination increases; thus the excess minority canier lifetime decreases.

Similarly, if we have a strongly extrinsic p-type material under low injection, we can assume that

$$p_0 \gg no, \quad p_0 \gg \delta n, \quad p_0 \gg n', \quad p_0 \gg p'$$

The lifetime then becomes that of the excess minority carrier electron lifetime, or

$$\tau_{n0} = \frac{1}{C_n N_t} \tag{6.104}$$

Again note that for the n-type material, the lifetime is a function of C_p, which is related to the capture rate of the minority carrier hole. And for the p-type material, the lifetime is a function of C_n, which is related to the capture rate of the minority carrier electron. The excess-carrier lifetime for an extrinsic material under low injection reduces to that of the minority carrier.

Objective

EXAMPLE 6.7

To determine the excess-carrier lifetime in an intrinsic semiconductor.

If we substitute the definitions of excess-carrier lifetimes from Equations (6.103) and (6.104) into Equation (6.99). the recumbination rate can be written as

$$R = \frac{(np - n_i^2)}{\tau_{p0}(n + n') + \tau_{n0}(p + p')} \tag{6.105}$$

Consider an intrinsic semiconductor containing excess carriers. Then $n = n_i + \delta n$ and $p = n_i + \delta n$. Also assume that $n' = p' = n_i$.

■ Solution

Equation (6 105) now become*

$$R = \frac{2n_i \delta n + (\delta n)^2}{(2n_i + \delta n)(\tau_{p0} + \tau_{n0})}$$

If we also assume very low injection, so that $\delta n \ll 2n_i$, then we can write

$$R = \frac{}{\tau_{p0} + \tau_{n0}} = \frac{\delta n}{\tau}$$

where τ is the excess carrier lifetime. We see that $\tau = \tau_{p0} + \tau_{n0}$ in the intrinsic material

■ Comment

The excess-canier lifetime increases as we change from an extrinsic to an intrinsic semiconductor.

Intuitively, we can see that the number of majority carriers that are available for recombining with excess minority carriers decreases as the extrinsic semiconductor

becomes intrinsic. Since there are fewer carriers available for recombining in the intrinsic material, the mean lifetime of an excess carrier increases.

TEST YOUR UNDERSTANDING

E6.12 Consider silicon at $T = 300$ K doped at concentrations of $N_d = 10^{15}$ cm^{-3} and $N_a = 0$. Assume that $n' = p' = n_i$ in the excess carrier recombination rate equation and assume parameter values of $\tau_{n0} = \tau_{p0} = 5 \times 10^{-7}$ s. Calculate the recombination rate of excess carriers if $\delta n = \delta p = 10^{14}$ cm^{-3}. (Ans. 1.83×10^{20} cm^{-3} s^{-1})

*6.6 | SURFACE EFFECTS

In all previous discussions. we have implicitly assumed the semiconductors were infinite in extent; thus, we were not concerned with any boundary conditions at a semiconductor surface. In any real application of semiconductors, the material is not infinitely large and therefore surfaces do exist between the semiconductor and an adjacent medium.

6.6.1 Surface States

When a semiconductor is abruptly terminated, the perfect periodic nature of the idealized single-crystal lattice ends abruptly at the surface. The disruption of the periodic-potential function results in allowed electronic energy states within the energy bandgap. In the previous section, we argued that simple defects in the semiconductor would create discrete energy states within the bandgap. The abrupt termination of the periodic potential at the surface results in a distribution of allowed energy states within the bandgap, shown schematically in Figure 6.17 along with the discrete energy states in the bulk semiconductor.

The Shockley–Read–Hall recombination theory shows that the excess minority carrier lifetime is inversely proportional to the density of trap states. We may argue

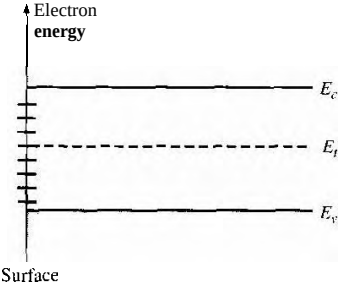


Figure 6.17 | Distribution of surface states within the forbidden bandgap.

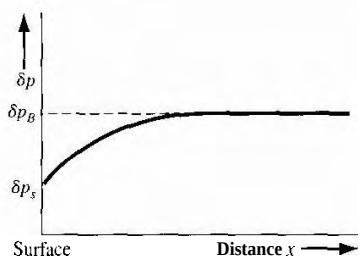


Figure 6.18 | Steady-state excess hole concentration versus distance from a semiconductor surface.

that, since the density of traps at the surface is larger than in the bulk, the excess minority carrier lifetime at the surface will he smaller than the corresponding lifetime in the bulk material. If we consider an extrinsic n-type semiconductor, for example, the recombination rate of excess carriers in the bulk. given by Equation (6.102), is

$$R = \frac{\delta p}{\tau_{p0}} \equiv \frac{\delta p_B}{\tau_{p0}} \tag{6.106}$$

where δp_B is the concentration of excess minority carrier holes in the bulk material. We may write a similar expression for the recombination rate of excess carriers at the surface as

$$R_s = \frac{\delta p_s}{\tau_{p0s}} \tag{6.107}$$

where δp_s is the excess minority carrier hole concentration at the surface and τ_{p0s} is the excess minority carrier hole lifetime at the surface.

Assume that excess carriers are being generated at a constant rate throughout the entire semiconductor material. We showed that, in steady state, the generation rate is equal to the recombination rate for the case of a homogeneous, infinite semiconductor. Using this argument, the recombination rates at the surface and in the hulk material must be equal. Since $\tau_{p0s} < \tau_{p0}$, then the excess minority carrierconcentration *at* the surface is smaller than the excess minority carrier concentration in the bulk region, or $\delta p_s < \delta p_B$. Figure 6.18 shows an example of the excess-carrier concentration plotted as a function of distance from the semiconductor surface.

Objective | **EXAMPLE *6.8***

To determine the steady-state excess-carrier concentration as a function of distance from the surface of a semiconductor.

Consider Figure 6.18, in which the surface is at $x = 0$. Assume that in the n-type semiconductor $\delta p_B = 10^{14}$ cm^{-3} and $\tau_{p0} = 10^{-6}$ s in the bulk, and $\tau_{p0s} = 10^{-7}$ s at the surface. Assume zero applied electric field and let $D_p = 10$ cm^2/s.

Solution

From Equations (6.106) and (6.107), we have

$$\frac{\delta p_B}{\tau_{p0}} = \frac{\delta p_s}{\tau_{p0s}}$$

so that

$$\delta p_s = \delta p_B \left(\frac{\tau_{p0s}}{\tau_{p0}}\right) = (10^{14})\left(\frac{10^{-7}}{10^{-6}}\right) = 10^{13}\ \text{cm}^{-3}$$

From Equation (6.56), we can write

$$D_p \frac{d^2(\delta p)}{dx^2} + g' - \frac{\delta p}{\tau_{p0}} = 0 \tag{6.108}$$

The generation rate can be determined from the steady-state conditions in the bulk, or

$$g' = \frac{\delta p_B}{\tau_{p0}} = \frac{10^{14}}{10^{-6}} = 10^{20}\ \text{cm}^{-3}\text{-s}^{-1}$$

The solution to Equation (6.107) is of the form

$$\delta p(x) = g'\tau_{p0} + Ae^{x/L_p} + Be^{-x/L_p} \tag{6.109}$$

As $x \to +\infty$, $\delta p(x) = \delta p_B = g'\tau_{p0} = 10^{14}\ \text{cm}^{-3}$, which implies that $A = 0$. At $x = 0$, we have

$$\delta p(0) = \delta p_s = 10^{14} + B = 10^{13}\ \text{cm}^{-3}$$

so that $B = -9 \times 10^{13}$. The entire solution for the minority carrier hole concentration as a function of distance from the surface is

$$\delta p(x) = 10^{14}(1 - 0.9e^{-x/L_p})$$

where

$$L_p = \sqrt{D_p\tau_{p0}} = \sqrt{(10)(10^{-6})} = 31.6\ \mu\text{m}$$

■ **Comment**

The excess carrier concentration is smaller at the surface than in the bulk.

6.6.2 Surface Recombination Velocity

A gradient in the excess-carrier concentration exists near the surface as shown in Figure 6.18; excess carriers from the bulk region diffuse toward the surface where they recombine. This diffusion toward the surface can be described by the equation

$$-D_p\left[\hat{n}\cdot\frac{d(\delta p)}{dx}\right]\Bigg|_{\text{surf}} = s\delta p|_{\text{surf}} \tag{6.110}$$

where each side of the equation is evaluated at the surface. The parameter $\hat{n}$ is the unit outward vector normal to the surface. Using the geometry of Figure 6.18,

$d(\delta p)/dx$ is a positive quantity and $\hat{n}$ is negative, so that the parameters is a positive quantity.

Adimensional analysis of Equation (6.110) shows that the parameters has units of cm/sec, or velocity. The parameters is called the ***surface recombination velocity*** If the excess concentrations at the surface and in the bulk region were equal, then the gradient term would be zero and the surface recombination velocity would be zero. As the excess concentration at the surface becomes smaller, the gradient term becomes larger, and the surface recombination velocity increases. The surface recombination velocity gives some indication of the surface characteristics as compared with the bulk region.

Equation (6.110) may be used as a boundary condition to the general solution given by Equation (6.109) in Example 6.8. Using Figure 6.18, we have that $\hat{n} = -1$, and Equation (6.110) becomes

$$D_p \left.\frac{d(\delta p)}{dx}\right|_{\text{surf}} = s\delta p|_{\text{surf}} \tag{6.111}$$

We have argued that the coefficient A is zero in Equation (6.109). Then, from Equation (6.109), we can write that

$$\delta p_{\text{surf}} = \delta p(0) = g'\tau_{p0} + B \tag{6.112a}$$

and

$$\left.\frac{d(\delta p)}{dx}\right|_{\text{surf}} = \left.\frac{d(\delta p)}{dx}\right|_{x=0} = -\frac{B}{L_p} \tag{6.112b}$$

Substituting Equations (6.112a) and (6.112b) into Equation (6.111) and solving for the coefficient B, we obtain

$$B = \frac{-sg'\tau_{p0}}{(D_p/L_p) + s} \tag{6.113}$$

The excess minority carrier hole concentration can then be written as

$$\boxed{\delta p(x) = g'\tau_{p0}\left(1 - \frac{sL_p e^{-x/L_p}}{D_p + sL_p}\right)} \tag{6.114}$$

EXAMPLE 6.9

Objective

To determine the steady-state excess concentration versus distance from the surface of a semiconductor as a function of surface recombination velocity.

Consider, initially, the case when the surface recombination velocity is zero, or $s = 0$.

Solution

Substituting $s = 0$ into Equation (6.114), we obtain

$$\delta p(x) = g'\tau_{p0}$$

Now consider the case when the surface recombination velocity is infinite, or $s = \infty$.

■ Solution

Substituting $s = \infty$ into Equation (6.114), we obtain

$$\delta p(x) = g'\tau_{p0}(1 - e^{-x/L_p})$$

Comment

For the case when $s = 0$, the surface has no effect and the excess minority carrier concentration at the surface is the same as in the bulk. In the other extreme when $s = \infty$, the excess minority carrier hole concentration at the surface is zero.

An infinite surface recombination velocity implies that the excess minority carrier concentration and lifetime at the surface are zero.

EXAMPLE 6.10

Objective

To determine the value of surface recombinaion velocity corresponding to the parameten given in Example 6.8.

From Example 6.8, we have that $g'\tau_{p0} = 10^{14}$ cm^{-3}, $D_p = 10$ cm^2/s, $L_p = 31.6$ μm, and $\delta p(0) = 10^{13}$ cm^{-3}.

■ Solution

Writing Equation (6.114) at the surface, we have

$$\delta p(0) = g'\tau_{p0}\left[1 - \frac{s}{(D_p/L_p) + s}\right]$$

Solving for the surface recombination velocity, we find that

$$s = \frac{D_p}{L_p}\left(\frac{g'\tau_{p0}}{\delta p(0)} - 1\right)$$

which becomes

$$s = \frac{10}{31.6 \times 10^{-4}}\left[\frac{10^{14}}{10^{13}} - 1\right] = 2.85 \times 10^4 \text{ cm/s}$$

■ Comment

This example shows that a surface recombination velocity of approximately $s = 3 \times 10^4$ cm/s could seriously degrade the performance of semiconductor devices, such as solar cells, since these devices tend to be fabricated close to a surface.

In the above example, the surface influences the excess-carrier concentration to the extent that, even at a distance of $L_p = 31.6$ μm from the surface, the excess-carrier concentration is only two-thirds of the value in the bulk. We will see in later chapters that device performance is dependent in large part on the properties of excess carriers.

6.7 | SUMMARY

- The processes of excess electron and hole generation and recombination were discussed. The excess carrier generation rate and recombination rate were defined.
- Excess electrons and holes do not move independently of each other, but move together. This common movement is called ambipolar transport.
- The ambipolar transport equation was derived and limits of low injection and extrinsic doping were applied to the coefficients. Under these conditions, the excess electrons and holes diffuse and drift together with the characteristics of the minority carrier, a result that is fundamental to the behavior of semiconductor devices
- The conceot of excess carrier lifetime was developed.
- Examples of excess carrier behavior as a function of time, as a function of space, and as afunction of both time and space were examined.
- The quasi-Fermi level for electrons and the quasi-Fermi level for holes were defined. These parameters characterize the total electron and hole concentrations in a semiconductor in nanequilibrium.
  The Shockley–Read–Hall theory of recombination was considered. Expressions for the excess minority carrier lifetime were developed.
- The effect of a semiconductor surface influences the behavior of excess electrons and holes. The surface recombination velocity was defined.

GLOSSARY OF IMPORTANT TERMS

ambipolar diffusion coefficient The effective diffusion coefficient of excess camers.

amhipolar **mobility** The effective mobility of excess camers.

amhipolar transport The process whereby excess electrons and holes diffuse, drift, and recombine with the same effective diffusion coefficient, mobility, and lifetime.

amhipolar transport equation The equation describing the behavior of excess carriers as a function of time and space coordinates.

carrier generation The process of elevating electrons from the valence band into the conduction band, creating an electron–hole pair.

carrier **recombination** The process whereby an electron "falls" into an empty state in the valence hand (a hole) so that an electron-hole pair are annihilated.

excess carriers The term describing both excess electrons and excess holes.

excess electrons The concentration of electrons in the conduction band over and above the thermal-equilibrium concentration.

excess holes The concentration of holes in the valence band over and above the thermal-equilibrium concentration.

excess minority carrier lifetime The average time that an excess minority carrier exists before it recombines.

generation rate The rate (#/cm^3-s) at which electron–hole pairs are created.

low-level injection The condition in which the excess-carrier concentration is much smaller than the thermal-equilibrium majority carrier concentration.

minority carrier ***diffusion*** length The average distance a minority carrier diffuses before recombining: a parameter equal to $\sqrt{D\tau}$ where D and τ are the minority carrier diffusion coefficient and lifetime, respectively.

quasi-Fermi level The quasi-Fermi level for electrons and the quasi-Fermi level for holes relate the nonequilibrium electron and hole concentrations, respectively, to the intrinsic carrier concentration and the intrinsic Fermi level.

recombination **rate** The rate (#/cm^3-s) at which electron–hole pairs recombine.

surface **recombination** velocity A parameter that relates the gradient of the excess carrier concentration at a surface to the surface concentration of excess carriers.

surface states The electronic energy states that exist within the bandgap at a semiconductor surface.

CHECKPOINT

After studying this chapter, the reader should have the ability to:

- Describe the concept of excess generation and recombination.
- Describe the concept of an cxcess carrier lifetime.
- Describe how the time-dependent diffusion equations for holes and electrons are derived.
- Describe how the ambipolar transport equation is derived.
- Understand the consequence of the coefficients in the ambipolar transport equation reducing to the minority carrier values under low injection and extrinsic semiconductors.
- Apply the ambipolar transport equation to various problems.
- Understand the conceot of the dielectric relaxation time constant.
- Calculate the quasi-Fermi levels lor electrons and holes.
- Calculate the cxcess carrier recombination rate for a given concentration of excess carriers.
- Understand the effect of a surface on the excess carrier concentrations.

REVIEW QUESTIONS

1. Why are the electron generation rate and recombination rate equal in thermal equilibrium?
2. Explain how the density of holes, for example, can change as a result of a change in the **flux** of particles.
3. Why is the general ambipolar transport equation nonlinear?
4. Explain qualitatively why a pulse of excess electrons and hales would move together in the presence of an applied electric field.
5. Explain qualitatively why the excess carrier lifetime reduces to that of the minority carrier under low injection.
6. What is the time dependence of the density of excess carriers when the generation rate becomes zero?
7. In the presence of an external force, why doesn't the density of excess carriers continue to increase with time?
8. When a concentration of one type of excess carrier is suddenly created in a semiconductor, what is the mechanism by which the net charge density quickly becomes zero?
9. State the definition of the quasi-Fermi level for electrons. Repeat for holes.
10. Why, in general. is the concentration of excess carriers less at the surface of a semiconductor than in the bulk?

PROBLEMS

(*Note:* Use the semiconductor parameters listed in Appendix B if they are not specifically given in a problem. Assume $T = 300$ K.)

Section 6.1 Carrier Generation and Recombination

6.1 Consider a semiconductor in which $n_0 = 10^{15}$ cm^{-3} and n, $= 10^{10}$ cm^{-3}. Assume that the excess-carrier lifetime is 10^{-6} s. Determine the electron–hole recombination rate if the excess-hole concentration is $\delta p = 5 \times 10^{13}$ cm^{-3}.

6.2 A semiconductor, in thermal equilibrium, has a hole concentration of $p_0 = 10^{16}$ cm^{-3} and an intrinsic concentration of n, $= 10^{10}$ cm^{-3}. The minority carrier lifetime is 2×10^{-7} s. ***(a)*** Determine the thermal-equilibrium recombination rate of electrons. (*b*) Determine the change in the recombination rate of electrons if an excess electron concentration of $\delta n = 10^{12}$ cm^{-3} exists.

6.3 An n-type silicon sample contains a donor concentration of $N_d = 10^{16}$ cm^{-3}. The minority carrier hole lifetime is found to be $\tau_{p0} = 20$ μs. ***(a)*** What is the lifetime of the majority carrier electrons? *(h)* Determine the thermal equilibrium generation rate for electrons and holes in this material. (*c*) Determine the thermal equilibrium recombination rate for electrons and holes in this material.

6.4 ***(a)*** A sample of semiconductor has a cross-sectional area of 1 cm^2 and a thickness of 0.1 cm. Determine the number of electron–hole pairs that are generated per unit volume per unit time by the uniform absorption of I watt of light at a wavelength of 6300 Å. Assume each photon creates one electron–hole pair. (b) If the excess minority carrier lifetime is 10 μs. what is the steady-state excess carrier concentration?

Section 6.2 Mathematical Analysis of Excess Carriers

6.5 Derive Equation (6.27) from Equations (6.18) and (6.20).

6.6 Consider a one-dimensional hole flux as shown in Figure 6.4. If the generation rate of holes in this differential volume is $g_p = 10^{20}$ cm^{-3}-s^{-1} and the recombination rate is 2×10^{19} cm^{-3}-s^{-1}, what must be the gradient in the particle current density to maintain a steddy-state hole concentrdtiun?

6.7 Repeat Problem 6.6 if the generation rate becomes zero.

Section 6.3 Ambipolar Transport

6.8 Starting with the continuity equations given by Equations (6.29) and (6.30), derive the ambipolar transport equation given by Equation (6.39).

6.9 A sample of Ge at T = 300 K has a uniform donor concentration of 2×10^{13} cm^{-3}. The excess carrier lifetime is found to be $\tau_{p0} = 24$ μs. Determine the ambipolar diffusion coefficient and the ambipolar mobility. What are the electron and hole lifetimes?

6.10 Assume that an n-type semiconductor is uniformly illuminated, producing a uniform excess generation rate g'. Show that in steady state the change in the semiconductor conductivity is given by

6.11 Light is incident on a silicon sample starting at $t = 0$ and generating excess carriers uniformly throughout the silicon for $t > 0$. The generation rate is $g' = 5 \times 10^{21}$ cm^{-3} s^{-1}. The silicon ($T = 300$ K) is n type with $N_d = 5 \times 10^{16}$ cm^{-3} and $N_a = 0$. Letn, $= 1.5 \times 10^{10}$ cm^{-3}, $\tau_{n0} = 10^{-6}$ s, and $\tau_{p0} = 10^{-7}$ s. Also let $\mu_n = 1000$ cm^2/V-s and $\mu_p = 420$ cm^2/V-s. Determine the conductivity of the silicon as a function of time for $t \geq 0$.

6.12 An n-type gallium arsenide semiconductor is doped with $N_d = 10^{16}$ cm^{-3} and $N_a = 0$. The minority carrier lifetime is $\tau_{p0} = 2 \times 10^{-7}$ s. Calculate the steady-state increase in conductivity and the steady-state excess carrier recombination rate if a uniform generation rate, $g' = 2 \times 10^{21}$ cm^{-3}-s^{-1}, is incident on the semiconductor

6.13 A silicon sample at T = 300 K is n type with $N_d = 5 \times 10^{16}$ cm^{-3} and $N_a = 0$. The sample has a length of 0.1 cm and a cross-sectional area of 10^{-4} cm^2. A voltage of 5 V is applied between the ends of the sample. Fort < 0. the sample has been illuminated with light. producing an excess-carrier generation rate of g' $= 5 \times 10^{21}$ cm^{-3}-s^{-1} uniformly throughout the entire silicon. The minority carrier lifetime is $\tau_{p0} = 3 \times 10^{-7}$ s. At t = 0, the light is turned off. Derive the expression for the current in the sample as a function of time $t \geq 0$. (Neglect surface effects.)

6.14 Consider a homogeneous gallium arsenide semiconductor at T = 300 K with $N_a = 10^{16}$ cm^{-3} and $N_d = 0$. A light source is turned on at $t = 0$ producing a uniform generation rate of $g' = 10^{20}$ cm^{-3}-s^{-1}. The electric field is zero. (*a*) Derive the expression for the excess-carrier concentration and excess carrier recombination rate as a function of time. (*b*) If the maximum, steady-state, excess-carrier concentration is to be 1×10^{14} cm^{-3}, determine the maximum value of the minority carrier lifetime. (*c*) Determine the times at which the excess minority carrier concentration will be equal to (i) three-fourths, (*ii*) one-half, and (*iii*) one-fourth of the steady-state value.

6.15 In a silicon semiconductor material at T = 300 K, the doping concentrations are $N_d = 10^{15}$ cm^{-3} and $N_a = 0$. The equilibrium recombination rate is $R_{p0} = 10^{11}$ cm^{-3}-s^{-1}. A uniform generation rate produces an excess-carrier concentration of $\delta n = \delta p = 10^{14}$ cm^{-3}. (*a*) By what factor does the total recombination rate increase? (*b*) What is the excess-carrier lifetime?

6.16 Consider a silicun material doped with 3×10^{16} cm^{-3} donor atoms. At $t = 0$, a light source is turned on, producing a uniform generation rate of $g' = 2 \times 10^{21}$ cm^{-3}-s^{-1} At $t = 10^{-7}$ s, the light source is turned off. Determine the excess minority carrier concentration as a function o f t for $0 \leq t \leq \infty$. Let $\tau_{p0} = 10^{-7}$ s. Plot the excess minority carrier concentration as a function of time.

6.17 A semiconductor has the following properties:

$$D_n = 25 \text{ cm}^2/\text{s} \qquad \tau_{n0} = 10^{-6} \text{ s}$$
$$D_p = 10 \text{ cm}^2/\text{s} \qquad \tau_{p0} = 10^{-7} \text{ s}$$

The semiconductor is a homogeneous. p-type ($N_a = 10^{17}$ cm^{-3}) material in thermal equilibrium for $t \leq 0$. At $t = 0$, an external source is turned on which produces excess carriers uniformly at the rate of g' $= 10^{20}$ cm^{-3}-s^{-1}. At t $= 2 \times 10^{-6}$ s, the external source is turned off. (*a*) Derive the expression for the excess-electron concentretion as a function of time for $0 \leq t \leq$ m. (*b*) Determine the value of the excess-electron concentration at (*i*) $t = 0$, (ii) t $= 2 \times 10^{-6}$ s, and (*iii*) $t = \infty$. (*c*) Plot the excess-electron concentration as a function of time.

6.18 Consider a bar of p-type silicon material that is homogeneously doped to a value of 3×10^{15} cm^{-3} at T = 300 K. The applied electric field is zero. A light source is

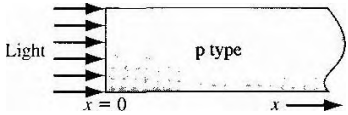


Figure 6.19 | Figure for Problems 6.18 and 6.20.

incident on the end of the semiconductor as shown in Figure 6.19. The excess-carrier concentration generated at $x = 0$ is $\delta p(0) = \delta n(0) = 10^{13}$ cm^{-3}. Assume the following parameters (neglect surface effects):

$$\mu_n = 1200 \text{ cm}^2/\text{V-s} \qquad \tau_{n0} = 5 \times 10^{-7} \text{ s}$$
$$\mu_p = 400 \text{ cm}^2/\text{V-s} \qquad \tau_{p0} = 1 \times 10^{-7} \text{ s}$$

(a) Calculate the steady-state excess electron and hole concentrations as a function of distance into the semiconductor. (*b*) Calculate the electron diffusion current density as a function of x.

6.19 The $x = 0$ end of an $N_a = 1 \times 10^{14}$ cm^{-3} doped semi-infinite ($x \geq 0$) bar of silicon maintained at T = 300 K is attached tu a "minority carrier digester" which makes $n_p = 0$ at $x = 0$ (n_p is the minority carrier electron concentration in a p-type semiconductor). The electric field is zero. (*a*) Determine the thermal-equilibrium values of n_{p0} and p_{p0}. (*b*) What is the excess minority carrier concentration at $x = 0$? (*c*) Derive the expression for the steady-state excess minority carrier concentration as a function of x.

6.20 In a p-type silicon semiconductor, excess carriers are being generated at the end of the semiconductor bar at $x = 0$ as shown in Figure 6.19. The doping concentration is $N_a = 5 \times 10^{16}$ cm^{-3} and $N_d = 0$. The steady-state excess-carrier concentration at $x = 0$ is 10^{15} cm^{-3}. (Neglect surface effects.) The applied electric field is zero. Assume that $\tau_{n0} = \tau_{p0} = 8 \times 10^{-7}$ s. (*a*) Calculate δn, and the electron and hole diffusion current densities at $x = 0$. (*b*) Repeat part (*a*) for $x = L_a$.

6.21 Consider an n-type silicon sample. Excess carriers are generated at $x = 0$ such as shown in Figure 6.6. A constant electric field E_0 is applied in the $+x$ direction. Show that the steady-state excess carrier concentration is given by

$$\delta p(x) = A \exp(s_- x) \qquad x > 0 \qquad \text{and} \qquad \delta p(x) = A \exp(s_+ x) \qquad x < 0$$

where

$$s_\mp = \frac{1}{L_p}\left[\beta \mp \sqrt{1 + \beta^2}\right]$$

and

$$\beta = \frac{\mu_p L_p E_0}{2 D_p}$$

6.22 Plot the excess carrier concentration $\delta p(x)$ versus x from Problem 6.21 for *(a)* $E_0 = 0$ and *(b)* $E_0 = 10$ V/cm.

***6.23** Consider the semiconductor described in Problem 6.18. Assume a constant electric field E_0 is applied in the $+x$ direction. (*a*) Derive the expression for the steady-state excess-electron concentration. (Assume the solution is of the form e^{-ax}.) (h) Plot δn

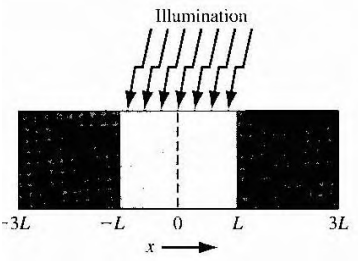


Figure 6.20 | Figure for Problem 6.25

versus x for (i) $E_0 = 0$ and (ii) $E_0 = 12$ V/cm. (c) Explain the general characteristics of the two curves plotted in pan (b).

6.24 Assume that a p-type semiconductor is in thermal equilibrium fort < 0 and has an infinite minority carrier lifetime. Also assume that the semiconductor is uniformly illuminated, resulting in a uniform generation rate, $g'(t)$, which is given by

$$g'(t) = G'_0 \qquad \text{for } 0 < t < T$$
$$g'(t) = 0 \qquad \text{for } t < 0 \text{ and } t > T$$

where G'_0 is a constant. Find the excess minority carrier concentration as a function of time.

***6.25** Consider the n-type semiconductor shown in Figure h.20. Illumination produces a constant excess-carrier generation rate, G'_0, in the region $L < x < +L$. Assume that the minority carrier lifetime is infinite and assume that the excess minority carrier hole concentration is zero at $x = -3L$ and at $x = +3L$. Find the steadv-state excess minority carrier concentration versus x, for the case of low injection and for zero applied electric field.

6.26 An n-type germanium sample is used in the Haynes–Shockley experiment. The length of the sample is 1 cm and the applied voltage is $V_1 = 2.5$ V The contacts A and B are seoarated bv 0.75 cm. The oeak of the pulse arrives at contact B 160 us after carrier injection at contact A. The width of the pulse is Ar $= 75.5$ μs. Determine the hole mobility and diffusion coefficient. Compare the results with the Einstein relation.

6.27 Consider the function $f(x, t) = (4\pi Dt)^{-1/2} \exp(-x^2/4Dt)$. (a) Show that this function is a solution to the differential equation $D(\partial^2 f/\partial x^2) = \partial f/\partial t$. (b) Show that the integral of the function $f(x, t)$ over x from $-\infty$ to $+\infty$ is unity for all values of time. *(r)* Show that this function approaches a δ function as t approaches zero.



6.28 The basic equation in the Haynes–Shockley experiment is given by Equation (6.70). (a) Plot $\delta p(x, t)$ versus x for various values of t and for $E_0 = 0$ as well as for $E_0 \neq 0$. *(b)* Plot $\delta p(x, t)$ versus t for various values of x and for $E_0 = 0$ as well as for $E_0 \neq 0$.

Section 6.4 Quasi-Fermi Energy Levels

6.29 An n-type silicon sample with $N_d = 10^{16}$ cm^{-3} is steadily illuminated such that $g' = 10^{21}$ cm^{-3}-s^{-1}. If $\tau_{n0} = \tau_{p0} = 10^{-6}$ s, calculate the position of the quasi-Fermi levels for electrons and hales with respect to the intrinsic level (assume that $n_i = 1.5 \times 10^{10}$ cm^{-3}). Plot these levels on an energy-hand diagram.

6.30 Consider a p-type silicon semiconductor at $T = 300$ K doped at $N_a = 5 \times 10^{15}$ cm^{-3}. *(a)* Determine the position of the Fermi level with respect to the intrinsic Fermi level. (b) Excess carriers are generated such that the excess-carrier concentration is 10 percent of the thermal-equilibrium majority carrier concentration. Determine the quasi-Fermi levels with respect to the intrinsic Fermi level. (c) Plot the Fermi level and quasi-Fermi levels with respect to the intrinsic level.

6.31 Consider an n-type gallium arsenide semiconductor at T = 300 K doped at $N_d = 5 \times 10^{16}$ cm^{-3}. *(a)* Determine $E_{Fn} - E_F$ if the excess-carrier concentration is 0.1 N_d. (b) Determine $E_{Fi} - E_{Fp}$.

6.32 Ap-type gallium arsenide semiconductor at T = 300 K is doped at $N_a = 10^{16}$ cm^{-3}. The excess-carrier concentration varies linearly from 10^{14} cm^{-3} to rero over a distance of 50 μm. Plot the position of the quasi-Fermi levels with respect to the intrinsic Fermi level versus distance.

6.33 Consider p-type silicon at T = 300 K duped to $N_a = 5 \times 10^{14}$ cm^{-3}. Assume excess carriers arc present and assume that $E_F - E_{Fp} = (0.01)kT$. *(a)* Does this condition correspond to low injection? Why or why not? *(b)* Determine $E_{Fn} - E_{Fi}$

6.34 An n-type silicon sample is doped with donors at a concentration of $N_d = 10^{16}$ cm^{-3}. Excess carriers are generated such that the excess hole concentration is given by $\delta p(x) = 10^{14} \exp(-x/10^{-4})$ cm^{-3}. Plot the function $E_{Fi} - E_{Fp}$ versusx over the range $0 \le x \le 4 \times 10^{-4}$.

6.35 For a p-type silicon material doped at $N_a = 10^{16}$ cm^{-3}, plot $E_{Fn} - E_F$ versus δn over the range $0 \le \delta n \le 10^{14}$ cm^{-3}. Use a log scale for δn.

Section 6.5 Excess Carrier Lifetime

6.36 Consider Equation (6.99) and the definitions of τ_{p0} and τ_{n0} hy Equations (6.103) and (6.104). Let $n' = p' = n_i$. Assume that in a particular region of a semiconductor, $n = p = 0$. (a) Determine the recomhination rate R. (h) Explain what this result means physically.

6.37 Again consider Equation (6.99) and the definitions of τ_{p0} and τ_{n0} given by Equations (6.103) and (6.104). Let $\tau_{p0} = 10^{-7}$ s and $\tau_{n0} = 5 \times 10^{-7}$ s. Also let n' = p' = n; $= 10^{10}$ cm^{-3}. Assume very low injection so that $\delta n \ll n_i$. Calculate $R/\delta n$ for a semiconductor which is (o) n-type ($n_0 \gg p_0$), (h) intrinsic ($n_0 = p_0 = n_i$), and (*c*) p-type ($p_0 \gg n_0$).

Section 6.6 Surface Effects

***6.38** Consider an n-type semiconductor as shown in Figure 6.21. doped at $N_d = 10^{16}$ cm^{-3} and with a uniform excess-carrier generation rate equal tog' $= 10^{21}$ cm^{-3}-s^{-1}. Assume that $D_p = 10$ cm^2/s and $\tau_{p0} = 10^{-7}$ s. The electric field is rero.

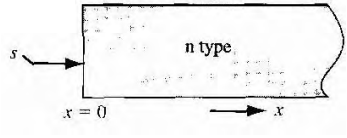


Figure 6.21 | Figure for Problem 6.38.

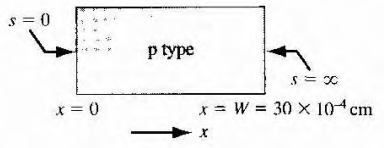


Figure 6.22 | Figure for Problem 6.39.

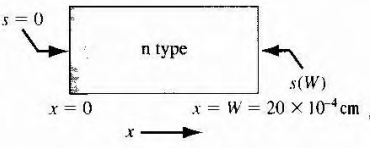


Figure 6.23 | Figure for Problem 6.40.

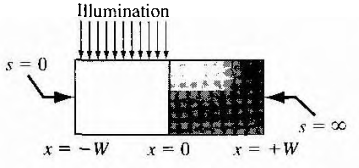


Figure 6.24 | Figure for Problem 6.41.

(a) Determine the steady-state excess minority carrier concentration versus x if the surface recombination velocity at $x = 0$ is (i) $s = 0$. (ii) $s = 2000$ cm/s, and (iii) $s = \infty$. (*b*) Calculate the excess minority carrier concentration at $\mathbf{x} = 0$ for (*i*) $s = 0$, (*ii*) $s = 2000$ cm/s, and (iii) $s = \infty$.

***6.39** **(a)** Consider the p-type semiconductor shown in Figure 6.22 with the following parameters: $N_a = 5 \times 10^{16}$ cm^{-3}, $D_n = 25$ cm^2/s, and $\tau_{n0} = 5 \times 10^{-7}$ s. The surface recombination velocities at the two surfaces are shown. The electric field is zero. The semiconductor is illuminated at $x = 0$ with an excess-canier generation rate equal to $g' = 2 \times 10^{\prime\prime}$ cm^{-3}-s^{-1}. Determine the excess minority carrier electron concentration versus x in steady state. (*b*) Repeat part (a) for $\tau_{n0} = \infty$.

***6.40** Consider the n-type semiconductor shown in Figure 6.23. Assume that $D_p = 10$ cm^2/s and $\tau_{p0} = \infty$. The electric field is zero. Assume that a flux of excess electrons and holes is incident at $x = 0$. Let the flux of each carrier type be 10^{19} carriers/cm^2-s. Determine the minority carrier hole current versus x if the surface recombination velocity is (a) $s(W) = \infty$ and (*b*) $s(W) = 2000$ cm/s.

***6.41** A p-type semiconductor is shown in Figure 6.24. The surface recombination velocities are shown. The semiconductor is uniformly illuminated for $-W < x < 0$ producing a constant excess-carrier generation rate G_0'. Determine the steady-state excess-carrier concentration versus x if the minority carrier lifetime is infinite and if the electric field is zero.



6.42 Plot $\delta p(x)$ versus x for various values of s using Equation (6.113). Choose reasonable parameter values.

Summary and Review

***6.43** Consider an n-type semiconductor as shown in Figure 6.21. The material is doped at $N_d = 3 \times 10^{16}$ cm^{-3} and $N_a = 0$. Assume that $D_p = 12$ cm^2/s and $\tau_{p0} = 2 \times 10^{-7}$ s. The electric field is zero. "Design" the surface recombination velocity so that the minority carrier diffusion current density at the surface is no grealer than $J_p = -0.18$ A/cm^2 with a uniform excess-carrier generation rate equal to $g' = 3 \times 10^{21}$ cm^{-3}-s^{-1}.

6.44 Consider a semiconductor with excess carriers present. From the definition of carrier lifetimes and recombination rates, determine the average time that an electron stays in the conduction band and the average time that a hole stays in the valence band. Discuss these relations for (a) an intrinsic semiconductor and *(h)* an n-type semiconductor.

6.45 Design a gallium arsenide photoconductor that is 5 μm thick. Assume that $\tau_{n0} = \tau_{p0} = 10^{-7}$ s and $N_d = 5 \times 10^{15} cm^{-3}$. With an excitation of $g' = 10^{21} cm^{-3}\text{-}s^{-1}$ a photocurrent of at least 1 μA is desired with an applied voltage of I V.

READING LIST

1. Bube, R. H. *Photoelectronic Properties of Semiconductors,* New York: Cambridge University Press, 1992.
2. deCogan, D. *Solid State Devices: A Quantum Physics Appmach.* New York: Springer-Verlag, 1987.
3. Hall. R. H. "Electron-Hole Recombination." *Physical Review* 87, no. 2 (July 15, 1952), p. 387.
4. Haynes, J. R., and W. Shockley. "The Mobility and Life of Injected Holes and Electrons in Germanium." *Physical Review* 81, no. 5 (March 1, 1951), pp. 835–843.
5. Hess, K. *Advanced Theory of Semiconductor Devices.* Englewood Cliffs, NJ: Prentice Hall, 1988.
6. Kano, K. *Semiconducror Devices.* Upper Saddle River, NJ: Prentice Hall, 1998.
7. Kingston, R. H. *Semiconducror Surface Physics.* Philadelphia: University of Pennsylvania Press, 1957.
8. McKelvey, J. P. *Solid State Physics for Engineering and Materials Science.* Malabar, FL.: Krieger Publishing, 1993.
9. Pierret, R. F. *Semiconductor Device Fundamentals.* Reading, MA: Addison-Wesley, 1996.
10. Shockley, W., and W. T. Read, Jr. "Statistics of the Recombinations of Holes and Electrons." *Physical Review* 87, no. 5 (September 1, 1952), pp. 835–842.
11. Singh, J. *Semiconducror Devices: An Introduction.* New York: McGraw-Hill, 1994.
12. Singh, J. *Semiconducror Devices: Basic Principles.* New York: John Wiley and Sons, 2001.
13. Streetman, B. G., and S. Banerjee. *Solid State Electronic Devices.* 5th ed. Upper Saddle River, NJ: Prentice Hall, 2000.
14. Wang, S. *Fundamentals of Semiconductor Theory and Device Physics.* Englewood Cliffs. *NJ:* Prentice Hall, 1989.

CHAPTER 7

The pn Junction

PREVIEW

Up to this point in the text, we have been considering the properties of the semiconductor material. We calculated electron and hole concentrations in thermal equilibrium and determined the position of the Fermi level. We then considered the nonequilibrium condition in which excess electrons and holes are present in the semiconductor. We now wish to consider the situation in which a p-type and an n-type semiconductor are brought into contact with one another to form a pn junction.

Most semiconductor devices contain at least one junction between p-type and n-type semiconductor regions. Semiconductor device characteristics and operation are intimately connected to these pn junctions, so considerable attention is devoted initially to this basic device. The pn junction diode itself provides characteristics that are used in rectifiers and switching circuits. In addition, the analysis of the pn junction device establishes some basic terminology and concepts that are used in the discussion of other semiconductor devices. The fundamental analysis techniques used for the pnjunction will also be applied to other devices. Understanding the physics of the pn junction is, therefore, an important step in the study of semiconductor devices.

The electrostatics of the pn junction is considered in this chapter and the current–voltage characteristics of the pn junction diode are developed in the next chapter. ■

7.1 | BASIC STRUCTURE OF THE pn JUNCTION

Figure 7.1a schematically shows the pn junction. It is important to realize that the entire semiconductor is a single-crystal material in which one region is doped with acceptor impurity atoms to form the p region and the adjacent region is doped with donor atoms to form the n region. The interface separating the n and p regions is referred to as the *metallurgical junction.*

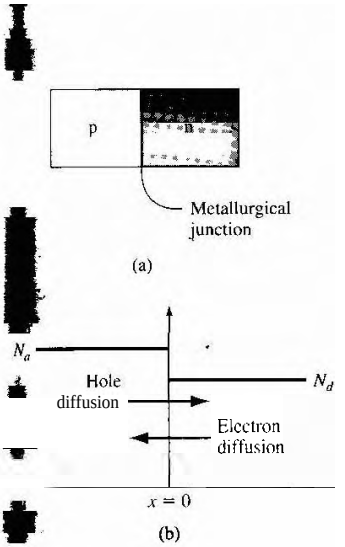


Figure 7.1 | (a) Simplified geometry of a pn junction; (b) doping profile of an ideal uniformly doped pn junction.

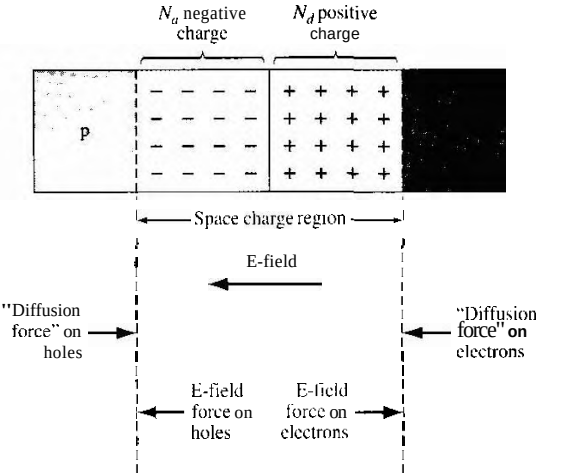


Figure 7.2 | The space charge region, the electric field, and the forces acting on the charged carriers.

The impurity doping concentrations in the p and n regions are shown in Figure 7.1b. For simplicity, we will consider a *step* junction in which the doping concentration is uniform in each region and there is an abrupt change in doping at the junction. Initially, at the metallurgical junction, there is a very large density gradient in both the electron and hole concentrations. Majority carrier electrons in the n region will begin diffusing into the p region and majority carrier holes in the p region will begin diffusing into the n region. If we assume there are no external connections to the semiconductor, then this diffusion process cannot continue indefinitely. As electrons diffuse from the n region, positively charged donor atoms are left behind. Similarly, as holes diffuse from the p region, they uncover negatively charged acceptor atoms. The net positive and negative charges in the n and p regions induce an electric field in the region near the metallurgical junction, in the direction from the positive to the negative charge, or from the n to the p region.

The net positively and negatively charged regions are shown in Figure 7.2. These two regions are referred to as the space charge region. Essentially all electrons and holes are swept out of the space charge region by the electric field. Since the space charge region is depleted of any mobile charge, this region is also referred to as the depletion region: these two terms will be used interchangeably. Density gradients still exist in the majority carrier concentrations at each edge of the space charge region. We can think of a density gradient as producing a "diffusion force" that acts on the majority carriers. These diffusion forces, acting on the electrons and holes at the edges of the space charge region, are shown in the figure. The electric field in the space charge region produces another force on the electrons and holes which is in the

opposite direction to the diffusion force for each type of particle. In thermal equilibrium, the diffusion force and the E-field force exactly balance each other.

7.2 | ZERO APPLIED BIAS

We have considered the basic pn junction structure and discussed briefly how the space charge region is formed. In this section we will examine the properties of the step junction in thermal equilibrium, where no currents exist and no external excitation is applied. We will determine the space charge region width, electric field, and potential through the depletion region.

7.2.1 Built-in Potential Barrier

If we assume that no voltage is applied across the pn junction, then the junction is in thermal equilibrium — the Fermi energy level is constant throughout the entire system. Figure 7.3 shows the energy-band diagram for the pn junction in thermal equilibrium. The conduction and valance band energies must bend as we go through the space charge region, since the relative position of the conduction and valence hands with respect to the Fermi energy changes between p and n regions.

Electrons in the conduction band of the n region see a potential barrier in trying to move into the conduction hand of the p region. This potential barrier is referred to as the ***built-in potential barrier*** and is denoted by V_{bi}. The built-in potential barrier maintains equilibrium between majority carrier electrons in then region and minority carrier electrons in the p region, and also between majority carrier holes in the p region and minority carrier holes in then region. This potential difference across the junction cannot be measured with a voltmeter because new potential barriers will be formed between the probes and the semiconductor that will cancel V_{bi}. The potential V_{bi} maintains equilibrium, so no current is produced by this voltage.

The intrinsic Fermi level is equidistant from the conduction hand edge through the junction, thus the built-in potential barrier can be determined as the difference

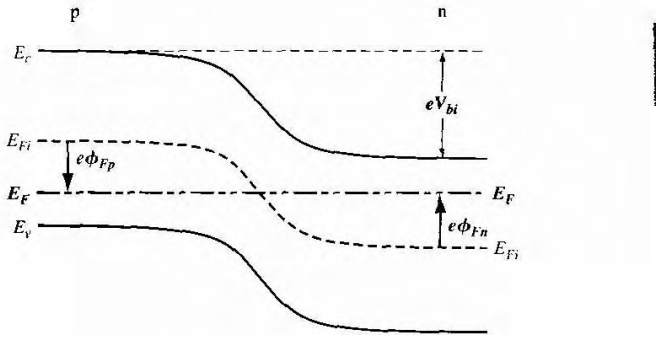


Figure **7.3** | Energy-band diagram of a pn junction in thermal equilibrium.

between the intrinsic Fermi levels in the p and n regions. We can define the potentials ϕ_{Fn} and ϕ_{Fp} as shown in Figure 7.3, so we have

$$V_{bi} = |\phi_{Fn}| + |\phi_{Fp}| \tag{7.1}$$

In the n region, the electron concentration in the conduction band is given by

$$n_0 = N_c \exp\left[\frac{-(E_c - E_F)}{kT}\right] \tag{7.2}$$

which can also be written in the form

$$n_0 = n_i \exp\left[\frac{E_F - E_{Fi}}{kT}\right] \tag{7.3}$$

where n_i and E_{Fi} are the intrinsic carrier concentration and the intrinsic Fermi energy, respectively. We may define the potential ϕ_{Fn} in the n region as

$$e\phi_{Fn} = E_{Fi} - E_F \tag{7.4}$$

Equation (7.3) may then be written as

$$n_0 = n_i \exp\left[\frac{-(e\phi_{Fn})}{kT}\right] \tag{7.5}$$

Taking the natural log of both sides of Equation (7.5), setting $n_0 = N_d$, and solving for the potential, we obtain

$$\phi_{Fn} = \frac{-kT}{e} \ln\left(\frac{N_d}{n_i}\right) \tag{7.6}$$

Similarly, in the p region, the hole concentration is given by

$$p_0 = N_a = n_i \exp\left[\frac{E_{Fi} - E_F}{kT}\right] \tag{7.7}$$

where N_a is the acceptor concentration. We can define the potential ϕ_{Fp} in the p region as

$$e\phi_{Fp} = E_{Fi} - E_F \tag{7.8}$$

Combining Equations (7.7) and (7.8), we find that

$$\phi_{Fp} = +\frac{kT}{e} \ln\left(\frac{N_a}{n_i}\right) \tag{7.9}$$

Finally, the built-in potential barrier for the step junction is found by substituting Equations (7.6) and (7.9) into Equation (7.1), which yields

$$\boxed{V_{bi} = \frac{kT}{e} \ln\left(\frac{N_a N_d}{n_i^2}\right) = V_t \ln\left(\frac{N_a N_d}{n_i^2}\right)} \tag{7.10}$$

where $V_t = kT/e$ and is defined as the thermal voltage.

At this time, we should note a subtle but important point concerning notation. Previously in the discussion of a semiconductor material, N_d and N_a denoted donor, and acceptor impurity concentrations in the same region, thereby forming a compensated semiconductor. From this point on in the text, N_d and N_a will denote the net donor and acceptor concentrations in the individual n and p regions, respectively. If the p region, for example, is a compensated material, then N_a will represent the difference between the actual acceptor and donor impurity concentrations. The parameter N_d is defined in a similar manner for the n region.

EXAMPLE 7.1

Objective

To calculate the built-in potential barrier in a pn junction.

Consider a silicon pn junction at T = 300 K with doping densities $N_a = 1 \times 10^{18}$ cm^{-3} and $N_d = 1 \times 10^{15}$ cm^{-3}. Assume that $n_i = 1.5 \times 10^{10}$ cm^{-3}

■ Solution

The built-in potential barrier is determined from Equation (7.10) as

$$V_{bi} = (0.0259) \ln \left[\frac{(10^{18})(10^{15})}{(1.5 \times 10^{10})^2} \right] = 0.754 \text{ V}$$

If we change the acceptor doping from $N_a = 1 \times 10^{18}$ c m ^{3} to $N_a = 1 \times 10^{16}$ c m ', but keep all other parameter values constant, then the built-in potential barrier becomes $V_{bi} = 0.635$V.

■ Comment

The built-in potential barrier changes only slightly as the doping concentrations change by orders of magnitude because of the logarithmic dependence.

TEST YOUR UNDERSTANDING

E7.1 Calculate the built-in potential barrier in a silicon pn junction at T = 300 K for (*a*) $N_a = 5 \times 10^{17}$ cm^{-3}, $N_d = 10^{16}$ cm^{-3} and (*b*) $N_a = 10^{15}$ cm^{-3}, $N_d = 2 \times 10^{16}$ c m ^{3}. [Ans. (*a*) 0.653 V, (*b*) 0.796 V]

E7.2 Repeat E7.1 for a GaAs pn junction. [Ans. (*a*) 1.26 V, (*b*) 1.21 V]

7.2.2 Electric Field

An electric field is created in the depletion region by the separation of positive and negative space charge densities. Figure 7.4 shows the volume charge density distribution in the pn junction assuming uniform doping and assuming an abrupt junction approximation. We will assume that the space charge region abruptly ends in the n region at $x = +x_n$ and abruptly ends in the p region at $x = x_p$, (x_p is a positive quantity).

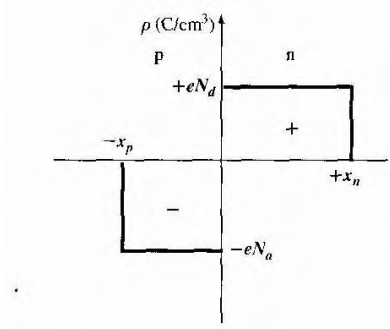


Figure 7.41 The space charge density in a uniformly doped pn junction assuming the abrupt junction approximation.

The electric field is determined from Poisson's equation which, for a one-dimensional analysis, is

$$\frac{d^2\phi(x)}{dx^2} = \frac{-\rho(x)}{\epsilon_s} = -\frac{dE(x)}{dx} \tag{7.11}$$

where $\phi(x)$ is the electric potential, $E(x)$ is the electric field, $\rho(x)$ is the volume charge density, and ϵ_s is the permittivity of the semiconductor. From Figure 7.4, the charge densities are

$$\rho(x) = -eN_a \qquad -x_p < x < 0 \tag{7.12a}$$

and

$$\rho(x) = eN_d \qquad 0 < x < x_n, \tag{7.12b}$$

The electric field in the p region is found by integrating Equation (7.11). We have that

$$E = \int \frac{\rho(x)}{\epsilon_s}\,dx = -\int \frac{eN_a}{\epsilon_s}\,dx = \frac{-eN_a}{\epsilon_s}x + C_1 \tag{7.13}$$

where C_1 is a constant of integration. The electric field is assumed to be zero in the neutral p region for $x < -x_p$ since the currents are zero in thermal equilibrium. As there are no surface charge densities within the pn junction structure. the electric field is a continuous function. The constant of integration is determined by setting $E = 0$ at $x = -x_p$. The electric field in the p region is then given by

$$E = \frac{-eN_a}{\epsilon_s}(x + x_p) \qquad -x_p \le x \le 0 \tag{7.14}$$

In the n region, the electric field is determined from

$$E = \int \frac{(eN_d)}{\epsilon_s}\,dx = \frac{eN_d}{\epsilon_s}x + C_2 \tag{7.15}$$

where C_2 is again a constant of integration. The constant C_2 is determined by setting $E = 0$ at $x = x_n$, since the E-field is assumed to be zero in the n region and is a continuous function. Then

$$E = \frac{-eN_d}{\epsilon_s}(x_n - x) \qquad 0 \le x \le x_n \tag{7.16}$$

The electric field is also continuous at the metallurgical junction, or at $x = 0$. Setting Equations (7.14) and (7.16) equal to each other at $x = 0$ gives

$$N_a x_p = N_d x_n \tag{7.17}$$

Equation (7.17) states that the number of negative charges per unit area in the p region is equal to the number of positive charges per unit area in the n region.

Figure 7.5 is a plot of the electric field in the depletion region. The electric field direction is from the n to the p region, or in the negative x direction for this geometry. For the uniformly doped pn junction, the E-field is a linear function of distance through the junction, and the maximum (magnitude) electric field occurs at the metallurgical junction. An electric field exists in the depletion region even when no voltage is applied between the p and n regions.

The potential in the junction is found by integrating the electric field. In the p region then, we have

$$\phi(x) = -\int E(x)\,dx = \int \frac{eN_a}{\epsilon_s}(x + x_p)\,dx \tag{7.18}$$

or

$$\phi(x) = \frac{eN_a}{\epsilon_s}\left(\frac{x^2}{2} + x_p \cdot x\right) + C_1' \tag{7.19}$$

where C_1' is again a constant of integration. The potential difference through the pn junction is the important parameter, rather than the absolute potential, so we may arbitrarily set the potential equal to zero at $x = -x_p$. The constant of integration is

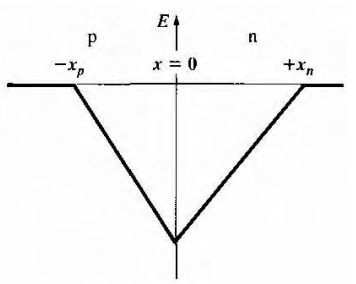


Figure 7.5 | Electric field in the space charge region of a uniformly doped pn junction.

then found as

$$C_1' = \frac{eN_a}{2\epsilon_s} x_p^2 \tag{7.20}$$

so that the potential in the p region can now be written as

$$\phi(x) = \frac{eN_a}{2\epsilon_s}(x + x_p)^2 \qquad (-x_p \le x \le 0) \tag{7.21}$$

The potential in then region is determined by integrating the electric field in the n region, or

$$\phi(x) = \int \frac{eN_d}{\epsilon_s}(x_n - x)\,dx \tag{7.22}$$

Then

$$\phi(x) = \frac{eN_d}{\epsilon_s}\left(x_n \cdot x - \frac{x^2}{2}\right) + C_2' \tag{7.23}$$

where C_2' is another constant of integration. The potential is a continuous function, so setting Equation (7.21) equal to Equation (7.23) at the metallurgical junction, or at $x = 0$, gives

$$C_2' = \frac{eN_a}{2\epsilon_s} x_p^2 \tag{7.24}$$

The potential in then region can thus be written as

$$\phi(x) = \frac{eN_d}{\epsilon_s}\left(x_n \cdot x - \frac{x^2}{2}\right) + \frac{eN_a}{2\epsilon_s} x_p^2 \qquad (0 \le x \le x_n) \tag{7.25}$$

Figure 7.6 is a plot of the potential through the junction and shows the quadratic dependence on distance. The magnitude of the potential at $x = x_n$ is equal to the

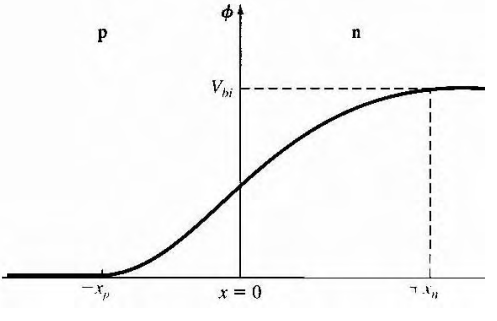


Figure 7.6 | Electric potential through the space charge region of a uniformly doped pn junction.

built-in potential barrier. Then from Equation (7.25), we have

$$V_{bi} = |\phi(x = x_n)| = \frac{e}{2\epsilon_s}\left(N_d x_n^2 + N_a x_p^2\right) \tag{7.26}$$

The potential energy of an electron is given by $E = -e\phi$, which means that the electron potential energy also varies as a quadratic function of distance through the space charge region. The quadratic dependence on distance was shown in the energy-band diagram of Figure 7.3, although we did not explicitly know the shape of the curve at that time.

7.2.3 Space Charge Width

We can determine the distance that the space charge region extends into the p and n regions from the metallurgical junction. This distance is known as the space charge width. From Equation (7.17). we may write, for example,

$$x_p = \frac{N_d x_n}{N_a} \tag{7.27}$$

Then, substituting Equation (7.27) into Equation (7.26) and solving for x_n, we obtain

$$\boxed{x_n = \left\{\frac{2\epsilon_s V_{bi}}{e}\left[\frac{N_a}{N_d}\right]\left[\frac{1}{N_a + N_d}\right]\right\}^{1/2}} \tag{7.28}$$

Equation (7.28) gives the space charge width, or the width of the depletion region, x_n extending into the n-type region for the case of zero applied voltage.

Similarly, if we solve for x_n from Equation (7.17) and substitute into Equation (7.26), we find

$$\boxed{x_p = \left\{\frac{2\epsilon_s V_{bi}}{e}\left[\frac{N_d}{N_a}\right]\left[\frac{1}{N_a + N_d}\right]\right\}^{1/2}} \tag{7.29}$$

where x_p is the width of the depletion region extending into the p region for the case of zero applied voltage.

The total depletion or space charge width W is the sum of the two components, or

$$W = x_n + x_p \tag{7.30}$$

Using Equations (7.28) and (7.29), we obtain

$$W = \left\{\frac{2\epsilon_s V_{bi}}{e}\left[\frac{N_a + N_d}{N_a N_d}\right]\right\}^{1/2} \tag{7.31}$$

The built-in potential barrier can he determined from Equation (7.10), and then the total space charge region width is obtained using Equation (7.31).

Objective

EXAMPLE 7.2

To calculate the space charge width and electric field in a *pn* junction.

Consider a silicon pn junction at T = 300 K with doping concentrations of $N_a = 10^{16}\ cm^{-3}$ and $N_d = 10^{15}\ cm^{-3}$.

■ Solution

In Example 7.1. we determined the built-in potential barrier as $V_{bi} = 0.635$ V. From Equation (7.31), the space charge width is

$$W = \left\{\frac{2\epsilon_s V_{bi}}{e}\left[\frac{N_a + N_d}{N_a N_d}\right]\right\}^{1/2}$$

$$= \left\{\frac{2(11.7)(8.85 \times 10^{-14})(0.635)}{1.6 \times 10^{-19}}\left[\frac{10^{16} + 10^{15}}{(10^{16})(10^{15})}\right]\right\}^{1/2}$$

$$= 0.951 \times 10^{-4}\ cm = 0.951\ \mu m$$

Using Equations (7.28) and (7.29). we can find $r_{,,} = 0.864\ \mu m$, and $x_p = 0.086\ \mu m$.

The peak electric field at the metallurgical junction, using Equation (7.16) for example, is

$$E_{max} = \frac{-eN_d x_n}{\epsilon_s} = \frac{-(1.6 \times 10^{-19})(10^{15})(0.864 \times 10^{-4})}{(11.7)(8.85 \times 10^{-14})} = 1.34 \times 10^4\ V/cm$$

■ Comment

The peak electric field in the space charge region of a pn junction is quite large. We must keep in mind. however, that there is no mobile charge in this region; hence there will be no drift current. We may also note, from this example, that the width of each space charge region is a reciprocal function of the doping concentration: The depletion region will extend further into the lower-doped region.

TEST YOUR UNDERSTANDING

E7.3 A silicon pn junction at T = 300 K with zero applied bias has doping concentrations of $N_d = 5 \times 10^{16}\ cm^{-3}$ and $N_a = 5 \times 10^{15}\ cm^{-3}$. Determine x_n, x_p, W, and $|E_{max}|$.

(Ans. $x_n = 4.11 \times 10^{-5}$ cm, $W = 4.52 \times 10^{-5}$ cm, $|E_{max}| = 3.18 \times 10^4$ V/cm, $x_p = 4.11 \times 10^{-6}$ cm)

E7.4 Repeat E7.3 for a GaAs pn junction. ($|E_{max}| = 3.98 \times 10^4$ V/cm, $x_n = 5.60 \times 10^{-6}$ cm, $x_p = 5.60 \times 10^{-5}$ cm, $W = 6.16 \times 10^{-5}$ cm)

7.3 | REVERSE APPLIED BIAS

If we apply a potential between the p and n regions, we will no longer be in an equilibrium condition — the Fermi energy level will no longer he constant through the system. Figure 7.7 shows the energy-band diagram of the pn junction for the case

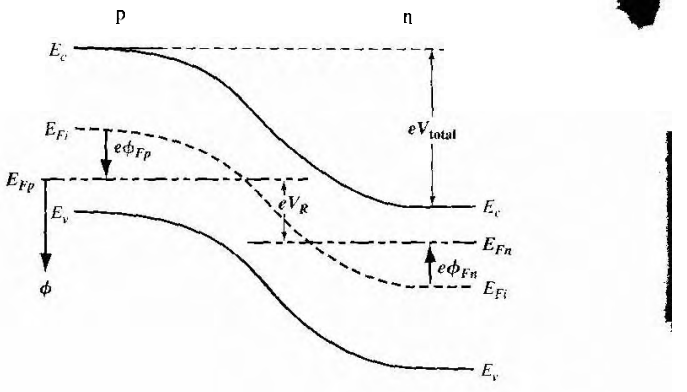


Figure 7.7 ! Energy-band diagram of a pn junction under reverse bias.

when a positive voltage is applied to the n region with respect to the p region. As the positive potential is downward, the Fermi level on then side is below the Fermi level on the p side. The difference between the two is equal to the applied voltage in units of energy.

The total potential barrier, indicated by V_{total}, has increased. The applied potential is the reverse-bias condition. The total potential harrier is now given by

$$V_{total} = |\phi_{Fn}| + |\phi_{Fp}| + V_R \tag{7.32}$$

where V_R is the magnitude of the applied reverse-bias voltage. Equation (7.32) can he rewritten as

$$V_{total} = V_{bi} + V_R \tag{7.33}$$

where V_{bi} is the same built-in potential bamer we had defined in thermal equilibrium.

7.3.1 Space Charge Width and Electric Field

Figure 7.8 shows a pn junction with an applied reverse-bias voltage V_R. Also indicated in the figure are the electric field in the space charge region and the electric field E_{app}, induced by the applied voltage. The electric fields in the neutral p and n regions are essentially zero, or at least very small, which means that the magnitude of the electric field in the space charge region must increase above the thermal-equilibrium value due to the applied voltage. The electric field originates on positive charge and terminates on negative charge; this means that the number of positive and negative charges must increase if the electric field increases. For given impurity dopingconcentrations, the number of positive and negative charges in the depletion region can be increased only if the space charge width W increases. The space charge width W increases,

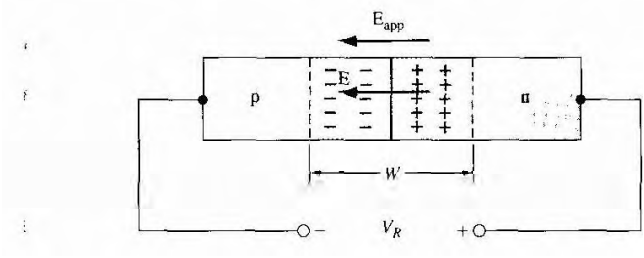


Figure 7.8 | A pn junction, with an applied reverse-bias voltage, showing the directions of the electric field induced by V_R and the space charge electric field.

therefore, with an increasing reverse-bias voltage V_R. We are assuming that the electric field in the hulk n and p regions is zero. This assumption will become clearer in the next chapter when we discuss the current-voltage characteristics.

In all of the previous equations, the built-in potential barrier can be replaced by the total potential barrier. The total space charge widrh can he written from Equation (7.31) as

$$W = \left\{ \frac{2\epsilon_s (V_{bi} + V_R)}{e} \left[\frac{N_a + N_d}{N_a N_d} \right] \right\}^{1/2} \tag{7.34}$$

showing that the total space charge widrh increases as we apply a reverse-bias voltage. By substituting the total potential barrier V_{total} into Equations **(7.28)** and (7.29), the space charge widths in the n and p regions, respectively, can be found as a function of applied reverse-bias voltage.

Objective

EXAMPLE 7.3

To calculate the width of the space charge region in a pn junction when a reverse-bias voltage is applied.

Again consider a silicon pn junction at $T = 300$ K with doping concentrations of $N_a = 10^{16}$ cm^{-3} and $N_d = 10^{15}$ c m^{-3}. Assume that $n_i = 1.5 \times 10^{10}$ cm^{-3} and let $V_R = 5$ V.

■ Solution

The built-in potential barrier was calculated in Example 7.1 for this case and is $V_{bi} = 0.635$ V. The space charge width is determined from Equation (7.34). We have

$$W = \left\{ \frac{2(11.7)(8.85 \times 10^{-14})(0.635 + 5)}{1.6 \times 10^{-19}} \left[\frac{10^{16} + 10^{15}}{(10^{16})(10^{15})} \right] \right\}^{1/2}$$

so that

$$W = 2.83 \times 10^{-4} \text{ cm} = 2.83 \ \mu\text{m}$$

■ **Comment**

The space charge width has increased from 0.951 μm at zero bias to 2.83 μm at a reverse bias of 5 V.

The magnitude of the electric field in the depletion region increases with an applied reverse-bias voltage. The electric field is still given by Equations (7.14) and (7.16) and is still a linear function of distance through the space charge region. Since x_n and x_p increase with reverse-bias voltage. the magnitude of the electric field also increases. The maximum electric field still occurs at the metallurgical junction.

The maximum electric field at the metallurgical junction, from Equations (7.14) and (7.16), is

$$E_{max} = \frac{-eN_d x_n}{\epsilon_s} = \frac{-eN_a x_p}{\epsilon_s} \tag{7.35}$$

If we use either Equation (7.28) or (7.29) in conjunction with the total potential barrier, $V_{bi} + V_R$, then

$$\boxed{E_{max} = -\left\{\frac{2e(V_{bi} + V_R)}{\epsilon_s}\left(\frac{N_a N_d}{N_a + N_d}\right)\right\}^{1/2}} \tag{7.36}$$

We can show that the maximum electric field in the pn junction can also be written as

$$\boxed{E_{max} = \frac{-2(V_{bi} + V_R)}{W}} \tag{7.37}$$

where W is the total space charge width

DESIGN EXAMPLE 7.4



Objective

To design a pn junction to meet maximum electric field and voltage specifications.

Consider a silicon pn junction at T = 300 K with a p-type doping concentration of $N_a = 10^{18}$ cm^{-3}. Determine the n-type doping concentration such that the maximum electric field is $|E_{max}| = 3 \times 10^5$ V/cm at a reverse-bias voltage of $V_R = 25$ V.

■ **Solution**

The maximum electric field is given by Equation (7.36). Neglecting V_{bi} compared to V_R, we can write

$$|E_{max}| \approx \left\{\frac{2eV_R}{\epsilon_s}\left(\frac{N_a N_d}{N_a + N_d}\right)\right\}^{1/2}$$

or

$$3 \times 10^5 = \left\{\frac{2(1.6 \times 10^{-19})(25)}{(11.7)(8.85 \times 10^{-14})}\left(\frac{10^{18} \cdot N_d}{10^{18} + N_d}\right)\right\}^{1/2}$$

which yields

$$N_d = 1.18 \times 10^{16}\ \text{cm}^{-3}$$

■ Conclusion

A smaller value of N_d results in a smaller value of $|E_{max}|$ for a given reverse-bias voltage. The value of N_d determined in this example, then, is the maximum value that will meet the specifications.

TEST YOUR UNDERSTANDING

E7.5 (*a*) A silicon pn junction at $T = 300$ K is reverse-biased at $V_R = 8$ V. The doping concentrations are $N_a = 5 \times 10^{16}\ \text{cm}^{-3}$ and $N_d = 5 \times 10^{15}\ \text{cm}^{-3}$. Determine x_n, x_p, W, and $|E_{max}|$. (*b*) Repeat pan (a) for a reverse bias voltage of $V_R = 12$ V.

[(a) $x_n = 1.43 \times 10^{-4}$ cm, $x_p = 1.43 \times 10^{-5}$ cm, $W = 1.573 \times 10^{-4}$ cm, $|E_{max}| = 1.11 \times 10^5$ V/cm; (b) $x_n = 1.73 \times 10^{-4}$ cm, $x_p = 1.73 \times 10^{-5}$ cm, $W = 1.90 \times 10^{-4}$ cm, $|E_{max}| = 1.34 \times 10^5$ V/cm]

E7.6 The maximum electric field in a reverse-biased GaAs pn junction at $T = 300$ K is to be $|E_{max}| = 2.5 \times 10^5$ V/cm. The doping concentrations are $N_d = 5 \times 10^{15}\ \text{cm}^{-3}$ and $N_a = 8 \times 10^{15}\ \text{cm}^{-3}$. Determine the reverse-hias voltage that will produce this maximum electric field. (Ans. 72.5 V)

7.3.2 Junction Capacitance

Since we have a separation of positive and negative charges in the depletion region, a capacitance is associated with the pn junction. Figure 7.9 shows the charge densities in the depletion region for applied reverse-bias voltages of V_R and $V_R + dV_R$. An increase in the reverse-bias voltage dV_R will uncover additional positive charges in then region and additional negative charges in the p region. The junction capacitance is defined as

$$C' = \frac{dQ'}{dV_R} \tag{7.38}$$

where

$$dQ' = eN_d\,dx, = eN_a\,dx_p \tag{7.39}$$

The differential charge dQ' is in units of C/cm^2 so that the capacitance C' is in units of farads per square centimeter (F/cm^2), or capacitance per unit area.

For the total potential barrier, Equation (7.28) may be written as

$$x_n = \left\{\frac{2\epsilon_s(V_{bi} + V_R)}{e}\left[\frac{N_a}{N_d}\right]\left[\frac{1}{N_a + N_d}\right]\right\}^{1/2} \tag{7.40}$$

The junction capacitance can be written as

$$C' = \frac{dQ'}{dV_R} = eN_d\,\frac{dx_n}{dV_R} \tag{7.41}$$

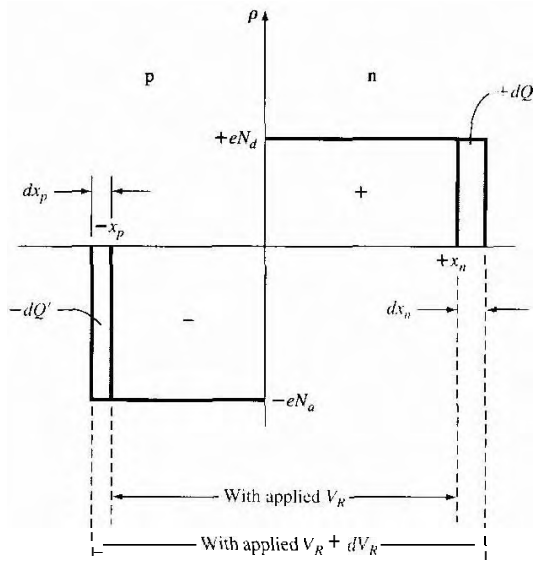


Figure 7.9 | Differential change in the space charge width with a differential change in reverse-bias voltage for a uniformly doped pn junction.

so that

$$C' = \left\{ \frac{e\epsilon_s N_a N_d}{2(V_{bi} + V_R)(N_a + N_d)} \right\}^{1/2} \tag{7.42}$$

Exactly the same capacitance expression is obtained by considering the space charge region extending into the p region x_p. The junction capacitance is also referred to as the depletion *layer capacitance.*

EXAMPLE 7.5

Objective

To calculate the junction capacitance of a pn junction.

Consider the same pn junction as that in Example 7.3. Again assume that $V_R = 5$ V.

■ **Solution**

The junction capacitance is found from Equatiun (7.42) as

$$C' = \left\{ \frac{(1.6 \times 10^{-19})(11.7)(8.85 \times 10^{-14})(10^{16})(10^{15})}{2(0.635 + 5)(10^{16} + 10^{15})} \right\}^{1/2}$$

If the cross-sectional area of the pn junction is, for example, $A = 10^{-4}$ cm^2, then the total junction capacitance is

$$C = C' \cdot A = 0.366 \times 10^{-12} \text{ F} = 0.366 \text{ pF}$$

Comment

The value of junction capacitance is usually in the pF, or smaller, range.

If we compare Equation (7.34) for the total depletion width W of the space charge region under reverse bias and Equation (7.42) for the junction capacitance C', we find that we can write

$$C' = \frac{\epsilon_s}{W} \tag{7.43}$$

Equation (7.43) is the same as the capacitance per unit area of a parallel plate capacitor. Considering Figure 7.9, we may have come to this same conclusion earlier. Keep in mind that the space charge width is a function of the reverse bias voltage so that the junction capacitance is also a function of the reverse bias voltage applied to the pn junction.

73.3 One-Sided Junctions

Consider a special pn junction called the one-sided junction. If, for example, $N_a \gg N_d$, this junction is referred to as a p^+n junction. The total space charge width, from Equation (7.34), reduces to

$$W \approx \left\{ \frac{2\epsilon_s (V_{bi} + V_R)}{eN_d} \right\}^{1/2} \tag{7.44}$$

Considering the expressions for x_n and x_p, we have for the p^+n junction

$$x_p \ll x_n \tag{7.45}$$

and

$$W \approx x_n \tag{7.46}$$

Almost the entire space charge layer extends into the low-doped region of the junction. This effect can be seen in Figure 7.10.

The junction capacitance of the p^+n junction reduces to

$$C' \approx \left\{ \frac{e\epsilon_s N_d}{2(V_{bi} + V_R)} \right\}^{1/2} \tag{7.47}$$

The depletion layer capacitance of a one-sided junction is a function of the doping concentration in the low-doped region. Equation (7.47) may be manipulated to give

$$\left(\frac{1}{C'} \right)^2 = \frac{2(V_{bi} + V_R)}{e\epsilon_s N_d} \tag{7.48}$$

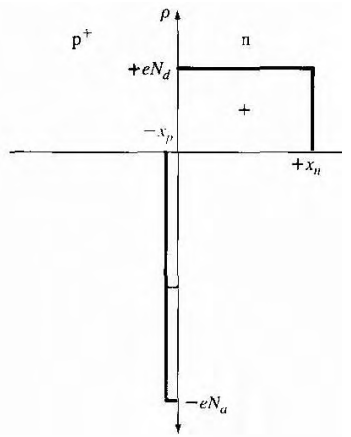


Figure 7.10 | Space charge density of a one-sided p^+n junction.

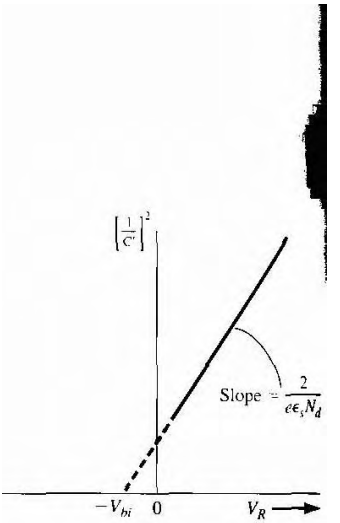


Figure 7.11 | $(1/C')^2$ versus V_R of a uniformly doped pn junction.

which shows that the inverse capacitance squared is a linear function of applied reverse-bias voltage.

Figure **7.11** shows a plot of Equation **(7.48).** The built-in potential of the junction can be determined by extrapolating the curve to the point where $(1/C')^2 = 0$. The slope of the curve is inversely proportional to the doping concentration of the low-doped region in the junction; thus, this doping concentration can be experimentally determined. The assumptions used in the derivation of this capacitance include uniform doping in both semiconductor regions, the abrupt junction approximation, and a planar junction.

EXAMPLE 7.6 Objective

To determine the impurity doping concentrations in a p^+n junction given the parameters from Figure 7.11.

Assume a silicon p+n junction at $T = 300\ K$ with $n_i = 1.5 \times 10^{10}\ cm^{-3}$. Assume that the intercept of the curve in Figure 7.11 gives $V_{bi} = 0.855$ V and that the slope is $1.32 \times 10^{15}\ (F/cm^2)^{-2}\ (V)^{-1}$.

■ Solution

The slope of the curve in Figure 7.11 is given by $2/e\epsilon_s N_d$, so we may write

$$N_d = \frac{2}{e\epsilon_s(\text{slope})} = \frac{2}{(1.6 \times 10^{-19})(11.7)(8.85 \times 10^{-14})(1.32 \times 10^{15})}$$

or

rom the expression for V_{bi}, which is

$$V_{bi} = V_t \ln\left(\frac{N_a N_d}{n_i^2}\right) = \frac{kT}{e}\ln\left(\frac{N_a N_d}{n_i^2}\right)$$

we can solve for N_a as

$$N_a = \frac{n_i^2}{N_d}\exp\left(\frac{eV_{bi}}{kT}\right) = \frac{(1.5 \times 10^{10})^2}{9.15 \times 10^{15}}\exp\left(\frac{0.855}{0.0259}\right)$$

ch yields

$$N_a = 5.34 \times 10^{18}\ \text{cm}^{-3}$$

Comment

The results of this example show that $N_a \gg N_d$; therefore the assumption of a one-sided junction was valid.

A one-sided pn junction is useful for experimentally determining the doping concentrations and built-in potential.

TEST YOUR UNDERSTANDING

E7.7 A silicon pn junction at T = 300 K has doping concentrations of $N_d = 3 \times 10^{16}\ \text{cm}^{-3}$ and $N_a = 8 \times 10^{15}\ \text{cm}^{-3}$, and has a cross-sectional area of $A = 5 \times 10^{-5}\ \text{cm}^2$. Determine the junction capacitance at (a) $V_R = 2$ V and (b) $V_R = 5$ V.

[Ans. (a) 0.478 pF, (b) 0.469 pF]

E7.8 The experimentally measured junction capacitance of a one-sided silicon n^+p junction biased at $V_R = 4$ V at T = 300 K is C = 1.10 pF. The built-in potential barrier is found to be $V_{bi} = 0.782$ V. The cross-sectional area is $A = 10^{-4}\ \text{cm}^2$. Find the doping concenlrations. (Ans. $N_a = 7 \times 10^{17}\ \text{cm}^{-3}$, $N_d = 4.17 \times 10^{15}\ \text{cm}^{-3}$)

*7.4 NONUNIFORMLY DOPED JUNCTIONS

In the pn junctions considered so far, we have assumed that each semiconductor region has been uniformly doped. In actual pn junction structures, this is not always true. In some electronic applications, specific nonuniform doping profiles are used to obtain special pn junction capacitance characteristics.

7.4.1 Linearly Graded Junctions

If we start with a uniformly duped n-type semiconductor, for example, and diffuse acceptor atoms through the surface, the impurity concentrations will tend to be like those shown in Figure 7.12. The point x = x' on the figure corresponds to the metallurgical junction. The depletion region extends into the p and n regions from the metallurgical junction as we have discussed previously. The net p-type doping

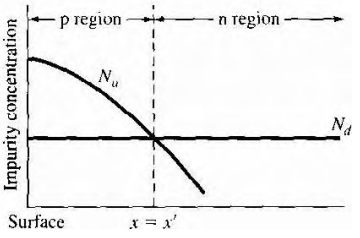


Figure 7.12 | Impurity concentrations of a pn junction with a nonuniformly doped p region.

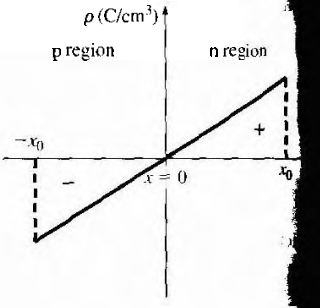


Figure 7.13 | Space charge density i linearly graded pn junction.

concentration near the metallurgical junction may be approximated as a linear func tion of distance from the metallurgical junction. Likewise, as a first approximatio the net n-type doping concentration is also a linear function of distance extendi into the n region from the metallurgical junction. This effective doping profile referred to as a linearly graded junction.

Figure 7.13 shows the space charge density in the depletion region of the li early graded junction. For convenience, the metallurgical junction is placed at x = The space charge density can he written as

$$\rho(x) = eax \tag{7.4}$$

where a is the gradient of the net impurity concentration.

The electric field and potential in the space charge region can be determine from Poisson's equation. We can write

$$\frac{dE}{dx} = \frac{\rho(x)}{\epsilon_s} = \frac{eax}{\epsilon_s} \tag{7.}$$

so that the electric field can be found by integration as

$$E = \int \frac{eax}{\epsilon_s}\,dx = \frac{ea}{2\epsilon_s}\left(x^2 - x_0^2\right) \tag{7.}$$

The electric field in the linearly graded junction is a quadratic function of distanc rather than the linear function found in the uniformly doped junction. The maximum electric field again occurs at the metallurgical junction. We may note that the electric field is zero at both $x = +x_0$ and at $x = -x_0$. The electric field in a nonuniformly doped semiconductor is not exactly zero, but the magnitude of this field is small, so setting $E = 0$ in the bulk regions is still a good approximation.

The potential is again found by integrating the electric field as

$$\phi(x) = -\int E\,dx \tag{7.52}$$

If we arbitrarily set $\phi = 0$ at $x = -x_0$, then the potential through the junction is

$$\phi(x) = \frac{-ea}{2\epsilon_s}\left(\frac{x^3}{3} - x_0^2 x\right) + \frac{ea}{3\epsilon_s}x_0^3 \tag{7.53}$$

The magnitude of the potential at $x = +x_0$ will equal the built-in potential barrier for this function. We then have that

$$\phi(x_0) = \frac{2}{3}\cdot\frac{eax_0^3}{\epsilon_s} = V_{bi} \tag{7.54}$$

Another expression for the built-in potential barrier for a linearly graded junction can be approximated from the expression used for a uniformly doped junction. We can write

$$V_{bi} = V_t \ln\left[\frac{N_d(x_0)N_a(-x_0)}{n_i^2}\right] \tag{7.55}$$

where $N_d(x_0)$ and $N_a(-x_0)$ are the doping concentrations at the edges of the space charge region. We can relate these doping concentrations to the gradient, so that

$$N_d(x_0) = ax_0 \tag{7.56a}$$

and

$$N_a(-x_0) = ax_0 \tag{7.56b}$$

Then the built-in potential barrier for the linearly graded junction becomes

$$V_{bi} = V_t \ln\left(\frac{ax_0}{n_i}\right)^2 \tag{7.57}$$

There may be situations in which the doping gradient is not the same on either side of the junction, but we will not consider that condition here.

If a reverse-bias voltage is applied to the junction, the potential barrier increases. The built-in potential barrier V_{bi} in the above equations is then replaced by the total potential barrier $V_{bi} + V_R$. Solving for x_0 from Equation (7.54) and using the total potential barrier, we obtain

$$x_0 = \left\{\frac{3}{2}\cdot\frac{\epsilon_s}{ea}(V_{bi} + V_R)\right\}^{1/3} \tag{7.58}$$

The junction capacitance per unit area can be determined by the same method as we used for the uniformly doped junction. Figure 7.14 shows the differential charge dQ' which is uncovered as a differential voltage dV_R is applied. The junction capacitance is then

$$C' = \frac{dQ'}{dV_R} = (eax_0)\frac{dx_0}{dV_R} \tag{7.59}$$

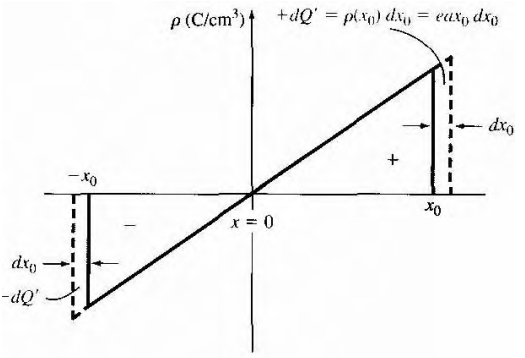


Figure 7.14 | Differential change in space charge width with a differential change in reverse-bias voltage for a linearly graded pn junction.

Using Equation (7.58), we obtain[1]

$$C' = \left\{ \frac{ea\epsilon_s^2}{12(V_{bi} + V_R)} \right\}^{1/3} \tag{7.60}$$

We may note that C' is proportional to $(V_{bi} + V_R)^{-1/3}$ for the linearly graded junction as compared to $C' \alpha (V_{bi} + V_R)^{-1/2}$ for the uniformly doped junction. In the linearly graded junction, the capacitance is less dependent on reverse-bias voltage than in the uniformly doped junction.

7.4.2 Hyperabrupt Junctions

The uniformly doped junction and linearly graded junction are not the only possible doping profiles. Figure 7.15 shows a generalized one-sided p^+n junction where the generalized n-type doping concentration for $x > 0$ is given by

$$N = Bx^m \tag{7.61}$$

The case of $m = 0$ corresponds to the uniformly doped junction and $m = +1$ corresponds to the linearly graded junction just discussed. The cases of $m = +2$ and $m = +3$ shown would approximate a fairly low-doped epitaxial n-type layer grown on a much more heavily doped n^+ substrate layer. When the value of m is negative, we have what is referred to as a *hyperabrupt junction.* In this case, the n-type doping is larger near the metallurgical junction than in the bulk semiconductor. Equation (7.61) is used to approximate the n-type doping over a small region near $x = x_0$ and does not hold at $x = 0$ when m is negative.

[1]In a more exact analysis, V_{bi} in Equation (7.60) is replaced by a gradient voltage. However. this analysis is beyond the scope of this text.

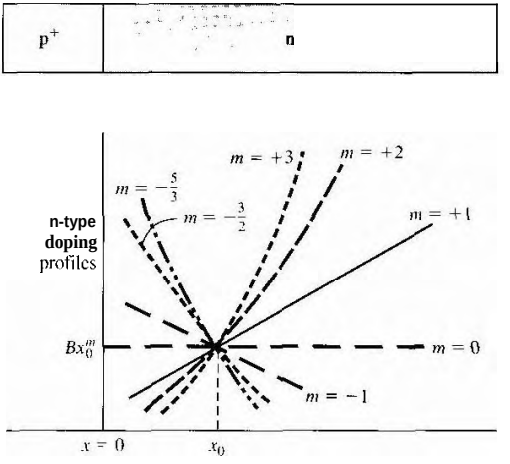


Figure 7.15 | Generalized doping profiles of a one-sided p+n junction.
(From Sze [14].)

The junction capacitance can be derived using the same analysis method as before and is given by

$$C' = \left\{ \frac{eB\epsilon_s^{(m+1)}}{(m+2)(V_{bi} + V_R)} \right\}^{1/(m+2)} \tag{7.62}$$

When m is negative, the capacitance becomes a very strong function of reverse-bias voltage, a desired characteristic in *varactor* **diodes.** The term *varactor* comes from the words *var*iable reuctor and means a device whose reactance can be varied in a controlled manner with bias voltage.

If a varactor diode and an inductance are in parallel, the resonant frequency of the LC circuit is

$$f_r = \frac{1}{2\pi\sqrt{LC}} \tag{7.63}$$

The capacitance of the diode, from Equation (7.62). can be written in the form

$$C = C_0(V_{bi} + V_R)^{-1/(m+2)} \tag{7.64}$$

In a circuit application. we would, in general, like to have the resonant frequency be a linear function of reverse-bias voltage V_R, so we need

$$C \propto V^{-2} \tag{7.65}$$

From Equation (7.64). the parameter m required is found from

$$\frac{I}{m+2} = 2 \tag{7.66a}$$

or

$$m = -\frac{3}{2} \tag{7.66b}$$

A specific doping profile will yield the desired capacitance characteristic.

7.5 | SUMMARY

- A uniformly doped pn junction was initially considered. in which one region *of* a semiconductor is uniformly duped with acceptor impurities and the adjacent region is uniformly doped with donor impurities. This type of junction is called a homojunction.
- A space charge region, or depletion region, is formed on either side of the metallurgical junction separating the n and p regions. This region is essentially depleted of any mobile electrons or holes. A net positive charge density, due to the positively charged donor impurity ions, exists in the n region and a net negative charge density, due to the negatively charged acceptor impurity ions, exists in the p region.
- An electric field exists in the depletion region due to the net space charge density. The direction of the electric field is from the n region to the p region.
- A potential difference exists across the space-charge region. Under zero applied bias, this potential difference. known as the built-in potential barrier, maintains thermal equilibrium and holds back majority carrier electrons in the n-region and majority carrier holes in the p region.
- An applied reverse bias voltage (n region positive with respect to the p region) increases the potential barrier. increases the space charge width, and increases the magnitude of the electric field.
- As the reverse bias voltage changes, the amount of charge in the depletion region changes. This change in charge with voltage defines the junction capacitance.
- The linearly graded junction represents a nonuniformly doped pn junction. Expressions for the electric field, built-in potential harrier, and junction capacitance were derived. The functional relationships differ from those of the uniformly doped junction.
- Specific doping profiles can be used to obtain specific capacitance characteristics. A hyperabrupt junction is one in which the doping decreases away from the metallurgical junction. This type ofjunction is advantageous in varactor diodes that are used in resonant circuits.

GLOSSARY OF IMPORTANT TERMS

abrupt junction approximation The assumption that there is an abrupt discontinuity in space charge density between the space charge region and neutral semiconductor region.

built-in potential harrier The electrostatic potential difference between the p and n regions of a pn junction in thermal equilibrium.

depletion layer capacitance Another term for junction capacitance.

depletion region Another term for space charge region.

hyperabrupt junction **A** pn junction in which the doping concentration on one side decreases away from the metallurgical junction to achieve n specific capacitance-voltage characteristic.

junction capacitance The capacitance of the pn junction under reverse bias.

linearly **graded junction** **A** pn junction in which the doping concentrations on either side of the metallurgical junction are approximated by a linear distribution.

metallurgical **junction** The interface between the p- and n-doped regions of a pn junction.

une-sided junction **A** pn junction in which one side of the junction is much more heavily doped than the adjacent side.

reverse bias The condition in which a positive voltage is applied to the n region with respect to the p region of a pn junction so that the potential barrier between the two regions increases above the thcrmal-equilibrium built-in potential barrier.

space charge region The region on either side of the metallurgical junction in which there is a net charge density due to ionized donors in the n-region and ionized acceptors in the p region.

space charge width The width of the space charge region, a function of doping concentrations and applied voltage.

raractor diode **A** diode whose reactance can be varied in a controlled manner with bias voltage.

CHECKPOINT

After studying this chapter, the reader should have the ability to:

- Describe why and how the space charge region is formed.
- Draw the energy band diagram of a zero-biased and reverse-biased pn junction.
- Define and derive the expression of the built-in potential barrier voltage.
- Derive the expression for the electric field in space charge region of the pn junction.
- Describe what happens to the parameters of the space charge region when a reverse bias voltage is applied.
- Define and explain the junction capacitance.
- Describe the characteristics and properties of a one-sided pn junction.
- Describe how a linearly graded junction is formed.
- Define a hyperabrupt junction.

REVIEW QUESTIONS

Define the built-in potential voltage and describe how it maintains thermal equilibrium.

Why is an electric field formed in the space charge region? Why is the electric field a linear function of distance in a uniformly doped pn junction?

3. Where does the maximum electric field occur in the space charge region?
4. Why is the space charge width larger in the lower doped side of a pn junction?
5. What is the functional dependence of the space charge width on reverse bias voltage?
6. Why does the space charge width increase with reverse bias voltage?
7. Why does a capacitance exist in a reverse-biased pn junction? Why dues the capacitance decrease with increasing reverse bias voltage?

8. What is a one-sided pn junction? What parameters can be determined in a one-sided pn junction?
9. What is **a** linearly graded junction?
10. What is a hyperabrupt junction and what is one advantage or characteristic of such a junction'?

PROBLEMS

Section 7.2 Zero Applied Bias

7.1 (*a*) Calculate V_{bi} in a silicon pn junction at $T = 300$ K for (*a*) $N_d = 10^{15}$ cm^{-3} and $N_a = (i)\ 10^{15}$, (ii) 10^{16}, (*iii*) 10^{17}, (*iv*) 10^{18} cm^{-3}. (*b*) Repeat part (*a*) for $N_d = 10^{18}$ cm^{-3}.

7.2 Calculate the built-in potential harrier, V_{bi}, for Si. Ge, and GaAs pn junctions if they each have the following dopant concentrations at $T = 300$ K:

(*a*) $N_d = 10^{14}$ cm^{-3} $N_a = 10^{17}$ cm^{-3}

(*b*) $N_d = 5 \times 10^{16}$ $N_a = 5 \times 10^{16}$

(*c*) $N_{,,} = 10^{17}$ $N_a = 10^{17}$



7.3 (*a*) Plat the built-in potential barrier for a symmetrical ($N_{,} = N_d$) silicon pn junction at $T = 300$ K over the range $10^{14} \leq N_a = N_d \leq 10^{19}$ cm^{-3}. (b) Repeat pan (*a*) for a GaAs pn junction.



7.4 Consider a uniformly doped GaAs pn junction with doping concentrations of $N_a = 5 \times 10^{18}$ cm^{-3} and $N_d = 5 \times 10^{16}$ cm^{-3}. Plot the built-in potential barrier voltage. V_{bi}, versus temperature for $200 \leq T \leq 500$ K.

7.5 An abrupt silicon pn junction at zero bias has dopant concentrations of $N_a = 10^{17}$ cm^{-3} and $N_d = 5 \times 10^{15}$ cm^{-3}. T = 300 K. (a) Calculate the Fermi level on each side of the junction with respect to the intrinsic Fermi level. (*b*) Sketch the equilibrium energy-band diagram for the junction and determine V_{bi} from the diagram and the results of pan (*a*). (*c*) Calculate V_{bi} using Equation (7.10). and compare the results to pan (*b*). (*d*) Determiner,,, r,,, **and** the peakelectric field for thisjunction.

7.6 Repent problem 7.5 for the case when the doping concentratinns are $N_{,} = N_d = 2 \times 10^{16}$ cm^{-3}.

7.7 A silicon abrupt junction in thermal equilibrium at $T = 300$K is doped such that $E_{,.} - E_f = 0.21$ eV in then region and $E_f - E_{,} = 0.18$ eV in the p region. (*a*) Draw the energy band diagram of the pn junction. (b) Determine the impurity doping concentrations in each region. (*c*) Determine V_{bi}.

7.8 Consider the uniformly doped GaAs junction at T = 300 K. At zero bias, only 20 percent of the total space charge region is to he in the p region. The built-in potential barrier is $V_{bi} = 1.20$ V. For zero bias, determine (*a*) N_a, (*b*) N_d, (*c*) x_n, (*d*) x_p, and (*e*) E_{max}.

7.9 Consider the impurity doping profile shown in Figure 7.16 in a silicon pn junction. For rero applied voltage, (*a*) determine V_{bi}, (*b*) calculate $x_{,,}$ and x_p, (*c*) sketch the thermal cquilihrium energy band diagram, and (d) plot the electric field versus distance through the junction.

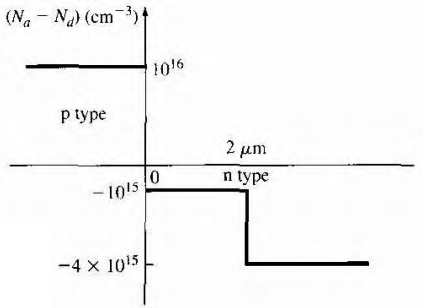


Figure 7.16 | Figure for Problem 7.9.

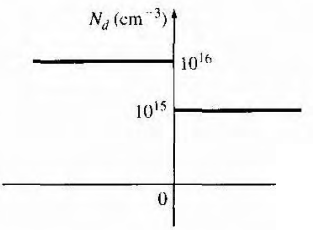


Figure 7.17 | Figure for Problem 7.12.

***7.10** **A** uniformly doped silicon pn junction is doped lo levels of $N_d = 5 \times 10^{15}$ cm^{-3} and $N_a = 10^{16}$ cm^{-3}. The measured built-in potential barrier is $V_{bi} = 0.40$ V. Determine the temperature at which this result occurs. (You may have to use trial and error to solve thir problem.)

7.11 Consider a uniformly doped silicon pn junction with duping concentrations N, $= 5 \times 10^{17}$ cm^{-3} and $N_d = 10^{17}$ cm^{-3}. (a) Calculate V_{bi} at T = 300 K. (*b*) Determine the temperature at which V_{bi} decreases by 1 percent.

7.12 An "isotype" step junction is one in which the same impurity type doping changes from one concentration value to another value. **An** n-n isotype doping profile is shown in Figure 7.17. (a) Sketch the thermal equilibrium energy band diagram of the isotype junction. *(b)* Using the energy band diagram. determine the built-in potential barrier. (c) Discuss the charge distribution through the junction.

7.13 A particular type of junction is an n region adjacent to an intrinsic region. This junction can he modeled as an n-type region to a lightly doped p-type region. Assume the doping concentrations in silicon at T = 300K are $N_d = 10^{16}$ cm^{-3} and $N_a = 10^{12}$ cm^{-3}. For zero applied bias. determine (*a*) V_{bi}, (b) $\mathbf{x_{n}}$. (*c*) x_p, and (d) E_{max}. Sketch the electric field versus distance through the junction.

7.14 We are assuming an abrupt depletion approximation for the space charge region. That is, no free carriers exist within the depletion region and the semiconductor abruptly changes to a neutral region outside the space charge region. This approximation is adequate for most applications. but the abrupt transition does not exist. The space charge region changes over a distance of a few Debye lengths, where the Debye length in then rcgion is given by

$$L_D = \left[\frac{\epsilon_s kT}{e^2 N_d}\right]^{1/2}$$

Calculate L_D and find the ratio of L_D/x_n for the following conditions. The p-type doping concentration is $N_a = 8 \times 10^{17}$ cm^{-3} and the n-type doping concentration is (*a*) $N_d = 8 \times 10^{14}$ cm^{-3}, *(b)* $N_d = 2.2 \times 10^{16}$ cm^{-3} and (*c*) $N_d = 8 \times 10^{17}$ cm^{-3}.

7.15 Examine how the electric field versus distance through a uniformly doped pn junction varies as the doping concentrations vary. For example, consider $N_d = 10^{18}$ cm^{-3} and let $10^{14} \leq N_a \leq 10^{18}$ cm^{-3}. then consider $N_d = 10^{14}$ cm^{-3} and let

$10^{14} \le N_a \le 10^{18}$ cm^{-3}, and finally consider $N_d = 10^{16}$ cm^{-3} and let $10^{14} \le N_a \le 10^{18}$ cm^{-3}. What can be said about the results for $N_a \ge 100N_d$ or $N_d \ge 100N_a$? Assume zero applied bias.

Section 7.3 Reverse Applied Bias

7.16 An abrupt silicon pnjunction has dopant concentrations of N„ = 2 x 10^{16} cm^{-3} and $N_d = 2 \times 10^{15}$ cm^{-3} at $T = 300$ K. Calculate *(a)* V_{bi}, *(h)* W at $V_R = 0$ and $V_R = 8$ V, and *(c)* the maximum electric field in the space charge region at $V_R = 0$ and $V_R = 8$ V.

7.17 Consider the junction described in Problem 7.11. The junction has a cross-sectional area of 10^{-4} cm^2 and has an applied reverse-bias voltage of $V_R = 5$ V. Calculate *(a)* V_{bi}, *(b)* x_n, x_p, W. *(c)* E_{max}, and *id)* the total junction capacitance.

7.18 An ideal one-sided silicon n$^+$p junction has uniform doping on both sides of the abrupt junction. The doping relation is $N_d = 50N_a$. The built-in potential barrier is $V_{bi} = 0.752$ V. The maximum electric field in the junction is $E_{max} = 1.14 \times 10^4$ V/cm for a reverse-bias voltage of 10 V. $T = 300$ K. Determine *(a)* N_a, N_d *(b)* x_p for $V_R = 10$, and *(c)* C_j' for $V_R = 10$.

7.19 A silicon n$^+$p junction is biased at $V_R = 10$ V. Determine the percent change in (a)junction capacitance and *(b)* built-in potential if the doping in the p region increases by a factor of 2.

7.20 Consider two p$^+$n silicon junctions at $T = 300$ K reverse biased at $V_R = 5$ V. The impurity doping concentrations in junction *A* are $N_a = 10^{18}$ cm^{-3} and $N_d = 10^{15}$ cm^{-3}, and those in junction B are N„ = 10^{18} cm^{-3} and $N_d = 10^{16}$ cm^{-3}. Calculate the ratio of the following parameters for junction *A* to junction *B*: *(a)* W. *(b)* $|E_{max}|$, and *(c)* C_j'.

7.21 *(a)*The peak electric field in a reverse-biased silicon pn junction is $|E_{max}| = 3 \times 10^5$ Vlcm. The doping concentrations are N, = 4 x 10^{15} cm^{-3} and $N_a = 4 \times 10^{17}$ cm^{-3}. Find the magnitude of the reverse-bias voltage. *(b)* Repeat pan *(a)* for $N_d = 4 \times 10^{16}$ cm^{-3} and N„ = 4 × 10^{17} c m . **(r)** Repeat part *(n)* for N_d = N, = 4 x 10^{17} cm^{-3}

7.22 Consider a uniformly doped GaAs pn junction at $T = 300$ K. The junction capacitance at zero bias is C, (0) and the junction capacitance with a 10-V reverse-bias voltage is $C_j(10)$. The ratio of the capacitances is

$$\frac{C_j(0)}{C_j(10)} = 3.13$$

Also under reverse bias, the space charge width into the p region is 0.2 of the total space charge width. Determine *(a)* V_{bi} and *(b)* N„, N_d.

7.23 GaAs pn junction at T = 300 K has impurity doping concentratians of $N_a = 10^{16}$ cm^{-3} and $N_d = 5 \times 10^{16}$ cm^{-3}. For a particular device application, the ratio ofjunction capacitances at two values of reverse bias voltage must he $C_j'(V_{R1})/C_j'(V_{R2}) = 3$ where the reverse bias voltage $V_{R1} = 1$ V. Determine V_{R2}.

7.24 An abrupt silicon pn junction at T = 300 K is uniformly doped with N„ = 10^{18} cm^{-3} and $N_d = 10^{15}$ cm^{-3}. The pn junction area is 6 x 10^{-4} cm^2. An inductance of 2.2 millihenry is placed in parallel with the pn junction. Calculate the resonant frequency of the circuit for reverse-bias voltages of *(a)* $V_R = 1$ V and *(b)* $V_R = 10$ V.

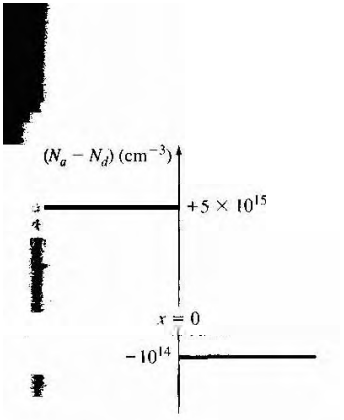


Figure 7.18 | Figure for Problem 7.27.

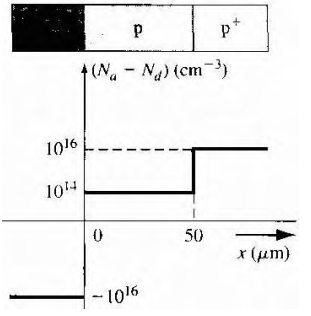


Figure 7.19 | Figure for Problem 7.28

7.25 A uniformly doped silicon p+n junction at $T = 300$ K is to bc designed such that at a reverse-bias voltage of $V_R = 10$ V, the maximum electric field is limited to $E_{max} = 10^6$ V/cm. Determine the maximum doping concentration in the n region.

7.26 A silicon pn junction is to be designed which meets the following specifications at $T = 300$ K. At a reverse-bias voltage of 1.2 V, 10 percent of the total space charge region is to be in then region and the total junction capacitance is to be 3.5×10^{-12} F with a cross-sectional area of 5.5×10^{-4} cm^2. Detennine (*a*) N_a, (*b*) N_d, and (*c*) V_{bi}.

7.27 A silicon pn junction at $T = 300$ K has the doping profile shown in Figure 7.18. Calculate (a) V_{bi}. (*b*) x_n and x_p at zero bias. and (*c*) the applied bias required so that $x_n = 30$ μm.

7.28 Consider a silicon pn junction with the doping profile shown in Figure 7.19. $T = 300$ K. (*a*) Calculate the applied reverse-bias voltage required so that the space charge region extends entirely through the p region. (*h*) Determine the space charge width into the n$^+$-region with the reverse-bias voltage calculated in pan (*a*). (*c*) Calculate the peak electric field for this applied voltage.

7.29 (a) A silicon p$^+$n junction has doping concentrations of $N_a = 10^{18}$ cm^{-3} and $N_d = 5 \times 10^{15}$ cm^{-3}. The cross-sectional area of the junction is $A = 5 \times 10^{-5}$ cm^2. Calculate the junction capacitance for (*i*) $V_R = 0$, (*ii*) $V_R = 3$ V, and (*iii*) $V_R = 6$ V. Plot $1/C^2$ versus V_R. Show that the slope uf the curve can be used to find N_d and that the intersection with the voltage axis yields V_{bi}. (*b*) Repeat part (*a*) if the n-type doping concentration changes to $N_d = 6 \times 10^{16}$ c m^{-3}.

7.30 The total junction capacitance of a one-sided silicon pn junction at $T = 300$ K is measured at $V_R = 50$ mV and found to be 1.3 pF. The junction area is 10^{-5} cm^2 and $V_{bi} = 0.95$ V. (*a*) Find the impurity doping concentration of the low-doped side of the junction. (*b*) Find the impurity doping concentration of the higher-doped region.

7.31 Examine how the capacitance C' and the function $(1/C')^2$ vary with reverse-bias voltage V_R as the doping concentrations change. In particular, consider these plots versus N_a for $N_a \geq 100 N_d$ and versus N_d for $N_d \geq 100 N_a$

***7.32** A pn junction has the doping profile shown in Figure 7.20. Assume that $x_n > x_u$ for all reverse-bias voltages. (*a*) What is the built-in potential across the junction? (*b*) For

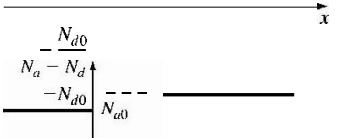


Figure 7.20 | Figure for Problem 7.32.

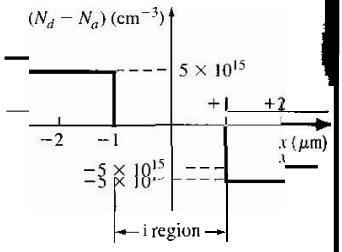


Figure 7.21 | Figure for Problem 7.33.

the abrupt junction approximation, cketch the charge density through the junction. (*b*) Derive the expression for the electric field through the space charge region.

***7.33** A silicon PIN junction has the doping profile shown in Figure 7.21. The "I" corresponds to an ideal intrinsic region in which there is no impurity doping concentration. A reverse-hias voltage is applied to the PIN junction so that the total depletion width extends from -2 μm to $+2$ μm. (a) Using Poisson's equation, calculate the magnitude of the electric field at $x = 0$. (*b*) Sketch the electric field through the PIN junction. (*c*) Calculate the reverse-bias voltage that must be applied.

Section 7.4 Nonuniformly Doped Junctions

7.34 Consider a linearly graded junction. (a) Starting with Equation (7.49), derive the expression for the electric field given in Equation (7.51). (*b*) Derive the expression for the potential through the space charge region given hy Equation (7.53).

7.35 The built-in potential barrier of a linearly graded silicon pn junction at T = 300 K is $V_{bi} = 0.70$ V The junction capacitance measured at $V_R = 3.5$ V is C' = 7.2 x F/cm^2. Find the gradient, a, of the net impurity concentration.

Summary and Review



7.36 Anne-sided p+n silicon diode at $T = 300$ K is doped at $N_a = 10^{18}$ cm^{-3}. Design the junction so that C, = 0.95 pF at $V_R = 3.5$ V. Calculate the junction capacitance when $V_R = 1.5$ V.

7.37 A one-sided p^+n junction with a cross-sectional **area** of 10^{-5} cm^2 has a measured built-in potential of $V_{bi} = 0.8$ V at T = 300 K. A plot of $(1/C_j)^2$ versus V_R is linear for $V_R < 1$ V and is essentially constant for $V_R > 1$ V. The capacitance is C, = 0.082 pF at $V_R = 1$ V Determine the doping concentrations on either side of the metallurgical junction that will produce this capacitance characteristic.

***7.38** Silicon, at T = 300 K, is doped at $N_{d1} = 10^{15}$ cm^{-3} for $x < 0$ and $N_{d2} = 5 \times 10^{16}$ cm^{-3} for $x > 0$ to form an $n - n$ step junction. (*a*) Sketch the energy-band diagram. (*b*) Derive an expression for V_{bi}. (*c*) Sketch the charge density, electric held, and potential through the junction. (*d*) Explain where the charge density came from and is located.

***7.39** A diffused silicon pn junction has a linearly graded junction on the p side with $a = 2 \times 10^{19}$ cm^{-4}, and a uniform doping of 10^{15} cm^{-3} on the n side. (a) If the

depletion width on the p side is 0.7 μm at zero bias, find the total depletion width, built-in potential, and maximum electric field at zero bias. *(b)* Plot the potential function through the junction.

READING LIST

1. `Dimitrijev, S. *Understanding Semiconductor Devices.* New York: Oxford University Press, 2000.
2. Kano, K. *Semiconductor Devices.* Upper Saddle River, NJ: Prentice Hall, 1998.
*3. Li, S. S. *Semiconductor Physical Electronics.* New York: Plenum Press, 1993.
4. Muller, R. S., and T. I. Kamins. *Device Electronics for Integrated Circuits.* 2nd ed. New York: Wiley, 1986.
5. Navon, D. H. *Semiconductor Microdevices and Materials.* New York: Holt, Rinehnrt &Winston. 1986.
6. Neudeck, G. W. *The PN Junction Diode.* Vol. *2* of the *Modular Series* on *Solid State Devices.* 2nd ed. Reading, MA: Addison-Wesley, 1989.
*7. Ng, K. K. *Complete Guide to Semiconductor Devices.* New York: McGraw-Hill. 1995.
8. Pierret, R. F. *Semiconductor Device Fundamentals.* Reading, MA: Addison-Wesley. 1996.
*9. Roulston, D. J. *An Introduction to the Physics of Semiconductor Devices.* New York: Oxford University Press, 1999.
10. Shur, M. *Introduction to Electronic Devices.* New York: John Wiley and Sons, 1996.
*11. Shur, M. *Physics of Semiconductor Devices.* Englewood Cliffs, NJ: Prentice Hall, 1990.
12. Singh, J. *Semiconductor Devices; Basic Principles.* New York: John Wiley and Sons, 2001.
13. Streetman, B. G., and S. Bancrjee. *Solid State Electronic Devices.* 5th ed. Upper Saddle River, NJ: Prentice-Hall, 2000.
14. Sze, S. M. *Physics of Semiconductor Devices.* 2nd ed. New York: Wiley, 1981.
15. Sze. S. M. *Semiconductor Devices: Physics and Technology,* 2nd ed. New York: John Wiley and Sons, Inc., 2001
*16. Wang, S. *Fundamentals of Semiconducror Theory and Device Physics.* Englewood Cliffs. NJ: Prentice Hall. 1989.
17. Yang, E. S. *Microelectronic Devices.* New York: McGraw-Hill, 1988.

CHAPTER 6

The pn Junction Diode

PREVIEW

In the last chapter, we discussed the electrostatics of the pn junction in thermal equilibrium and under reverse bias. We determined the built-in potential barrier at thermal equilibrium and calculated the electric field in the space charge region. We also considered the junction capacitance. In this chapter, we will consider the pn junction with a forward-bias voltage applied and will determine the current–voltage characteristics. The potential barrier of the pn junction is lowered when a forward-bias voltage is applied, allowing electrons and holes to flow across the space charge region. When holes flow from the p region across the space charge region into the n region, they become excess minority carrier holes and are subject to the excess minority carrier diffusion, drift, and recombination processes discussed in Chapter 6. Likewise, when electrons from the n region flow across the space charge region into the p region, they become excess minority carrier electrons and are subject to these same processes.

When semiconductor devices with pn junctions are used in linear amplifiers, for example, time-varying signals are superimposed on the dc currents and voltages. A small sinusoidal voltage superimposed on a dc voltage applied across a pn junction will generate a small-signal sinusoidal current. The ratio of the sinusoidal current to voltage yields the small-signal admittance of the pn junction. The admittance of a forward-biased pn junction contains both conductance and capacitance terms. The capacitance, called a diffusion capacitance, differs from the junction capacitance discussed in the last chapter. Using the admittance function, the small-signal equivalent circuit of the pn junction will be developed.

The last three topics considered in this chapter are junction breakdown, switching transients. and the tunnel diode. When a sufficiently large reverse-bias voltage is applied across a pn junction, breakdown can occur, producing a large reverse-bias current in the junction, which can cause heating effects and catastrophic failure of the diode. Zener diodes, however, are designed to operate in the breakdown region.

Breakdown puts limits on the amount of voltage that can he applied across a pn junction. When a pn junction is switched from one conducting state to the other, transients in the diode current and voltage occur. The switching time of the pn junction will he discussed here, and again in later chapters which deal with the switching of transistors..

8.1 | pn JUNCTION CURRENT

When a forward-bias voltage is applied to a pn junction, a current will he induced in the device. We initially consider a qualitative discussion of how charges flow in the pn junction and then consider the mathematical derivation of the current–voltage relationship.

8.1.1 Qualitative Description of Charge Flow in a pn Junction

We can qualitatively understand the mechanism of the current in a pn junction by again considering the energy hand diagrams. Figure 8.1a shows the energy band diagram of a pn junction in thermal equilibrium that was developed in the last chapter. We argued that the potential barrier seen by the electrons, for example, holds hack the large concentration of electrons in the n region and keeps them from flowing into the p region. Similarly, the potential harrier seen by the holes holds hack the large

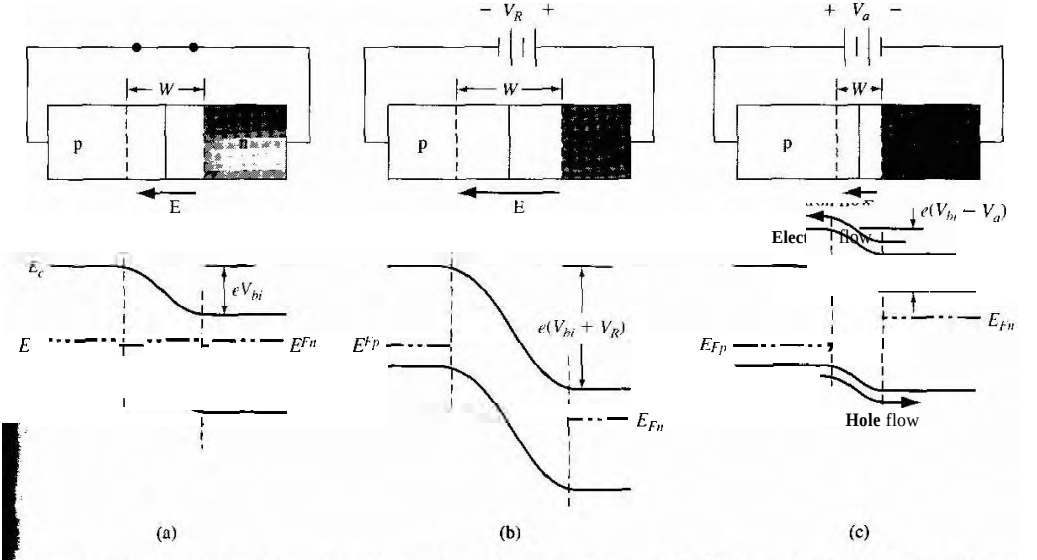


Figure 8.1 | A pn junction and its associated energy band diagram for (a) zero bias, (b) reverse bias, and (c) forward bias.

concentration of holes in the p region and keeps them from flowing into then region. The potential barrier, then, maintains thermal equilibrium.

Figure 8.1b shows the energy band diagram of a reverse-biased pn junction. The potential of the n region is positive with respect to the p region so the Fermi energy in the n region is lower than that in the p region. The total potential harrier is now larger than for the zero-bias case. We argued in the last chapter that the increased potential barrier continues to hold back the electrons and holes so that there is still essentially no charge flow and hence essentially no current.

Figure 8.1c now shows the energy band diagram for the case when a positive voltage is applied to the p region with respect to the n region. The Fermi level in the p region is now lower than that in the n region. The total potential barrier is now reduced. The smaller potential barrier means that the electric field in the depletion region is also reduced. The smaller electric field means that the electrons and holes are no longer held hack in then and p regions, respectively. There will be a diffusion of holes from the p region across the space-charge region where they now will How into the n region. Similarly, there will he a diffusion of electrons from the n region across the space-charge region where they will flow into the p region. The flow of charge generates a current through the pn junction.

The injection of holes into the n region means that these holes are minority carriers. Likewise, the injection of electrons into the p-region means thar these electrons are minority carriers. The behavior of these minority carriers is described by the ambipolar transport equations that were discussed in Chapter 6. There will be diffusion as well as recombination of excess carriers in these regions. The diffusion of carriers implies that there will be diffusion currents. The mathematical derivation of the pn junction current–voltage relationship is considered in the next section.

8.1.2 Ideal Current–Voltage Relationship

The ideal current–voltage relationship of a pn junction is derived on the basis of four assumptions. (The last assumption has three parts, but each part deals with current.) They are:

1. The abrupr depletion layer approximation applies. The space charge regions have abrupt boundaries and the semiconductor is neutral outside of the depletion region.
2. The Maxwell–Boltzmann approximation applies to carrier statistics.
3. The concept of low injection applies.

4a. The total current is a constant throughout the entire pn structure.

4b. The individual electron and hole currents are continuous functions through the pn structure.

4c. The individual electron and hole currents are constant throughout the depletion region.

Notation can sometimes appear to be overwhelming in the equations in this chapter. Table 8.1 lists some of the various electron and hole concentration terms that

Table 8.1 | Commonly used terms and notation for this chapter

Term	Meaning
N_a	Acceptor concentration in the p region of the pn junction
N_d	Donor concentration in then region of the pn junction
$n_{n0} = N_d$	Thermal equilibrium majority carrier electron concentration in the n region
$p_{p0} = N_a$	Thermal equilibrium majority carrier hole concentration in the p region
$n_{p0} = n_i^2/N_a$	Thermal equilibrium minority carrier electron concentration in the p region
$p_{n0} = n_i^2/N_d$	Thermal equilibrium minority carrier hole concentration in then region
n_p	Total minority carrier electron concentration in the p region
Pn	Total minority carrier hole concentration in the n region
$n_p(-x_p)$	Minority carrier electron concentration in the p region at the space-charge edge
$p_n(x_n)$	Minority carrier hole concentration in then region at the space charge edge
$\delta n_p = n_p - n_{p0}$	Excess minority carrier electron concentration in the p region
$\delta p_n = p_n - p_{n0}$	Excess minority carrier hole concentration in the n region

appear. Many terms have already been used in previous chapters but are repeated here for convenience.

8.1.3 Boundary Conditions

Figure 8.2 shows the conduction-band energy through the pn junction in thermal equilibrium. Then region contains many more electrons in the conduction band than the p region; the built-in potential barrier prevents this large density of electrons from flowing into the p region. The built-in potential barrier maintains equilibrium between the carrier distributions on either side of the junction.

An expression for the built-in potential barrier was derived in the last chapter and was given by Equation (7.10) as

$$V_{bi} = V_t \ln\left(\frac{N_a N_d}{n_i^2}\right)$$

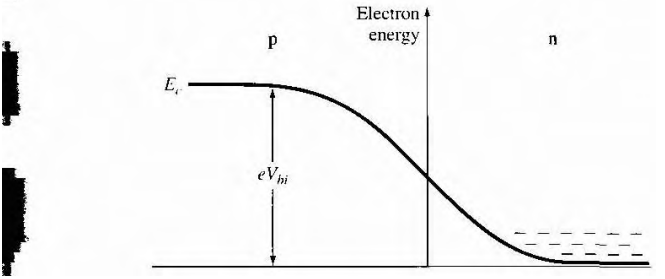


Figure 8.2 | Conduction-band energy through a pn junction.

If we divide the equation by $V_t = kT/e$, take the exponential of both sides, and then take the reciprocal, we obtain

$$\frac{n_i^2}{N_a N_d} = \exp\left(\frac{-eV_{bi}}{kT}\right) \tag{8.1}$$

If we assume complete ionization, we can write

$$n_{n0} \approx N_d \tag{8.2}$$

where n_{n0} is the thermal-equilibrium concentration of majority carrier electrons in the n region. In the p region, we can write

$$n_{p0} \approx \frac{n_i^2}{N_a} \tag{8.3}$$

where n_{p0} is the thermal-equilibrium concentration of minority carrier electrons. Substituting Equations (8.2) and (8.3) into Equation (8.1) yields

$$n_{p0} = n_{n0} \exp\left(\frac{-eV_{bi}}{kT}\right) \tag{8.4}$$

This equation relates the minority carrier electron concentration on the p side of the junction to the majority camer electron concentration on the n side of the junction in thermal equilibrium.

If a positive voltage is applied to the p region with respect to the n region, the potential barrier is reduced. Figure 8.3a shows a pn junction with an applied voltage V_a. The electric field in the bulk p and n regions is normally very small. Essentially all of the applied voltage is across the junction region. The electric field E_{app} induced by the applied voltage is in the opposite direction to the thermal equilibrium space charge electric field, so the net electric field in the space charge region is reduced below the equilibrium value. The delicate balance between diffusion and the E-field force

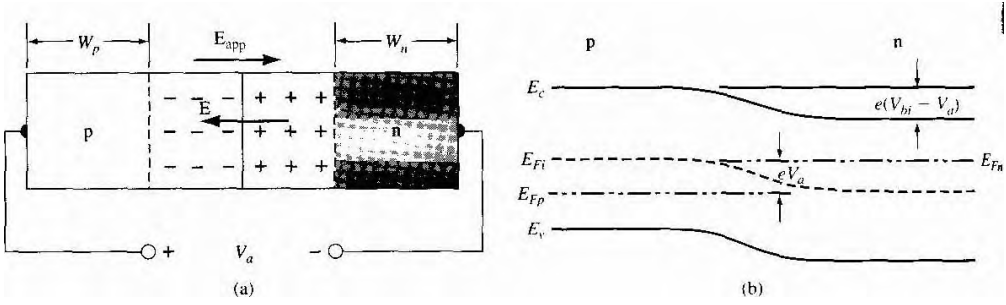


Figure 8.3 | (a) A pn junction with an applied forward-bias voltage showing the directions of the electric field induced by V_a and the space charge electric field. (b) Energy-band diagram of the forward-biased pn junction.

achieved at thermal equilibrium is upset. The electric field force that prevented majority carriers from crossing the space charge region is reduced; majority carrier electrons from the n side are now injected across the depletion region into the p material, and majority carrier holes from the p side are injected across the depletion region into then material. As long as the bias V_a is applied, the injection of carriers across the space charge region continues and a current is created in the pn junction. This bias condition is known as forward bias; the energy-band diagram of the forward-biased pn junction is shown in Figure 8.3b.

The potential barrier V_{bi} in Equation (8.4) can be replaced by $(V_{bi} - V_a)$ when the junction is forward biased. Equation (8.4) becomes

$$n_p = n_{n0} \exp\left(\frac{-e(V_{bi} - V_a)}{kT}\right) = n_{n0} \exp\left(\frac{-eV_{bi}}{kT}\right) \exp\left(\frac{-eV_a}{kT}\right) \tag{8.5}$$

If we assume low injection, the majority carrier electron concentration n_{n0}, for example, does not change significantly. However, the minority carrier concentration, n_p, can deviate from its thermal-equilibrium value n_{p0} by orders of magnitude. Using Equation (8.4), we can write Equation (8.5) as

$$\boxed{n_p = n_{p0} \exp\left(\frac{eV_a}{kT}\right)} \tag{8.6}$$

When a forward-bias voltage is applied to the pn junction, the junction is no longer in thermal equilibrium. The left side of Equation (8.6) is the total minority carrier electron concentration in the p region, which is now greater than the thermal equilibrium value. The forward-bias voltage lowers the potential bamer so that majority carrier electrons from the n region are injected across the junction into the p region, thereby increasing the minority carrier electron concentration. We have produced excess minority carrier electrons in the p region.

When the electrons are injected into the p region, these excess carriers are subject to the diffusion and recombination processes we discussed in Chapter 6. Equation (8.6). then, is the expression for the minority carrier electron concentration at the edge of the space charge region in the p region.

Exactly the same process occurs for majority carrier holes in the p region which are injected across the space charge region into the n region under a forward-bias voltage. We can write that

$$\boxed{p_n = p_{n0} \exp\left(\frac{eV_a}{kT}\right)} \tag{8.7}$$

where p_n is the concentration of minority carrier holes at the edge of the space charge region in the n region. Figure 8.4 shows these results. By applying a forward-bias voltage, we create excess minority carriers in each region of the pn junction.

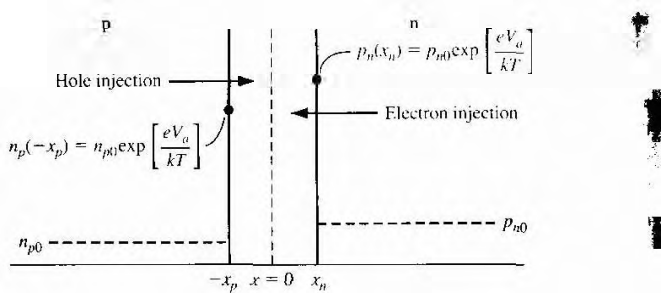


Figure 8.4 | Excess minority carrier concentrations at the space charge edges generated by the forward-bias voltage.

EXAMPLE 8.1

Objective

To calculate the minority carrier hole concentration at the edge of the space charge region of a pn junction when a forward bias is applied.

Consider a silicon pn junction at T = 300 K so that $n_i = 1.5 \times 10^{10}$ cm^{-3}. Assume the n-type doping is 1×10^{16} cm^{-3} and assume that a forward bias of 0.60 V is applied to the pn junction. Calculate the minority carrier hole concentration at thr edge of the space charge region.

■ Solution

From Equation **(8.7)** we have

$$p_n = p_{n0} \exp\left(\frac{eV_a}{kT}\right)$$

The thermal-equilibrium minority carrier hole concentration is

$$p_{n0} = \frac{n_i^2}{N_d} = \frac{(1.5 \times 10^{10})^2}{10^{16}} = 2.25 \times 10^4 \text{ cm}^{-3}$$

We then have

$$p_n = 2.25 \times 10^4 \exp\left(\frac{0.60}{0.0259}\right) = 2.59 \times 10^{14} \text{ cm}^{-3}$$

■ Comment

The minority carrier concentration can increase by many orders of magnitude when a forward-bias voltage is applied. Low injection still applies. however, since the excess-electron concentration (equal to the excess-hole concentration in order to maintain charge neutrality) is much less than the thermal-equilibrium electron concentration.

TEST YOUR UNDERSTANDING

E8.1 A silicon pn junction at T = 300 K is doped with impurity concentrations of $N_d = 5 \times 10^{16}$ cm^{-3} and $N_a = 2 \times 10^{16}$ cm^{-3}. The junction is forward biasedat $V_a = 0.610$ V.

Determine the minority carrier concentrations at the space charge edges.
[Ans. $n_p(x_p) = 7.62 \times 10^{15}\ cm^{-3}$, $p_n(-x_n) = 1.90 \times 10^{15}\ cm^{-3}$]

E8.2 The impurity doping concentrations in a silicon pn junction at T = 300 K are $N_d = 5 \times 10^{15}\ cm^{-3}$ and $N_a = 5 \times 10^{16}\ cm^{-3}$. The minority carrier concentration at either space charge edge is to be no larger than 10 percent of the respective majority carrier concentration. Calculate the maximum forward bias voltage that can be applied to this junction and still meet the required specifications.
[Ans. $V_a(\max) = 0.559$ V]

E8.3 Repeat E8.2 for a GaAs pn junction with the same doping concentrations.
[Ans. $V_a(\max) = 1.067$ V]

The minority carrier concentrations at the space charge edges, given by Equations (8.6) and (8.7), were derived assuming a forward-bias voltage ($V_a > 0$) was applied across the pn junction. However, nothing in the derivation prevents V_a from being negative (reverse bias). If a reverse-bias voltage greater than a few tenths of a volt is applied to the pn junction, then we see from Equations (8.6) and (8.7) that the minority carrier concentrations at the space charge edge are essentially zero. The minority carrier concentrations for the reverse-bias condition drop below the thermal-eauilibnum values.

8.1.4 Minority Carrier Distribution

We developed, in Chapter 6, the ambipolar transport equation for excess minority carrier holes in an n region. This equation, in one dimension, is

$$D_p \frac{\partial^2(\delta p_n)}{\partial x^2} - \mu_p E \frac{\partial(\delta p_n)}{\partial x} + g' - \frac{\delta p_n}{\tau_{p0}} = \frac{\partial(\delta p_n)}{\partial t} \tag{8.8}$$

where $\delta p_n = p_n - p_{n0}$ is the excess minority carrier hole concentration and is the difference between the total and thermal equilibrium minority carrier concentrations. The ambipolar transport equation describes the behavior of excess carriers as a function of time and spatial coordinates.

In Chapter 5, we calculated drift current densities in a semiconductor. We determined that relatively large currents could be created with fairly small electric fields. As a first approximation, we will assume that the electric field is zero in both the neutral p and n regions. In the n region for $x > x_n$, we have that E = 0 and g' = 0. If we also assume steady state so $\partial(\delta p_n)/\partial t = 0$, then Equation (8.8) reduces to

$$\frac{d^2(\delta p_n)}{dx^2} - \frac{\delta p_n}{L_p^2} = 0 \qquad (x > x_n) \tag{8.9}$$

where $L_p^2 = D_p \tau_{p0}$. For the same set of conditions, the excess minority carrier electron concentration in the p region is determined from

$$\frac{d^2(\delta n_p)}{dx^2} - \frac{\delta n_p}{L_n^2} = 0 \qquad (x < x_p) \tag{8.10}$$

The boundary conditions for the total minority carrier concentrations are

$$p_n(x_n) = p_{n0}\exp\left(\frac{eV_a}{kT}\right) \tag{8.11a}$$

$$n_p(-x_p) = n_{p0}\exp\left(\frac{eV_a}{kT}\right) \tag{8.11b}$$

$$p_n(x \to +\infty) = p_{n0} \tag{8.11c}$$

$$n_p(x \to -\infty) = n_{p0} \tag{8.11d}$$

As minority carriers diffuse from the space charge edge into the neutral semiconductor regions, they will recombine with majority carriers. We will assume that the lengths W_n and W, shown in Figure 8.3a are very long, meaning in particular that $W_n \gg L_p$ and $W_p \gg L_n$. The excess minority carrier concentrations must approach zero at distances far from the space charge region. The structure is referred to as a long pn junction.

The general solution to Equation (8.9) is

$$\delta p_n(x) = p_n(x) - p_{n0} = Ae^{x/L_p} + Be^{-x/L_p} \qquad (x \geq x_n) \tag{8.12}$$

and the general solution to Equation (8.10) is

$$\delta n_p(x) = n_p(x) - n_{p0} = Ce^{x/L_n} + De^{-x/L_n} \qquad (x \leq -x_p) \tag{8.13}$$

Applying the boundary conditions from Equations (8.11c) and (8.11d), the coefficients A and D must be zero. The coefficients B and C may be determined from the boundary conditions given by Equations (8.11a) and (8.11b). The excess carrier concentrations are then found to be, for $(x \geq x_n)$,

$$\delta p_n(x) = p_n(x) - p_{n0} = p_{n0}\left[\exp\left(\frac{eV_a}{kT}\right) - 1\right]\exp\left(\frac{-(x - x_n)}{L_p}\right) \tag{8.14}$$

and, for $(x \leq -x_p)$,

$$\delta n_p(x) = n_p(x) - n_{p0} = n_{p0}\left[\exp\left(\frac{eV_a}{kT}\right) - 1\right]\exp\left(\frac{x_p + x}{L_n}\right) \tag{8.15}$$

The minority carrier concentrations decay exponentially with distance away from the junction to their thermal-equilibrium values. Figure 8.5 shows these results. Again, we have assumed that both the n-region and the p-region lengths are long compared to the minority carrier diffusion lengths.

To review, a forward-bias voltage lowers the built-in potential barrier of a pn junction so that electrons from the n region are injected across the space charge region, creating excess minority carriers in the p region. These excess electrons begin diffusing into the bulk p region where they can recombine with majority carrier holes. The excess minority carrier electron concentration then decreases with

distance from the junction. The same discussion applies to holes injected across the space charge region into the n region.

8.1.5 Ideal pn Junction Current

The approach we use to determine the current in a pn junction is based on the three parts of the fourth assumption stated earlier in this section. The total current in the junction is the sum of the individual electron and hole currents which are constant through the depletion region. Since the electron and hole currents are continuous functions through the pn junction, the total pn junction current will be the minority carrier holediffusion current at $x = x_n$ plus the minority carrier electron diffusion current at $x = -x_p$. The gradients in the minority carrier concentrations, as shown in Figure 8.5, produce diffusion currents, and since we are assuming the electric field to be zero at the space charge edges, we can neglect any minority carrier drift current component. This approach in determining the pn junction current is shown in Figure 8.6.

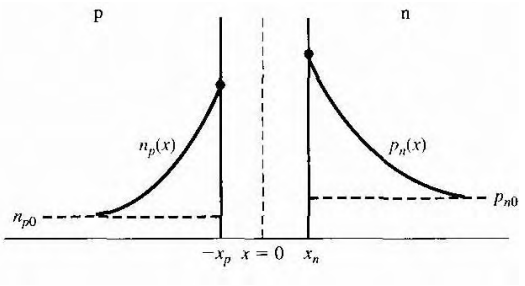


Figure 8.5 | Steady-state minority carrier concentrations in a pn junction under forward bias.

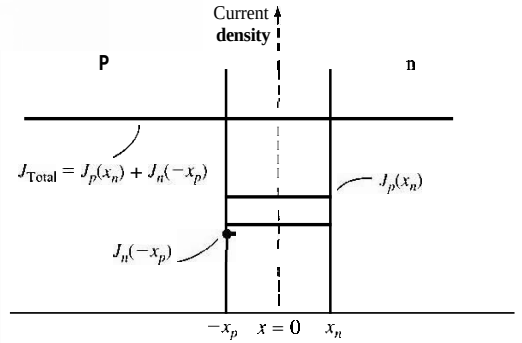


Figure 8.6 | Electron and hole current densities through the space charge region of a pn junction.

We can calculate the minority carrier hole diffusion current density at x = from the relation

$$J_p(x_n) = -eD_p \frac{dp_n(x)}{dx}\bigg|_{x=x_n} \quad (8.1$$

Since we are assuming uniformly doped regions, the thermal-equilibrium carr concentration is constant, so the hole diffusion current density may be written as

$$J_p(x_n) = -eD_p \frac{d(\delta p_n(x))}{dx}\bigg|_{x=x_n} \quad (8.17$$

Taking the derivative of Equation (8.14) and substituting into Equation (8.17). w obtain

$$\boxed{J_p(x_n) = \frac{eD_p p_{n0}}{L_p}\left[\exp\left(\frac{eV_a}{kT}\right) - 1\right]} \quad (8.1$$

The hole current density for this forward-bias condition is in the $+x$ direction, which is from the p to the n region.

Similarly, we may calculate the electron diffusion current density at x = $-x_p$. This may be written as

$$J_n(-x_p) = eD_n \frac{d(\delta n_p(x))}{dx}\bigg|_{x=-x_p} \quad (8.19$$

Using Equation (8.15). we obtain

$$\boxed{J_n(-x_p) = \frac{eD_n n_{p0}}{L_n}\left[\exp\left(\frac{eV_a}{kT}\right) - 1\right]} \quad (8.2$$

The electron current density is also in the $+x$ direction.

An assumption we made at the beginning was that the individual electron and hole currents were continuous functions and constant through the space charge region The total current is the sum of the electron and hole currents and is constant throu the entire junction. Figure 8.6 again shows a plot of the magnitudes of these curren

The total current density in the pn junction is then

$$J = J_p(x_n) + J_n(-x_p) = \left[\frac{eD_p p_{n0}}{L_p} + \frac{eD_n n_{p0}}{L_n}\right]\left[\exp\left(\frac{eV_a}{kT}\right) - 1\right] \quad (8.21)$$

Equation (8.21) is the ideal current–voltage relationship of a pn junction.

We may define a parameter J_s as

$$\frac{eD_p p_{n0}}{} \quad \frac{eD_n n_{p0}}{} \quad (8.2$$

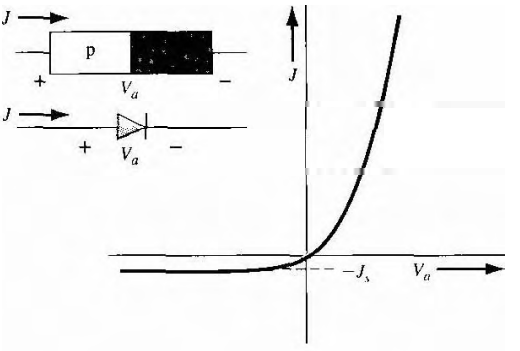


Figure 8.7 | Ideal *I–V characteristic* of a pn junction diode

so that Equadon (8.21) may he written as

$$J = J_s\left[\exp\left(\frac{eV_a}{kT}\right) - 1\right] \tag{8.23}$$

Equation (8.23), known as the ideal-diode equation, gives a good description of the current–voltage characteristics of the pn junction over a wide range of currents and voltages. Although Equation (8.23) was derived assuming a forward-bias voltage ($V_a > 0$), there is nothing to prevent V_a from being negative (reverse bias). Equation (8.23) is plotted in Figure 8.7 as a function of forward-bias voltage V_a. If the voltage V_a becomes negative (reverse bias) by a few kT/e V, then the reverse-bias current density becomes independent of the reverse-bias voltage. The parameter J_s is then referred to as the reverse saturation current density. The current–voltage characteristics of the pn junction diode are obviously not bilateral.

Objective | EXAMPLE 8.2

To determine the ideal reverse saturation current density in a silicon pn junction at T = 300 K.

Consider the following parameters in a silicon *pn* junction:

$N_a = N_d = 10^{16}$ cm^{-3} $\quad n_i = 1.5 \times 10^{10}$ cm^{-3}

$D_n = 25$ cm^2/s $\quad \tau_{p0} = \tau_{n0} = 5 \times 10^{-7}$ s

$D_p = 10$ cm^2/s $\quad \epsilon_r = 11.7$

Solution

The ideal reverse saturation current density is given by

$$J_s = \frac{eD_n n_{p0}}{L_n} + \frac{eD_p p_{n0}}{L_p}$$

which may be rewritten as

$$J_s = en_i^2 \left[\frac{1}{N_a}\sqrt{\frac{D_n}{\tau_{n0}}} + \frac{1}{N_d}\sqrt{\frac{D_p}{\tau_{p0}}} \right]$$

Substituting the parameters, we obtain $J_s = 4.15 \times 10^{-11}$ Alcm^2

■ Comment

The ideal reverse-bias saturation current density is very small. If the pn junction cross-sectional area were A = 10^{-4} cm^2, for example, then the ideal reverse-bias diode current would be $I_s = 4.15 \times 10^{-15}$ **A.**

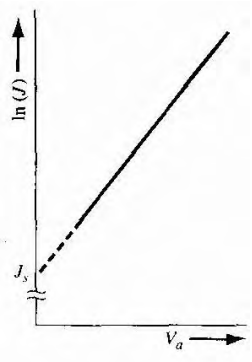


Figure 8.8 | Ideal *I–V* characteristic of *a* pn junction diode with the current plotted on a log scale.

If the forward-bias voltage in Equation (8.23) is positive by more than a few kT/evolts, then the (-1) term in Equation (8.23) becomes negligible. Figure 8.8 shows the forward-bias current–voltage characteristic when the current is plotted on a log scale. Ideally, this plot yields a straight line when V, is greater than a few kT/evolts. The forward-bias current is an exponential function of the forward-bias voltage.

DESIGN EXAMPLE 8.3



Objective

To design a pn junction diode to produce particular electron and hole current densities at a given forward-bias voltage.

Consider a silicon pn junction diode at T = 300 K. Design the diode such that J_n = 20 A/cm^2 and J_p = 5 A/cm^2 at $V_a = 0.65$ V. Assume the remaining semiconductor parameters are as given in Example 8.2.

Solution

The electron diffusion current density is given by Equation (8.20) as

$$J_n = \frac{eD_n n_{p0}}{L_n}\left[\exp\left(\frac{eV_a}{kT}\right) - 1\right] = e\sqrt{\frac{D_n}{\tau_{n0}}} \cdot \frac{n_i^2}{N_a}\left[\exp\left(\frac{eV_a}{kT}\right) - 1\right]$$

Substituting the numbers, we have

$$20 = (1.6 \times 10^{-19})\sqrt{\frac{25}{5 \times 10^{-7}}} \cdot \frac{(1.5 \times 10^{10})^2}{N_a}\left[\exp\left(\frac{0.65}{0.0259}\right) - 1\right]$$

which yields

$$N_a = 1.01 \times 10^{15}\ \text{cm}^{-3}$$

The hole diffusion current density is given by Equation (8.18) as

$$J_p = \frac{eD_p p_{n0}}{L_p}\left[\exp\left(\frac{eV_a}{kT}\right) - 1\right] = e\sqrt{\frac{D_p}{\tau_{p0}}} \cdot \frac{n_i^2}{N_d}\left[\exp\left(\frac{eV_a}{kT}\right) - 1\right]$$

Substituting the numbers, we have

$$5 = (1.6 \times 10^{-19})\sqrt{\frac{10}{5 \times 10^{-7}}} \cdot \frac{(1.5 \times 10^{10})^2}{N_d}\left[\exp\left(\frac{0.65}{0.0259}\right) - 1\right]$$

which yields

$$N_d = 2.55 \times 10^{15}\ \text{cm}^{-3}$$

Comment

The relative magnitude of the electron and hole current densities through a diode can be varied by changing the doping concentrations in the device.

TEST YOUR UNDERSTANDING

E8.4 A silicon pn junction at T = 300 K has the following parameters: $N_a = 5 \times 10^{16}\ \text{cm}^{-3}$, $N_d = 1 \times 10^{16}\ \text{cm}^{-3}$, $D_n = 25\ \text{cm}^2/\text{s}$, $D_p = 10\ \text{cm}^2/\text{s}$, $\tau_{n0} = 5 \times 10^{-7}$ s, and $\tau_{p0} = 1 \times 10^{-7}$ s. The cross-sectional area is $A = 10^{-3}\ \text{cm}^2$ and the forward-bias voltage is $V_a = 0.625$ V. Calculate the (*a*) minority electron diffusion current at the space charge edge, (b) minority hole diffusion current at the space charge edge, and (*c*) total current in the pn junction diode. [Ans. (*a*) 0.154 mA, (*b*) 1.09 mA, (*c*) 1.24 mA]

E8.5 Repeat **E8.4** for a GaAs pn junction diode biased at $V_a = 1.10$ V. [Ans. (*a*) 0.204 mA, (*b*) 1.44 mA, (*c*) 1.65 mA]

8.1.6 Summary of Physics

We have been considering the case of a forward-bias voltage being applied to a pn junction. The forward-bias voltage lowers the potential barrier so that electrons and

holes are injected across the space charge region. The injected carriers become minority carriers which then diffuse from the junction and recombine with majority carriers.

We calculated the minority carrier diffusion current densities at the edge of the space charge region. We can reconsider Equations (8.14) and (8.15) and determine the minority carrier diffusion current densities as a function of distance through the p- and n-regions. These results are

$$J_p(x) = \frac{eD_p p_{n0}}{L_p}\left[\exp\left(\frac{eV_a}{kT}\right) - 1\right]\exp\left(\frac{x_n - x}{L_p}\right) \qquad (x \geq x_n) \qquad (8.24)$$

and

$$J_n(x) = \frac{eD_n n_{p0}}{L_n}\left[\exp\left(\frac{eV_a}{kT}\right) - 1\right]\exp\left(\frac{x_p + x}{L_n}\right) \qquad (x \leq -x_p) \qquad (8.25)$$

The minority carrier diffusion current densities decay exponentially in each region. However, the total current through the pn junction is constant. The difference between total current and minority carrier diffusion current is a majority carrier current. Figure 8.9 shows the various current components through the pn structure. The drift of majority carrier holes in the p region far from the junction, for example, is to supply holes that are being injected across the space charge region into the n region and also to supply holes that are lost by recombination with excess minority carrier electrons. The same discussion applies to the drift of electrons in the n region.

We have seen that excess caniers are created in a forward-biased pn junction. From the results of the ambipolar transport theory derived in Chapter 6, the behavior of the excess carriers is determined by the minority carrier parameters for low injection. In determining the current–voltage relationship of the pn junction, we consider the Row of minority carriers since we know the behavior and characteristics of these

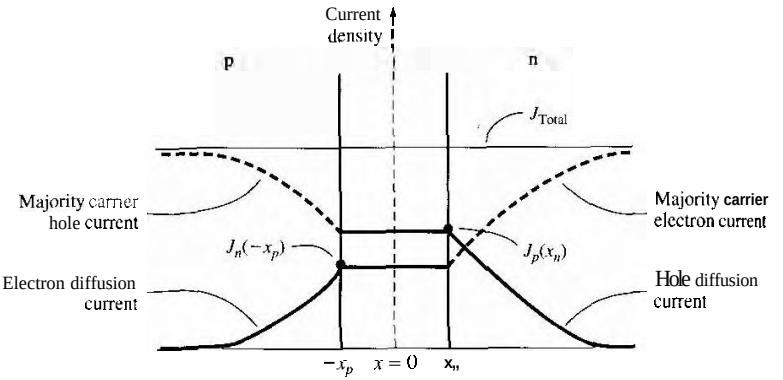


Figure 8.9 | Ideal electron and hole current components through a pn junction under forward bias.