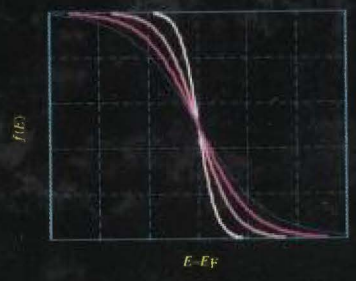
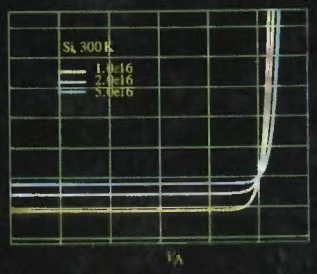
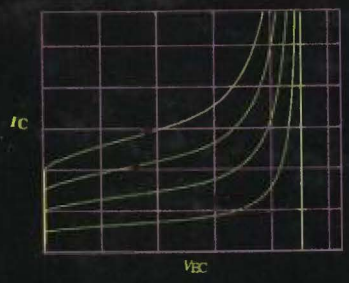
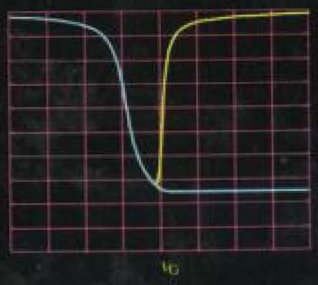
Robert F. Pierret





Semiconductor Device Fundamentals





with Computer-Based Exercises

and Homework Problems

HANWHA 1033

SEMICONDUCTOR DEVICE FUNDAMENTALS

Robert F. Pierret
School of Electrical and Computer Engineering
Purdue University

Addison Wesley Longman

Reading, Massachusetts · Menlo Park, California · New York
Don Mills, Ontario · Wokingham, England · Amsterdam · Bonn
Sydney · Singapore · Tokyo · Madrid · San Juan · Milan · Paris

Katherine Harutunian *Associate Editor*
Helen Wythe *Senior Production Supervisor*
Hugh Crawford *Manufacturing Supervisor*
Barbara Atkinson *Associate Cover Design Supervisor*
Peter Blaiwas *Cover design*
Kenneth J. Wilson *Text design*
Joyce Grandy *Copyeditor*
Sandra Rigney *Production Packaging Services*
G & S Typesetters, Inc. *Composition*
Publishers' Design and Production Services, Inc. *Illustrations*

Library of Congress Cataloging-in-Publication Data
Pierret, Robert F.
Semiconductor device fundamentals / Robert F. Pierret.
p. cm.
Includes index.
ISBN 0-201-54393-1
1. Semiconductors. I. Title.
TK7871.85.P484 1996
621.3815'2–dc20 95-17387
CIP

Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and Addison-Wesley was aware of a trademark claim, the designations have been printed in initial caps or all caps.

MATLAB is a registered trademark of The MathWorks, Inc.,
24 Prime Park Way, Natick, MA 01760-1500.
Phone: (508) 653-1415. Fax: (508) 653-2997
E-mail: info@mathworks.com
WWW:http://www.mathworks.com

Access the latest information about Addison-Wesley books from our Internet gopher site or our World Wide Web page:
gopher aw.com
http://www.aw.com/cseng/

Copyright: © 1996 by Addison-Wesley Publishing Company, Inc.

Reprinted with corrections March, 1996

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the written permission of the publisher.

Printed in the United States of America

27 28 V092 16 15

"The little voice inside never grows any older."

Frank Pierret (1906–1994)

PREFACE

Why another text on solid state devices? The author is aware of at least 14 undergraduate texts published on the subject during the past decade. Although several motivating factors could be cited, a very significant factor was the desire to write a book for the next millennium (a Book 2000 so to speak) that successfully incorporates computer-assisted learning. In a recent survey, members of the Undergraduate Curriculum Committee in the School of Electrical and Computer Engineering at Purdue University listed integration of the computer into the learning process as the number one priority. Nationally, university consortiums have been formed which emphasize computer-assisted learning. In January 1992, distribution began of the *Student Edition of MATLAB*, essentially a copy of the original MATLAB manual bundled with a low-cost version of the math-tools software. Over 37,000 copies of the book/software were sold in the first year! Texts and books on a variety of topics from several publishers are now available that make specific use of the MATLAB software. The direction is clear as we proceed into the second millennium: Computer-assisted learning will become more and more prevalent. In dealing with solid state devices, the computer allows one to address more realistic problems, to more readily experiment with "what-if" scenarios, and to conveniently obtain a graphical output. An entire device characteristic can often be computer generated with less time and effort than a small set of manually calculated single-point values.

It should be clarified that the present text is not a totally new entry in the field, but is derived in part from Volumes I–IV of the Addison-Wesley Modular Series on Solid State Devices. Lest there be a misunderstanding, the latest versions of the volumes in the Modular Series were not simply glued together. To the contrary, more than half of the material coverage in the four volumes was completely rewritten. Moreover, several supplemental sections and two additional chapters were added to the Volumes I–IV outline. The new text also contains computer-based text exercises and end-of-chapter problems, plus a number of other special features that are fully described in the General Introduction.

In just about any engineering endeavor there are tradeoffs. Device design is replete with tradeoffs. Tradeoffs also enter into the design of a book. For example, a few topics can be covered in detail (depth) or lesser coverage can be given to several topics (breadth). Similarly one can emphasize the understanding of concepts or optimize the transmission of factual information. Volumes I–IV in the Modular Series are known for their pedantic depth of coverage emphasizing concepts. While retaining the same basic depth of coverage, four "read-only" chapters have been specifically added herein to broaden the coverage and enhance the transmission of factual information. In the read-only chapters the emphasis is more on describing the exciting world of modern-day devices. Compound semiconductor devices likewise receive increased coverage throughout the text. There is also a natural

tradeoff between the effort devoted to developing qualitative insight and the implementation of a quantitative analysis. Careful attention has been given to avoid slighting the development of "intuition" in light of the greatly enhanced quantitative capabilities arising from the integrated use of the computer. Lastly, we have not attempted to be all-inclusive in the depth and breadth of coverage—many things are left for later (another course, other books). Hopefully, the proper tradeoffs have been achieved whereby the reader is reasonably knowledgeable about the subject matter and acceptably equipped to perform device analyses after completing the text.

The present text is intended for undergraduate juniors or seniors who have had at least an introductory exposure to electric field theory. Chapters are grouped into three major divisions or "parts," with Part II being further subdivided into IIA and IIB. With some deletions, the material in each of the three parts is covered during a five-week segment of a one-semester, three-credit-hour, junior-senior course in Electrical and Computer Engineering at Purdue University. A day-by-day course outline is supplied on the Instructor's Disk accompanying the Solutions Manual. If necessary to meet time constraints, read-only Chapters 4, 9, 13, and 19 could be deleted from the lecture schedule. (An instructor might preferably assign the chapters as independent readings and reward compliant students by including extra-credit examination questions covering the material.) Standard Chapters 12, 14, and 15, except for the general field-effect introduction in Section 15.1, may also be omitted with little or no loss in continuity.

Although a complete listing of special features is given in the General Introduction, instructors should take special note of the Problem Information Tables inserted prior to the end-of-chapter problems. These tables should prove useful in assigning problems and in dealing with homework graders. When faced with constructing a test, instructors may also be interested in examining the Review Problem Sets found in the mini-chapters (identified by a darkened thumb tab) at the end of the three book parts. The Review Problem Sets are derived from old "open-book" and "closed-book" tests. Concerning the computer-based exercises and problems, the use of either the student or professional version of MATLAB is recommended but not required. The in-text exercise solutions and the problem answers supplied to the instructor, however, do make use of MATLAB. Although it would be helpful, the user need not be familiar with the MATLAB program at the beginning of the book. The MATLAB problems in successive chapters make increasingly sophisticated use of the program. In other words, the early exercises and homework problems provide a learning MATLAB by using MATLAB experience. It is critical, however, that the user complete a large percentage of the computer-based exercises and problems in the first three chapters. The exercises and problems found in later chapters not only assume a reasonably competent use of MATLAB, but also build upon the programs developed in the earlier chapters.

The author gratefully acknowledges the assistance of associates, EE305 students, the respondents to an early marketing survey, the manuscript reviewers, and Addison-Wesley personnel in making Book 2000 a reality. Deserving of special thanks is Ali Keshavarzi for arranging the author's sabbatical at Intel Corporation and for providing photographs of equipment inside the Albuquerque fabrication facility. Prof. Mark Lundstrom at Purdue University was also most helpful in supplying key information and figures for several book sections. Of the undergraduate students asked to examine the manuscript for readability

and errors, Eric Bragg stands out as especially perceptive and helpful. The very conscientious manuscript reviewers were Prof. Kenneth A. James, California State University, Long Beach; Prof. Peter Lanyon, Worcester Polytechnic Institute; Prof. Gary S. May, Georgia Institute of Technology; Prof. Dieter K. Schroder, Arizona State University; and Prof. G. W. Stillman, University of Illinois at Urbana-Champaign. In recognition of a fruitful association, a special thanks to Don Fowley, the former editor at Addison-Wesley who enticed the author into writing the book. Last but not least, editor Katherine Harutunian is to be credited with smoothly implementing the project, and executive assistant Anita Devine with cheerfully handling many of the early details.

Prof. Robert F. Pierret
School of Electrical and Computer Engineering
Purdue University

CONTENTS

General Introduction xxi

Part I Semiconductor Fundamentals 1

Chapter 1 Semiconductors: A General Introduction 3

1.1 General Material Properties 3
1.1.1 Composition 3
1.1.2 Purity 5
1.1.3 Structure 6
1.2 Crystal Structure 6
1.2.1 The Unit Cell Concept 7
1.2.2 Simple 3-D Unit Cells 8
1.2.3 Semiconductor Lattices 9
1.2.4 Miller Indices 12
1.3 Crystal Growth 16
1.3.1 Obtaining Ultrapure Si 16
1.3.2 Single-Crystal Formation 17
1.4 Summary 19
Problems 19

Chapter 2 Carrier Modeling 23

2.1 The Quantization Concept 23
2.2 Semiconductor Models 25
2.2.1 Bonding Model 26
2.2.2 Energy Band Model 26
2.2.3 Carriers 29
2.2.4 Band Gap and Material Classification 31
2.3 Carrier Properties 32
2.3.1 Charge 32
2.3.2 Effective Mass 32
2.3.3 Carrier Numbers in Intrinsic Material 34

2.3.4 Manipulation of Carrier Numbers—Doping 35
2.3.5 Carrier-Related Terminology 40
2.4 State and Carrier Distributions 40
2.4.1 Density of States 41
2.4.2 The Fermi Function 42
2.4.3 Equilibrium Distribution of Carriers 46
2.5 Equilibrium Carrier Concentrations 49
2.5.1 Formulas for n and p 49
2.5.2 Alternative Expressions for n and p 52
2.5.3 n_i and the np Product 53
2.5.4 Charge Neutrality Relationship 57
2.5.5 Carrier Concentration Calculations 59
2.5.6 Determination of E_F 61
2.5.7 Carrier Concentration Temperature Dependence 65
2.6 Summary and Concluding Comments 67
Problems 69

Chapter 3 Carrier Action 75

3.1 Drift 75
3.1.1 Definition–Visualization 75
3.1.2 Drift Current 76
3.1.3 Mobility 79
3.1.4 Resistivity 85
3.1.5 Band Bending 89
3.2 Diffusion 94
3.2.1 Definition–Visualization 94
3.2.2 Hot-Point Probe Measurement 97
3.2.3 Diffusion and Total Currents 98
Diffusion Currents 98
Total Currents 99
3.2.4 Relating Diffusion Coefficients/Mobilities 99
Constancy of the Fermi Level 99
Current Flow Under Equilibrium Conditions 101
Einstein Relationship 101
3.3 Recombination–Generation 105
3.3.1 Definition–Visualization 105
Band-to-Band Recombination 105

R–G Center Recombination 105
Auger Recombination 107
Generation Processes 107
3.3.2 Momentum Considerations 107
3.3.3 R–G Statistics 110
Photogeneration 110
Indirect Thermal Recombination–Generation 112
3.3.4 Minority Carrier Lifetimes 116
General Information 116
A Lifetime Measurement 116
3.4 Equations of State 120
3.4.1 Continuity Equations 121
3.4.2 Minority Carrier Diffusion Equations 122
3.4.3 Simplifications and Solutions 124
3.4.4 Problem Solving 124
Sample Problem No. 1 124
Sample Problem No. 2 128
3.5 Supplemental Concepts 131
3.5.1 Diffusion Lengths 131
3.5.2 Quasi-Fermi Levels 132
3.6 Summary and Concluding Comments 136
Problems 138

Chapter 4 Basics of Device Fabrication 149

4.1 Fabrication Processes 149
4.1.1 Oxidation 149
4.1.2 Diffusion 152
4.1.3 Ion Implantation 155
4.1.4 Lithography 159
4.1.5 Thin-Film Deposition 162
Evaporation 162
Sputtering 162
Chemical Vapor Deposition (CVD) 164
4.1.6 Epitaxy 164
4.2 Device Fabrication Examples 165
4.2.1 *pn* Junction Diode Fabrication 166
4.2.2 Computer CPU Process Flow 166
4.3 Summary 174

R1 Part I Supplement and Review 175

Alternative/Supplemental Reading List 175

Figure Sources/Cited References 177

Review List of Terms 178

Part I—Review Problem Sets and Answers 179

Part IIA *pn* Junction Diodes 193

Chapter 5 *pn* Junction Electrostatics 195

5.1 Preliminaries 195

5.1.1 Junction Terminology/Idealized Profiles 195

5.1.2 Poisson's Equation 197

5.1.3 Qualitative Solution 198

5.1.4 The Built-in Potential (V_{bi}) 203

5.1.5 The Depletion Approximation 206

5.2 Quantitative Electrostatic Relationships 209

5.2.1 Assumptions/Definitions 209

5.2.2 Step Junction with $V_A = 0$ 210

Solution for ρ 210

Solution for $\mathscr{E}$ 210

Solution for V 212

Solution for x_n and x_p 213

5.2.3 Step Junction with $V_A \neq 0$ 215

5.2.4 Examination/Extrapolation of Results 219

5.2.5 Linearly Graded Junctions 223

5.3 Summary 226

Problems 227

Chapter 6 *pn* Junction Diode: *I–V* Characteristics 235

6.1 The Ideal Diode Equation 235

6.1.1 Qualitative Derivation 235

6.1.2 Quantitative Solution Strategy 241

General Considerations 241

Quasineutral Region Considerations 242

Depletion Region Considerations 243

Boundary Conditions 244

"Game Plan" Summary 246
6.1.3 Derivation Proper 247
6.1.4 Examination of Results 249
Ideal I–V 249
The Saturation Current 250
Carrier Currents 254
Carrier Concentrations 255
6.2 Deviations from the Ideal 260
6.2.1 Ideal Theory Versus Experiment 260
6.2.2 Reverse-Bias Breakdown 263
Avalanching 264
Zener Process 268
6.2.3 The R–G Current 270
6.2.4 $V_A \rightarrow V_{bi}$ High-Current Phenomena 277
Series Resistance 278
High-Level Injection 279
6.3 Special Considerations 281
6.3.1 Charge Control Approach 282
6.3.2 Narrow-Base Diode 284
Current Derivation 284
Limiting Cases/Punch-Through 286
6.4 Summary and Concluding Comments 288
Problems 289

Chapter 7 *pn* Junction Diode: Small-Signal Admittance 301
7.1 Introduction 301
7.2 Reverse-Bias Junction Capacitance 301
7.2.1 General Information 301
7.2.2 C–V Relationships 305
7.2.3 Parameter Extraction/Profiling 309
7.2.4 Reverse-Bias Conductance 313
7.3 Forward-Bias Diffusion Admittance 315
7.3.1 General Information 315
7.3.2 Admittance Relationships 318
7.4 Summary 323
Problems 324

Chapter 8 *pn* Junction Diode: Transient Response 327

8.1 Turn-Off Transient 327

8.1.1 Introduction 327

8.1.2 Qualitative Analysis 329

8.1.3 The Storage Delay Time 333

Quantitative Analysis 333

Measurement 334

8.1.4 General Information 338

8.2 Turn-On Transient 338

8.3 Summary 343

Problems 344

Chapter 9 Optoelectronic Diodes 347

9.1 Introduction 347

9.2 Photodiodes 349

9.2.1 *pn* Junction Photodiodes 349

9.2.2 *p-i-n* and Avalanche Photodiodes 352

p-i-n Photodiodes 352

Avalanche Photodiodes 355

9.3 Solar Cells 356

9.3.1 Solar Cell Basics 356

9.3.2 Efficiency Considerations 357

9.3.3 Solar Cell Technology 360

9.4 LEDs 361

9.4.1 General Overview 361

9.4.2 Commercial LEDs 362

9.4.3 LED Packaging and Photon Extraction 366

Part IIB BJTs and Other Junction Devices 369

Chapter 10 BJT Fundamentals 371

10.1 Terminology 371

10.2 Fabrication 374

10.3 Electrostatics 378

10.4 Introductory Operational Considerations 380

10.5 Performance Parameters 382

Emitter Efficiency 382

Base Transport Factor 383
Common Base d.c. Current Gain 383
Common Emitter d.c. Current Gain 384
10.6 Summary 385
Problems 385

Chapter 11 BJT Static Characteristics 389

11.1 Ideal Transistor Analysis 389
11.1.1 Solution Strategy 389
Basic Assumptions 389
Notation 390
Diffusion Equations/Boundary Conditions 390
Computational Relationships 392
11.1.2 General Solution (W Arbitrary) 393
Emitter/Collector Region Solutions 393
Base Region Solution 394
Performance Parameters/Terminal Currents 395
11.1.3 Simplified Relationships ($W \ll L_B$) 397
$\Delta p_B(x)$ in the Base 398
Performance Parameters 398
11.1.4 Ebers–Moll Equations and Model 403
11.2 Deviations from the Ideal 407
11.2.1 Ideal Theory/Experiment Comparison 407
11.2.2 Base Width Modulation 410
11.2.3 Punch-Through 412
11.2.4 Avalanche Multiplication and Breakdown 414
Common Base 414
Common Emitter 414
11.2.5 Geometrical Effects 420
Emitter Area ≠ Collector Area 420
Series Resistances 421
Current Crowding 421
11.2.6 Recombination–Generation Current 422
11.2.7 Graded Base 423
11.2.8 Figures of Merit 424
11.3 Modern BJT Structures 426
11.3.1 Polysilicon Emitter BJT 426
11.3.2 Heterojunction Bipolar Transistor (HBT) 429

11.4 Summary 432
Problems 433

Chapter 12 BJT Dynamic Response Modeling 443

12.1 Small-Signal Equivalent Circuits 443
12.1.1 Generalized Two-Port Model 443
12.1.2 Hybrid-Pi Models 446
12.2 Transient (Switching) Response 449
12.2.1 Qualitative Observations 449
12.2.2 Charge Control Relationships 452
12.2.3 Quantitative Analysis 454
Turn-on Transient 454
Turn-off Transient 456
12.2.4 Practical Considerations 457
12.3 Summary 458
Problems 459

Chapter 13 PNPN Devices 463

13.1 Silicon Controlled Rectifier (SCR) 463
13.2 SCR Operational Theory 465
13.3 Practical Turn-on/Turn-off Considerations 470
13.3.1 Circuit Operation 470
13.3.2 Additional Triggering Mechanisms 471
13.3.3 Shorted-Cathode Configuration 471
13.3.4 *di/dt* and *dv/dt* Effects 472
13.3.5 Triggering Time 473
13.3.6 Switching Advantages/Disadvantages 473
13.4 Other PNPN Devices 474

Chapter 14 MS Contacts and Schottky Diodes 477

14.1 Ideal MS Contacts 477
14.2 Schottky Diode 483
14.2.1 Electrostatics 483
Built-in Voltage 483
ρ, $\mathscr{E}$, V 485
Depletion Width 486
14.2.2 I–V Characteristics 487
14.2.3 a.c. Response 493

14.2.4 Transient Response 496
14.3 Practical Contact Considerations 497
14.3.1 Rectifying Contacts 497
14.3.2 Ohmic Contacts 498
14.4 Summary 500
Problems 501

R2 Part II Supplement and Review 505
Alternative/Supplemental Reading List 505
Figure Sources/Cited References 506
Review List of Terms 507
Part II—Review Problem Sets and Answers 508

Part III Field Effect Devices 523

Chapter 15 Field Effect Introduction—The J-FET and MESFET 525
15.1 General Introduction 525
15.2 J-FET 530
15.2.1 Introduction 530
15.2.2 Qualitative Theory of Operation 531
15.2.3 Quantitative I_D–V_D Relationships 536
15.2.4 a.c. Response 547
15.3 MESFET 550
15.3.1 General Information 550
15.3.2 Short-Channel Considerations 552
Variable Mobility Model 553
Saturated Velocity Model 554
Two-Region Model 555
15.4 Summary 557
Problems 557

Chapter 16 MOS Fundamentals 563
16.1 Ideal Structure Definition 563
16.2 Electrostatics—Mostly Qualitative 565
16.2.1 Visualization Aids 565
Energy Band Diagram 565
Block Charge Diagrams 566

16.2.2 Effect of an Applied Bias 567
General Observations 567
Specific Biasing Regions 568
16.3 Electrostatics—Quantitative Formulation 571
16.3.1 Semiconductor Electrostatics 571
Preparatory Considerations 571
Delta-Depletion Solution 576
16.3.2 Gate Voltage Relationship 580
16.4 Capacitance–Voltage Characteristics 584
16.4.1 Theory and Analysis 584
Qualitative Theory 584
Delta-Depletion Analysis 590
16.4.2 Computations and Observations 591
Exact Computations 591
Practical Observations 595
16.5 Summary and Concluding Comments 599
Problems 600

Chapter 17 MOSFETs—*The Essentials* 611

17.1 Qualitative Theory of Operation 611
17.2 Quantitative I_D–V_D Relationships 617
17.2.1 Preliminary Considerations 617
Threshold Voltage 617
Effective Mobility 618
17.2.2 Square-Law Theory 620
17.2.3 Bulk-Charge Theory 625
17.2.4 Charge-Sheet and Exact-Charge Theories 628
17.3 a.c. Response 630
17.3.1 Small-Signal Equivalent Circuits 630
17.3.2 Cutoff Frequency 633
17.3.3 Small-Signal Characteristics 634
17.4 Summary 637
Problems 638

Chapter 18 Nonideal MOS 645

18.1 Metal–Semiconductor Workfunction Difference 645

18.2 Oxide Charges 650
18.2.1 General Information 650
18.2.2 Mobile Ions 653
18.2.3 The Fixed Charge 658
18.2.4 Interfacial Traps 662
18.2.5 Induced Charges 668
Radiation Effects 668
Negative-Bias Instability 669
18.2.6 ΔV_G Summary 670
18.3 MOSFET Threshold Considerations 674
18.3.1 V_T Relationships 675
18.3.2 Threshold, Terminology, and Technology 676
18.3.3 Threshold Adjustment 678
18.3.4 Back Biasing 680
18.3.5 Threshold Summary 681
Problems 684

Chapter 19 Modern FET Structures 691

19.1 Small Dimension Effects 691
19.1.1 Introduction 691
19.1.2 Threshold Voltage Modification 694
Short Channel 694
Narrow Width 697
19.1.3 Parasitic BJT Action 698
19.1.4 Hot-Carrier Effects 700
Oxide Charging 700
Velocity Saturation 700
Velocity Overshoot/Ballistic Transport 701
19.2 Select Structure Survey 702
19.2.1 MOSFET Structures 702
LDD Transistors 702
DMOS 703
Buried-Channel MOSFET 704
SiGe Devices 704
SOI Structures 705
19.2.2 MODFET (HEMT) 707
Problems 710

R3 Part III Supplement and Review 713

Alternative/Supplemental Reading List 713

Figure Sources/Cited References 714

Review List of Terms 717

Part III—Review Problem Sets and Answers 718

Appendices 733

Appendix A Elements of Quantum Mechanics 733

A.1 The Quantization Concept 733

A.1.1 Blackbody Radiation 733

A.1.2 The Bohr Atom 735

A.1.3 Wave-Particle Duality 737

A.2 Basic Formalism 739

A.3 Electronic States in Atoms 741

A.3.1 The Hydrogen Atom 741

A.3.2 Multi-Electron Atoms 744

Appendix B MOS Semiconductor Electrostatics—Exact Solution 749

Definition of Parameters 749

Exact Solution 750

Appendix C MOS *C–V* Supplement 753

Appendix D MOS *I–V* Supplement 755

Appendix E List of Symbols 757

Appendix M MATLAB Program Script 771

Exercise 10.2 (BJT_Eband) 771

Exercise 11.7 (BJT) and Exercise 11.10 (BJTplus) 774

Exercise 16.5 (MOS_CV) 778

Index 781

GENERAL INTRODUCTION

Coincident with the writing of this book, there has been considerable media discussion about the "Information Superhighway." The envisioned highway itself, the physical link between points supporting the information traffic, is fiber optic cable. Relative to the topic of this book, the on and off ramps, which insert and extract the information from the highway, are semiconductor (solid state) devices. Traffic control, the information processing and the conversion to and from the human interface, is performed by computers. The central processing unit (CPU), memory, and other major components inside the computer are again semiconductor devices. In the modern world, semiconductor devices are incorporated in just about every major system from automobiles to washing machines.

Although roughly a half-century old, the field of study associated with semiconductor devices continues to be dynamic and exciting. New and improved devices are being developed at an almost maddening pace. While the device count in complex integrated circuits increases through the millions and the side-length of chips is measured in centimeters, the individual devices are literally being shrunk to atomic dimensions. Moreover, semiconductor properties desired for a given device structure but not available in nature are being produced artificially; in essence, the semiconductor properties themselves are now being engineered to fit device specifications.

This book should be viewed as a gateway to what the reader will hopefully agree is the fascinating realm of semiconductor devices. It was written for junior- or senior-level students who have at least an introductory exposure to electric field theory. The coverage includes a representative sampling of information about a wide variety of devices. Primary emphasis, however, is placed on developing a fundamental understanding of the internal workings of the more basic device structures. As detailed below, this book contains a number of unique features to assist the reader in learning the material. Alerted at an early stage to their existence, the reader can plan to take full advantage of the cited features.

- *Computer-Based Exercises and End-of Chapter (Homework) Problems.* The majority of chapters contain one or more MATLAB-based exercises requiring the use of a computer. MATLAB is a math-tools software program that has been adapted to run on most computer platforms. A low-cost student edition of MATLAB, which can be used to run all of the files associated with this book, is available in both IBM-compatible and Macintosh versions. The MATLAB program scripts yielding exercise answers are listed in the text and are available in electronic form as detailed below. Computer-based problems, identified by a bullet (●) before the problem number, make up approximately 25% of the

problem total. Although other math-tools programs could be employed, the use of MATLAB is recommended in answering computer-based problems. Because computer-based exercises and problems in the early chapters are specifically designed to progressively enhance MATLAB-use proficiency, the user need not be familiar with the MATLAB program at the beginning of the book. It is very important, however, to complete a large percentage of the computer-based exercises and problems in the first three chapters. The exercises and problems found in later chapters not only assume a reasonably competent use of MATLAB, but also build upon the programs developed in the earlier chapters.

- *Computer Program Files.* Program files of the MATLAB scripts associated with computer-based exercises are available via the Internet (`ftp.mathworks.com` in the directory `pub/books/pierret`) or on a floppy disk distributed free of charge by MathWorks, Inc. A pull-out card is provided herein for obtaining the free program disk which is formatted for use with either an IBM-compatible or Macintosh computer. Each floppy disk contains two sets of "m-files" to be used respectively with the pre-4.0 (student 1st edition) or post-4.0 (student 2nd edition) versions of MATLAB. The listings in the text are specifically derived from the Macintosh post-4.0 version, but they are identical to the corresponding IBM-compatible version except for the occasional appearance of a Greek letter.
- *Supplement and Review Mini-Chapters.* The book is divided into three parts. At the end of each part is a Supplement and Review mini-chapter. The mini-chapters, identified by a darkened thumb tab, contain an alternative/supplemental reading list and information table, reference citations for the preceding chapters, an extensive review-list of terms, and review problem sets with answers. The review problem sets are derived from "closed-book" and "open-book" examinations.
- *Read-Only Chapters.* Chapters 4, 9, 13, and 19 have been classified as "read-only." Chapters with the read-only designation contain mostly qualitative information of a supplemental nature. Two of the chapters survey some of the latest device structures. Intended to be fun-reading, the read-only chapters are strategically placed to provide a change of pace. The chapters contain only a small number of equations, no exercises, and few, if any, end-of-chapter problems. In a course format, the chapters could be skipped with little loss in continuity or preferably assigned as independent readings.
- *Problem Information Tables.* A compact table containing information about the end-of-chapter problems in a given chapter is inserted just before the problems. The information provided is (i) the text section or subsection after which the problem can be completed, (ii) the estimated problem difficulty on a scale of 1 (easy or straightforward) to 5 (very difficult or extremely time consuming), (iii) suggested credit or point weighting, and (iv) a short problem description. A bullet before the problem number identifies a computer-based problem. An asterisk indicates computer usage for part of the problem.
- *Equation Summaries.* The very basic carrier modeling equations in Chapter 2 and the carrier action equations in Chapter 3, equations referenced throughout the text, are organized and repeated in Tables 2.4 and 3.3, respectively. These tables would be ideal as "crib sheets" for closed-book examinations covering the material in Part I of the text.

- *Measurements and Data.* Contrary to the impression sometimes left by the sketches and idealized plots often found in introductory texts, device characteristics are real, seldom perfect, and are routinely recorded in measurement laboratories. Herein a sampling of measurement details and results, derived from an undergraduate EE laboratory administered by the author, is included in an attempt to convey the proper sense of reality. For added details on the described measurements, and for a description of additional measurements, the reader is referred to R. F. Pierret, *Semiconductor Measurements Laboratory Operations Manual*, Supplement A in the Modular Series on Solid State Devices, Addison-Wesley Publishing Company, Reading MA, © 1991.
- *Alternative Treatment.* Section 2.1 provides the minimum required treatment on the topic of energy quantization in atomic systems. Appendix A, which contains a more in-depth introduction to the quantization concept and related topics, has been included for those desiring supplemental information. Section 2.1 may be totally replaced by Appendix A with no loss in continuity.

PART I

SEMICONDUCTOR FUNDAMENTALS

1 Semiconductors: A General Introduction

1.1 GENERAL MATERIAL PROPERTIES

The vast majority of all solid state devices on the market today are fabricated from a class of materials known as semiconductors. It is therefore appropriate that we begin the discussion by examining the general nature of semiconducting materials.

1.1.1 Composition

Table 1.1 lists the atomic compositions of semiconductors that are likely to be encountered in the device literature. As noted, the semiconductor family of materials includes the elemental semiconductors Si and Ge, compound semiconductors such as GaAs and ZnSe, and alloys like $Al_xGa_{1-x}As$.† Due in large part to the advanced state of its fabrication technology, Si is far and away the most important of the semiconductors, totally dominating the present commercial market. The vast majority of discrete devices and integrated circuits (ICs), including the central processing unit (CPU) in microcomputers and the ignition module in modern automobiles, are made from this material. GaAs, exhibiting superior electron transport properties and special optical properties, is employed in a significant number of applications ranging from laser diodes to high-speed ICs. The remaining semiconductors are utilized in "niche" applications that are invariably of a high-speed, high-temperature, or optoelectronic nature. Given its present position of dominance, we will tend to focus our attention on Si in the text development. Where feasible, however, GaAs will be given comparable consideration and other semiconductors will be featured as the discussion warrants.

Although the number of semiconducting materials is reasonably large, the list is actually quite limited considering the total number of elements and possible combinations of elements. Note that, referring to the abbreviated periodic chart of the elements in Table 1.2, only a certain group of elements and elemental combinations typically give rise to semiconducting materials. Except for the IV-VI compounds, all of the semiconductors listed in Table 1.1 are composed of elements appearing in Column IV of the Periodic Table or are a combination of elements in Periodic Table columns an equal distance to either side of

† The x (or y) in alloy formulas is a fraction lying between 0 and 1. $Al_{0.3}Ga_{0.7}As$ would indicate a material with 3 Al and 7 Ga atoms per every 10 As atoms.

Table 1.1 Semiconductor Materials.

General Classification	*Semiconductor Symbol*	*Semiconductor Name*
(1) Elemental	Si	Silicon
	Ge	Germanium
(2) Compounds		
(a) IV-IV	SiC	Silicon carbide
(b) III-V	AlP	Aluminum phosphide
	AlAs	Aluminum arsenide
	AlSb	Aluminum antimonide
	GaN	Gallium nitride
	GaP	Gallium phosphide
	GaAs	Gallium arsenide
	GaSb	Gallium antimonide
	InP	Indium phosphide
	InAs	Indium arsenide
	InSb	Indium antimonide
(c) II-VI	ZnO	Zinc oxide
	ZnS	Zinc sulfide
	ZnSe	Zinc selenide
	ZnTe	Zinc telluride
	CdS	Cadmium sulfide
	CdSe	Cadmium selenide
	CdTe	Cadmium telluride
	HgS	Mercury sulfide
(d) IV-VI	PbS	Lead sulfide
	PbSe	Lead selenide
	PbTe	Lead telluride
(3) Alloys		
(a) Binary	$Si_{1-x}Ge_x$	
(b) Ternary	$Al_xGa_{1-x}As$	(or $Ga_{1-x}Al_xAs$)
	$Al_xIn_{1-x}As$	(or $In_{1-x}Al_xAs$)
	$Cd_{1-x}Mn_xTe$	
	$GaAs_{1-x}P_x$	
	$Ga_xIn_{1-x}As$	(or $In_{1-x}Ga_xAs$)
	$Ga_xIn_{1-x}P$	(or $In_{1-x}Ga_xP$)
	$Hg_{1-x}Cd_xTe$	
(c) Quaternary	$Al_xGa_{1-x}As_ySb_{1-y}$	
	$Ga_xIn_{1-x}As_{1-y}P_y$	

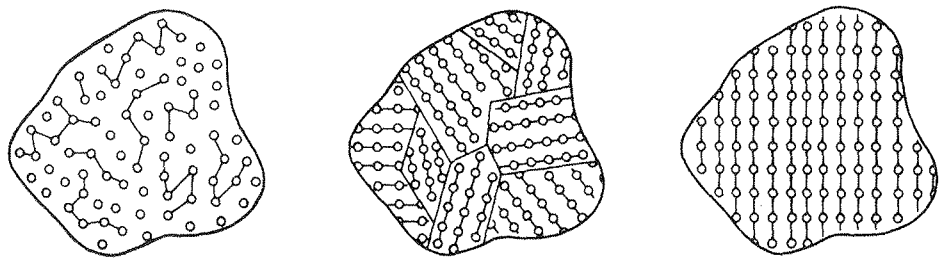
Table 1.2 Abbreviated Periodic Chart of the Elements.

II	III	IV	V	VI
4 Be	5 B	6 C	7 N	8 O
12 Mg	13 **Al**	14 **Si**	15 P	16 S
30 **Zn**	31 **Ga**	32 Ge	33 **As**	34 **Se**
48 Cd	49 In	50 Sn	51 Sb	52 Te
80 Hg	81 Tl	82 Pb	83 Bi	84 Po

Column IV. The Column III element Ga plus the Column V element As yields the III-V compound semiconductor GaAs; the Column II element Zn plus the Column VI element Se yields the II-VI compound semiconductor ZnSe; the fractional combination of the Column III elements Al and Ga plus the Column V element As yields the alloy semiconductor $Al_xGa_{1-x}As$. This very general property is related to the chemical bonding in semiconductors, where, on the average, there are four valence electrons per atom.

1.1.2 Purity

As will be explained in Chapter 2, extremely minute traces of impurity atoms called "dopants" can have a drastic effect on the electrical properties of semiconductors. For this reason, the compositional purity of semiconductors must be very carefully controlled and, in fact, modern semiconductors are some of the purest solid materials in existence. In Si, for example, the unintentional content of dopant atoms is routinely less than one atom per 10^9 Si atoms. To assist the reader in attempting to comprehend this incredible level of purity, let us suppose a forest of maple trees was planted from coast to coast, border to border, at 50-ft centers across the United States (including Alaska). An impurity level of one part per 10^9 would correspond to finding about 25 crabapple trees in the maple tree forest covering the United States! It should be emphasized that the cited material purity refers to *unintentional* undesired impurities. Typically, dopant atoms at levels ranging from one part per 10^8 to one impurity atom per 10^3 semiconductor atoms will be *purposely* added to the semiconductor to control its electrical properties.



(a) Amorphous
No recognizable long-range order

(b) Polycrystalline
Completely ordered in segments

(c) Crystalline
Entire solid is made up of atoms in an orderly array

Figure 1.1 General classification of solids based on the degree of atomic order: (a) amorphous, (b) polycrystalline, and (c) crystalline.

1.1.3 Structure

The spatial arrangement of atoms within a material plays an important role in determining the precise properties of the material. As shown schematically in Fig. 1.1, the atomic arrangement within a solid causes it to be placed into one of three broad classifications; namely, amorphous, polycrystalline, or crystalline. An amorphous solid is a material in which there is no recognizable long-range order in the positioning of atoms within the material. The atomic arrangement in any given section of an amorphous material will look different from the atomic arrangement in any other section of the material. Crystalline solids lie at the opposite end of the "order" spectrum; in a crystalline material the atoms are arranged in an orderly three-dimensional array. Given any section of a crystalline material, one can readily reproduce the atomic arrangement in any other section of the material. Polycrystalline solids comprise an intermediate case in which the solid is composed of crystalline subsections that are disjointed or misoriented relative to one another.

Upon examining the many solid state devices in existence, one readily finds examples of all three structural forms. An amorphous-Si thin-film transistor is used as the switching element in liquid crystal displays (LCDs); polycrystalline Si gates are employed in Metal-Oxide-Semiconductor Field-Effect Transistors (MOSFETs). In the vast majority of devices, however, the active region of the device is situated within a *crystalline* semiconductor. The overwhelming number of devices fabricated today employ *crystalline* semiconductors.

1.2 CRYSTAL STRUCTURE

The discussion at the end of the preceding section leads nicely into the topic of this section. Since in-use semiconductors are typically crystalline in form, it seems reasonable to seek out additional information about the crystalline state. Our major goal here is to present a more detailed picture of the atomic arrangement within the principal semiconductors. To

achieve this goal, we first examine how one goes about describing the spatial positioning of atoms within crystals. Next, a bit of "visualization" practice with simple three-dimensional lattices (atomic arrangements) is presented prior to examining semiconductor lattices themselves. The section concludes with an introduction to the related topic of Miller indices. Miller indices are a convenient shorthand notation widely employed for identifying specific planes and directions within crystals.

1.2.1 The Unit Cell Concept

Simply stated, a unit cell is a small portion of any given crystal that could be used to reproduce the crystal. To help establish the unit cell (or building-block) concept, let us consider the two-dimensional lattice shown in Fig. 1.2(a). To describe this lattice or to totally specify the physical characteristics of the lattice, one need only provide the unit cell shown in Fig. 1.2(b). As indicated in Fig. 1.2(c), the original lattice can be readily reproduced by merely duplicating the unit cell and stacking the duplicates next to each other in an orderly fashion.

In dealing with unit cells there often arises a misunderstanding, and hence confusion, relative to two points. First, unit cells are not necessarily unique. The unit cell shown in Fig. 1.2(d) is as acceptable as the Fig. 1.2(b) unit cell for specifying the original lattice of

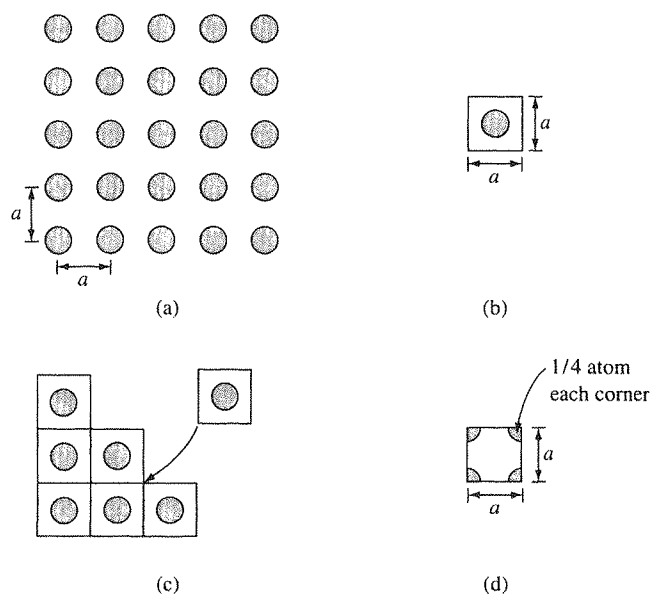


Figure 1.2 Introduction to the unit cell method of describing atomic arrangements within crystals. (a) Sample two-dimensional lattice. (b) Unit cell corresponding to the part (a) lattice. (c) Reproduction of the original lattice. (d) An alternative unit cell.

Fig. 1.2(a). Second, a unit cell need not be primitive (the smallest unit cell possible). In fact, it is usually advantageous to employ a slightly larger unit cell with orthogonal sides instead of a primitive cell with nonorthogonal sides. This is especially true in three dimensions where noncubic unit cells are quite difficult to describe and visualize.

1.2.2 Simple 3-D Unit Cells

Semiconductor crystals are three-dimensional and are therefore described in terms of three-dimensional (3-D) unit cells. In Fig. 1.3(a) we have pictured the simplest of all three-dimensional unit cells, the simple cubic unit cell. The simple cubic cell is an equal-sided box or cube with an atom positioned at each corner of the cube. The simple cubic lattice associated with this cell is constructed in a manner paralleling the two-dimensional case. In doing so, however, it should be noted that only 1/8 of each corner atom is actually *inside* the cell, as pictured in Fig. 1.3(b). Duplicating the Fig. 1.3(b) cell and stacking the duplicates like blocks in a nursery yields the simple cubic lattice. Specifically, the procedure generates planes of atoms like that previously shown in Fig. 1.2(a). Planes of atoms parallel to the base plane are separated from one another by a unit cell side length or *lattice con-*

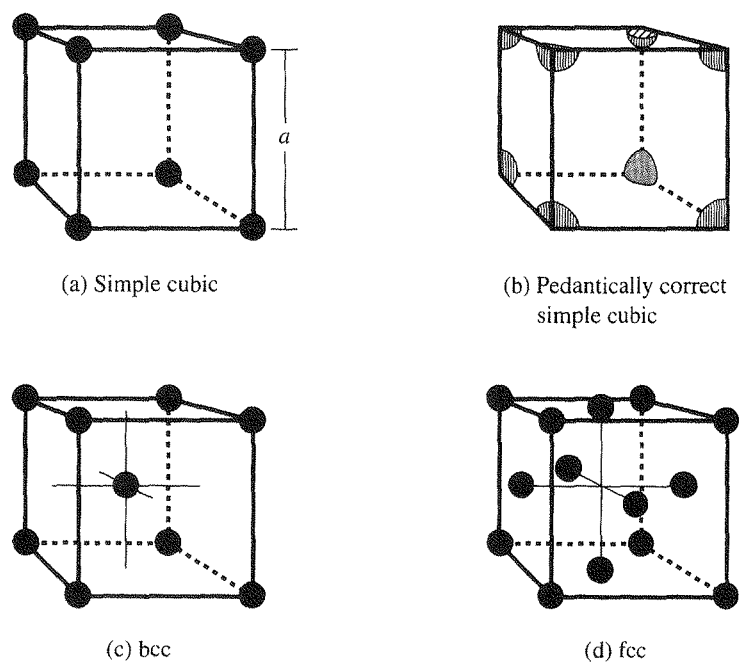


Figure 1.3 Simple three-dimensional unit cells. (a) Simple cubic unit cell. (b) Pedantically correct simple cubic unit cell including only the fractional portion (1/8) of each corner atom actually within the cell cube. (c) Body centered cubic unit cell. (d) Face centered cubic unit cell.

stant, a. When viewed normal to the base plane, atoms in a given plane are positioned directly over atoms in a lower-lying plane.

Figures 1.3(c) and 1.3(d) display two common 3-D cells that are somewhat more complex but still closely related to the simple cubic cell. The unit cell of Fig. 1.3(c) has an atom added at the center of the cube; this configuration is appropriately called the body centered cubic (bcc) unit cell. The face centered cubic (fcc) unit cell of Fig. 1.3(d) contains an atom at each face of the cube in addition to the atoms at each corner. (Note, however, that only one-half of each face atom actually lies inside the fcc unit cell.) Whereas the simple cubic cell contains one atom (1/8 of an atom at each of the eight cube corners), the somewhat more complex bcc and fcc cells contain two and four atoms, respectively. The reader should verify these facts and visualize the lattices associated with the bcc and fcc cells.

(C) Exercise 1.1

The text-associated software distributed via the Internet or on a floppy disk contains a directory or folder named MacMolecule. Files stored inside this directory or folder are ASCII input files generated by the author for use with a computer program by the same name. The MacMolecule program, copyrighted by the University of Arizona, is distributed free of charge to academic users and is available at many academic sites. As the name implies, the program runs only on Macintosh personal computers. The input files supplied to IBM-compatible users, however, can be readily converted for use on a public-access Macintosh.

MacMolecule generates and displays a "ball-and-stick" 3-D color rendering of molecules, unit cells, and lattices. The input files supplied by the author can be used to help visualize the simple cubic, bcc, and fcc unit cells, plus the diamond and zinc-blende unit cells described in the next subsection. The About-MacMolecule file that is distributed with the MacMolecule program provides detailed information about the use of the program and the generation/modification of input files.

The reader is urged to investigate and play with the MacMolecule software. The initial display is always a z-direction view and is typically pseudo-two-dimensional in nature. A more informative view is obtained by rotating the model. Rotation is best accomplished by "grabbing and dragging" the edges of the model. The more adventuresome may wish to try their hand at modifying the existing input files or generating new input files.

1.2.3 Semiconductor Lattices

We are finally in a position to supply details relative to the positioning of atoms within the principal semiconductors. In Si (and Ge) the lattice structure is described by the unit cell pictured in Fig. 1.4(a). The Fig. 1.4(a) arrangement is known as the *diamond lattice* unit cell because it also characterizes diamond, a form of the Column IV element carbon. Examining the diamond lattice unit cell, we note that the cell is cubic, with atoms at each

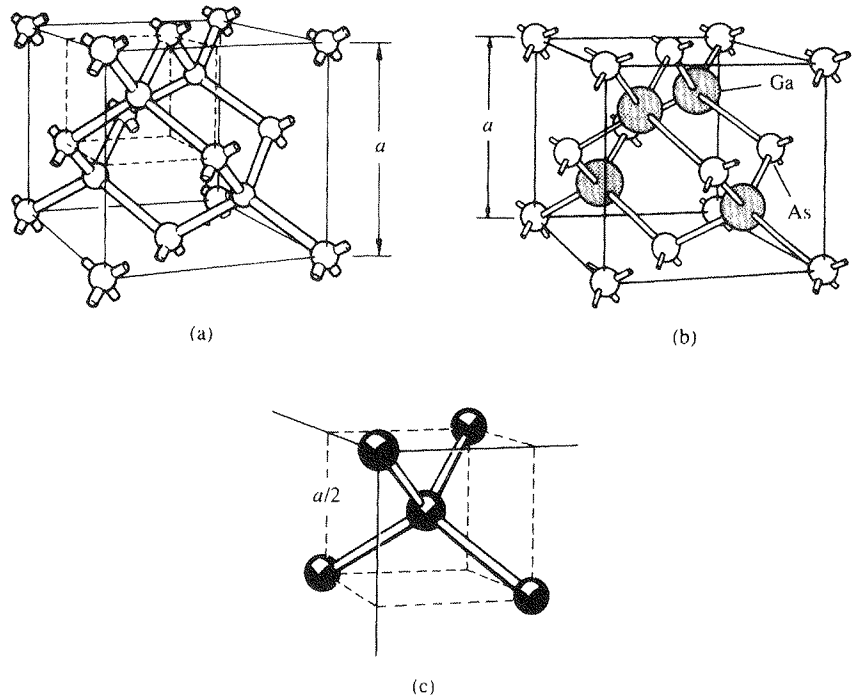


Figure 1.4 (a) Diamond lattice unit cell. (b) Zincblende lattice unit cell (GaAs used for illustration). (c) Enlarged top corner of the part (a) diamond lattice emphasizing the four-nearest-neighbor bonding within the structure. The cube side length, a, is 5.43 Å and 5.65 Å at $T = 300$ K for Si and GaAs, respectively. [(a) Adapted from Shockley[1]. (b) From Sze[2], ©1981 by John Wiley & Sons, Inc. Reprinted with permission.]

corner and at each face of the cube similar to the fcc cell. The interior of the Fig. 1.4(a) cell, however, contains four additional atoms. One of the interior atoms is located along a cube body diagonal exactly one-quarter of the way down the diagonal from the top front left-hand corner of the cube. The other three interior atoms are displaced one-quarter of the body diagonal length along the previously noted body diagonal direction from the front, top, and left-side face atoms, respectively. Although it may be difficult to visualize from Fig. 1.4(a), the diamond lattice can also be described as nothing more than two interpenetrating fcc lattices. The corner and face atoms of the unit cell can be viewed as belonging to one fcc lattice, while the atoms totally contained within the cell belong to the second fcc lattice. The second lattice is displaced one-quarter of a body diagonal along a body diagonal direction relative to the first fcc lattice.

Most of the III-V semiconductors, including GaAs, crystallize in the *zincblende* structure. The zincblende lattice, typified by the GaAs unit cell shown in Fig. 1.4(b), is essen-

tially identical to the diamond lattice, except that lattice sites are apportioned equally between two different atoms. Ga occupies sites on one of the two interpenetrating fcc sublattices; arsenic (As) populates the other fcc sublattice.

Now that the positioning of atoms within the principal semiconductors has been established, the question may arise as to the practical utilization of such information. Although several applications could be cited, geometrical-type calculations constitute a very common and readily explained use of the unit cell formalism. For example, in Si at room temperature the unit cell side length (a) is 5.43 Å (where 1 Å = 10^{-8} cm). Since there are eight Si atoms per unit cell and the volume of the unit cell is a^3, it follows that there are $8/a^3$ or almost exactly 5×10^{22} atoms/cm^3 in the Si lattice. Similar calculations could be performed to determine atomic radii, the distance between atomic planes, and so forth. For the purposes of the development herein, however, a major reason for the discussion of semiconductor lattices was to establish that, as emphasized in Fig. 1.4(c), *atoms in the diamond and zincblende lattices have four nearest neighbors.* The chemical bonding (or crystalline glue) within the major semiconductors is therefore dominated by the attraction between any given atom and its four closest neighbors. This is an important fact that should be filed away for future reference.

Exercise 1.2

P: If the lattice constant or unit cell side-length in Si is $a = 5.43 \times 10^{-8}$ cm, what is the distance (d) from the center of one Si atom to the center of its nearest neighbor?

S: As noted in the Fig. 1.4 caption, the atom in the upper front corner of the Si unit cell and the atom along the cube diagonal one-fourth of the way down the diagonal are nearest neighbors. Since the diagonal of a cube is equal to $\sqrt{3}$ times a side length, one concludes $d = (1/4)\sqrt{3}a = (\sqrt{3}/4)(5.43 \times 10^{-8}) = \boxed{2.35 \times 10^{-8}\,\text{cm}}$.

(C) Exercise 1.3

P: Construct a MATLAB program that computes the number of atoms/cm^3 in cubic crystals. Use the MATLAB `input` function to enter the number of atoms/unit-cell and the unit cell side length (a) for a specific crystal. Make a listing of your program and record the program result when applied to silicon.

S: MATLAB program script . . .

```
%Exercise 1.3
%Computation of the number of atoms/cm3 in a cubic lattice
N=input('input number of atoms/unit cell, N = ');
a=input('lattice constant in angstrom, a = ');
atmden=N*(1.0e24)/(a^3)    %number of atoms/cm3
```

```
Program output for Si . . .
   input number of atoms/unit cell, N = 8
   lattice constant in angstrom, a = 5.43
   atmden =
      4.9968e+22
```

1.2.4 Miller Indices

Single crystals of silicon used in device processing normally assume the thin, round form exhibited in Fig. 1.5. The pictured plate-like single crystals, better known as Si wafers, are typical of the starting substrates presently employed by major manufacturers. Of particular interest here is the fact that the surface of a wafer is carefully preoriented to lie along a specific crystallographic plane. Moreover, a "flat" or "notch" is ground along the periphery of the wafer to identify a reference direction within the surface plane. Precise surface orientation is critical in certain device processing steps and directly affects the characteristics exhibited by a number of devices. The flat or notch is routinely employed, for example, to orient device arrays on the wafer so as to achieve high yields during device

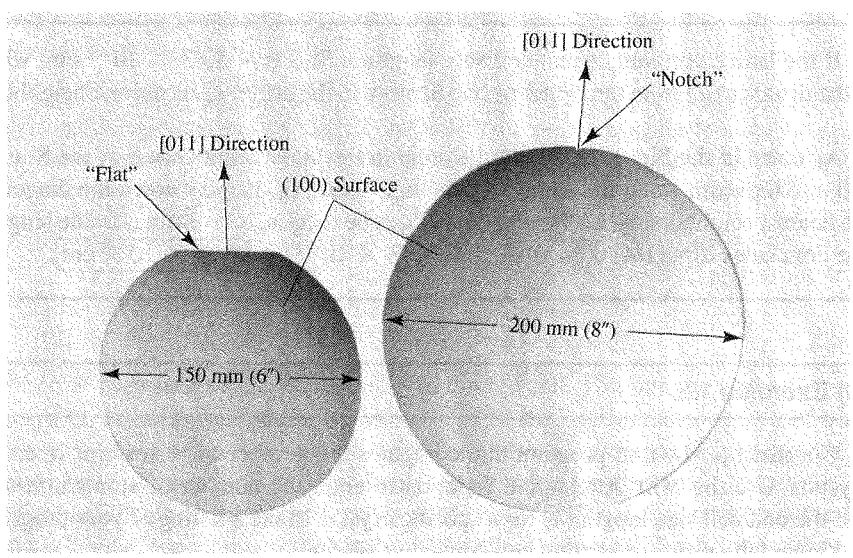


Figure 1.5 Single-crystal silicon wafers typical of the starting substrates presently employed by major device manufacturers. The 150 mm (6 inch) and 200 mm (8 inch) wafers are nominally 0.625 mm and 0.725 mm thick, respectively. The facing surface is polished and etched yielding a damage-free, mirror-like finish. The figure dramatizes the utility of Miller indices exemplified by the (100) plane and [011] direction designations. (Photograph courtesy of Intel Corporation.)

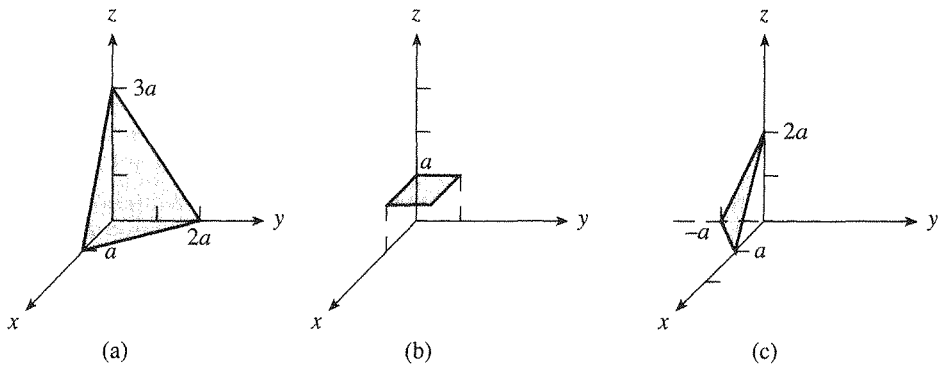


Figure 1.6 Sample cubic crystal planes. (a) The (632) plane used in explaining the Miller indexing procedure. (b) The (001) plane. (c) The $(2\bar{2}1)$ plane.

separation. The point we wish to make is that the specification of crystallographic planes and directions is of practical importance. Miller indices, exemplified by the (100) plane and [011] direction designations used in Fig. 1.5, constitute the conventional means of identifying planes and directions within a crystal.

The Miller indices for any given plane of atoms within a crystal are obtained by following a straightforward four-step procedure. The procedure is detailed below, along with the simultaneous sample indexing of the plane shown in Fig. 1.6(a).

Indexing Procedure	Sample Implementation
After setting up coordinate axes along the edges of the unit cell, note where the plane to be indexed intercepts the axes. Divide each intercept value by the unit cell length along the respective coordinate axis. Record the resulting normalized (pure-number) intercept set in the order *x, y, z.*	1, 2, 3
Invert the intercept values; that is, form [1/intercept]s.	1, ½, ⅓
Using an appropriate multiplier, convert the 1/intercept set to the smallest possible set of whole numbers.	6, 3, 2
Enclose the whole-number set in curvilinear brackets.	(632)

To complete the description of the plane-indexing procedure, the user should also be aware of the following special facts:

(i) If the plane to be indexed is parallel to a coordinate axis, the "intercept" along the axis is taken to be at infinity. Thus, for example, the plane shown in Fig. 1.6(b) intercepts the coordinate axes at ∞, ∞, 1, and is therefore a (001) plane.

(ii) If the plane to be indexed has an intercept along the negative portion of a coordinate axis, a minus sign is placed *over* the corresponding index number. Thus the Fig. 1.6(c) plane is designated a $(2\bar{2}1)$ plane.

(iii) Referring to the diamond lattice of Fig. 1.4(a), note that the six planes passing through the cube faces contain identical atom arrangements; that is, because of crystal symmetry, it is impossible to distinguish between the "equivalent" (100), (010), (001), $(\bar{1}00)$, $(0\bar{1}0)$, and $(00\bar{1})$ planes; or, it is impossible to distinguish between {100} planes. A group of equivalent planes is concisely referenced in the Miller notation through the use of { } braces.

(iv) Miller indices cannot be established for a plane passing through the origin of coordinates. The origin of coordinates must be moved to a lattice point not on the plane to be indexed. This procedure is acceptable because of the equivalent nature of parallel planes.

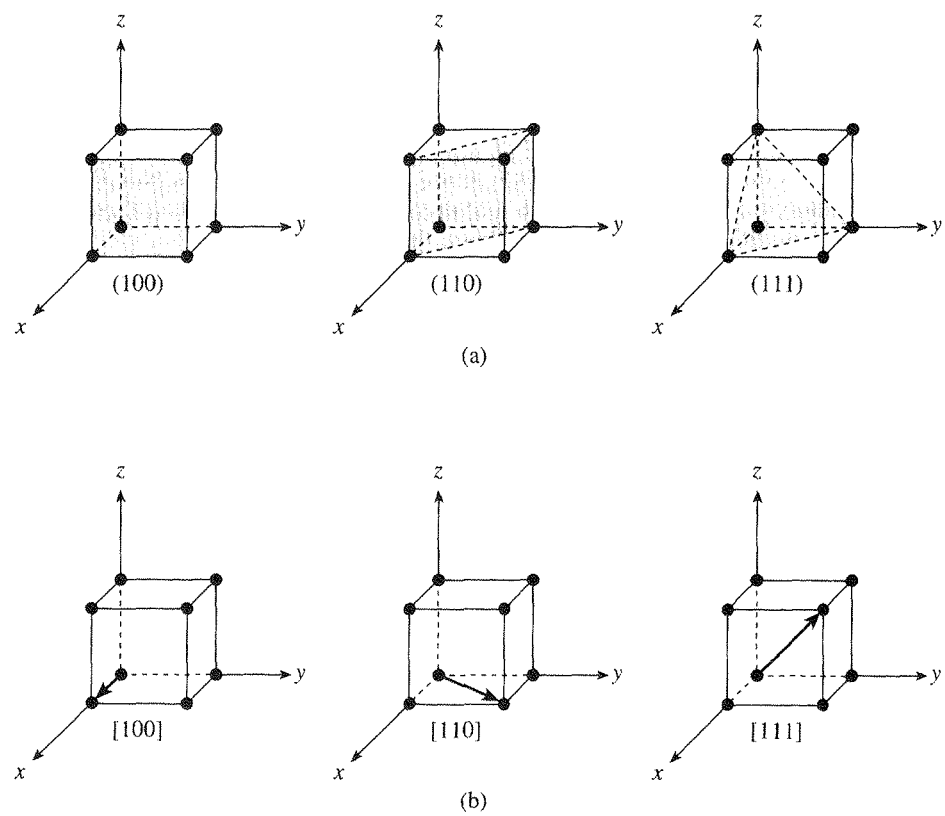


Figure 1.7 Visualization and Miller indices of commonly encountered (a) crystalline planes and (b) direction vectors.

Table 1.3 Miller Convention Summary.

Convention	*Interpretation*
(hkl)	Crystal plane
$\{hkl\}$	Equivalent planes
$[hkl]$	Crystal direction
$\langle hkl \rangle$	Equivalent directions

Using the simple cubic cell as a reference, the three most commonly encountered crystalline planes are visualized in Fig. 1.7(a).

The Miller indices for *directions* are established in a manner analogous to the well-known procedure for finding the components of a vector. First, set up a vector of arbitrary length in the direction of interest. Next, decompose the vector into its components by noting the projections of the vector along the coordinate axes. Using an appropriate multiplier, convert the component values into the smallest possible whole-number set. This of course changes the length of the original vector but not its direction. Finally, enclose the whole-number set in brackets. Square brackets, [ ], are used to designate specific directions within a crystal; triangular brackets, ⟨ ⟩, designate an equivalent set of directions. Common direction vectors and their corresponding Miller indices are displayed in Fig. 1.7(b). A summary of the Miller bracketing convention for planes and directions is given in Table 1.3.

In the foregoing discussion we presented the procedure for progressing from a given plane or direction in a crystal to the corresponding Miller indices. More often than not, one is faced with the inverse process—visualizing the crystalline plane or direction corresponding to a given set of indices. Fortunately, one seldom encounters other than low-index planes and directions such as (111), (110), [001]. Thus it is possible to become fairly adept at the inverse process by simply memorizing the orientations for planes and directions associated with small-number indices. It is also helpful to note that, *for cubic crystals, a plane and the direction normal to the plane have precisely the same indices*—e.g., the [110] direction is normal to the (110) plane. Of course, any plane or direction can always be deduced by reversing the indexing procedure.

Exercise 1.4

P: For a cubic crystal lattice:

(a) Determine the Miller indices for the plane and direction vector shown in Fig. E1.4(a).

(b) Sketch the plane and direction vector characterized by (011) and [011], respectively.

S: (a) The intercepts of the plane are -1, 1, and 2 along the x, y, and z coordinate axes, respectively. Taking [1/intercept]s yields -1, 1, 1/2. Multiplying by 2 to obtain the lowest whole-number set, and enclosing in parentheses, gives $\boxed{(\bar{2}21)}$ for the Miller indices of the plane. The direction vector has projections of $2a$, a, and 0 along the x, y, and z coordinate axes, respectively. The Miller indices for the direction are then $\boxed{[210]}$.

(b) For the (011) plane, [1/intercept]s = 0, 1, 1. The normalized x, y, z intercepts of the plane are therefore ∞, 1, 1; the plane intersects the y-axis at a, intersects the z-axis at a, and is parallel to the x-axis. In a cubic crystal the [011] direction is normal to the (011) plane. The deduced plane and direction are as sketched in Fig. E1.4(b).

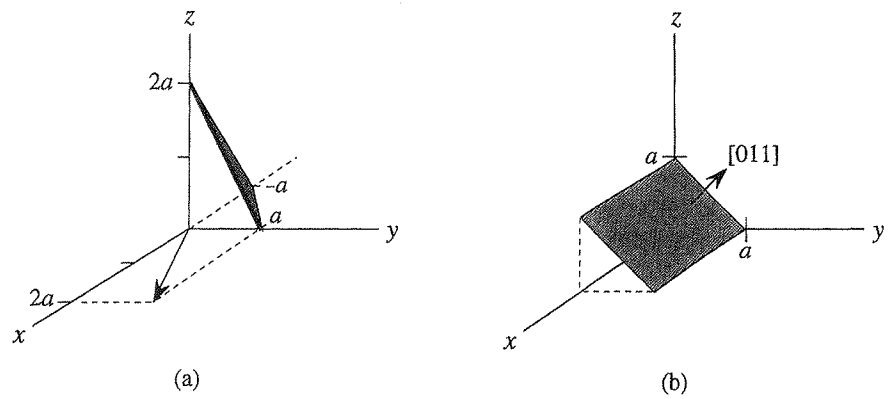


Figure E1.4

1.3 CRYSTAL GROWTH

1.3.1 Obtaining Ultrapure Si

Given the obvious general availability and widespread use of semiconductor materials, particularly Si, it is reasonable to wonder about the origin of the single-crystal Si used in modern-day device production. Could it be that Si is readily available in sandstone deposits?—No. Perhaps, as a sort of by-product, Si single crystals come from South African diamond mines?—Wrong again. As suggested recently in a low-budget science fiction movie, maybe Si is scooped from the ocean bottom by special submarines?—Sorry, no. Although Si is the second most abundant element in the earth's crust and a component in numerous compounds, chief of which are silica (impure SiO_2) and the silicates (Si + O + another element), silicon never occurs alone in nature as an element. Single-crystal Si used in device production, it turns out, is a man-made material.

Given the preceding introduction, it should be clear that the initial step in producing device-quality silicon must involve separating Si from its compounds and purifying the

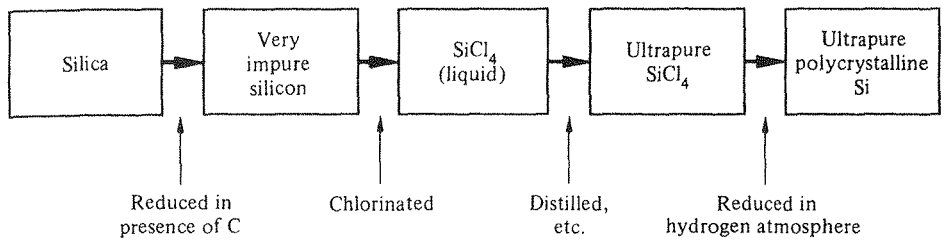
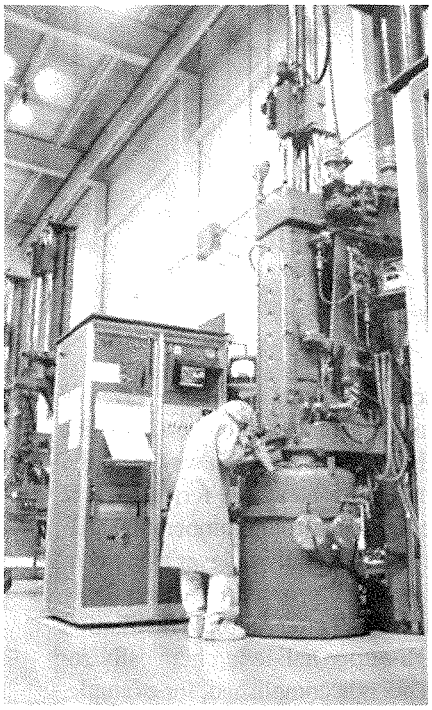


Figure 1.8 Summary of the process employed to produce ultrapure silicon.

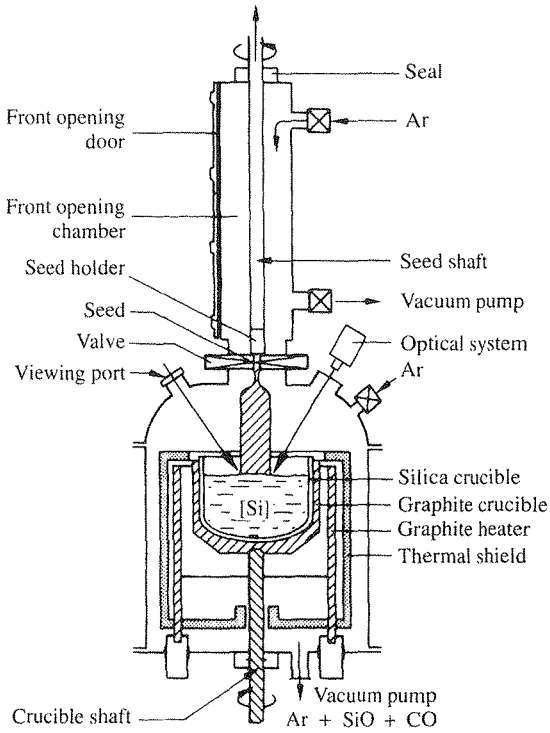
separated material. The ingenious separation and purification process that has evolved is summarized in Fig. 1.8. Low-grade silicon or ferrosilicon is first produced by heating silica with carbon in an electric furnace. The carbon essentially pulls the oxygen away from the impure SiO_2 (i.e., *reduces* the SiO_2), leaving behind impure Si. Next, the ferrosilicon is chlorinated to yield $SiCl_4$ or $SiHCl_3$, both of which are liquids at room temperature. Although appearing odd at first glance, the liquefaction process is actually a clever maneuver. Whereas solids are very difficult to purify, a number of standard procedures are available for purifying liquids. An ultrapure $SiCl_4$ (or $SiHCl_3$) is the result after multiple distillation and other liquid purification procedures. Lastly, the high-purity halide is chemically reduced, yielding the desired ultrapure elemental silicon. This is accomplished, for example, by heating $SiCl_4$ in a hydrogen atmosphere [$SiCl_4 + 2H_2 \rightarrow 4\ HCl + Si$].

1.3.2 Single-Crystal Formation

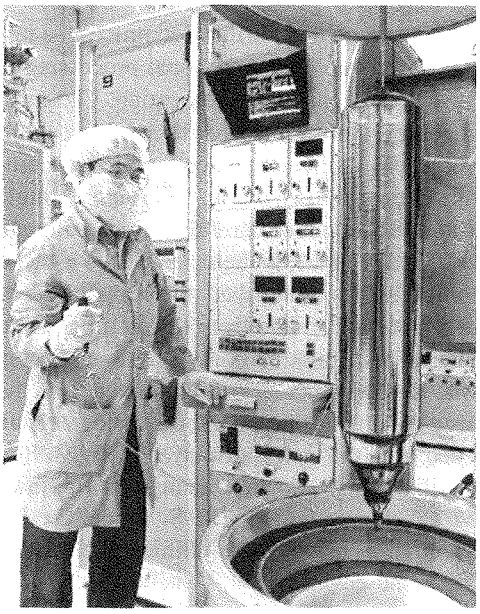
Although ultrapure, the silicon derived from the separation and purification process as just described is not a single crystal, but is instead polycrystalline. Additional processing is therefore required to form the large single crystals used in device fabrication. The most commonly employed method yielding large single crystals of Si is known as the Czochralski method. In this method the ultrapure polycrystalline silicon is placed in a quartz crucible and heated in an inert atmosphere to form a melt, as shown in Fig. 1.9. A small single crystal, or Si "seed" crystal, with the normal to its bottom face carefully aligned along a predetermined direction (typically a $\langle 111 \rangle$ or $\langle 100 \rangle$ direction), is then clamped to a metal rod and dipped into the melt. Once thermal equilibrium is established, the temperature of the melt in the vicinity of the seed crystal is reduced, and silicon from the melt begins to freeze out onto the seed crystal, the added material being a structurally perfect extension of the seed crystal. Subsequently, the seed crystal is slowly rotated and withdrawn from the melt; this permits more and more silicon to freeze out on the bottom of the growing crystal. A photograph showing a crystal after being pulled from the melt is displayed in Fig. 1.9(c). The large, cylindrically shaped single crystal of silicon, also known as an ingot, is routinely up to 200 mm (8 inches) in diameter and 1 to 2 meters in length. The Si wafers used in device processing, such as those previously pictured in Fig. 1.5, are ultimately produced by cutting the ingot into thin sections using a diamond-edged saw.



(a)



(b)



(c)

Figure 1.9 Silicon crystal-pulling apparatus and resulting single crystal. (a) Photograph of a computer-controlled Czochralski crystal puller. (b) Simplified schematic drawing of the puller. (c) Silicon ingot. [(a) and (c) Photographs courtesy of Wacker Siltronic. (b) From Zuhlehner and Huber[3]; reprinted with permission.]

1.4 SUMMARY

This chapter provides basic information about semiconductors in general and silicon in particular. Examining the elemental compositions of semiconductors, one finds that there are, on average, four valence electrons per atom. Moreover, device-quality semiconductors are typically single crystals of high compositional purity. Si crystallizes in the diamond lattice, GaAs crystallizes in the zincblende lattice, and the lattice atoms in both cases have four nearest neighbors. Miller indices, introduced herein as an adjunct to the crystal structure discussion, are the accepted means for identifying planes and directions within a crystal. Finally, large single crystals of device-grade Si are commonly produced by the Czochralski method.

PROBLEMS

CHAPTER 1 PROBLEM INFORMATION TABLE				
Problem	*Complete After*	*Difficulty Level*	*Suggested Point Weighting*	*Short Description*
1.1	1.4	1	10 (a::h-1, i-2)	Short answer review
1.2	1.2.3	1	8	Unit cell for alloy
● 1.3	"	1	10 (5 each part)	Ge atomic density
1.4	"	1	8 (4 each part)	Nearest-neighbor distances
● 1.5	1.2.4	2–3	15 (a-2, b-2, c-3, d-3, e-5)	Atoms/cm^2 on planes
1.6	"	1	8 (2 each subpart)	Determine Miller indices
1.7	"	1	16 (2 each part)	Given (*hkl*), sketch plane
1.8	"	1	16 (2 each part)	Given [*hkl*], sketch direction
1.9	"	1	8 (4 each part)	Perpendicular directions
1.10	"	2	10	Direction on wafer surface
1.11	"	2	10 (a-3, b-2, c-3, d-2)	Combination question
1.12	"	2	12 (2 each part)	Equivalent planes/directions
1.13	1.2.3	3	20 (5 each part)	% of volume occupied

1.1 Quick Quiz. Answer the following questions as concisely as possible.

(a) Name (i) one elemental semiconductor and (ii) one III-V compound semiconductor.

(b) What is the difference between a crystalline and a polycrystalline material?

(c) Give a word definition of *unit cell.*

(d) How many atoms are there in a simple cubic unit cell? in a bcc unit cell? in a fcc unit cell? in the unit cell characterizing the diamond lattice?

(e) 1Å = ? cm.

(f) In terms of the lattice constant a, what is the distance between nearest-neighbor atoms in a simple cubic lattice?

(g) How many nearest-neighbor atoms are there in the diamond and the zincblende lattices?

(h) What is being indicated by the bracket sets ( ), [ ], { }, and ⟨ ⟩ as employed in the Miller indexing scheme?

(i) Describe the Czochralski method for obtaining large single crystals of silicon.

1.2 The GaAs unit cell is pictured in Fig. 1.4(b). Describe (or sketch) the unit cell for $Al_{0.5}Ga_{0.5}As$.

1.3 (a) The lattice constant of Ge at room temperature is $a = 5.65 \times 10^{-8}$ cm. Determine the number of Ge atoms/cm^3.

● (b) Copies of computer-required exercises employing MATLAB are included on disk. Run the Exercise 1.3 program to verify your result from part (a).

1.4 In terms of the lattice constant a, what is the distance between nearest-neighbor atoms in

(a) a bcc lattice?

(b) an fcc lattice?

1.5 The surface of a Si wafer is a (100) plane.

(a) Sketch the placement of Si atoms on the surface of the wafer.

(b) Determine the number of atoms per cm^2 at the surface of the wafer.

(c)/(d) Repeat parts (a) and (b), this time taking the surface of the Si wafer to be a (110) plane.

● (e) Primarily for practice in utilizing MATLAB, construct a MATLAB program that computes the surface density of atoms on (100) planes of cubic crystals. The number of atoms with centers on the (100) face of the unit cell, and the lattice constant in Å, are to be the input variables. Confirm the result obtained manually in part (b).

1.6 Record all intermediate steps in answering the following questions.

(a) As shown in Fig. P1.6(a), a crystalline plane has intercepts of $1a$, $3a$, and $1a$ on the x, y, and z axes, respectively. a is the cubic cell side length.

(i) What is the Miller index notation for the plane?

(ii) What is the Miller index notation for the direction normal to the plane?

(b) Assuming the crystal structure to be cubic, determine the Miller indices for (i) the plane and (ii) the vector pictured in Fig. P1.6(b).

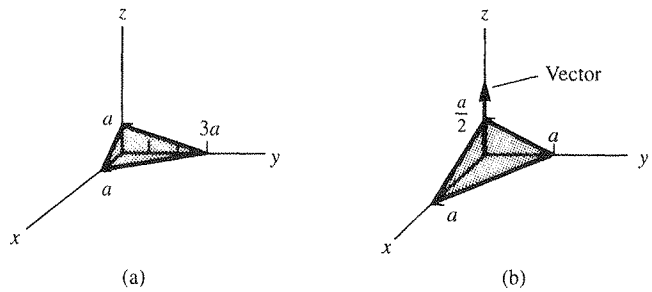


Figure P1.6

1.7 Assuming a cubic crystal system, make a sketch of the following planes.

(a) (001), (b) (111), (c) (123), (d) $(\bar{1}10)$,
(e) (010), (f) $(\bar{1}\bar{1}\bar{1})$, (g) (221), (h) $(0\bar{1}0)$.

1.8 Assuming a cubic crystal system, use an appropriately directed arrow to identify each of the following directions:

(a) [010], (b) [101], (c) $[00\bar{1}]$, (d) [111],
(e) [001], (f) [110], (g) $[0\bar{1}0]$, (h) [123].

1.9 Identify two crystalline directions in a cubic crystal which are perpendicular to

(a) the [100] direction,

(b) the [111] direction.

NOTE: The cosine of the angle θ between two arbitrary directions, $[h_1 k_1 l_1]$ and $[h_2 k_2 l_2]$, in a cubic crystal is

$$\cos(\theta) = \frac{h_1 h_2 + k_1 k_2 + l_1 l_2}{[(h_1^2 + k_1^2 + l_1^2)(h_2^2 + k_2^2 + l_2^2)]^{1/2}}$$

Consequently, for two directions to be perpendicular, $\cos(\theta) = 0$ and one must have $h_1 h_2 + k_1 k_2 + l_1 l_2 = 0$.

1.10 As pictured in Fig. 1.5, [011] is the direction in the surface plane normal to the major flat when the surface of the Si wafer is a (100) plane. To construct a particular device structure, a parallel set of grooves must be etched in the (100) surface plane along the [010] direction. Make a sketch indicating how the grooves are to be oriented on the wafer's surface. Explain how you arrived at your sketch.

1.11 A crystalline lattice is characterized by the cubic unit cell pictured in Fig. P1.11. The cell has a single atom positioned at the center of the cube.

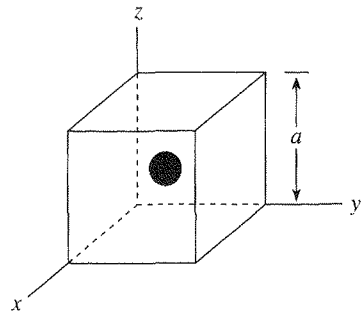


Figure P1.11

(a) What is the name of the lattice generated by the given unit cell?

(b) Determine the number of atoms per unit volume in the crystal. (Your answer should be in terms of the lattice constant, a.)

(c) Suppose the crystal has a (110) surface plane. Determine the number of atoms per unit area whose centers lie on the (110) plane.

(d) A direction vector is drawn through the center of the atom in the unit cell. Specify the Miller indices of the direction vector.

1.12 Given a cubic lattice, indicate how many equivalent planes [parts (a) to (c)] or equivalent directions [parts (d) to (f)] are associated with each of the following designations:

(a) $\{100\}$, (b) $\{110\}$, (c) $\{111\}$,

(d) $\langle 100 \rangle$, (e) $\langle 110 \rangle$, (f) $\langle 111 \rangle$.

1.13 Treating atoms as rigid spheres with radii equal to one-half the distance between nearest neighbors, show that the ratio of the volume occupied by the atoms to the total available volume in the various crystal structures is

(a) $\pi/6$ or 52% for the simple cubic lattice.

(b) $\sqrt{3}\pi/8$ or 68% for the body centered cubic lattice.

(c) $\sqrt{2}\pi/6$ or 74% for the face centered cubic lattice.

(d) $\sqrt{3}\pi/16$ or 34% for the diamond lattice.

2 Carrier Modeling

Carriers are the entities that transport charge from place to place inside a material and hence give rise to electrical currents. In everyday life the most commonly encountered type of carrier is the electron, the subatomic particle responsible for charge transport in metallic wires. Within semiconductors one again encounters the familiar electron, but there is also a second equally important type of carrier—the hole. Electrons and holes are the focal point of this chapter, wherein we examine carrier related concepts, models, properties, and terminology.

Although reminders will be periodically interjected, it should be emphasized from the very start that the development throughout this chapter assumes *equilibrium* conditions exist within the semiconductor. "Equilibrium" is the term used to describe the unperturbed state of a system. Under equilibrium conditions there are no external voltages, magnetic fields, stresses, or other perturbing forces acting on the semiconductor. All observables are invariant with time. This "rest" condition, as we will see, provides an excellent *frame of reference*. Being able to characterize the semiconductor under equilibrium conditions permits one to extrapolate and ascertain the semiconductor's condition when a perturbation has been applied.

The reader should be forewarned that a number of formulas and facts will be presented without justification in this chapter. We would like to provide a complete discussion, properly developing every concept and deriving every formula. Unfortunately, this is not possible. It is our underlying philosophy, moreover, that being able to derive a result is secondary to knowing how to *interpret* and *use* a result. One can, of course, fill in any information gap through supplemental readings. For suggested references see the supplemental reading list at the end of Part I.

Lastly, please note that Appendix A provides a more in-depth introduction to the quantization concept and related topics considered in Section 2.1. Section 2.1 may be replaced by Appendix A with no loss in continuity.

2.1 THE QUANTIZATION CONCEPT

Instead of attempting to deal immediately with electrons in crystalline Si, where there are 14 electrons per atom and 5×10^{22} atoms/cm^3, we will take a more realistic approach and first establish certain ground rules by examining much simpler atomic systems. We begin

with the simplest of all atomic systems, the isolated hydrogen atom. This atom, as the reader may recall from a course in modern physics, was under intense scrutiny at the start of the twentieth century. Scientists of that time knew the hydrogen atom contained a negatively charged particle in orbit about a more massive positively charged nucleus. What they could not explain was the nature of the light emitted from the system when the hydrogen atom was heated to an elevated temperature. Specifically, the emitted light was observed at only certain discrete wavelengths; according to the prevailing theory of the time, scientists expected a continuum of wavelengths.

In 1913 Niels Bohr proposed a solution to the dilemma. Bohr hypothesized that the H-atom electron was restricted to certain well-defined orbits; or equivalently, Bohr assumed that the orbiting electron could take on only certain values of angular momentum. This "quantization" of the electron's angular momentum was, in turn, coupled directly to energy quantization. As can be readily established, if the electron's angular momentum is assumed to be $\mathbf{n}\hbar$, then

$$E_H = -\frac{m_0 q^4}{2(4\pi\varepsilon_0\hbar\mathbf{n})^2} = -\frac{13.6}{\mathbf{n}^2}\ \text{eV}, \qquad \mathbf{n} = 1, 2, 3, \ldots \tag{2.1}$$

where (also see Fig. 2.1) E_H is the electron binding energy within the hydrogen atom, m_0 is the mass of a free electron, q is the magnitude of the electronic charge, ε_0 is the permittivity of free space, h is Planck's constant, $\hbar = h/2\pi$, and $\mathbf{n}$ is the energy quantum number or orbit identifier. The *electron volt* (eV) is a unit of energy equal to 1.6×10^{-19} joules. Now, with the electron limited to certain energies inside the hydrogen atom, it follows from the Bohr model that the transition from a higher $\mathbf{n}$ to a lower $\mathbf{n}$ orbit will release quantized energies of light; this explains the observation of emitted light at only certain discrete wavelengths.

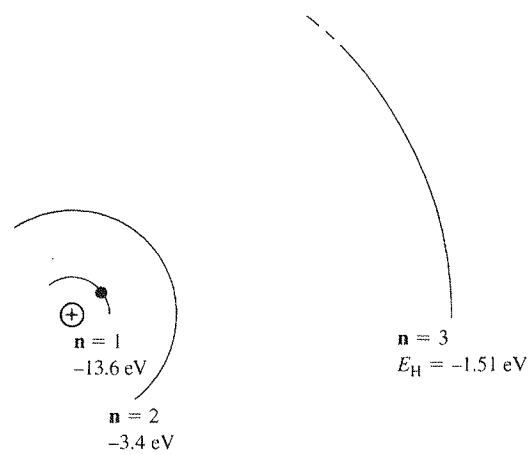


Figure 2.1 The hydrogen atom—idealized representation showing the first three allowed electron orbits and the associated energy quantization.

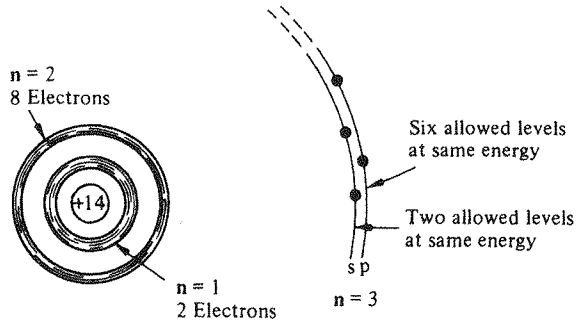


Figure 2.2 Schematic representation of an isolated Si atom.

For our purposes, the most important idea to be obtained from the Bohr model is that the energy of electrons in atomic systems is restricted to a limited set of values.† Relative to the hydrogen atom, the energy level scheme in a multi-electron atom like silicon is, as one might intuitively expect, decidedly more complex. It is still a relatively easy task, however, to describe the salient energy-related features of an isolated silicon atom. As pictured in Fig. 2.2, ten of the 14 Si-atom electrons occupy very deep-lying energy levels and are tightly bound to the nucleus of the atom. The binding is so strong, in fact, that these ten electrons remain essentially unperturbed during chemical reactions or normal atom–atom interactions, with the ten-electron-plus-nucleus combination often being referred to as the "core" of the atom. The remaining four Si-atom electrons, on the other hand, are rather weakly bound and are collectively called the *valence electrons* because of their strong participation in chemical reactions. As emphasized in Fig. 2.2, the four valence electrons, if unperturbed, occupy four of the eight allowed slots (or states) having the next highest energy above the deep-lying core levels. Finally, we should note, for completeness, that the electronic configuration in the 32-electron Ge-atom (germanium being the other elemental semiconductor) is essentially identical to the Si-atom configuration except that the Ge-core contains 28 electrons.

2.2 SEMICONDUCTOR MODELS

Building on the information presented in previous sections, we introduce and describe in this section two very important models or visualization aids that are used extensively in the analysis of semiconductor devices. The inclusion of semiconductor models in a chapter devoted to carrier modeling may appear odd at first glance but is actually quite appropriate. We are, in effect, modeling the carrier "container," the semiconductor crystal.

† Actually, not only energy but many other observables relating to atomic-sized particles are quantized. An entire field of study, Quantum Mechanics, has been developed to describe the properties and actions of atomic-sized particles and systems.

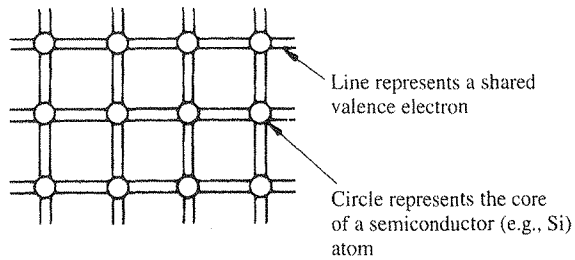


Figure 2.3 The bonding model.

2.2.1 Bonding Model

The isolated Si atom, or a Si atom not interacting with other atoms, was found to contain four valence electrons. Si atoms incorporated in the diamond lattice, on the other hand, exhibit a bonding that involves an attraction between each atom and its four nearest neighbors (refer to Fig. 1.4c). The implication here is that, in going from isolated atoms to the collective crystalline state, Si atoms come to share one of their valence electrons with each of the four nearest neighbors. This covalent bonding, or equal sharing of valence electrons with nearest neighbors, and the mere fact that atoms in the diamond lattice have four nearest neighbors, give rise to the idealized semiconductor representation, the bonding model, shown in Fig. 2.3. Each circle in the Fig. 2.3 bonding model represents the core of a semiconductor atom, while each line represents a shared valence electron. (There are eight lines connected to each atom because any given atom not only *contributes* four shared electrons but must also *accept* four shared electrons from adjacent atoms.) The two-dimensional nature of the model is, of course, an idealization that facilitates mental visualizations.

Although considerable use will be made of the bonding model in subsequent discussions, it is nevertheless worthwhile at this point to examine sample applications of the model to provide some indication of the model's utility. Two sample applications are presented in Fig. 2.4. In Fig. 2.4(a) we use the bonding model to picture a point defect, a missing atom, in the lattice structure. In Fig. 2.4(b) we visualize the breaking of an atom-to-atom bond and the associated release or freeing of an electron. [Bond breaking (at $T > 0$ K) and defects occur naturally in all semiconductors, and hence (if we may be somewhat overcritical) the basic model of Fig. 2.3 is strictly valid for an entire semiconductor only at $T \simeq 0$ K when the semiconductor in question contains no defects and no impurity atoms.]

2.2.2 Energy Band Model

If our interests were restricted to describing the spatial aspects of events taking place within a semiconductor, the bonding model alone would probably be adequate. Quite often, however, one is more interested in the energy-related aspects of an event. In such instances the bonding model, which says nothing about electron energies, is of little value and the energy band model becomes the primary visualization aid.

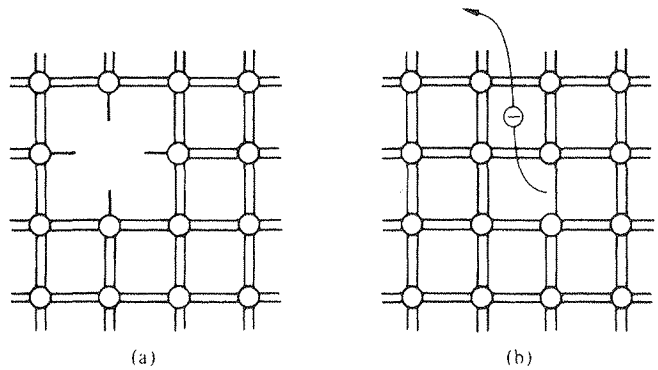


Figure 2.4 Sample utilization of the bonding model. (a) Visualization of a missing atom or point defect. (b) Breaking of an atom-to-atom bond and freeing of an electron.

Let us begin the conceptual path leading to the energy band model by recalling the situation inside an isolated Si atom. Reviewing the Section 2.1 discussion, ten of the 14 electrons inside an isolated Si atom are tightly bound to the nucleus and are unlikely to be significantly perturbed by normal atom–atom interactions. The remaining four electrons are rather weakly bound and, if unperturbed, occupy four of the eight allowed energy states immediately above the last core level. Moreover, it is implicitly understood that the electronic energy states within a group of Si atoms, say N Si atoms, are all identical—as long as the atoms are isolated, that is, far enough apart so that they are noninteracting.

Given the foregoing knowledge of the isolated atom situation, the question next arises as to whether we can use the knowledge to deduce information about the crystalline state. Assuredly we can omit any further mention of the core electrons because these electrons are not significantly perturbed by normal interatomic forces. The opposite, however, is true of the valence electrons. If N atoms are brought into close proximity (the case in crystalline Si), it is quite reasonable to expect a modification in the energy states of the valence electrons.

The modification in the valence-electron energy states actually known to take place is summarized in Fig. 2.5. Starting with N-isolated Si atoms, and conceptually bringing the atoms closer and closer together, one finds the interatomic forces lead to a progressive spread in the allowed energies. The spread in energies gives rise to closely spaced sets of allowed states known as *energy bands.* At the interatomic distance corresponding to the Si lattice spacing, the distribution of allowed states consists of two bands separated by an intervening energy gap. The upper band of allowed states is called the *conduction band;* the lower band of allowed states, the *valence band;* and the intervening energy gap, the *forbidden gap* or *band gap.* In filling the allowed energy band states, electrons tend to gravitate to the lowest possible energies. Noting that electrons are restricted to single occupancy in allowed states (the Pauli exclusion principle) and remembering that the $4N$ valence band states can just accommodate what were formerly $4N$ valence electrons, we typically find that the valence band is almost completely filled with electrons and the

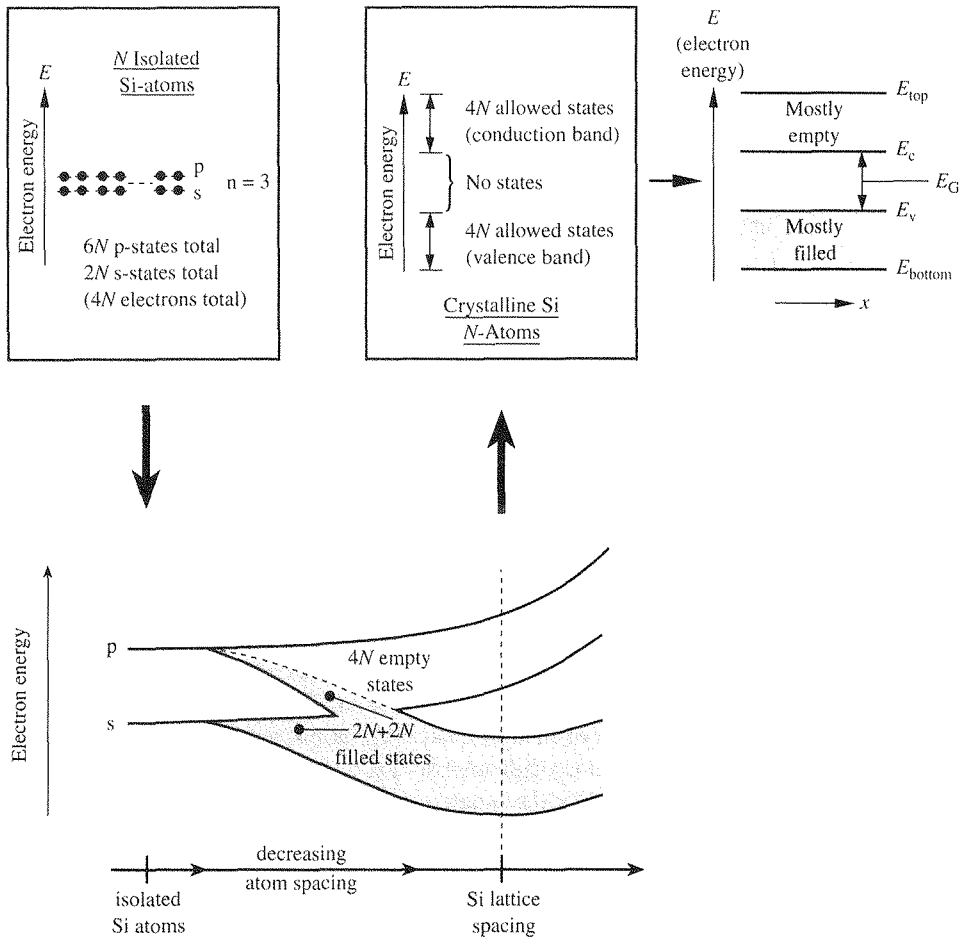


Figure 2.5 Conceptual development of the energy band model starting with N isolated Si atoms on the top left and concluding with a "dressed-up" version of the energy band model on the top right.

conduction band is all but devoid of electrons. Indeed, the valence band is completely filled and the conduction band completely empty at temperatures approaching $T = 0$ K.

To complete our plausibilization of the energy band model, we need to introduce and utilize one additional fact: Unlike the valence electrons in the isolated-atom case, the band electrons in crystalline silicon are not tied to or associated with any one particular atom. True, on the average, one will typically find four electrons being shared between any given Si atom and its four nearest neighbors (as in the bonding model). However, the identity of the shared electrons changes as a function of time, with the electrons moving around from point to point in the crystal. In other words, the allowed electronic states are no longer atomic states but are associated with the crystal as a whole; independent of the point ex-

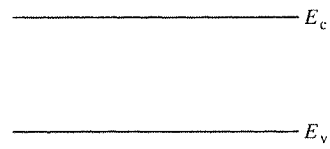


Figure 2.6 The energy band diagram—widely employed simplified version of the energy band model.

amined in a perfect crystal, one sees the same allowed-state configuration. We therefore conclude that for a perfect crystal under equilibrium conditions a plot of the allowed electron energies versus distance along any preselected crystalline direction (always called the x-direction) is as shown on the right-hand side of Fig. 2.5. The cited plot, a plot of allowed electron energy states as a function of position, is the basic energy band model. E_c introduced in the Fig. 2.5 plot is the lowest possible conduction band energy, E_v is the highest possible valence band energy, and $E_G = E_c - E_v$ is the band gap energy.

Finally, Fig. 2.6 displays the form of the basic energy band model encountered in practice. In this widely employed "shorthand" version, the line to indicate the top energy in the conduction band, the line to indicate the bottom energy in the valence band, the fill pattern drawing attention to mostly filled states in the valence band, the labeling of the y- or electron-energy axis, and the labeling of the x- or position axis are all understood to exist implicitly, but are not shown explicitly.

2.2.3 Carriers

With the semiconductor models having been firmly established, we are now in a position to introduce and to visualize the current-carrying entities within semiconductors. Referring to Fig. 2.7, we note first from part (a) that there are no carriers or possible current flow if the bonding model has no broken bonds. Equivalently, in the energy band model, if the valence band is completely filled with electrons and the conduction band is devoid of electrons, there are no carriers or possible current flow. This lack of carriers and associated current flow is easy to understand in terms of the bonding model where the shared electrons are viewed as being tied to the atomic cores. As accurately portrayed in the energy band model, however, the valence band electrons actually move about in the crystal. How is it then that no current can arise from this group of electrons? As it turns out, the momentum of the electrons is quantized in addition to their energy. Moreover, for each and every possible momentum state in a band, there is another state with an oppositely directed momentum of equal magnitude. Thus, if a band is completely filled with electrons, the *net* momentum of the electrons in the band is always identically zero. It follows that no current can arise from the electrons in a completely filled energy band.

The electrons that do give rise to charge transport are visualized in Fig. 2.7(b). When a Si–Si bond is broken and the associated electron is free to wander about the lattice, the released electron is a carrier. Equivalently, in terms of the energy band model, excitation of valence band electrons into the conduction band creates carriers; that is, *electrons in the*

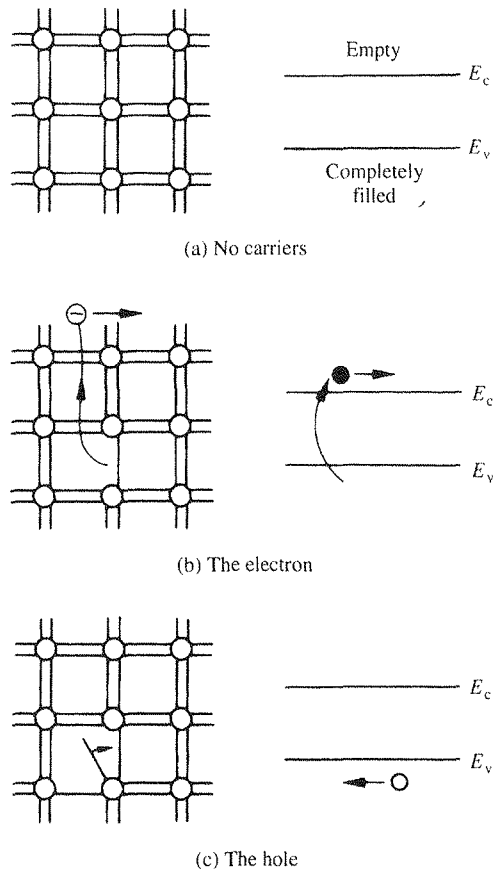


Figure 2.7 Visualization of carriers using the bonding model (left) and the energy band model (right). (a) No-carrier situation; (b) visualization of an electron; (c) visualization of a hole.

conduction band are carriers. Note that the energy required to break a bond in the bonding model and the band gap energy, E_G, are one and the same thing. Likewise, freed bonding-model electrons and conduction band electrons are just different names for the same electrons; in subsequent discussions the word "electrons," when used without a modifier, will be understood to refer to these conduction band electrons.

In addition to releasing an electron, the breaking of a Si–Si bond also creates a missing bond or void in the bonding structure. Thinking in terms of the bonding model, one can visualize the movement of this missing bond from place to place in the lattice as a result of nearby bound electrons jumping into the void (see Fig. 2.7c). Alternatively, one can think in terms of the energy band model where the removal of an electron from the valence band creates an empty state in an otherwise vast sea of filled states. The empty state, like a bubble in a liquid, moves about rather freely in the lattice because of the cooperative motion of the valence band electrons. What we have been describing, the missing bond in the bonding

scheme, the empty state in the valence band, is the second type of carrier found in semiconductors—the hole.

Although perhaps not obvious from the preceding introductory description, the valence band hole is an entity on an equal footing with the conduction band electron. Both electrons and holes participate in the operation of most semiconductor devices. Holes are even the primary carrier in some devices. As one's familiarity with carrier modeling grows, the comparable status of electrons and holes becomes more and more apparent. Ultimately, it is commonplace to conceive of the hole as a type of subatomic particle.

2.2.4 Band Gap and Material Classification

We conclude this section by citing a very important tie between the band gap of a material, the number of carriers available for transport in a material, and the overall nature of a material. As it turns out, although specifically established for semiconductors, the energy band model of Fig. 2.6 can be applied with only slight modification to all materials. The major difference between materials lies not in the nature of the energy bands, but rather in the magnitude of the energy gap between the bands.

Insulators, as illustrated in Fig. 2.8(a), are characterized by very wide band gaps, with E_G for the insulating materials diamond and SiO_2 being $\simeq$5 eV and $\simeq$8 eV, respectively. In these wide band gap materials the thermal energy available at room temperature excites very few electrons from the valence band into the conduction band; thus very few carriers exist inside the material and the material is therefore a poor conductor of current. The band

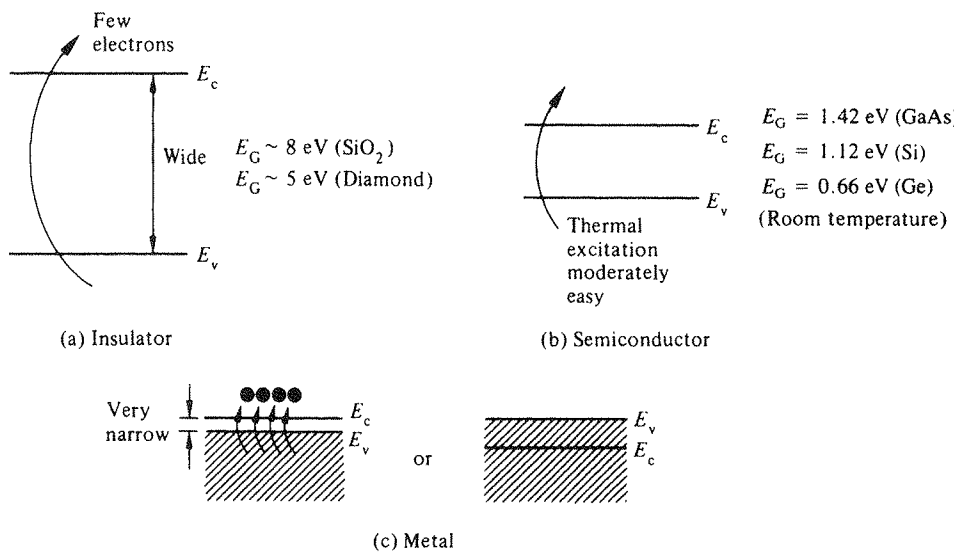


Figure 2.8 Explanation of the distinction between (a) insulators, (b) semiconductors, and (c) metals using the energy band model.

gap in metals, by way of comparison, is either very small, or no band gap exists at all due to an overlap of the valence and conduction bands (see Fig. 2.8c). There is always an abundance of carriers in metals, and hence metals are excellent conductors. Semiconductors present an intermediate case between insulators and metals. At room temperature ($T = 300$ K), $E_G = 1.42$ eV in GaAs, $E_G = 1.12$ eV in Si, and $E_G = 0.66$ eV in Ge. Thermal energy, by exciting electrons from the valence band into the conduction band, creates a moderate number of carriers in these materials, giving rise in turn to a current-carrying capability intermediate between poor and excellent.

2.3 CARRIER PROPERTIES

Having formally introduced the electron and the hole, we next seek to learn as much as possible about the nature of these carriers. In this particular section we examine a collage of carrier-related information, information of a general nature including general facts, properties, and terminology.

2.3.1 Charge

Both electrons and holes are charged entities. Electrons are negatively charged, holes are positively charged, and the *magnitude* of the carrier charge, q, is the same for the two types of carriers. To three-place accuracy in MKS units, $q = 1.60 \times 10^{-19}$ coul. Please note that, under the convention adopted herein, the electron and hole charges are $-q$ and $+q$, respectively; i.e., the sign of the charge is displayed explicitly.

2.3.2 Effective Mass

Mass, like charge, is another very basic property of electrons and holes. Unlike charge, however, the carrier mass is not a simple property and cannot be disposed of simply by quoting a number. Indeed, the apparent or "effective" mass of electrons within a crystal is a function of the semiconductor material (Si, Ge, etc.) and is different from the mass of electrons within a vacuum.

Seeking to obtain insight into the effective mass concept, let us first consider the motion of electrons in a vacuum. If, as illustrated in Fig. 2.9(a), an electron of rest mass m_0 is moving in a vacuum between two parallel plates under the influence of an electric field $\mathscr{E}$, then, according to Newton's second law, the force $\mathbf{F}$ on the electron will be

$$\mathbf{F} = -q\mathscr{E} = m_0 \frac{d\boldsymbol{v}}{dt} \tag{2.2}$$

where $\boldsymbol{v}$ is the electron velocity and t is time. Next consider electrons (conduction band electrons) moving between the two parallel end faces of a semiconductor crystal under the influence of an applied electric field, as envisioned in Fig. 2.9(b). Does Eq. (2.2) also describe the overall motion of electrons within the semiconductor crystal? The answer is *no*.

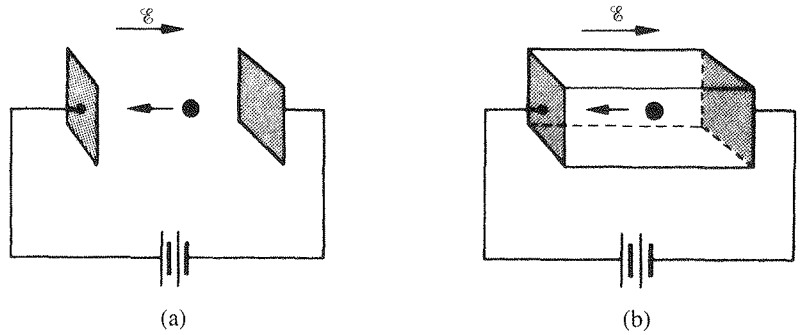


Figure 2.9 An electron moving in response to an applied electric field (a) within a vacuum, and (b) within a semiconductor crystal.

Electrons moving inside a semiconductor crystal will collide with semiconductor atoms, thereby causing a periodic deceleration of the carriers. However, should not Eq. (2.2) apply to the portion of the electronic motion occurring *between* collisions? The answer is again *no*. In addition to the applied electric field, electrons in a crystal are also subject to complex crystalline fields not specifically included in Eq. (2.2).

The foregoing discussion delineated certain important differences between electrons in a crystal and electrons in a vacuum, but it left unresolved the equally important question as to how one properly describes the motion of carriers in a crystal. Strictly speaking, the motion of carriers in a crystal can be described only by using Quantum Mechanics, the formalism appropriate for atomic-sized systems. Fortunately, if the dimensions of the crystal are large compared to atomic dimensions, the complex quantum mechanical formulation for the carrier motion between collisions simplifies to yield an equation of motion identical to Eq. (2.2), except that m_0 is replaced by an effective carrier mass. In other words, for the Fig. 2.9(b) electrons one can write

$$\mathbf{F} = -q\mathscr{E} = m_n^* \frac{d\mathbf{v}}{dt} \tag{2.3}$$

where m_n^* is the electron effective mass. A similar equation can be written for holes with $-q \rightarrow q$ and $m_n^* \rightarrow m_p^*$. In each case the internal crystalline fields and quantum-mechanical effects are all suitably lumped into the effective mass factor multiplying $d\mathbf{v}/dt$. This is a very significant result. *It allows us to conceive of electrons and holes as quasi-classical particles and to employ classical particle relationships in most device analyses.*

Although the effective mass formulation is a significant simplification, it should be mentioned that the carrier acceleration can vary with the direction of travel in a crystal; i.e., the effective masses can have multiple components. Moreover, depending on how a macroscopic observable is related to the carrier motion, a different grouping of mass components can lead to a different m^* being utilized in a particular relationship. There are, for example, cyclotron resonance effective masses, conductivity effective masses, density of

Table 2.1 Density of States Effective Masses at 300 K.

Material	m_n^*/m_0	m_p^*/m_0
Si	1.18	0.81
Ge	0.55	0.36
GaAs	0.066	0.52

states effective masses, among others. It is also probably not too surprising that the effective masses vary somewhat with temperature. Herein we make direct use of only the density of states effective masses (see Subsection 2.4.1). The density of states effective masses for electrons and holes in Si, Ge, and GaAs at 300 K are listed in Table 2.1.

2.3.3 Carrier Numbers in Intrinsic Material

The term *intrinsic semiconductor* in common usage refers to an extremely pure semiconductor sample containing an insignificant amount of impurity atoms. More precisely, an intrinsic semiconductor is a semiconductor whose properties are native to the material (that is, not caused by external additives). The number of carriers in an intrinsic semiconductor fits into the scheme of things in that it is an identifiable intrinsic property of the material.

Defining, quite generally,

$$n = \text{number of electrons/cm}^3$$

$$p = \text{number of holes/cm}^3$$

existing inside a semiconductor, then, given an intrinsic semiconductor under equilibrium conditions, one finds

$$n = p = n_i \tag{2.4}$$

and

$$\left.\begin{aligned} n_i &\simeq 2 \times 10^6/\text{cm}^3 && \text{in GaAs} \\ &\simeq 1 \times 10^{10}/\text{cm}^3 && \text{in Si} \\ &\simeq 2 \times 10^{13}/\text{cm}^3 && \text{in Ge} \end{aligned}\right\} \text{at room temperature}$$

The electron and hole concentrations in an intrinsic semiconductor are equal because carriers within a very pure material can be created only in pairs. Referring to Fig. 2.7, if a semiconductor bond is broken, a free electron and a missing bond or hole are created simultaneously. Likewise, the excitation of an electron from the valence band into the con-

duction band automatically creates a valence band hole along with the conduction band electron. Also note that the intrinsic carrier concentration, although large in an absolute sense, is relatively small compared with the number of bonds that could be broken. For example, in Si there are 5×10^{22} atoms/cm^3 and four bonds per atom, making a grand total of 2×10^{23} bonds or valence band electrons per cm^3. Since $n_i \simeq 10^{10}$/cm^3, one finds less than one bond in 10^{13} broken in Si at room temperature. To accurately represent the situation inside intrinsic Si at room temperature, we could cover all the university chalkboards in the world with the bonding model and possibly show only *one* broken bond.

2.3.4 Manipulation of Carrier Numbers—Doping

Doping, in semiconductor terminology, is the addition of controlled amounts of specific impurity atoms with the express purpose of increasing either the electron or the hole concentration. The addition of dopants in controlled amounts to semiconductor materials occurs routinely in the fabrication of almost all semiconductor devices. Common Si dopants are listed in Table 2.2. To increase the electron concentration, one can add either phosphorus, arsenic, or antimony atoms to the Si crystal, with phosphorus followed closely by arsenic being the most commonly employed donor (electron-increasing) dopants. To increase the hole concentration, one adds either boron, gallium, indium, or aluminum atoms to the Si crystal, with boron being the most commonly employed acceptor (hole-increasing) dopant.

To understand how the addition of impurity atoms can lead to a manipulation of carrier numbers, it is important to note that the Table 2.2 donors are all from Column V in the Periodic Table of Elements, while all of the cited acceptors are from Column III in the Periodic Table of Elements. As visualized in Fig. 2.10(a) using the bonding model, when a Column V element with five valence electrons is substituted for a Si atom in the semiconductor lattice, four of the five valence electrons fit snugly into the bonding structure. The fifth donor electron, however, does not fit into the bonding structure and is weakly bound to the donor site. At room temperature this electron is readily freed to wander about the lattice and hence becomes a carrier. Please note that this donation (hence the name "donor") of carrier electrons does not increase the hole concentration. The donor ion left behind when the fifth electron is released cannot move, and there are no broken atom–atom bonds associated with the release of the fifth electron.

Table 2.2 Common Silicon Dopants. Arrows indicate the most widely employed dopants.

Donors (Electron-increasing dopants)		*Acceptors (Hole-increasing dopants)*	
P ←	Column V elements	B ←	Column III elements
As ←		Ga	
Sb		In	
		Al	

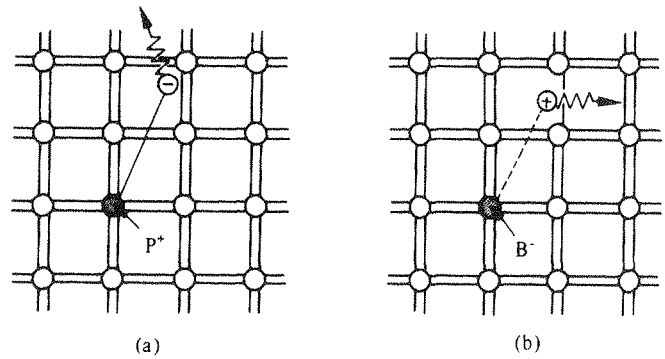


Figure 2.10 Visualization of (a) donor and (b) acceptor action using the bonding model. In (a) the Column V element P is substituted for a Si atom; in (b) the Column III element B is substituted for a Si atom.

The explanation of acceptor action follows a similar line of reasoning. The Column III acceptors have three valence electrons and cannot complete one of the semiconductor bonds when substituted for Si atoms in the semiconductor lattice (see Fig. 2.10b). The Column III atom, however, readily accepts (hence the name "acceptor") an electron from a nearby Si–Si bond, thereby completing its own bonding scheme and in the process creating a hole that can wander about the lattice. Here again there is an increase in only one type of carrier. The negatively charged acceptor ion (acceptor atom plus accepted electron) cannot move, and no electrons are released in the hole-creation process.

The foregoing bonding-model-based explanation of dopant action is reasonably understandable. There are, nevertheless, a few loose ends. For one, we noted that the fifth donor electron was rather weakly bound and readily freed at room temperature. How does one interpret the relative term "weakly bound"? It takes $\simeq 1$ eV to break Si–Si bonds and very few of the Si–Si bonds are broken at room temperature. Perhaps "weakly bound" means a binding energy $\simeq 0.1$ eV or less? The question also arises as to how one visualizes dopant action in terms of the energy band model. Both of the cited questions, as it turns out, involve energy considerations and are actually interrelated.

Let us concentrate first on the binding energy of the fifth donor electron. Crudely speaking, the positively charged donor-core-plus-fifth-electron may be likened to a hydrogen atom (see Fig. 2.11). Conceptually, the donor core replaces the hydrogen-atom nucleus and the fifth donor electron replaces the hydrogen-atom electron. In the real hydrogen atom the electron moves of course in a vacuum, can be characterized by the mass of a free electron, and, referring to Eq. (2.1), has a ground-state binding energy of -13.6 eV. In the pseudo-hydrogen atom, on the other hand, the orbiting electron moves through a sea of Si atoms and is characterized by an effective mass. Hence, in the donor or pseudo-atom case, the permittivity of free space must be replaced by the permittivity of Si and m_0 must be replaced by m_n^*. We therefore conclude that the binding energy (E_B) of the fifth donor electron is approximately

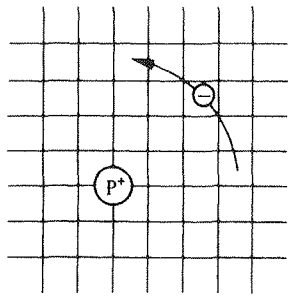


Figure 2.11 Pseudo-hydrogen atom model for the donor-site bond.

$$E_B \simeq -\frac{m_n^* q^4}{2(4\pi K_S \varepsilon_0 \hbar)^2} = \frac{m_n^*}{m_0}\frac{1}{K_S^2} E_{H|n=1} \simeq -0.1 \text{ eV} \tag{2.5}$$

where K_S is the Si dielectric constant ($K_S = 11.8$). Actual donor-site binding energies in Si are listed in Table 2.3. The observed binding energies are seen to be in good agreement with the Eq. (2.5) estimate and, confirming the earlier speculation, are roughly $\simeq 1/20$ the Si band gap energy.

Having established the strength of dopant-site bonds, we are now in a position to tackle the problem of how one visualizes dopant action using the energy band model. Working on the problem, we note first that when an electron is released from a donor it becomes a conduction band electron. If the energy absorbed at the donor is precisely equal to the electron binding energy, the released electron will moreover have the lowest possible energy in the conducting band—namely, E_c. Adding an energy $|E_B|$ to the bound electron, in other words, *raises* the electron's energy to E_c. Hence, we are led to conclude that the bound electron occupies an allowed electronic level an energy $|E_B|$ below the conduction band edge, or, as visualized in Fig. 2.12, donor sites can be incorporated into the energy band scheme by adding allowed electronic levels at an energy $E_D = E_c - |E_B|$. Note that the donor energy level is represented by a set of dashes, instead of a continuous line, because an electron bound to a donor site is localized in space; i.e., a bound electron does not leave the general Δx vicinity of the donor. The relative closeness of E_D to E_c of course reflects the fact that $E_c - E_D = |E_B| \simeq (1/20)E_G(\text{Si})$.

Table 2.3 Dopant-Site Binding Energies.

Donors	$\|E_B\|$	*Acceptors*	$\|E_B\|$
Sb	0.039 eV	B	0.045 eV
P	0.045 eV	Al	0.067 eV
As	0.054 eV	Ga	0.072 eV
		In	0.16 eV

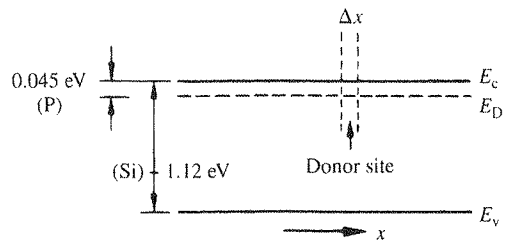


Figure 2.12 Addition of the $E = E_D$ donor levels to the energy band diagram. Dashes of width Δx emphasize the localized nature of the bound donor-site states.

The actual visualization of dopant action using the energy band model is pictured in Fig. 2.13. Examining the left-hand side of Fig. 2.13(a), one finds all the donor sites filled with bound electrons at temperatures $T \rightarrow 0$ K. This is true because very little thermal energy is available to excite electrons from the donor sites into the conduction band at these very low temperatures. The situation changes, of course, as the temperature is increased, with more and more of the weakly bound electrons being donated to the conduction band. At room temperature the ionization of the donors is all but total, giving rise to the situation pictured at the extreme right of Fig. 2.13(a). Although we have concentrated on donors, the situation for acceptors is completely analogous. As visualized in Fig. 2.13(b), acceptors introduce allowed electronic levels into the forbidden gap at an energy slightly above the valence band edge. At low temperatures, all of these sites will be empty—there is insuffi-

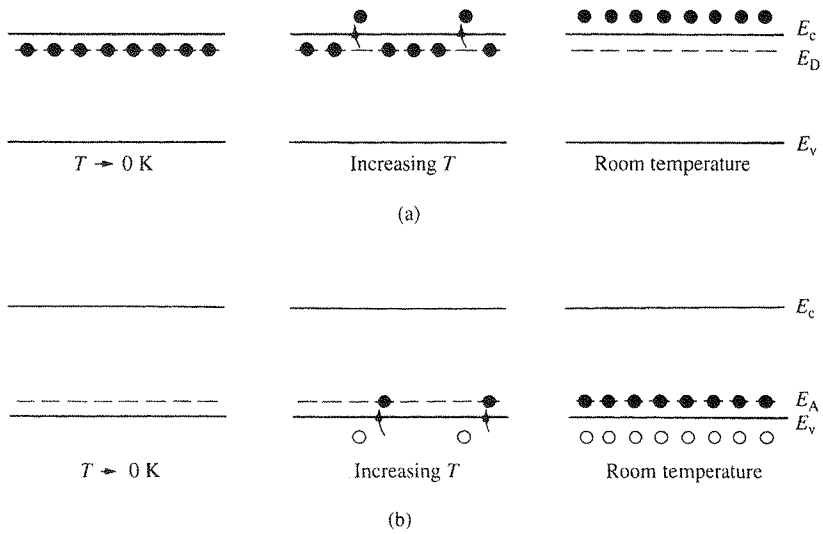


Figure 2.13 Visualization of (a) donor and (b) acceptor action using the energy band model.

cient energy at temperatures $T \rightarrow 0$ K for a valence band electron to make the transition to an acceptor site. Increasing temperature, implying an increased store of thermal energy, facilitates electrons jumping from the valence band onto the acceptor levels. The removal of electrons from the valence band of course creates holes. At room temperature, essentially all of the acceptor sites are filled with an excess electron and an increased hole concentration is effected in the material.

To complete this subsection, a few words are in order concerning the doping of non-elemental semiconductors such as GaAs. Dopant action in GaAs follows the same general principles but is slightly more involved because of the existence of two different lattice-site atoms. Completely analogous to the doping of a Si crystal, the Column VI elements S, Se, and Te act as donors when replacing the Column V element As in the GaAs lattice. Similarly, the Column II elements Be, Mg, and Zn act as acceptors when replacing the Column III element Ga. A new situation arises when Column IV elements such as Si and Ge are incorporated into GaAs. Si typically replaces Ga in the GaAs lattice and is a popular n-type dopant. However, under certain conditions Si can be made to replace As in the GaAs lattice, thereby functioning as an acceptor. In fact, GaAs pn junctions have been fabricated where Si was both the p-side and n-side dopant. An impurity that can act as either a donor or an acceptor is referred to as an *amphoteric* dopant.

Exercise 2.1

Energy Quiz

P: (a) 1 eV is equal to how many joules of energy?

(b) kT is equal to how many eV at 300 K?

(c) The ionization energy of acceptors and donors in Si is roughly equal to _____?

(d) $E_G(\text{Si}) =$ _____?

(e) $E_G(\text{SiO}_2) =$ _____?

(f) The energy required to ionize a hydrogen atom initially in the $\mathbf{n} = 1$ state is ___?

S: (a) 1 eV $= 1.60 \times 10^{-19}$ joules.

(b) $kT = (8.617 \times 10^{-5})(300) = 0.0259$ eV at 300 K.

(c) The ionization energy of dopants is equal to $|E_B|$ of the dopant sites. As discussed in Subsection 2.3.4, $|E_B| \cong 0.1$ eV for dopants in Si.

(d) $E_G(\text{Si}) = 1.12$ eV at 300 K (see Subsection 2.2.4).

(e) $E_G(\text{SiO}_2) \cong 8$ eV (see Subsection 2.2.4).

(f) Making use of Eq. (2.1) in Section 2.1, the hydrogen atom ionization energy is $|E_{\text{H}|\mathbf{n}=1}| = 13.6$ eV.

2.3.5 Carrier-Related Terminology

Since terminology is often a stumbling block to understanding and since this particular section is replete with specialized terms, it seems appropriate to conclude the section with an overview of carrier-related terminology. Approximately one half of the carrier-related terms that follow were introduced and defined earlier in this section; the remainder of the terms are being listed here for the first time. All of the terms are widely employed and their definitions should be committed to memory.

Dopants—specific impurity atoms that are added to semiconductors in controlled amounts for the express purpose of increasing either the electron or the hole concentration.

Intrinsic semiconductor—undoped semiconductor; extremely pure semiconductor sample containing an insignificant amount of impurity atoms; a semiconductor whose properties are native to the material.

Extrinsic semiconductor—doped semiconductor; a semiconductor whose properties are controlled by added impurity atoms.

Donor—impurity atom that increases the electron concentration; n-type dopant.

Acceptor—impurity atom that increases the hole concentration; p-type dopant.

n-type material—a donor-doped material; a semiconductor containing more electrons than holes.

p-type material—an acceptor-doped material; a semiconductor containing more holes than electrons.

Majority carrier—the most abundant carrier in a given semiconductor sample; electrons in an n-type material, holes in a p-type material.

Minority carrier—the least abundant carrier in a given semiconductor sample; holes in an n-type material, electrons in a p-type material.

2.4 STATE AND CARRIER DISTRIBUTIONS

Up to this point in the modeling process we have concentrated on carrier properties and information of a conceptual, qualitative, or, at most, semiquantitative nature. Practically speaking, there is often a need for more detailed information. For example, most semiconductors are doped, and the precise numerical value of the carrier concentrations inside doped semiconductors is of routine interest. Another property of interest is the distribution of carriers as a function of energy in the respective energy bands. In this section we begin the process of developing a more detailed description of the carrier populations. The development will eventually lead to relationships for the carrier distributions and concentrations within semiconductors under equilibrium conditions.

2.4.1 Density of States

When the energy band model was first introduced in Section 2.2 we indicated that the total number of allowed states in each band was four times the number of atoms in the crystal. Not mentioned at the time was how the allowed states were distributed in energy; i.e., how many states were to be found at any given energy in the conduction and valence bands. We are now interested in this energy distribution of states, or *density of states,* as it is more commonly known, because the state distribution is an essential component in determining carrier distributions and concentrations.

To determine the desired density of states, it is necessary to perform an analysis based on quantum-mechanical considerations. Herein we will merely summarize the results of the analysis; namely, for energies not too far removed from the band edges, one finds

$$g_c(E) = \frac{m_n^* \sqrt{2m_n^*(E - E_c)}}{\pi^2 \hbar^3}, \qquad E \geq E_c \tag{2.6a}$$

$$g_v(E) = \frac{m_p^* \sqrt{2m_p^*(E_v - E)}}{\pi^2 \hbar^3}, \qquad E \leq E_v \tag{2.6b}$$

where $g_c(E)$ and $g_v(E)$ are the density of states at an energy E in the conduction and valence bands, respectively.

What exactly should be known and remembered about the cited density of states? For one, it is important to grasp the general density of states concept. The density of states can be likened to the description of the seating in a football stadium, with the number of seats in the stadium a given distance from the playing field corresponding to the number of states a specified energy interval from E_c or E_v. Second, the general form of the relationships should be noted. As illustrated in Fig. 2.14, $g_c(E)$ is zero at E_c and increases as the square root of energy when one proceeds upward into the conduction band. Similarly, $g_v(E)$ is precisely zero at E_v and increases with the square root of energy as one proceeds downward from E_v into the valence band. Also note that differences between $g_c(E)$ and $g_v(E)$ stem from differences in the carrier effective masses. If m_n^* were equal to m_p^*, the seating (states) on both sides of the football field (the band gap) would be mirror images of each other. Finally, considering closely spaced energies E and $E + dE$ in the respective bands, one can state

$g_c(E)dE$ represents the number of conduction band states/cm^3 lying in the energy range between E and $E + dE$ (if $E \geq E_c$),

$g_v(E)dE$ represents the number of valence band states/cm^3 lying in the energy range between E and $E + dE$ (if $E \leq E_v$).

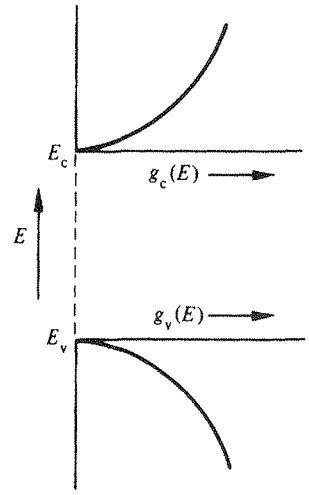


Figure 2.14 General energy dependence of $g_c(E)$ and $g_v(E)$ near the band edges. $g_c(E)$ and $g_v(E)$ are the density of states in the conduction and valence bands, respectively.

It therefore follows that $g_c(E)$ and $g_v(E)$ themselves are numbers/unit volume-unit energy, or typically, numbers/cm^3-eV.

2.4.2 The Fermi Function

Whereas the density of states tells one how many states exist at a given energy E, the Fermi function $f(E)$ specifies how many of the existing states at the energy E will be filled with an electron, or equivalently,

$f(E)$ specifies, under equilibrium conditions, the probability that an available state at an energy E will be occupied by an electron.

Mathematically speaking, the Fermi function is simply a probability distribution function. In mathematical symbols,

$$\boxed{f(E) = \frac{1}{1 + e^{(E - E_F)/kT}}} \tag{2.7}$$

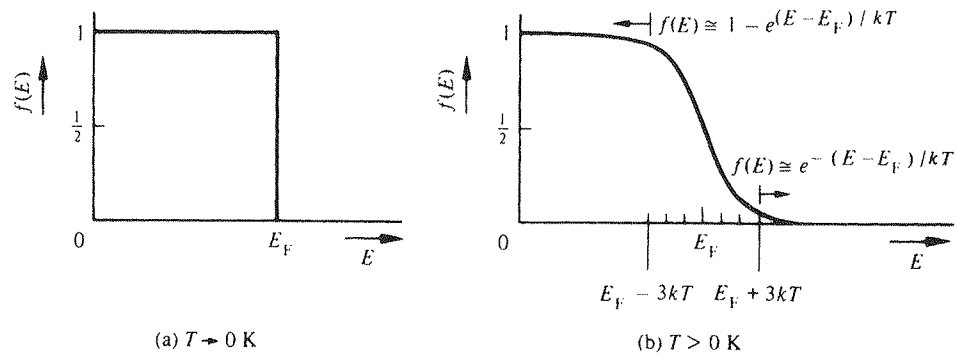


Figure 2.15 Energy dependence of the Fermi function. (a) $T \to 0$ K; (b) generalized $T > 0$ K plot with the energy coordinate expressed in kT units.

where

$$E_F = \text{Fermi energy or Fermi level}$$

$$k = \text{Boltzmann constant } (k = 8.617 \times 10^{-5} \text{ eV/K})$$

$$T = \text{temperature in Kelvin (K)}$$

Seeking insight into the nature of the newly introduced function, let us begin by investigating the Fermi function's energy dependence. Consider first temperatures where $T \to 0$ K. As $T \to 0$ K, $(E - E_F)/kT \to -\infty$ for all energies $E < E_F$ and $(E - E_F)/kT \to +\infty$ for all energies $E > E_F$. Hence $f(E < E_F) \to 1/[1 + \exp(-\infty)] = 1$ and $f(E > E_F) \to 1/[1 + \exp(+\infty)] = 0$. This result is plotted in Fig. 2.15(a) and is simply interpreted as meaning that all states at energies below E_F will be filled and all states at energies above E_F will be empty for temperatures $T \to 0$ K. In other words, there is a sharp cutoff in the filling of allowed energy states at the Fermi energy E_F when the system temperature approaches absolute zero.

Let us next consider temperatures $T > 0$ K. Examining the Fermi function, we make the following pertinent observations.

(i) If $E = E_F$, $f(E_F) = 1/2$.

(ii) If $E \geq E_F + 3kT$, $\exp[(E - E_F)/kT] \gg 1$ and $f(E) \simeq \exp[-(E - E_F)/kT]$. Consequently, above $E_F + 3kT$ the Fermi function or filled-state probability decays exponentially to zero with increasing energy. Moreover, most states at energies $3kT$ or more above E_F will be empty.

(iii) If $E \leq E_F - 3kT$, $\exp[(E - E_F)/kT] \ll 1$ and $f(E) \simeq 1 - \exp[(E - E_F)/kT]$. Below $E_F - 3kT$, therefore, $[1 - f(E)]$, the probability that a given state will be *empty*, decays exponentially to zero with decreasing energy. Most states at energies $3kT$ or more below E_F will be filled.

(iv) At room temperature ($T = 300$ K), $kT = 0.0259$ eV and $3kT = 0.0777$ eV $\ll E_G$(Si). Compared to the Si band gap, the $3kT$ energy interval that appears prominently in the $T > 0$ K formalism is typically quite small.

The properties just cited are reflected and summarized in the $T > 0$ K Fermi function plot displayed in Fig. 2.15(b).

Before concluding the discussion here, it is perhaps worthwhile to reemphasize that the Fermi function applies only under equilibrium conditions. The Fermi function, however, is universal in the sense that it applies with equal validity to all materials—insulators, semiconductors, and metals. Although introduced in relationship to semiconductors, the Fermi function is not dependent in any way on the special nature of semiconductors but is simply a statistical function associated with electrons in general. Finally, the relative positioning of the Fermi energy E_F compared to E_c (or E_v), an item of obvious concern, is treated in subsequent subsections.

Exercise 2.2

P: The probability that a state is filled at the conduction band edge (E_c) is precisely equal to the probability that a state is *empty* at the valence band edge (E_v). Where is the Fermi level located?

S: The Fermi function, $f(E)$, specifies the probability of electrons occupying states at a given energy E. The probability that a state is empty (not filled) at a given energy E is equal to $1 - f(E)$. Here we are told

$$f(E_c) = 1 - f(E_v)$$

Since

$$f(E_c) = \frac{1}{1 + e^{(E_c - E_F)/kT}}$$

and

$$1 - f(E_v) = 1 - \frac{1}{1 + e^{(E_v - E_F)/kT}} = \frac{1}{1 + e^{(E_F - E_v)/kT}}$$

we conclude

$$\frac{E_c - E_F}{kT} = \frac{E_F - E_v}{kT}$$

or

$$E_F = \frac{E_c + E_v}{2}$$

The Fermi level is positioned at midgap.

(C) Exercise 2.3

In this exercise we wish to examine in more detail how the Fermi function varies with temperature.

P: Successively setting $T = 100$, 200, 300, and then 400 K, compute and plot $f(E)$ versus $\Delta E = E - E_F$ for $-0.2\,\text{eV} \leq \Delta E \leq 0.2\,\text{eV}$. All $f(E)$ versus ΔE curves should be superimposed on a single set of coordinates.

S: MATLAB program script . . .

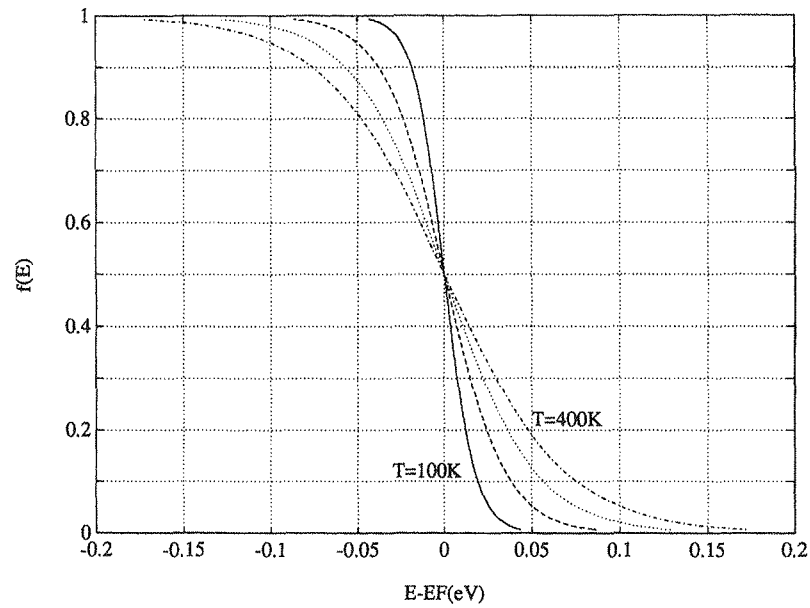
```
%Fermi Function Calculation, f(ΔE,T)

% Constant
k=8.617e-5;

%Computation proper
for ii=1:4;
   T=100*ii;
   kT=k*T;
   dE(ii,1)=-5*kT;
   for jj=1:101
        f(ii,jj)=1/(1+exp(dE(ii,jj)/kT));
        dE(ii,jj+1)=dE(ii,jj)+0.1*kT;
   end
end
dE=dE(:,1:jj);    %This step strips the extra dE value
```

```
%Plotting result
close
plot(dE',f'); grid;   %Note the transpose (') to form data columns
xlabel('E-EF(eV)');   ylabel('f(E)');
text(.05,.2,'T=400K');   text(-.03,.1,'T=100K');
```

Program output . . .



2.4.3 Equilibrium Distribution of Carriers

Having established the distribution of available band states and the probability of filling those states under equilibrium conditions, we can now easily deduce the distribution of carriers in the respective energy bands. To be specific, the desired distribution is obtained by simply multiplying the appropriate density of states by the appropriate occupancy factor—$g_c(E)f(E)$ yields the distribution of electrons in the conduction band and $g_v(E)\ [1 - f(E)]$ yields the distribution of holes (unfilled states) in the valence band. Sample carrier distributions for three different assumed positions of the Fermi energy (along with associated energy band diagram, Fermi function, and density of states plots) are pictured in Fig. 2.16.

Examining Fig. 2.16, we note in general that all carrier distributions are zero at the

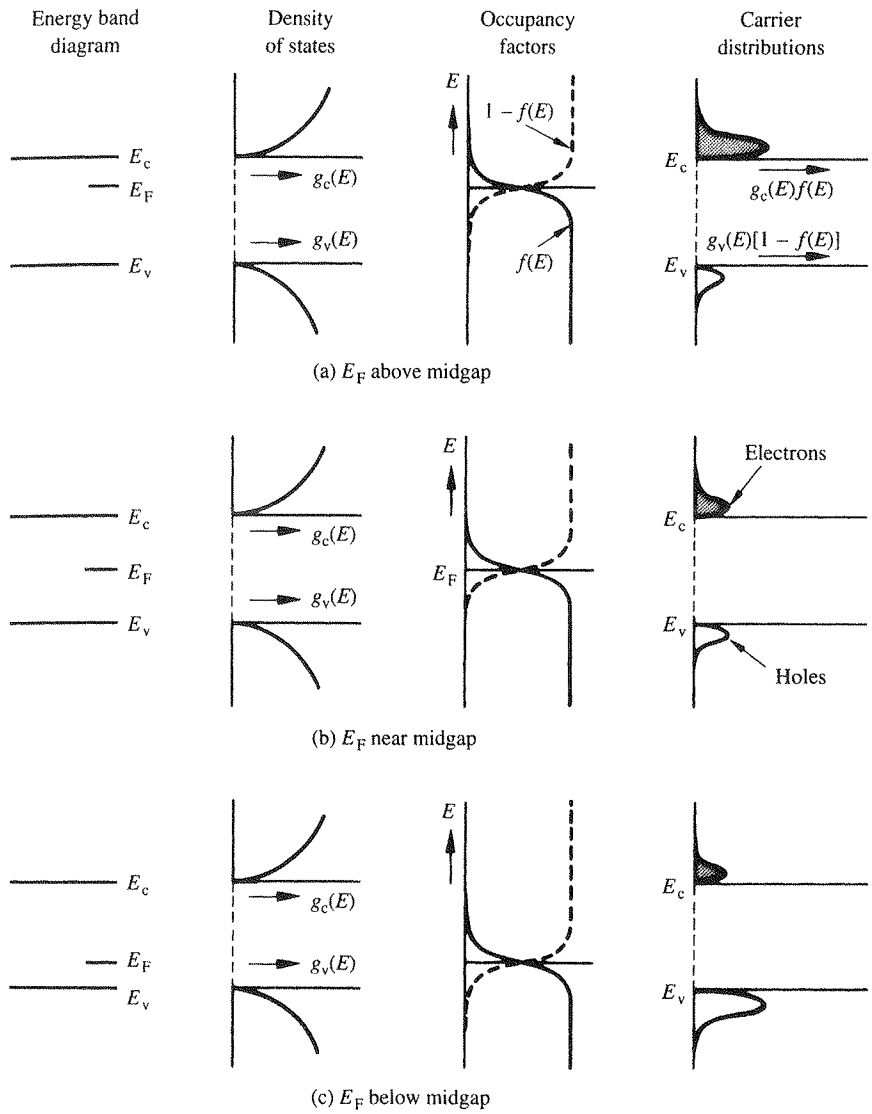


Figure 2.16 Carrier distributions (not drawn to scale) in the respective bands when the Fermi level is positioned (a) above midgap, (b) near midgap, and (c) below midgap. Also shown in each case are coordinated sketches of the energy band diagram, density of states, and the occupancy factors (the Fermi function and one minus the Fermi function).

band edges, reach a peak value very close to E_c or E_v, and then decay very rapidly toward zero as one moves upward into the conduction band or downward into the valence band. In other words, most of the carriers are grouped energetically in the near vicinity of the band edges. Another general point to note is the effect of the Fermi level positioning on the relative magnitude of the carrier distributions. When E_F is positioned in the upper half of the band gap (or higher), the electron distribution greatly outweighs the hole distribution. Although both the filled-state occupancy factor $[f(E)]$ and the empty-state occupancy factor $[1 - f(E)]$ fall off exponentially as one proceeds from the band edges deeper into the conduction and valence bands, respectively, $[1 - f(E)]$ is clearly much smaller than $f(E)$ at corresponding energies if E_F lies in the upper half of the band gap. Lowering E_F effectively slides the occupancy plots downward, giving rise to a nearly equal number of carriers when E_F is at the middle of the gap. Likewise, a predominance of holes results when E_F lies below the middle of the gap. The argument here assumes, of course, that $g_c(E)$ and $g_v(E)$ are of the same order of magnitude at corresponding energies (as is the case for Si and Ge). Also, referring back to the previous subsection, the statements concerning the occupancy factors falling off exponentially in the respective bands are valid provided $E_c - 3kT \geq E_F \geq E_v + 3kT$.

The information just presented concerning the carrier distributions and the relative magnitudes of the carrier numbers finds widespread usage. The information, however, is often conveyed in an abbreviated or shorthand fashion. Figure 2.17, for example, shows a common way of representing the carrier energy distributions. The greatest number of circles or dots are drawn close to E_c and E_v, reflecting the peak in the carrier concentrations near the band edges. The smaller number of dots as one progresses upward into the conduction band crudely models the rapid falloff in the electron density with increasing energy. An extensively utilized means of conveying the relative magnitude of the carrier numbers is displayed in Fig. 2.18. To represent an intrinsic material, a dashed line is drawn in approximately the middle of the band gap and labeled E_i. The near-midgap positioning of E_i, the intrinsic Fermi level, is of course consistent with the previously cited fact that the electron and hole numbers are about equal when E_F is near the center of the band gap. Similarly, a solid line labeled E_F appearing above midgap tells one at a glance that the semiconductor in question is n-type; a solid line labeled E_F appearing below midgap signifies that the semiconductor is p-type. Note finally that the dashed E_i line also typically appears on the

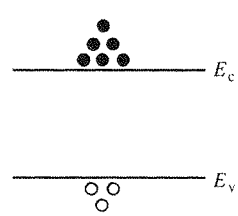


Figure 2.17 Schematic representation of carrier energy distributions.

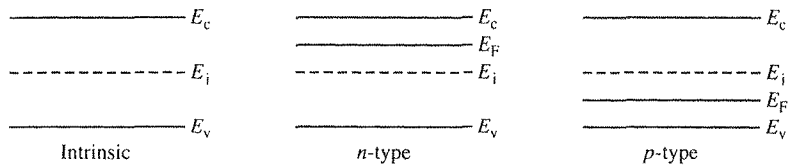


Figure 2.18 "At a glance" representation of intrinsic (left), n-type (middle), and p-type (right) semiconductor materials using the energy band diagram.

energy band diagrams characterizing extrinsic semiconductors. The E_i line in such cases represents the expected positioning of the Fermi level *if* the material were intrinsic, and it serves as a reference energy level dividing the upper and lower halves of the band gap.

2.5 EQUILIBRIUM CARRIER CONCENTRATIONS

We have arrived at an important point in the carrier modeling process. For the most part, this section simply embodies the culmination of our modeling efforts, with working relationships for the equilibrium carrier concentrations being established to complement the qualitative carrier information presented in previous sections. Unfortunately, the emphasis on the development of mathematical relationships makes the final assault on the carrier modeling summit a little more arduous (and perhaps just a bit boring). Hopefully, the reader can stay focused. A comment is also in order concerning the presentation herein of alternative forms for the carrier relationships. The alternative forms can be likened to the different kinds of wrenches used, for example, in home and automobile repairs. The open-end wrench, the box wrench, and the ratchet wrench all serve the same general purpose. In some applications one can use any of the wrenches. In other applications, however, a special situation restricts the type of wrench employed or favors the use of one wrench over another. The same is true of the alternative carrier relationships. Finally, boxes are drawn around expressions that find widespread usage. A single-walled box signifies a moderately important result; a double-walled box, a very important result.

2.5.1 Formulas for *n* and *p*

Since $g_c(E)\,dE$ represents the number of conduction band states/cm^3 lying in the E to $E + dE$ energy range, and $f(E)$ specifies the probability that an available state at an energy E will be occupied by an electron, it then follows that $g_c(E)f(E)dE$ gives the number of conduction band electrons/cm^3 lying in the E to $E + dE$ energy range, and $g_c(E)f(E)dE$ integrated over all conduction band energies must yield the total number of electrons in the conduction band. In other words, integration over the equilibrium distribution of electrons in the conduction band yields the equilibrium electron concentration. A similar statement can be made relative to the hole concentration. We therefore conclude

$$n = \int_{E_c}^{E_{top}} g_c(E) f(E) dE \tag{2.8a}$$

$$p = \int_{E_{bottom}}^{E_v} g_v(E)[1 - f(E)] dE \tag{2.8b}$$

Seeking explicit expressions for the carrier concentrations, let us focus our efforts on the n-integral. (The analogous p-integral manipulations are left to the reader as an exercise.) Substituting the Eq. (2.6a) expression for $g_c(E)$ and the Eq. (2.7) expression for $f(E)$ into Eq. (2.8a), one obtains

$$n = \frac{m_n^* \sqrt{2m_n^*}}{\pi^2 \hbar^3} \int_{E_c}^{E_{top}} \frac{\sqrt{E - E_c}\, dE}{1 + e^{(E - E_F)/kT}} \tag{2.9}$$

Now letting

$$\eta = \frac{(E - E_c)}{kT} \tag{2.10a}$$

$$\eta_c = \frac{(E_F - E_c)}{kT} \tag{2.10b}$$

$$E_{top} \to \infty \tag{2.10c}$$

yields

$$n = \frac{m_n^* \sqrt{2m_n^*}\, (kT)^{3/2}}{\pi^2 \hbar^3} \int_0^\infty \frac{\eta^{1/2} d\eta}{1 + e^{\eta - \eta_c}} \tag{2.11}$$

The (2.10c) simplification on the upper integration limit makes use of the fact that the integrand in question falls off rapidly with increasing energy and is essentially zero for energies only a few kT above E_c. Hence, extending the upper limit to ∞ has a totally negligible effect on the value of the integral.

Even with the cited simplification, the Eq. (2.11) integral cannot be expressed in a closed form containing simple functions. The integral itself is in fact a tabulated function that can be found in a variety of mathematical references. Identifying

$$F_{1/2}(\eta_c) \equiv \int_0^\infty \frac{\eta^{1/2} d\eta}{1 + e^{\eta - \eta_c}}, \quad \text{the Fermi-Dirac integral of order } 1/2 \tag{2.12}$$

and also defining

$$N_C = 2\left[\frac{m_n^* kT}{2\pi\hbar^2}\right]^{3/2}, \quad \text{the "effective" density of conduction band states} \tag{2.13a}$$

$$N_V = 2\left[\frac{m_p^* kT}{2\pi\hbar^2}\right]^{3/2}, \quad \text{the "effective" density of valence band states} \tag{2.13b}$$

one obtains

$$\boxed{n = N_C \frac{2}{\sqrt{\pi}} F_{1/2}(\eta_c)} \tag{2.14a}$$

and, by analogy,

$$\boxed{p = N_V \frac{2}{\sqrt{\pi}} F_{1/2}(\eta_v)} \tag{2.14b}$$

where $\eta_v \equiv (E_v - E_F)/kT$.

The Eq. (2.14) relationships are a very general result, valid for any conceivable positioning of the Fermi level. The constants N_C and N_V are readily calculated; $\boxed{\text{at 300 K, } N_{C,V} = (2.510 \times 10^{19}/\text{cm}^3)(m_{n,p}^*/m_0^*)^{3/2}}$. The value of the Fermi integral can be obtained from available tables, from plots, or by direct computation. The general-form relationships, nonetheless, are admittedly cumbersome and inconvenient to use in routine analyses. Fortunately, simplified closed-form expressions do exist that can be employed in the vast majority of practical problems. To be specific, if E_F is restricted to values $E_F \leq E_c - 3kT$, then $1/[1 + \exp(\eta - \eta_c)] \simeq \exp[-(\eta - \eta_c)]$ for all $E \geq E_c (\eta \geq 0)$, and

$$F_{1/2}(\eta_c) = \frac{\sqrt{\pi}}{2} e^{(E_F - E_c)/kT} \tag{2.15a}$$

Likewise, if $E_F \geq E_v + 3kT$, then

$$F_{1/2}(\eta_v) = \frac{\sqrt{\pi}}{2} e^{(E_v - E_F)/kT} \tag{2.15b}$$

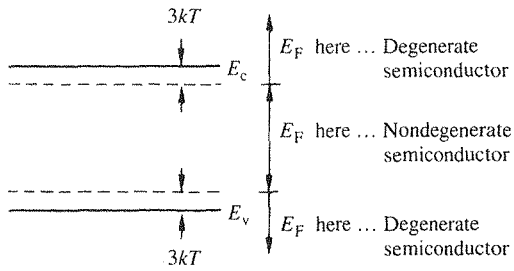


Figure 2.19 Definition of degenerate/nondegenerate semiconductors.

It therefore follows that, if $E_v + 3kT \leq E_F \leq E_c - 3kT$,

$$n = N_C e^{(E_F - E_c)/kT} \tag{2.16a}$$

$$p = N_V e^{(E_v - E_F)/kT} \tag{2.16b}$$

The mathematical simplification leading to Eqs. (2.16) is equivalent to approximating the occupancy factors, $f(E)$ and $1 - f(E)$, by simple exponential functions—an approximation earlier shown to be valid provided E_F was somewhere in the band gap no closer than $3kT$ to either band edge. Whenever E_F is confined, as noted, to $E_v + 3kT \leq E_F \leq E_c - 3kT$, instead of continually repeating the E_F restriction, the semiconductor is simply said to be *nondegenerate*. Whenever E_F lies in the band gap closer than $3kT$ to either band edge or actually penetrates one of the bands, the semiconductor is said to be *degenerate*. These very important terms are also defined pictorially in Fig. 2.19.

2.5.2 Alternative Expressions for *n* and *p*

Although in closed form, the Eq. (2.16) relationships are not in the simplest form possible, and, more often than not, it is the simpler alternative form of these relationships that one encounters in device analyses. The alternative-form relationships can be obtained by recalling that E_i, the Fermi level for an intrinsic semiconductor, lies close to midgap, and hence Eqs. (2.16) most assuredly apply to an intrinsic semiconductor. If this be the case, then specializing Eqs. (2.16) to an intrinsic semiconductor, i.e., setting $n = p = n_i$ and $E_i = E_F$, one obtains

$$n_i = N_C e^{(E_i - E_c)/kT} \tag{2.17a}$$

and

$$n_i = N_V e^{(E_v - E_i)/kT} \tag{2.17b}$$

Solving Eqs. (2.17) for N_C and N_V yields

$$N_C = n_i e^{(E_c - E_i)/kT} \tag{2.18a}$$

$$N_V = n_i e^{(E_i - E_v)/kT} \tag{2.18b}$$

Finally, eliminating N_C and N_V in the original Eq. (2.16) relationships using Eqs. (2.18) gives

$$\boxed{\begin{aligned} n &= n_i e^{(E_F - E_i)/kT} \\ p &= n_i e^{(E_i - E_F)/kT} \end{aligned}} \qquad \begin{aligned} &(2.19a) \\ &(2.19b) \end{aligned}$$

Like Eqs. (2.16), the Eq. (2.19) expressions are valid for any semiconductor in equilibrium whose doping is such as to give rise to a nondegenerate positioning of the Fermi level. Whereas two constants and three energy levels appear in the original relationships, however, only one constant and two energy levels appear in the alternative relationships. Because of their symmetrical nature, the alternative expressions are also easier to remember, requiring merely an interchange of E_F and E_i in going from the n-formula to the p-formula.

2.5.3 n_i and the np Product

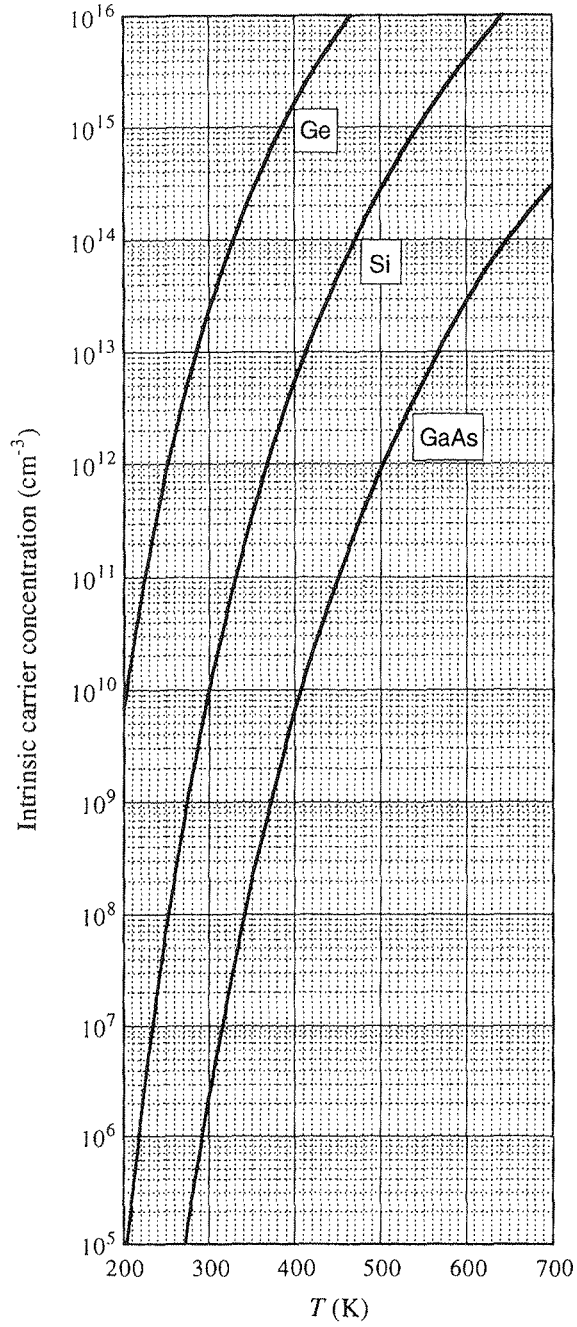
As can be inferred from its appearance in Eqs. (2.19), the intrinsic carrier concentration can figure prominently in the quantitative calculation of the carrier concentrations. Continuing to establish pertinent carrier concentration relationships, we next interject considerations specifically involving this important material parameter.

First, if the corresponding sides of Eqs. (2.17a) and (2.17b) are multiplied together, one obtains

$$n_i^2 = N_C N_V e^{-(E_c - E_v)/kT} = N_C N_V e^{-E_G/kT} \tag{2.20}$$

or

$$\boxed{n_i = \sqrt{N_C N_V}\, e^{-E_G/2kT}} \tag{2.21}$$



Si

T(°C)	n_i(cm^{-3})
0	8.86×10^8
5	1.44×10^9
10	2.30×10^9
15	3.62×10^9
20	5.62×10^9
25	8.60×10^9
30	1.30×10^{10}
35	1.93×10^{10}
40	2.85×10^{10}
45	4.15×10^{10}
50	5.97×10^{10}
300 K	1.00×10^{10}

GaAs

T(°C)	n_i(cm^{-3})
0	1.02×10^5
5	1.89×10^5
10	3.45×10^5
15	6.15×10^5
20	1.08×10^6
25	1.85×10^6
30	3.13×10^6
35	5.20×10^6
40	8.51×10^6
45	1.37×10^7
50	2.18×10^7
300 K	2.25×10^6

Figure 2.20 Intrinsic carrier concentrations in Ge, Si, and GaAs as a function of temperature.

Equation (2.21) expresses n_i as a function of known quantities and can be used to compute n_i at a specified temperature or as a function of temperature. Numerical values for the intrinsic carrier concentration in Si and Ge at room temperature were cited previously; the best available plots of n_i as functions of temperature in Si, Ge, and GaAs are displayed in Fig. 2.20.

A second very important n_i-based relationship follows directly from Eqs. (2.19). Multiplying the corresponding sides of Eqs. (2.19a) and (2.19b) together yields

$$\boxed{np = n_i^2} \tag{2.22}$$

Although appearing trivial, the np product relationship (Eq. 2.22) often proves to be extremely useful in practical computations. If one of the carrier concentrations is known, the remaining carrier concentration is readily determined using Eq. (2.22)—provided of course that the semiconductor is in equilibrium and nondegenerate.

(C) Exercise 2.4

Substituting the Eq. (2.13) definitions of N_C and N_V into Eq. (2.21), introducing normalization factors, and replacing known constants by their numerical values, one obtains

$$n_i = (2.510 \times 10^{19})\left(\frac{m_n^*}{m_0}\frac{m_p^*}{m_0}\right)^{3/4}\left(\frac{T}{300}\right)^{3/2} e^{-E_G/2kT}$$

E_G and the effective masses exhibit a weak but non-negligible temperature dependence. The E_G versus T variation can be modeled to four-place accuracy by the fit relationship noted in Problem 2.1(a). As deduced from the analysis by Barber [Solid-State Electronics, **10,** 1039 (1967)], the temperature dependence of the effective masses over the range 200 K $\leq T \leq$ 700 K can be approximated by

$$\frac{m_n^*}{m_0} = 1.028 + (6.11 \times 10^{-4})T - (3.09 \times 10^{-7})T^2$$

$$\frac{m_p^*}{m_0} = 0.610 + (7.83 \times 10^{-4})T - (4.46 \times 10^{-7})T^2$$

P: (a) Confirm that the n_i versus T curve for Si graphed in Fig. 2.20 is generated employing the relationships just cited, *provided* E_G in the n_i expression is replaced by $E_G - E_{ex}$, where $E_{ex} = 0.0074$ eV. (In the previously cited article, Barber suggested using an exciton correction factor of $E_{ex} = 0.007$ eV. The slightly larger value employed in the Fig. 2.20 computation was specifically chosen to give $n_i = 10^{10}/\text{cm}^3$ at 300 K.)

(b) The accepted value of n_i in Si at 300 K has been revised recently to be in agreement with the experimental n_i versus T data acquired by Sproul and Green [Journal of Applied Physics, **70,** 846 (July 1991)]. The authors concluded $n_i = 1.00 \pm 0.03 \times 10^{10}/\text{cm}^3$ in Si at 300 K. Over the probed temperature range of $275\ \text{K} \leq T \leq 375\ \text{K}$, their experimental data could be fitted by the relationship

$$n_i = (9.15 \times 10^{19})\left(\frac{T}{300}\right)^2 e^{-0.5928/kT}$$

Compare the n_i computed from this experimental-fit relationship to the part (a) output over the temperature range of mutual validity.

S: (a) MATLAB program script . . .

```
%ni vs. T calculation for Si (200K - 700K) used in Fig. 2.20

%Initialization
format short e

%Constants and T-range
k=8.617e-5;
A=2.510e19;
Eex=0.0074;    %Value was adjusted to match S&G ni(300K) value
T=200:25:700;

%Band Gap vs. T
EG0=1.17;
a=4.730e-4;
b=636;
EG=EG0-a.*(T.^2)./(T+b);

%Effective mass ratio (mnr=mn*/m0, mpr=mp*/m0)
mnr=1.028 + (6.11e-4).*T - (3.09e-7).*T.^2;
mpr=0.610 + (7.83e-4).*T - (4.46e-7).*T.^2;

%Computation of ni
ni=A.*((T./300).^(1.5)).*((mnr.*mpr).^(0.75)).*exp(-(EG-Eex)./(2 .*k.*T));

%Display output on screen
j=length(T);
fprintf('\n \n T        ni\n');   %There are ten spaces between T and ni.
for ii=1:j,
fprintf('%-10.f%-10.3e\n',T(ii),ni(ii));
end
```

(b) MATLAB program script . . .

```
%Experimental fit of Sproul-Green ni data (275K - 375K)

%ni calculation
T=275:25:375;
k=8.617e-5;
ni=(9.15e19).*(T./300).^2 .*exp(-0.5928./(k*T));

%Display result on screen
j=length(T);
fprintf('\n\n T       ni\n');   %There are ten spaces between T and ni.
for ii=1:j,
fprintf('%-10.f%-10.3e\n',T(ii),ni(ii));
end
```

Output from the two programs is reproduced below. Note the excellent agreement of the computational results (to within 2%) over the temperature range of mutual validity.

T (K)	(*a*) n_i (cm^{-3})	(*b*) n_i (cm^{-3})
200	5.246×10^{4}	—
275	1.059×10^{9}	1.051×10^{9}
300	1.000×10^{10}	1.006×10^{10}
325	6.798×10^{10}	6.887×10^{10}
350	3.565×10^{11}	3.623×10^{11}
375	1.518×10^{12}	1.542×10^{12}
400	5.449×10^{12}	—
500	2.716×10^{14}	—
600	3.988×10^{15}	—
700	2.865×10^{16}	—

2.5.4 Charge Neutrality Relationship

The relationships established to this point are devoid of an explicit dependence on the dopant concentrations introduced into a semiconductor. It is the charge neutrality relationship that provides the general tie between the carrier and dopant concentrations.

To establish the charge neutrality relationship, let us consider a *uniformly doped* semiconductor, a semiconductor where the number of dopant atoms/cm^3 is the same everywhere. Systematically examining little sections of the semiconductor far from any surfaces, and assuming equilibrium conditions prevail, one must invariably find that each and every section is charge-neutral, i.e., contains no net charge. If this were not the case, electric fields

would exist inside the semiconductor. The electric fields in turn would give rise to carrier motion and associated currents—a situation totally inconsistent with the assumed equilibrium conditions. There are, however, charged entities inside all semiconductors. Electrons, holes, ionized donors (donor atoms that have become positively charged by donating an electron to the conduction band) and negatively-charged ionized acceptors can all exist simultaneously inside any given semiconductor. For the uniformly doped material to be everywhere charge-neutral clearly requires

$$\frac{\text{charge}}{\text{cm}^3} = qp - qn + qN_D^+ - qN_A^- = 0 \tag{2.23}$$

or

$$\boxed{p - n + N_D^+ - N_A^- = 0} \tag{2.24}$$

where, by definition,

N_D^+ = number of ionized (positively charged) donors/cm^3,

N_A^- = number of ionized (negatively charged) acceptors/cm^3.

As previously discussed, there is sufficient thermal energy available in a semiconductor at room temperature to ionize almost all of the shallow-level donor and acceptor sites. Defining

N_D = total number of donors/cm^3,

N_A = total number of acceptors/cm^3,

and setting

$$N_D^+ = N_D$$
$$N_A^- = N_A$$

one then obtains

$$\boxed{p - n + N_D - N_A = 0} \quad \text{assumes total ionization of dopant atoms} \tag{2.25}$$

Equation (2.25) is the standard form of the charge neutrality relationship.

2.5.5 Carrier Concentration Calculations

We are finally in a position to calculate the carrier concentrations in a uniformly doped semiconductor under equilibrium conditions. In the computations to be presented we specifically make the assumptions of NONDEGENERACY (allowing us to use the *np* product relationship) and TOTAL IONIZATION of the dopant atoms. Note that n_i, which appears in the *np* product expression, has been calculated and plotted and must be considered a known quantity. Likewise, N_A and N_D, which appear in the charge neutrality relationship, are typically controlled and determined experimentally and should also be considered known quantities. The only other symbols used in the two equations are n and p. Thus, under the cited assumptions of nondegeneracy and total ionization of dopant atoms, we have two equations and two unknowns from which n and p can be deduced.

Starting with the *np* product expression, one can write

$$p = \frac{n_i^2}{n} \tag{2.26}$$

Eliminating p in Eq. (2.25) using Eq. (2.26) gives

$$\frac{n_i^2}{n} - n + N_D - N_A = 0 \tag{2.27}$$

or

$$n^2 - n(N_D - N_A) - n_i^2 = 0 \tag{2.28}$$

Solving the quadratic equation for n then yields

$$\boxed{n = \frac{N_D - N_A}{2} + \left[\left(\frac{N_D - N_A}{2}\right)^2 + n_i^2\right]^{1/2}} \tag{2.29a}$$

and

$$\boxed{p = \frac{n_i^2}{n} = \frac{N_A - N_D}{2} + \left[\left(\frac{N_A - N_D}{2}\right)^2 + n_i^2\right]^{1/2}} \tag{2.29b}$$

Only the plus roots have been retained in Eqs. (2.29) because physically the carrier concentrations must be greater than or equal to zero.

Equations (2.29) are general-case solutions. In the vast majority of practical computations it is possible to simplify these equations prior to substituting in numerical values for N_D, N_A, and n_i. Special cases of specific interest are considered next.

(1) *Intrinsic Semiconductor* ($N_A = 0$, $N_D = 0$). With $N_A = 0$ and $N_D = 0$, Eqs. (2.29) simplify to $\boxed{n = n_i}$ and $\boxed{p = n_i}$. $n = p = n_i$ is of course the expected result for the equilibrium carrier concentrations in an intrinsic semiconductor.

(2) *Doped Semiconductor where either* $N_D - N_A \simeq N_D \gg n_i$ *or* $N_A - N_D \simeq N_A \gg n_i$. This is the special case of greatest practical interest. The unintentional doping levels in Si are such that the controlled addition of dopants routinely yields $N_D \gg N_A$ or $N_A \gg N_D$. Moreover, the intrinsic carrier concentration in Si at room temperature is about $10^{10}/\text{cm}^3$, while the dominant doping concentration (N_A or N_D) is seldom less than $10^{14}/\text{cm}^3$. Thus the special case considered here is the usual case encountered in practice. If $N_D - N_A \simeq N_D \gg n_i$, the square root in Eq. (2.29a) reduces to $N_D/2$ and

$$\boxed{\begin{aligned} n &\simeq N_D \\ p &\simeq n_i^2/N_D \end{aligned}} \qquad \begin{aligned} &N_D \gg N_A, \quad N_D \gg n_i \\ &\text{(nondegenerate, total ionization)} \end{aligned} \qquad \begin{aligned} &(2.30a) \\ &(2.30b) \end{aligned}$$

Similarly

$$\boxed{\begin{aligned} p &\simeq N_A \\ n &\simeq n_i^2/N_A \end{aligned}} \qquad \begin{aligned} &N_A \gg N_D, \quad N_A \gg n_i \\ &\text{(nondegenerate, total ionization)} \end{aligned} \qquad \begin{aligned} &(2.31a) \\ &(2.31b) \end{aligned}$$

As a numerical example, suppose a Si sample maintained at room temperature is uniformly doped with $N_D = 10^{15}/\text{cm}^3$ donors. Using Eqs. (2.30), one rapidly concludes $n \simeq 10^{15}/\text{cm}^3$ and $p \simeq 10^5/\text{cm}^3$.

(3) *Doped Semiconductor where* $n_i \gg |N_D - N_A|$. Systematically increasing the ambient temperature causes a monotonic rise in the intrinsic carrier concentration (see Fig. 2.20). At sufficiently elevated temperatures, n_i will eventually equal and then exceed the net doping concentration. If $n_i \gg |N_D - N_A|$, the square roots in Eqs. (2.29) reduce to n_i and $\boxed{n \simeq p \simeq n_i}$. In other words, *all semiconductors become intrinsic at sufficiently high temperatures where* $n_i \gg |N_D - N_A|$.

(4) *Compensated Semiconductor.* As is evident from Eqs. (2.29), donors and acceptors tend to negate each other. Indeed, it is possible to produce intrinsic-like material by making $N_D - N_A = 0$. In some materials, such as GaAs, N_A may be comparable to N_D in the as-grown crystal. When N_A and N_D are comparable and nonzero, the material is said to be *compensated.* If the semiconductor is compensated, both N_A and N_D must be retained in all carrier concentration expressions.

In summary, Eqs. (2.29) can always be used to compute the carrier concentrations if the semiconductor is nondegenerate and the dopant atoms are totally ionized. In the vast majority of practical situations, however, it is possible to simplify these equations prior to performing numerical computations. Equations (2.29) must be used to compute the carrier concentrations only in those rare instances when $|N_D - N_A| \sim n_i$. The simplified relationships of greatest practical utility are Eqs. (2.30) and (2.31).

Exercise 2.5

P: A Si sample is doped with 10^{14} boron atoms per cm^3.

(a) What are the carrier concentrations in the Si sample at 300 K?

(b) What are the carrier concentrations at 470 K?

S: (a) Boron in Si is an acceptor (see Table 2.2). Thus $N_A = 10^{14}/cm^3$. At 300 K, $n_i = 1.00 \times 10^{10}/cm^3$ and the given N_A is clearly much greater than n_i. Moreover, since the N_D doping was omitted from the problem statement, we infer $N_D \ll N_A$. With $N_A \gg n_i$ and $N_A \gg N_D$, Eqs. (2.31) may be used to calculate the carrier concentrations: $p = N_A = 10^{14}/cm^3$; $n = n_i^2/N_A = 10^6/cm^3$.

(b) As deduced from Fig. 2.20, $n_i \simeq 10^{14}/cm^3$ at 470 K. Because n_i is comparable to N_A, Eqs. (2.29) must be used to calculate at least one of the carrier concentrations. (Once one of the carrier concentrations is known, the second is more readily computed using the np product expression.) Performing the indicated calculations gives $p = N_A/2 + [(N_A/2)^2 + n_i^2]^{1/2} = 1.62 \times 10^{14}/cm^3$; $n = n_i^2/p = 6.18 \times 10^{13}/cm^3$.

2.5.6 Determination of E_F

Knowledge concerning the exact position of the Fermi level on the energy band diagram is often of interest. For example, when discussing the intrinsic Fermi level, we indicated that E_i was located somewhere near the middle of the band gap. It would be useful to know the *precise* positioning of E_i in the band gap. Moreover, we have developed computational formulas for n and p appropriate for nondegenerate semiconductors. Whether a doped semiconductor is nondegenerate or degenerate depends, of course, on the value or positioning of E_F.

Before running though the mechanics of finding the Fermi level in selected cases of interest, it is useful to make a general observation; namely, Eqs. (2.19) or (2.16) [or even more generally, Eqs. (2.14)] provide a one-to-one correspondence between the Fermi energy and the carrier concentrations. Thus, having computed any one of the three variables—n, p, or E_F—one can always determine the remaining two variables under equilibrium conditions.

(1) *Exact positioning of* E_i. In an intrinsic material

$$n = p \tag{2.32}$$

Substituting for n and p in Eq. (2.32) using Eqs. (2.16), and setting $E_F = E_i$, yields

$$N_C e^{(E_i - E_c)/kT} = N_V e^{(E_v - E_i)/kT} \tag{2.33}$$

Solving for E_i, one obtains

$$E_i = \frac{E_c + E_v}{2} + \frac{kT}{2} \ln\left(\frac{N_V}{N_C}\right) \tag{2.34}$$

But

$$\frac{N_V}{N_C} = \left(\frac{m_p^*}{m_n^*}\right)^{3/2} \tag{2.35}$$

Consequently,

$$\boxed{E_i = \frac{E_c + E_v}{2} + \frac{3}{4} kT \ln\left(\frac{m_p^*}{m_n^*}\right)} \tag{2.36}$$

According to Eq. (2.36), E_i lies precisely at midgap only if $m_p^* = m_n^*$ or if $T = 0$ K. For the more practical case of silicon at room temperature, Table 2.1 gives $m_p^*/m_n^* = 0.69$, $(3/4)kT \ln(m_p^*/m_n^*) = -0.0073$ eV, and E_i therefore lies 0.0073 eV below midgap. Although potentially significant in certain problems, this small deviation from midgap is typically neglected in drawing energy band diagrams, etc.

(2) *Doped semiconductors (nondegenerate, dopants totally ionized).* The general positioning of the Fermi level in donor- and acceptor-doped semiconductors assumed to be nondegenerate, in equilibrium, and maintained at temperatures where the dopants are fully ionized, is readily deduced from Eqs. (2.19). Specifically, solving Eqs. (2.19) for $E_F - E_i$, one obtains

$$E_F - E_i = kT \ln(n/n_i) = -kT \ln(p/n_i) \tag{2.37}$$

Depending on the simplifications inherent in a particular problem, the appropriate carrier concentration solution [Eqs. (2.29), (2.30), (2.31), etc.] is then substituted into Eq. (2.37) to determine the positioning of E_F. For example, per Eqs. (2.30a) and (2.31a), $n \simeq N_D$ in typical donor-doped semiconductors and $p \simeq N_A$ in typical accep-

tor-doped semiconductors maintained at or near room temperature. Substituting into Eq. (2.37), we therefore conclude

$$\boxed{\begin{aligned} E_F - E_i &= kT \ln(N_D/n_i) \\ E_i - E_F &= kT \ln(N_A/n_i) \end{aligned}} \qquad \begin{aligned} &\ldots N_D \gg N_A, \quad N_D \gg n_i \qquad (2.38a) \\ &\ldots N_A \gg N_D, \quad N_A \gg n_i \qquad (2.38b) \end{aligned}$$

From Eqs. (2.38) it is obvious that the Fermi level moves systematically upward in energy from E_i with increasing donor doping and systematically downward in energy from E_i with increasing acceptor doping. The exact Fermi level positioning in Si at room temperature, nicely reinforcing the foregoing statement, is displayed in Fig. 2.21. Also note that for a given semiconductor material and ambient temperature, there exist maximum nondegenerate donor and acceptor concentrations, doping concentrations above which the material becomes degenerate. In Si at room temperature the maximum nondegenerate doping concentrations are $N_D \simeq 1.6 \times 10^{18}/\text{cm}^3$ and $N_A \simeq 9.1 \times 10^{17}/\text{cm}^3$. The large Si doping values required for degeneracy, we should interject, have led to the common usage of "highly doped" (or n^+-material/p^+-material) and "degenerate" as interchangeable terms.

Finally, the question may arise: What procedure should be followed in computing E_F when one is not sure whether a material is nondegenerate or degenerate? Unless a material is known to be degenerate, always assume nondegeneracy and compute E_F employing the appropriate nondegenerate relationship. If E_F derived from the nondegenerate formula lies in the degenerate zone, one must then, of course, recompute E_F using the more complex formalism valid for degenerate materials.

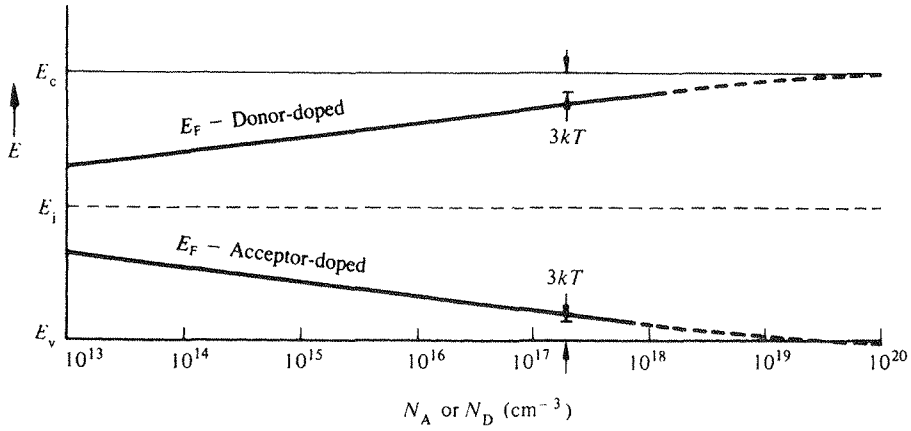


Figure 2.21 Fermi level positioning in Si at 300 K as a function of the doping concentration. The solid E_F lines were established using Eq. (2.38a) for donor-doped material and Eq. (2.38b) for acceptor-doped material ($kT = 0.0259$ eV, and $n_i = 10^{10}/\text{cm}^3$).

Exercise 2.6

P: For each of the conditions specified in Exercise 2.5, determine the position of E_i, compute $E_F - E_i$, and draw a carefully dimensioned energy band diagram for the Si sample. Note that $E_G(\text{Si}) = 1.08$ eV and $m_p^*/m_n^* = 0.71$ at 470 K.

S: (a) In part (a) of Exercise 2.5, the $N_A = 10^{14}/\text{cm}^3$ Si sample is mainted at 300 K. Using Eq. (2.36), we conclude E_i is located 0.0073 eV below midgap. (The positioning of E_i in Si at 300 K is also noted in the text following Eq. 2.36). Next applying Eq. (2.38b), we find

$$\begin{aligned} E_i - E_F &= kT \ln(N_A/n_i) \\ &= 0.0259 \ln(10^{14}/10^{10}) \\ &= 0.239 \text{ eV} \end{aligned}$$

The energy band diagram constructed from the deduced positioning of E_i and E_F is shown in Fig. E2.6(a).

(b) In part (b) of Exercise 2.5 the Si sample is heated to 470 K. With $m_p^*/m_n^* = 0.71$ and $kT = 0.0405$ eV at 470 K, $(3/4)kT \ln(m_p^*/m_n^*) = -0.0104$ eV and E_i is deduced to be located 0.0104 eV below midgap. Because N_A is comparable to n_i at 470 K, Eq. (2.37) must be used to compute the positioning of E_F. Specifically, with $n_i = 10^{14}/\text{cm}^3$ and $p = 1.62 \times 10^{14}/\text{cm}^3$,

$$\begin{aligned} E_i - E_F &= kT \ln(p/n_i) \\ &= 0.0405 \ln(1.62 \times 10^{14}/10^{14}) \\ &= 0.0195 \text{ eV} \end{aligned}$$

Here E_F is only slightly removed from E_i as pictured in Fig. E2.6(b).

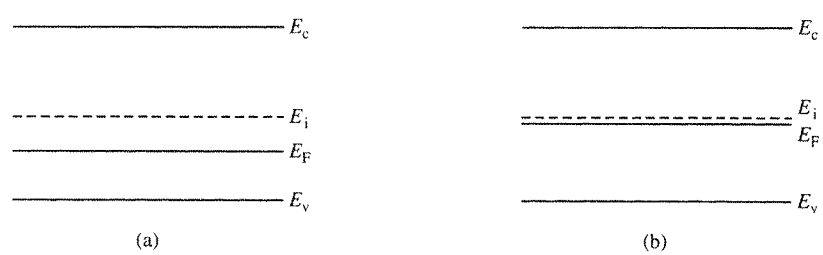


Figure E2.6

2.5.7 Carrier Concentration Temperature Dependence

A number of isolated facts about the carrier concentration temperature dependence have already been presented at various points in this chapter. The discussion in Section 2.3 concerned with dopant action, for example, described the increased ionization of dopant sites and the associated increase in the majority carrier concentration when the temperature of a semiconductor is raised from near $T = 0$ K toward room temperature. More recent subsections have included a plot of the intrinsic carrier concentration versus temperature (Fig. 2.20) and a calculation indicating that all semiconductors become intrinsic ($n \rightarrow n_i$, $p \rightarrow n_i$) at sufficiently high temperatures. In this subsection, which concludes the carrier concentration discussion, temperature-related facts are combined and embellished to provide a broader, more complete description of how the carrier concentrations vary with temperature.

Figure 2.22(a), a typical majority-carrier concentration-versus-temperature plot constructed assuming a phosphorus-doped $N_D = 10^{15}/\text{cm}^3$ Si sample, nicely illustrates the general features of the concentration-versus-temperature dependence. Examining Fig. 2.22(a), we find that n is fixed at approximately N_D over a broad temperature range extending roughly from 150 K to 450 K for the given Si sample. This $n \simeq N_D$ or "extrinsic temperature region" constitutes the normal operating range for most solid-state devices. Below 100 K or so, in the "freeze-out temperature region," n drops significantly below N_D and approaches zero as $T \rightarrow 0$ K. In the "intrinsic temperature region" at the opposite end of the temperature scale, n rises above N_D, asymptotically approaching n_i with increasing T.

To qualitatively explain the just-described concentration-versus-temperature dependence, it is important to recall that the equilibrium number of carriers within a material is affected by two separate mechanisms. Electrons donated to the conduction band from donor atoms and valence band electrons excited across the band gap into the conduction band (broken Si–Si bonds) both contribute to the majority-carrier electron concentration in a donor-doped material. At temperatures $T \rightarrow 0$ K the thermal energy available in the system is insufficient to release the weakly bound fifth electron on donor sites and totally insufficient to excite electrons across the band gap. Hence $n = 0$ at $T = 0$ K, as visualized on the left-hand side of Fig. 2.22(b). Slightly increasing the material temperature above $T = 0$ K "defrosts" or frees some of the electrons weakly bound to donor sites. Band-to-band excitation, however, remains extremely unlikely, and therefore the number of observed electrons in the freeze-out temperature region equals the number of ionized donors—$n = N_D^+$. Continuing to increase the system temperature eventually frees almost all of the weakly bound electrons on donor sites, n approaches N_D, and one enters the extrinsic temperature region. In progressing through the extrinsic temperature region, more and more electrons are excited across the band gap, but the number of electrons supplied in this fashion stays comfortably below N_D. Ultimately, of course, electrons excited across the band gap equal, then exceed, and, as pictured on the right-hand side of Fig. 2.22(b), finally swamp the fixed number of electrons derived from the donors.

As a practical note, it should be pointed out that the wider the band gap, the greater the energy required to excite electrons from the valence band into the conduction band, and

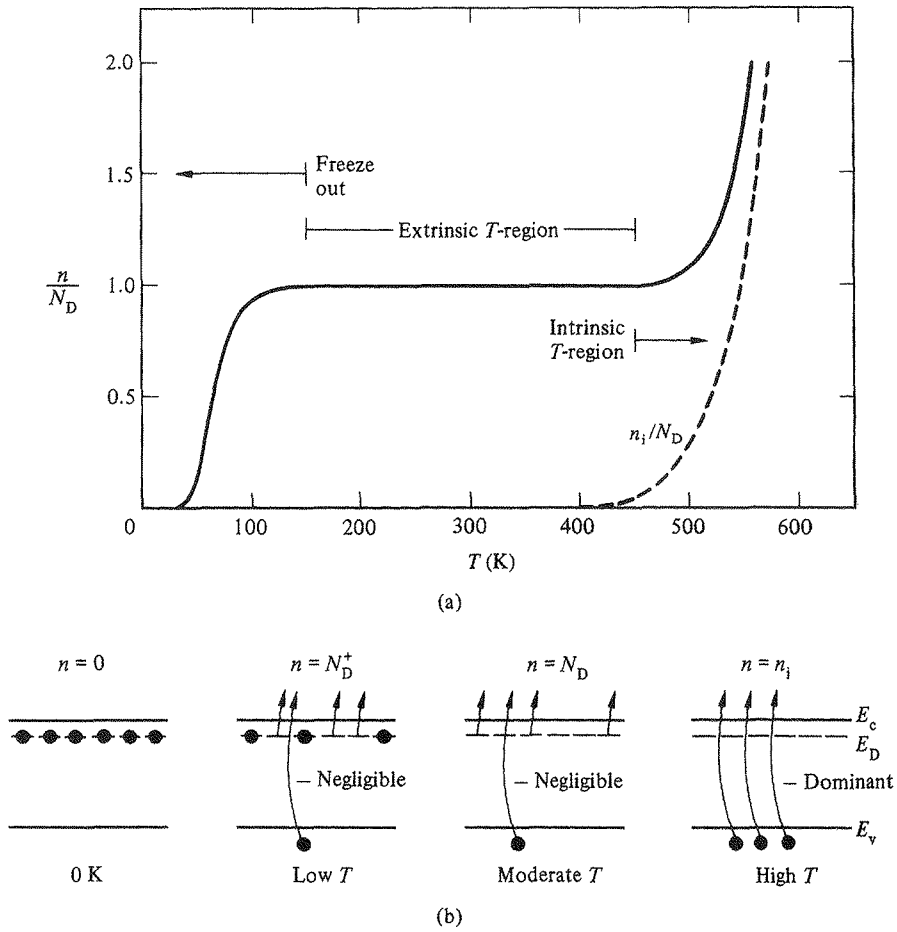


Figure 2.22 (a) Typical temperature dependence of the majority-carrier concentration in a doped semiconductor. The plot was constructed assuming a phosphorus-doped $N_D = 10^{15}/\text{cm}^3$ Si sample. n_i/N_D versus T (dashed line) has been included for comparison purposes. (b) Qualitative explanation of the concentration-versus-temperature dependence displayed in part (a).

the higher the temperature at the onset of the intrinsic temperature region. Since the temperature at the onset of the intrinsic temperature region corresponds to the upper end of the normal operating range for most solid-state devices, GaAs devices can inherently operate at a higher maximum temperature than similarly doped Si devices, which in turn can operate at a higher maximum temperature than similarly doped Ge devices. If we assume, for example, the critical doping concentration is $N_D = 10^{15}/\text{cm}^3$, and the onset of the intrinsic temperature region is approximated as the temperature where $n_i = N_D$, then from Fig. 2.20

the maximum operating temperatures are deduced to be 385 K, 540 K, and > 700 K for Ge, Si, and GaAs, respectively. Indeed, GaAs and SiC ($E_G > 2$ eV) devices continue to be under development for use in high temperature environments.

2.6 SUMMARY AND CONCLUDING COMMENTS

Under the general heading of carrier modeling we have described, examined, and characterized the carriers within a semiconductor under "rest" or equilibrium conditions. The many important topics addressed in this chapter included the introduction of two "visualization" models: the bonding model and the energy band model. The extremely useful energy band model is actually more than just a model—it is a sophisticated sign language providing a concise means of communicating on a nonverbal level. Relative to the carriers themselves, the reader by now has been successfully prodded into thinking of electrons and holes as classical ball-like "particles," where the charge on an electron is $-q$, the charge on a hole is $+q$, and the effective masses of the particles are m_n^* and m_p^*, respectively. The reader should also know that the carrier numbers in an intrinsic material are equal and relatively small; the carrier concentrations, however, can be selectively increased by adding special impurity atoms or dopants to the semiconductor.

In addressing the problem of determining the carrier concentrations in doped semiconductors, we developed or derived a number of useful mathematical relationships. The density of states functions (Eqs. 2.6), the Fermi function (Eq. 2.7), the symmetrical nondegenerate relationships for n and p (Eqs. 2.19), the np product (Eq. 2.22), the charge neutrality relationship (Eq. 2.25), and the simplified n and p expressions appropriate for a typical semiconductor maintained at room temperature (Eqs. 2.30 and 2.31) deserve special mention. The cited equations and a few others are collected in Table 2.4. With regard to the use of these relationships, the reader should be cautioned against "no-think plug and chug." Because semiconductor problems are replete with exceptions, special cases, and nonideal situations, it is imperative that the formula user be aware of derivational assumptions and the validity limits of any and all expressions used in an analysis or computation. In addition to the quantitative carrier relationships, the reader should also have a qualitative "feel" for the carrier distributions in the respective energy bands, the temperature dependence of the intrinsic concentration, and the typical temperature dependence of the majority carrier concentration in a doped semiconductor.

Finally, some attention should be given to the many technical terms and the key parametric values presented in this chapter. The terms extrinsic semiconductor, donor, acceptor, nondegenerate semiconductor, Fermi level, and so on, will be encountered again and again in the discussion of semiconductor devices. Likewise, a knowledge of typical numerical values for key parameters, such as $E_G = 1.12$ eV and $n_i = 10^{10}/\text{cm}^3$ in Si at 300 K, will be invaluable in subsequent work when performing both "back-of-the-envelope" and computer-assisted computations. Key parametric values also serve as yardsticks for gauging whether newly encountered quantities are relatively small or relatively large.

Table 2.4 Carrier Modeling Equation Summary.

Density of States and Fermi Function

$$g_c(E) = \frac{m_n^*\sqrt{2m_n^*(E - E_c)}}{\pi^2\hbar^3}, \quad E \geq E_c$$

$$g_v(E) = \frac{m_p^*\sqrt{2m_p^*(E_v - E)}}{\pi^2\hbar^3}, \quad E \leq E_v$$

$$f(E) = \frac{1}{1 + e^{(E - E_F)/kT}}$$

Carrier Concentration Relationships

$$n = N_C \frac{2}{\sqrt{\pi}} F_{1/2}(\eta_c) \qquad N_C = 2\left[\frac{m_n^* kT}{2\pi\hbar^2}\right]^{3/2}$$

$$p = N_V \frac{2}{\sqrt{\pi}} F_{1/2}(\eta_v) \qquad N_V = 2\left[\frac{m_p^* kT}{2\pi\hbar^2}\right]^{3/2}$$

$$n = N_C e^{(E_F - E_c)/kT}$$

$$p = N_V e^{(E_v - E_F)/kT}$$

$$n = n_i e^{(E_F - E_i)/kT}$$

$$p = n_i e^{(E_i - E_F)/kT}$$

n_i, np-Product, and Charge Neutrality

$$n_i = \sqrt{N_C N_V}\, e^{-E_G/2kT} \qquad np = n_i^2 \qquad p - n + N_D - N_A = 0$$

n, p, and Fermi Level Computational Relationships

$$n = \frac{N_D - N_A}{2} + \left[\left(\frac{N_D - N_A}{2}\right)^2 + n_i^2\right]^{1/2} \qquad E_i = \frac{E_c + E_v}{2} + \frac{3}{4}kT\ln\left(\frac{m_p^*}{m_n^*}\right)$$

$$n \simeq N_D \qquad N_D \gg N_A, N_D \gg n_i \qquad E_F - E_i = kT\ln(n/n_i) = -kT\ln(p/n_i)$$

$$p \simeq n_i^2/N_D$$

$$p \simeq N_A \qquad N_A \gg N_D, N_A \gg n_i \qquad E_F - E_i = kT\ln(N_D/n_i) \quad N_D \gg N_A, N_D \gg n_i$$

$$n \simeq n_i^2/N_A \qquad E_i - E_F = kT\ln(N_A/n_i) \quad N_A \gg N_D, N_A \gg n_i$$

PROBLEMS

CHAPTER 2 PROBLEM INFORMATION TABLE				
Problem	*Complete After*	*Difficulty Level*	*Suggested Point Weighting*	*Short Description*
● 2.1	2.2.4	2	10 (5 each part)	E_G vs. T computation
2.2	2.3.4	1	10 (2 each part)	Bonding model applications
2.3	2.3.4 (a–d) 2.5.7 all	1	24 (2 each part)	E-band model applications
2.4	2.3.4	2	10 (2 each part)	Si-doped GaAs
2.5	2.4.1	2	5	States/cm^3 in ΔE
2.6	2.4.2	2	8 (a-2, b-3, c-3)	Fermi function questions
2.7	2.5.1	2–3	10	Find distribution peak
2.8	"	2	5	Population at $E_c+\Delta E$
● 2.9	"	2	15	Plot distributions in bands
● 2.10	"	3	20 (a-3, b-17..discuss-2)	T variation of distribution
2.11	"	2	10	Derive Eqs. (2.14b), (2.16b)
2.12	"	3	15 (a-3, b-12)	Hypothetical g_c = constant
2.13	"	2	10 (a-7, b-3)	Compute N_C, N_V
2.14	2.5.3	1	8 (4 each part)	n_i comparison
● 2.15	"	1	5	Plot n_i vs. T for Ge
2.16	2.5.5	2	12 (a::d-2, e-4)	Tricky conc. questions
2.17	"	1–2	10 (2 each part)	Compute n and p
2.18	2.5.6	2	15 (3 each part)	Determine E_i, $E_F - E_i$, etc.
● 2.19	"	2	15 (a-12, b-3)	Computer check P2.17/2.18
2.20	"	2	8	Verify nondegenerate limits
● 2.21	"	2	10	Plot $E_F - E_i$ vs. N_A, N_D
2.22	"	2–3	12 (a-3, b-3, c-2, d-4)	GaAs considerations
● 2.23	2.5.7	4	25 (a-8, b-10, c-2, d-3, e-2)	E_F variation with T

● **2.1** E_G versus T Computation

With increasing temperature an expansion of the crystal lattice usually leads to a weakening of the interatomic bonds and an associated decrease in the band gap energy. For many semiconductors the cited variation of the band gap energy with temperature can be modeled by the empirical relationship

$$E_G(T) = E_G(0) - \frac{\alpha T^2}{(T + \beta)}$$

where α and β are constants chosen to obtain the best fit to experimental data and $E_G(0)$ is the limiting value of the band gap at 0 K. As far as Si is concerned, a fit accurate to four places is obtained by employing

$$\begin{aligned} E_G(0) &= 1.170 \text{ eV} \\ \alpha &= 4.730 \times 10^{-4} \text{ eV/K} \\ \beta &= 636 \text{ K} \\ T &\text{ in Kelvin} \end{aligned}$$

(a) Make a plot of E_G versus T for Si spanning the temperature range from $T = 0$ K to $T = 600$ K. Specifically note the value of E_G at 300 K.

(b) For $T > 300$ K, the temperature variation is nearly linear. Noting this fact, some authors have employed

$$E_G(T) = 1.205 - 2.8 \times 10^{-4} T \qquad \ldots T > 300 \text{ K}$$

How does this simplified relationship compare with the more precise relationship over the temperature region of mutual validity?

2.2 Using the *bonding* model for a semiconductor, indicate how one visualizes (a) a missing atom, (b) an electron, (c) a hole, (d) a donor, (e) an acceptor.

2.3 Using the *energy band* model for a semiconductor, indicate how one visualizes (a) an electron, (b) a hole, (c) donor sites, (d) acceptor sites, (e) freeze-out of majority carrier electrons at donor sites as the temperature is lowered toward 0 K, (f) freeze-out of majority carrier holes at acceptor sites as the temperature is lowered toward 0 K, (g) the energy distribution of carriers in the respective bands, (h) an intrinsic semiconductor, (i) an n-type semiconductor, (j) a p-type semiconductor, (k) a nondegenerate semiconductor, (l) a degenerate semiconductor.

2.4 The bonding model for GaAs is pictured in Fig. P2.4.

(a) Draw the bonding model for GaAs depicting the removal of the shaded Ga and As atoms in Fig. P2.4. HINT: Ga and As take their bonding electrons with them when they are removed from the lattice. Also see Fig. 2.4(a) showing the results of removing an atom from the Si lattice.

(b) Redraw the bonding model for GaAs showing the insertion of Si atoms into the missing Ga and As atom sites.

(c) Is the GaAs doped p- or n-type when Si atoms replace Ga atoms? Explain.

(d) Is the GaAs doped p- or n-type when Si atoms replace As atoms? Explain.

(e) Draw the *energy band diagram* for GaAs when the GaAs is doped with Si on (i) Ga sites and (ii) on As sites.

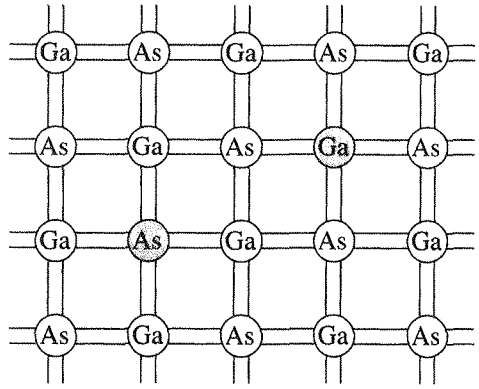


Figure P2.4

2.5 Develop an expression for the total number of available STATES/cm^3 in the conduction band between energies E_c and $E_c + \gamma kT$, where γ is an arbitrary constant.

2.6 (a) Under equilibrium conditions and $T > 0$ K, what is the probability of an electron state being occupied if it is located at the Fermi level?

(b) If E_F is positioned at E_c, determine (numerical answer required) the probability of finding electrons in states at $E_c + kT$.

(c) The probability a state is filled at $E_c + kT$ is equal to the probability a state is empty at $E_c + kT$. Where is the Fermi level located?

2.7 The carrier distributions or numbers of carriers as a function of energy in the conduction and valence bands were noted to peak at an energy very close to the band edges. (See the carrier distribution sketches in Fig. 2.16.) Taking the semiconductor to be nondegenerate, show that the energy at which the carrier distributions peak is $E_c + kT/2$ and $E_v - kT/2$ for the conduction and valence bands, respectively.

2.8 For a nondegenerate semiconductor, the peak in the electron distribution versus energy inside the conduction band noted in Fig. 2.16 occurs at $E_c + kT/2$. Expressed as a fraction of the electron population at the peak energy, what is the electron population in a nondegenerate semiconductor at $E = E_c + 5kT$?

• **2.9** The Fermi level in a Si sample maintained at $T = 300$ K is located at $E_c - E_G/4$. Compute and plot the electron and hole distributions (numbers/cm^3–eV) as a function of energy in the conduction and valence bands, respectively.

• **2.10** Let us investigate how the electron energy distribution in the conduction band varies as a function of temperature.

(a) Assuming the semiconductor to be nondegenerate and employing the Eq. (2.16a) expression for n, confirm that the electron distribution in the conduction band normalized to the total electron concentration is given by

$$\frac{g_c(E)f(E)}{n} = \frac{2\sqrt{E - E_c}}{\sqrt{\pi}\,(kT)^{3/2}}\, e^{-(E-E_c)/kT}$$

(b) Compute and plot the normalized electron distribution in the conduction band versus $E - E_c$ for temperatures $T = 300$ K, 600 K, and 1200 K. Plot the distribution values along the x-axis ($0 \leq g_c(E)f(E)/n \leq 20\ \text{eV}^{-1}$) and $E - E_c$ ($0 \leq E - E_c \leq 0.4$ eV) along the y-axis on a single set of coordinates. Discuss your results.

2.11 Starting with Eq. (2.8b) and following a procedure analogous to that outlined in the text, present the intermediate steps and arguments leading to Eqs. (2.14b) and (2.16b).

2.12 The density of states in the conduction band of a hypothetical semiconductor is

$$g_c(E) = \text{constant} = N_C/kT \qquad \ldots\ E \geq E_c$$

(a) Assuming $E_F < E_c - 3kT$, sketch the electron distribution in the conduction band of the hypothetical semiconductor.

(b) Following the procedure outlined in the text, establish relationships for the electron concentration in the hypothetical semiconductor analogous to Eqs. (2.14a) and (2.16a).

2.13 (a) Verify the statement in Subsection 2.5.1 that, at 300 K,

$$N_{C,V} = (2.510 \times 10^{19}/\text{cm}^3)(m^*/m_0)^{3/2}$$

where one sets $m^* = m_n^*$ in computing N_C and $m^* = m_p^*$ in computing N_V. $m_0 = 9.109 \times 10^{-31}$ kg; $h = 6.625 \times 10^{-34}$ joule-sec; and $q = 1.602 \times 10^{-19}$ coul.

(b) Using the effective mass values recorded in Table 2.1, construct a table that lists the numerical values of N_C and N_V for Si, Ge, and GaAs at 300 K.

2.14 (a) Determine the temperature at which the intrinsic carrier concentration in (i) Si and (ii) GaAs are equal to the room temperature (300 K) intrinsic carrier concentration of Ge.

(b) Semiconductor A has a band gap of 1 eV, while semiconductor B has a band gap of 2 eV. What is the ratio of the intrinsic carrier concentrations in the two materials (n_{iA}/n_{iB}) at 300 K. Assume any differences in the carrier effective masses may be neglected.

• **2.15** Confirm that the n_i versus T curve for Ge graphed in Fig. 2.20 can be generated employing the empirical-fit relationship

$$n_i(\text{Ge}) = (1.76 \times 10^{16})T^{3/2}e^{-0.392/kT}$$

2.16 Concentration questions with a twist.

(a) A silicon wafer is uniformly doped p-type with $N_A = 10^{15}/\text{cm}^3$. At $T \simeq 0$ K, what are the equilibrium hole and electron concentrations?

(b) A semiconductor is doped with an impurity concentration N such that $N \gg n_i$ and all the impurities are ionized. Also, $n = N$ and $p = n_i^2/N$. Is the impurity a donor or an acceptor? Explain.

(c) The electron concentration in a piece of Si maintained at 300 K under equilibrium conditions is $10^5/\text{cm}^3$. What is the hole concentration?

(d) For a silicon sample maintained at $T = 300$ K, the Fermi level is located 0.259 eV above the intrinsic Fermi level. What are the hole and electron concentrations?

(e) In a nondegenerate germanium sample maintained under equilibrium conditions near room temperature, it is known that $n_i = 10^{13}/\text{cm}^3$, $n = 2p$, and $N_A = 0$. Determine n and N_D.

2.17 Determine the equilibrium electron and hole concentrations inside a uniformly doped sample of Si under the following conditions:

(a) $T = 300$ K, $N_A \ll N_D$, $N_D = 10^{15}/\text{cm}^3$.

(b) $T = 300$ K, $N_A = 10^{16}/\text{cm}^3$, $N_D \ll N_A$.

(c) $T = 300$ K, $N_A = 9 \times 10^{15}/\text{cm}^3$, $N_D = 10^{16}/\text{cm}^3$.

(d) $T = 450$ K, $N_A = 0$, $N_D = 10^{14}/\text{cm}^3$.

(e) $T = 650$ K, $N_A = 0$, $N_D = 10^{14}/\text{cm}^3$.

2.18 (a to e) For each of the conditions specified in Problem 2.17, determine the position of E_i, compute $E_F - E_i$, and draw a carefully dimensioned energy band diagram for the Si sample. NOTE: $E_G(\text{Si}) = 1.08$ eV at 450 K and 1.015 eV at 650 K.

• **2.19** (a) Assuming a nondegenerate Si sample and total impurity atom ionization, construct a MATLAB program that computes n, p, and $E_F - E_i$ given acceptable input values of T (temperature in Kelvin), N_D (cm^{-3}), and N_A (cm^{-3}). Incorporate the program presented in part (a) of Exercise 2.4 to compute n_i at a specified T. Make use of the MATLAB `input` function to enter the input variables from the command window.

(b) Use your program to check the relevant answers to Problems 2.17 and 2.18.

2.20 According to the text, the maximum nondegenerate donor and acceptor doping concentrations in Si at room temperature are $N_D \simeq 1.6 \times 10^{18}/\text{cm}^3$ and $N_A \simeq 9.1 \times 10^{17}/\text{cm}^3$, respectively. Verify the text statement.

● **2.21** Construct a computer program to produce a plot of $E_F - E_i$ versus N_A or N_D similar to Fig. 2.21. Use the program to verify the accuracy of the cited figure. (You may find it convenient to employ the MATLAB `logspace` function in constructing your program.)

2.22 GaAs considerations.

(a) Make a sketch similar to Fig. 2.14 that is specifically appropriate for GaAs. Be sure to take into account the fact that $m_n^* \ll m_p^*$ in GaAs.

(b) Based on your answer to part (a), would you expect E_i in GaAs to lie above or below midgap? Explain.

(c) Determine the precise position of the intrinsic Fermi level in GaAs at room temperature (300 K).

(d) Determine the maximum nondegenerate donor and acceptor doping concentrations in GaAs at room temperature.

2.23 Given an $N_A = 10^{14}/\text{cm}^3$ doped Si sample:

(a) Present a *qualitative* argument that leads to the approximate positioning of the Fermi level in the material as $T \rightarrow 0$ K.

● (b) Construct a MATLAB (computer) program to calculate and plot $E_F - E_i$ in the material as a function of temperature for $200 \text{ K} \leq T \leq 500$ K.

(c) What do you conclude relative to the general behavior of the Fermi level positioning as a function of temperature?

(d) Run your part (b) program to determine what happens if N_A is progressively increased in decade steps from $N_A = 10^{14}/\text{cm}^3$ to $N_A = 10^{18}/\text{cm}^3$. Summarize your observations.

(e) How would your foregoing answers be modified if the Si sample was doped with donors instead of acceptors?

3 Carrier Action

Carrier modeling under "rest" or equilibrium conditions, considered in Chapter 2, is important because it establishes the proper frame of reference. From a device standpoint, however, the zero current observed under equilibrium conditions is rather uninteresting. Only when a semiconductor system is perturbed, giving rise to carrier action or a net carrier response, can currents flow within and external to the semiconductor system. Action, carrier action, is the general concern of this chapter.

Under normal operating conditions the three primary types of carrier action occurring inside semiconductors are *drift, diffusion,* and *recombination–generation.* In this chapter we first describe each primary type of carrier action qualitatively and then quantitatively relate the action to the current flowing within the semiconductor. Special emphasis is placed on characterizing the "constants of the motion" associated with each type of action, and wherever appropriate the discussion is extended to subsidiary topics of a relevant nature. Although introduced individually, the various types of carrier action are understood to occur simultaneously inside any given semiconductor. Mathematically combining the various carrier activities next leads to the culmination of our carrier-action efforts; we obtain the basic set of starting equations employed in solving device problems of an electrical nature. Finally, simple example problems are considered to illustrate solution approaches and to introduce key supplemental concepts.

3.1 DRIFT

3.1.1 Definition–Visualization

Drift, by definition, *is charged-particle motion in response to an applied electric field.* Within semiconductors the drifting motion of the carriers on a microscopic scale can be described as follows: When an electric field ($\mathscr{E}$) is applied across a semiconductor as visualized in Fig. 3.1(a), the resulting force on the carriers tends to accelerate the $+q$ charged holes in the direction of the electric field and the $-q$ charged electrons in the direction opposite to the electric field. Because of collisions with ionized impurity atoms and thermally agitated lattice atoms, however, the carrier acceleration is frequently interrupted (the carriers are said to be scattered). The net result, pictured in Fig. 3.1(b), is carrier motion generally along the direction of the electric field, but in a disjointed fashion involving repeated periods of acceleration and subsequent decelerating collisions.

The microscopic drifting motion of a single carrier is obviously complex and quite tedious to analyze in any detail. Fortunately, measurable quantities are *macroscopic*

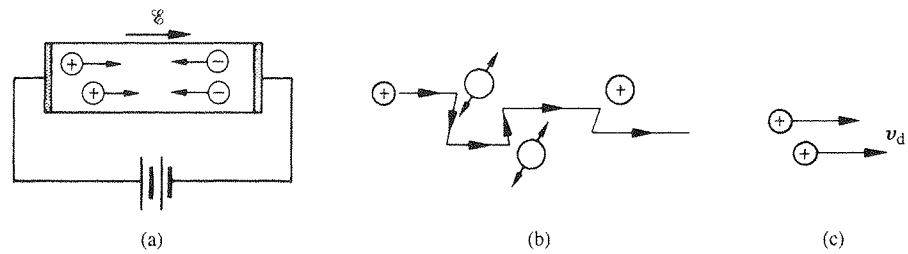


Figure 3.1 Visualization of carrier drift: (a) motion of carriers within a biased semiconductor bar; (b) drifting hole on a microscopic or atomic scale; (c) carrier drift on a macroscopic scale.

observables that reflect the average or overall motion of the carriers. Averaging over all electrons or holes at any given time, we find that the resultant motion of each carrier type can be described in terms of a constant drift velocity, v_d. In other words, on a macroscopic scale, drift may be visualized (see Fig. 3.1c) as nothing more than all carriers of a given type moving along at a constant velocity in a direction parallel or antiparallel to the applied electric field.

By way of clarification, it is important to point out that the drifting motion of the carriers arising in response to an applied electric field is actually superimposed upon the always-present thermal motion of the carriers. Electrons in the conduction band and holes in the valence band gain and lose energy via collisions with the semiconductor lattice and are nowhere near stationary even under equilibrium conditions. In fact, under equilibrium conditions the thermally related carrier velocities average ~1/1000 the speed of light at room temperature! As pictured in Fig. 3.2, however, the thermal motion of the carriers is completely random. Thermal motion therefore averages out to zero on a macroscopic scale, does not contribute to current transport, and can be conceptually neglected.

3.1.2 Drift Current

Let us next turn to the task of developing an analytical expression for the current flowing within a semiconductor as a result of carrier drift. By definition

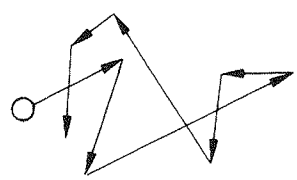


Figure 3.2 Thermal motion of a carrier.

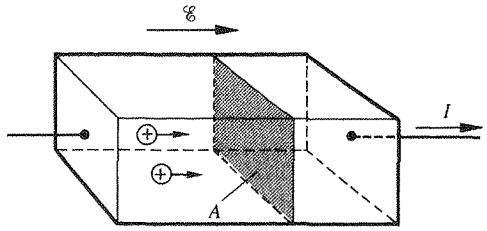


Figure 3.3 Expanded view of a biased p-type semiconductor bar of cross-sectional area A.

I (current) = the charge per unit time crossing an arbitrarily chosen plane of observation oriented normal to the direction of current flow.

Considering the p-type semiconductor bar of cross-sectional area A shown in Fig. 3.3, and specifically noting the arbitrarily chosen v_d-normal plane lying within the bar, we can argue:

$v_d t$	...	All holes this distance back from the v_d-normal plane will cross the plane in a time t.
$v_d tA$	...	All holes in this volume will cross the plane in a time t.
$pv_d tA$	...	Holes crossing the plane in a time t.
$qpv_d tA$	...	Charge crossing the plane in a time t.
$qpv_d A$	...	Charge crossing the plane per unit time.

The word definition of the last quantity is clearly identical to the formal definition of current. Thus

$$I_{\text{P|drift}} = qpv_d A \qquad \text{hole drift current} \tag{3.1}$$

As a practical matter, the cross-sectional area A appearing in Eq. (3.1) and other current formulas is often excess baggage. Current, moreover, is generally thought of as a scalar quantity, while in reality it is obviously a vector. These deficiencies are overcome by introducing a related parameter known as the current density, **J**. **J** has the same orientation as the direction of current flow and is equal in magnitude to the current per unit area (or $J = I/A$). By inspection, the current density associated with hole drift is simply

$$\mathbf{J}_{\text{P|drift}} = qpv_d \tag{3.2}$$

Since the drift current arises in response to an applied electric field, it is reasonable to proceed one step further and seek a form of the current relationship that explicitly relates $\mathbf{J}_{\text{P|drift}}$ to the applied electric field. To this end, we make reference to the representative drift velocity versus electric field data presented in Fig. 3.4. Note that v_d is proportional to $\mathscr{E}$ at

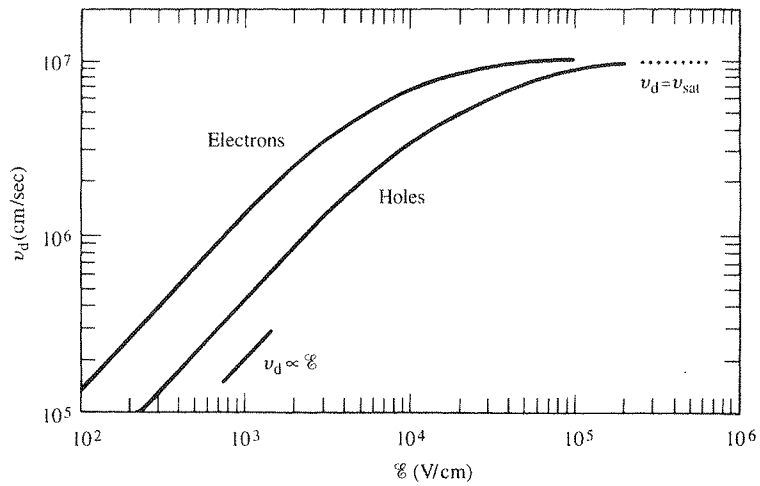


Figure 3.4 Measured drift velocity of the carriers in ultrapure silicon maintained at room temperature as a function of the applied electric field. Constructed from the data fits and the data respectively in Jacoboni et al.[4] and Smith et al.[5]

low electric fields, while at high electric fields v_d saturates and becomes independent of $\mathscr{E}$. To be more precise,

$$v_d = \frac{\mu_0 \mathscr{E}}{\left[1 + \left(\frac{\mu_0 \mathscr{E}}{v_{sat}}\right)^{\beta}\right]^{1/\beta}} = \begin{cases} \mu_0 \mathscr{E} & \ldots \mathscr{E} \to 0 \\ v_{sat} & \ldots \mathscr{E} \to \infty \end{cases} \tag{3.3}$$

where $\beta \cong 1$ for holes and $\beta \cong 2$ for electrons in silicon, μ_0 is the constant of proportionality between v_d and $\mathscr{E}$ at low to moderate electric fields, and v_{sat} is the limiting or *saturation velocity* approached at very high electric fields. Obviously, in the high-field limit v_d in Eq. (3.2) is simply replaced by v_{sat} and $\mathbf{J}_{P|drift}$ does not exhibit an $\mathscr{E}$-field dependence. In the low-field limit, which is of greatest practical interest and assumed herein unless specified otherwise, $v_d = \mu_p \mathscr{E}$ ($\mu_0 \to \mu_p$ for holes) and substitution into Eq. (3.2) yields

$$\boxed{\mathbf{J}_{P|drift} = q\mu_p p \mathscr{E}} \tag{3.4a}$$

Similarly, for electrons one obtains

$$\boxed{\mathbf{J}_{N|drift} = q\mu_n n \mathscr{E}} \tag{3.4b}$$

Respectively known as the *electron mobility* and *hole mobility*, μ_n and μ_p are always taken to be positive quantities. Note that, although electrons drift in the direction opposite to the applied electric field ($\boldsymbol{v}_d = -\mu_n \mathscr{E}$), the current transported by negatively charged particles is in turn counter to the direction of drift ($\mathbf{J}_{N|drift} = -qn\boldsymbol{v}_d$). The net result, as indicated in Eq. (3.4b), is an electron current in the direction of the applied electric field.

3.1.3 Mobility

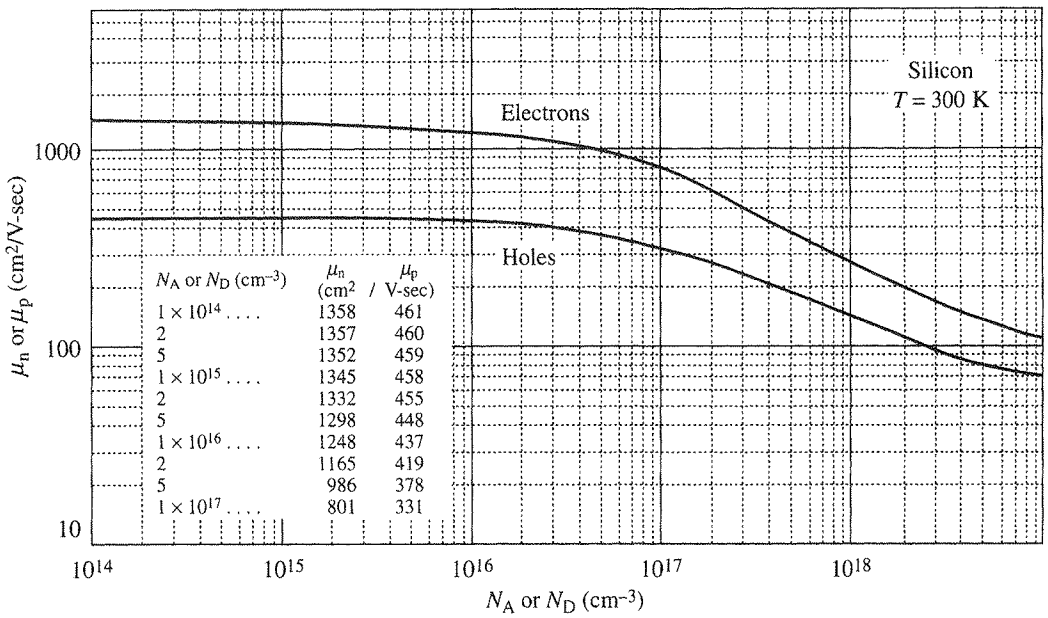
Mobility is obviously a central parameter in characterizing electron and hole transport due to drift. As further readings will reveal, the carrier mobilities also play a key role in characterizing the performance of many devices. It is reasonable therefore to examine μ_n and μ_p in some detail to enhance our general familiarity with the parameters and to establish a core of useful information for future reference.

Standard Units: $cm^2/V\text{-}sec$.

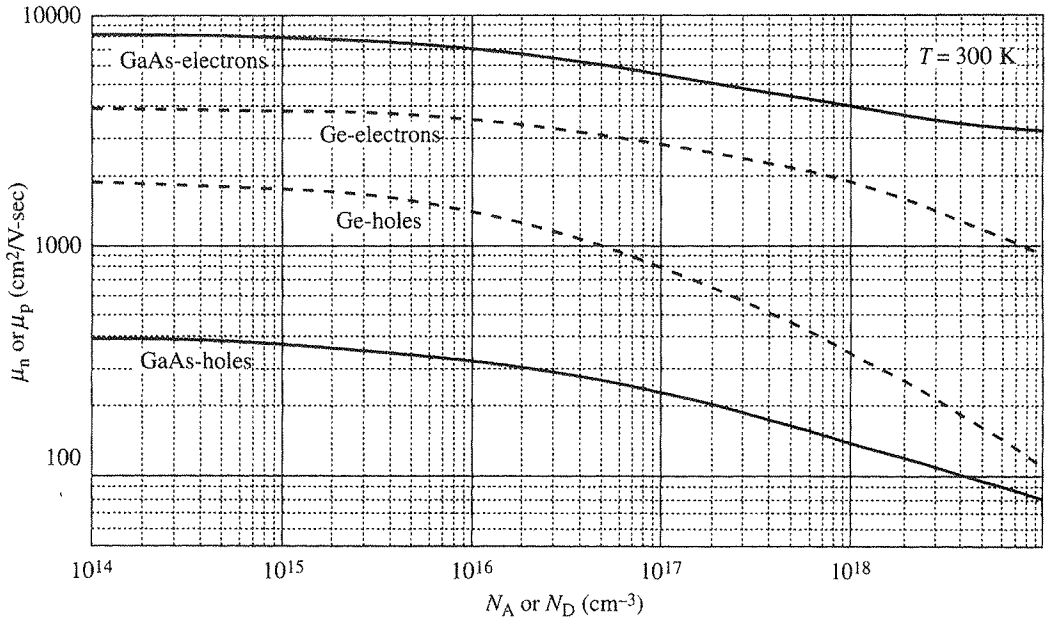
Sample Numerical Values: $\mu_n \simeq 1360\ cm^2/V\text{-}sec$ and $\mu_p \simeq 460\ cm^2/V\text{-}sec$ at 300 K in $N_D = 10^{14}/cm^3$ and $N_A = 10^{14}/cm^3$ doped Si, respectively. In uncompensated high-purity (N_D or $N_A \leq 10^{15}/cm^3$) GaAs, the room-temperature mobilities are $\mu_n \cong 8000\ cm^2/V\text{-}sec$ and $\mu_p \cong 400\ cm^2/V\text{-}sec$. The quoted values are useful for comparison purposes and when performing order-of-magnitude computations. Also note that $\mu_n > \mu_p$ for both Si and GaAs. In the major semiconductors, μ_n is consistently greater than μ_p for a given doping and system temperature.

Relationship to Scattering: The word *mobility* in everyday usage refers to a general freedom of movement. Analogously, in semiconductor work the mobility parameter is a measure of the ease of carrier motion in a crystal. Increasing the motion-impeding collisions within a crystal decreases the mobility of the carriers. In other words, the carrier mobility varies inversely with the amount of scattering taking place within the semiconductor. As visualized in Fig. 3.1(b), the dominant scattering mechanisms in nondegenerately doped materials of device quality are typically (i) lattice scattering involving collisions with thermally agitated lattice atoms, and (ii) ionized impurity (i.e., donor-site and/or acceptor-site) scattering. Relative to lattice scattering, it should be emphasized that it is the thermal vibration, the *displacement* of lattice atoms from their lattice positions, that leads to carrier scattering. The internal field associated with the stationary array of atoms in a crystal is already taken into account in the effective mass formulation.

Although quantitative relationships connecting the mobility and scattering can become quite involved, it is readily established that $\mu = q\langle\tau\rangle/m^*$, where $\langle\tau\rangle$ is the mean free time between collisions and m^* is the conductivity effective mass. Since increasing the number of motion-impeding collisions decreases the mean free time between collisions, we again conclude μ varies inversely with the amount of scattering. However, μ is also noted to vary inversely with the carrier effective mass—lighter carriers move more readily. The m_n^* in GaAs is significantly smaller than the m_n^* in Si, thereby explaining the higher mobility of the GaAs electrons.



(a)



(b)

Figure 3.5 Room temperature carrier mobilities as a function of the dopant concentration in (a) Si and (b) Ge and GaAs. μ_n is the electron mobility; μ_p is the hole mobility.

Doping Dependence: Figure 3.5 exhibits the observed doping dependence of the electron and hole mobilities in Si, Ge, and GaAs. All semiconductors exhibit the same general dependence. At low doping concentrations, below approximately $10^{15}/\text{cm}^3$ in Si, the carrier mobilities are essentially independent of the doping concentration. For dopings in excess of $\sim 10^{15}/\text{cm}^3$, the mobilities monotonically decrease with increasing N_A or N_D.

With the aid of Fig. 3.6 the explanation of the observed doping dependence is relatively straightforward. Invoking an electrical analogy, one can associate a resistance to motion with each of the scattering mechanisms. These resistances are in series. At sufficiently low doping levels, ionized impurity scattering can be neglected compared to lattice scattering, or in the analogy, $R_{\text{TOTAL}} = R_L + R_I \cong R_L$. When lattice scattering, which is not a function of N_A or N_D, becomes the dominant scattering mechanism, it automatically follows that the carrier mobilities will be likewise independent of N_A or N_D. For dopings in excess of $\sim 10^{15}/\text{cm}^3$ in Si, ionized impurity scattering and the associated resistance to motion can no longer be neglected. Increasing the number of scattering centers by adding more and more acceptors or donors progressively increases the amount of ionized impurity scattering and systematically decreases the carrier mobilities.

Temperature Dependence: The temperature dependence of the electron and hole mobilities in Si with doping as a parameter is displayed in Fig. 3.7. For dopings of N_A or $N_D \leq 10^{14}/\text{cm}^3$, the data merge into a single curve and there is a near power-law increase in the carrier mobilities as the temperature is decreased. Roughly, $\mu_n \propto T^{-2.3\pm0.1}$ and $\mu_p \propto T^{-2.2\pm0.1}$. For progressively higher dopings, the carrier mobilities still increase with decreasing temperature, but at a systematically decreasing rate. In fact, some $N_D \geq 10^{18}/\text{cm}^3$ experimental data (not shown) exhibit a *decrease* in μ_n as T is decreased below 200 K.

The general temperature dependence of the carrier mobilities is relatively easy to explain in the low doping limit. As noted in the doping dependence discussion, lattice scattering is the dominant scattering mechanism ($R_{\text{TOTAL}} \cong R_L$) in lightly doped samples. Decreasing the system temperature causes an ever-decreasing thermal agitation of the semiconductor atoms, which in turn decreases the lattice scattering. The decreased scattering

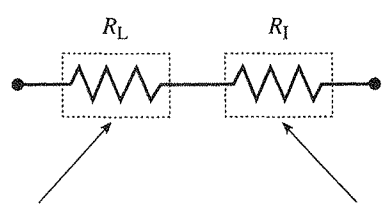


Figure 3.6 Electrical analogy for scattering in a semiconductor. R_L and R_I represent the impedance to motion due to lattice scattering and ionized impurity scattering, respectively.

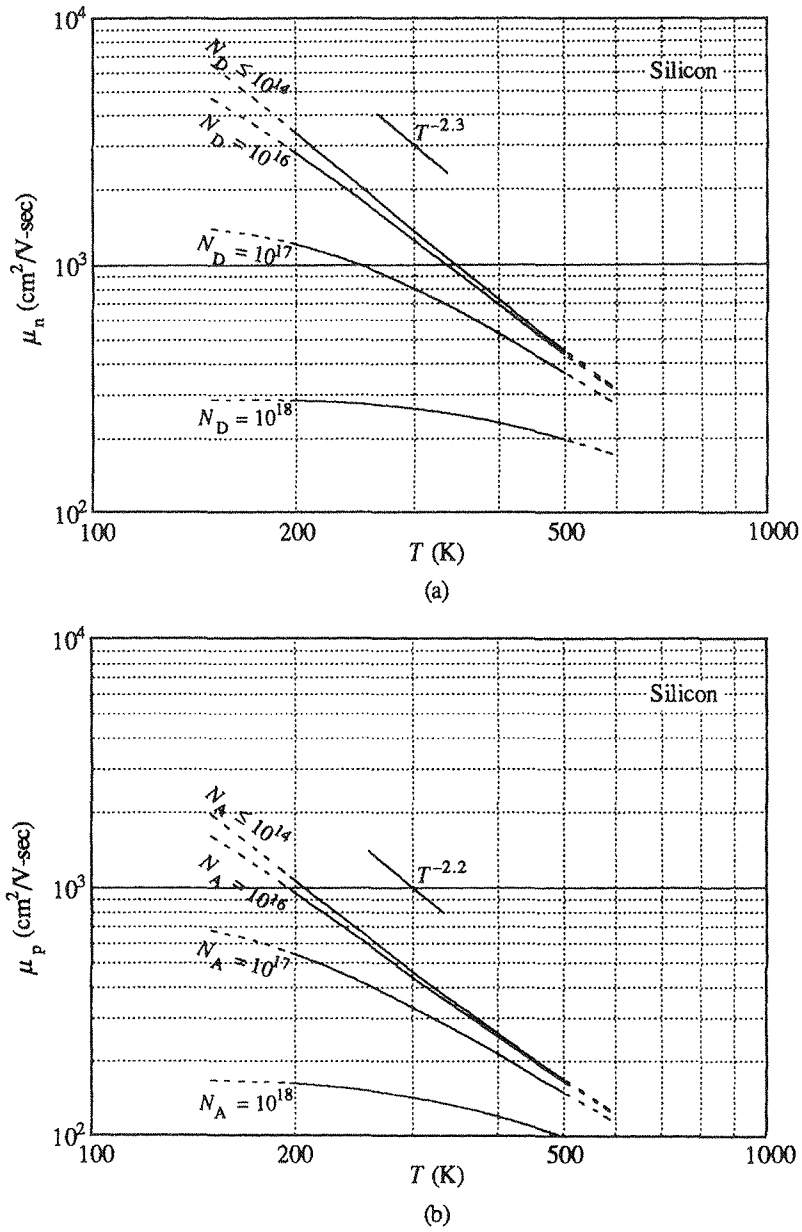


Figure 3.7 Temperature dependence of (a) electron and (b) hole mobilities in silicon for dopings ranging from $\leq 10^{14}/\text{cm}^3$ to $10^{18}/\text{cm}^3$. The curves were constructed using the empirical fit relationships and parameters presented in Exercise 3.1. The dashed line portion of the curves correspond to a slight extension of the fit beyond the verified 200 K $\leq T \leq$ 500 K range of validity.

roughly follows a simple power-law dependence. Being inversely proportional to the amount of lattice scattering, the mobility of lightly doped samples is therefore expected to increase with decreasing temperature, varying roughly as the temperature to a negative power.

The more complex dependence of the higher doped samples reflects the added effect of ionized impurity scattering. In the electrical analogy, R_I can no longer be neglected. Moreover, whereas lattice scattering (R_L) decreases with decreasing T, ionized impurity scattering (R_I) increases with decreasing T. Ionized impurities become more and more effective in deflecting the charged carriers as the temperature and hence the speed of the carriers decreases. Thus ionized impurity scattering becomes a larger and larger percentage of the overall scattering as the temperature is decreased ($R_{TOTAL} \rightarrow R_I$). Clearly, this explains the decreased slope of the mobility versus temperature dependence exhibited by the higher doped samples.

(C) Exercise 3.1

There exist surprisingly accurate "empirical-fit" relationships that are widely employed to compute the carrier mobilities at a given doping and temperature. Figures 3.5 and 3.7 were constructed using such relationships. The form of a computational relationship is typically established on an empirical basis by noting the general functional dependencies predicted theoretically and observed experimentally. Parameters in the relationship are then adjusted until an acceptable match is obtained to the best available experimental data.

The majority carrier mobility versus doping at room temperature is popularly computed from

$$\mu = \mu_{min} + \frac{\mu_0}{1 + (N/N_{ref})^{\alpha}}$$

where μ is the carrier mobility (μ_n or μ_p), N is the doping concentration (N_A or N_D), and all other quantities are fit parameters. To model the temperature dependence, one additionally employs

$$A = A_{300}\left(\frac{T}{300}\right)^{\eta}$$

A in the above equation represents μ_{min}, μ_0, N_{ref}, or α; A_{300} is the 300 K value of the parameter, T is temperature in Kelvin, and η is the temperature exponent. The fit parameters appropriate for Si are listed in the following table:

Parameter	*Value at 300 K* *Electrons*	*Holes*	*Temperature Exponent (η)*
N_{ref}(cm^{-3})	1.3×10^{17}	2.35×10^{17}	2.4
μ_{min}(cm^2/V-sec)	92	54.3	−0.57
μ_0(cm^2/V-sec)	1268	406.9	−2.33 electrons −2.23 holes
α	0.91	0.88	−0.146

P: (a) Construct a log-log plot of μ_n and μ_p versus N_A or N_D for $10^{14}/\text{cm}^3 \leq N_A$ or $N_D \leq 10^{19}/\text{cm}^3$ using the quoted fit relationship and the listed Si fit parameters. Compare your result with Fig. 3.5(a).

(b) Construct log-log plots of μ_n versus T and μ_p versus T for 200 K $\leq T \leq$ 500 K and N_D or N_A stepped in decade values from $10^{14}/\text{cm}^3$ to $10^{18}/\text{cm}^3$. Compare your results with Figs. 3.7(a) and 3.7(b), respectively.

S: (a) MATLAB program script . . .

```
%Mobility versus Dopant Concentration (Si,300K)

%Fit Parameters
NDref=1.3e17; NAref=2.35e17;
μnmin=92; μpmin=54.3;
μn0=1268; μp0=406.9;
an=0.91; ap=0.88

%Mobility Calculation
N=logspace(14,19);
μn=μnmin+μn0./(1+(N/NDref).^an);
μp=μpmin+μp0./(1+(N/NAref).^ap);

%Plotting results
close
loglog(N,μn,N,μp); grid;
axis([1.0e14 1.0e19 1.0e1 1.0e4]);
xlabel('NA or ND (cm-3)');
ylabel('Mobility (cm2/V-sec)');
text(1.0e15,1500,'Electrons');
text(1.0e15,500,'Holes');
text(1.0e18,2000,'Si,300K');
```

The results obtained by running the preceding program should be numerically identical to Fig. 3.5(a).

(b) Part (b) is left for the reader to complete.

3.1.4 Resistivity

Resistivity is an important material parameter that is closely related to carrier drift. Qualitatively, *resistivity* is a measure of a material's inherent resistance to current flow—a "normalized" resistance that does not depend on the physical dimensions of the material. Quantitatively, resistivity (ρ) is defined as the proportionality constant between the electric field impressed across a homogeneous material and the total particle current per unit area flowing in the material; that is,

$$\mathscr{E} = \rho \mathbf{J} \tag{3.5a}$$

or

$$\mathbf{J} = \sigma \mathscr{E} = \frac{1}{\rho}\mathscr{E} \tag{3.5b}$$

where $\sigma = 1/\rho$ is the material *conductivity*. In a homogeneous material, $\mathbf{J} = \mathbf{J}_{\text{drift}}$ and, as established with the aid of Eqs. (3.4),

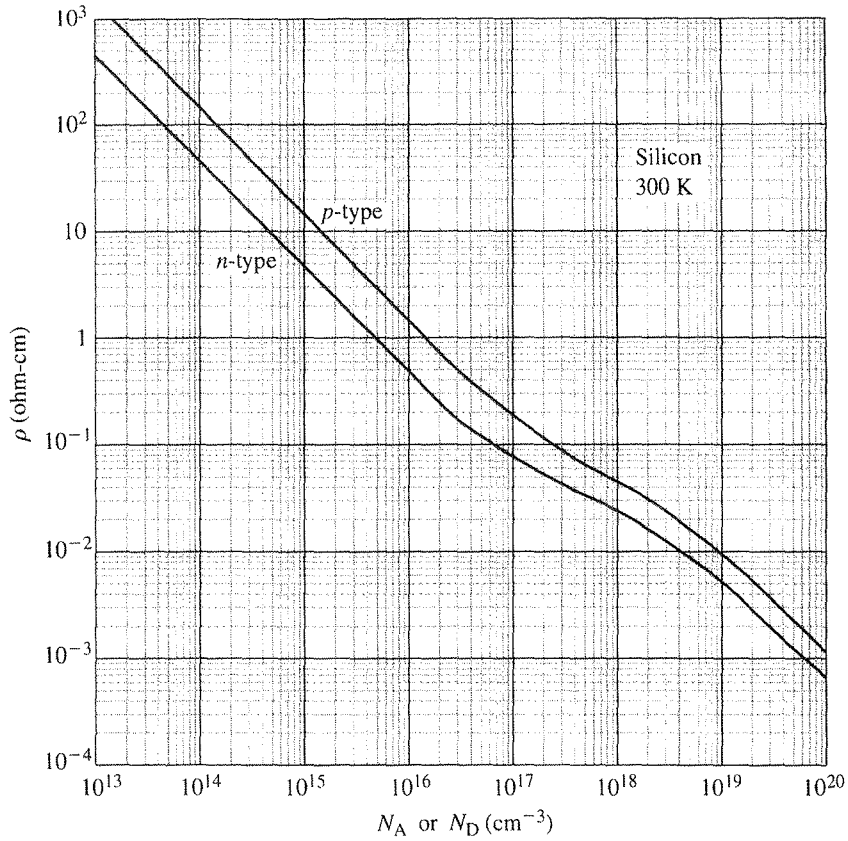
$$\mathbf{J}_{\text{drift}} = \mathbf{J}_{\text{N|drift}} + \mathbf{J}_{\text{P|drift}} = q(\mu_n n + \mu_p p)\mathscr{E} \tag{3.6}$$

It therefore follows that

$$\boxed{\rho = \frac{1}{q(\mu_n n + \mu_p p)}} \tag{3.7}$$

In a nondegenerate donor-doped semiconductor maintained in the extrinsic temperature region where $N_D \gg n_i$, $n \simeq N_D$ and $p \simeq n_i^2/N_D \ll n$. This result was established in Subsection 2.5.5. Thus, for typical dopings and mobilities, $\mu_n n + \mu_p p \simeq \mu_n N_D$ in an n-type semiconductor. Similar arguments yield $\mu_n n + \mu_p p \simeq \mu_p N_A$ in a p-type semiconductor. Consequently, under conditions normally encountered in Si samples maintained at or near room temperature, Eq. (3.7) simplifies to

$$\boxed{\rho = \frac{1}{q\mu_n N_D}} \quad \ldots\ n\text{-type semiconductor} \tag{3.8a}$$



(a)

Figure 3.8 Resistivity versus impurity concentration at 300 K in (a) Si and (b) other semiconductors. [(b) From Sze[2], © 1981 by John Wiley & Sons, Inc. Reprinted with permission.]

and

$$\boxed{\rho = \frac{1}{q\mu_p N_A}} \quad \ldots \; p\text{-type semiconductor} \tag{3.8b}$$

When combined with mobility-versus-doping data, Eqs. (3.8) provide a one-to-one correspondence between the resistivity, a directly measurable quantity, and the doping inside a semiconductor. In conjunction with plots of ρ versus doping (see Fig. 3.8), the measured resistivity is in fact routinely used to determine N_A or N_D.

The measured resistivity required in determining the doping can be obtained in a number of different ways. A seemingly straightforward approach would be to form the semiconductor into a bar, apply a bias V across contacts attached to the ends of the bar as in

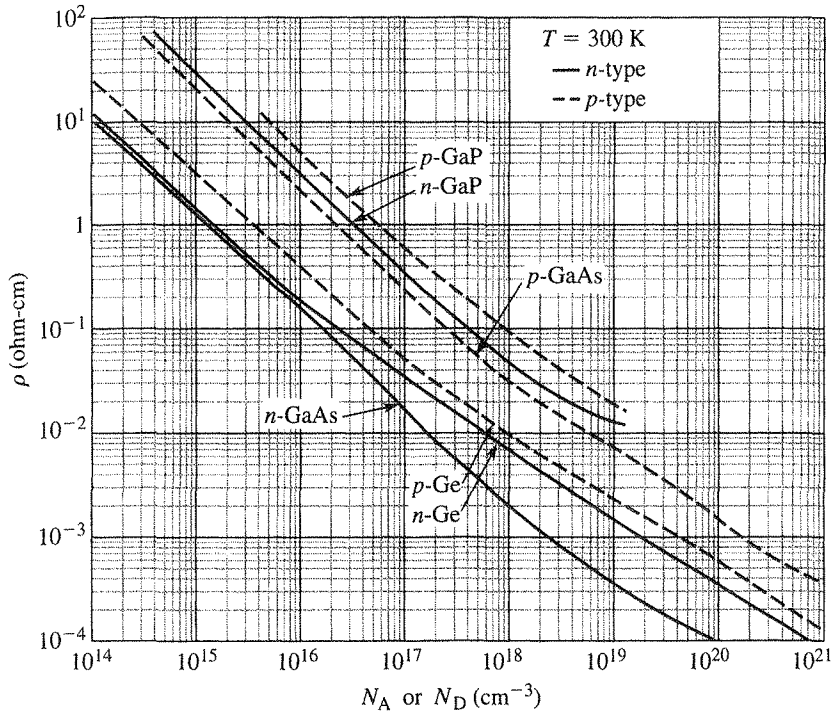


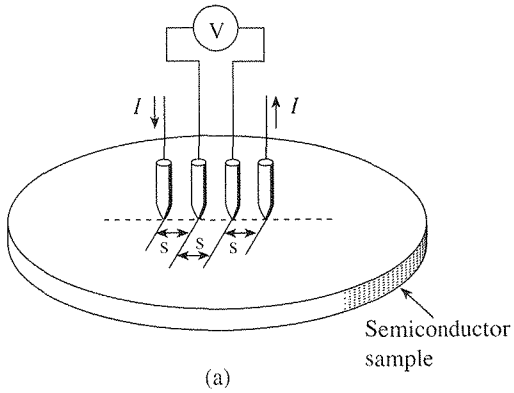
Figure 3.8 (b)

Fig. 3.1(a), measure the current I flowing in the circuit, and then deduce ρ from the measured resistance. [R(resistance) $= V/I = \rho l/A$, where l is the bar length and A is the cross-sectional area.] Unfortunately, the straightforward approach is deceptively difficult, is destructive (wastes semiconductor material), and is not readily adaptable to the wafers used in device processing.

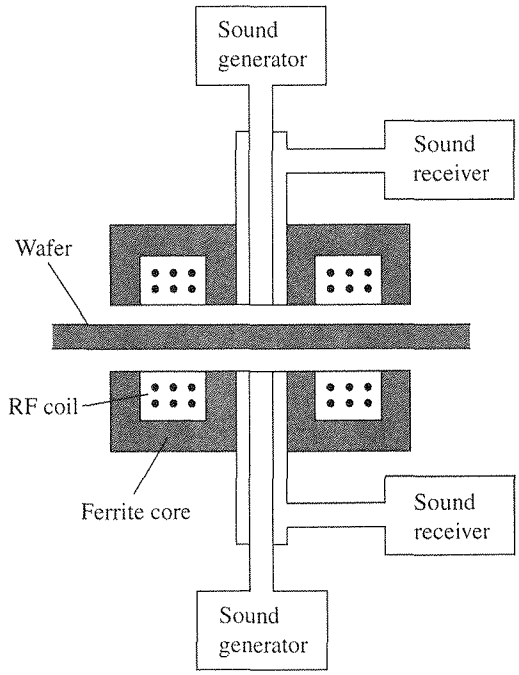
A measurement method finding widespread usage in practice is the four-point probe technique. In the standard four-point probe technique, four collinear, evenly spaced probes, as shown in Fig. 3.9(a), are brought into contact with the surface of the semiconductor. A known current I is passed through the outer two probes, and the potential V thereby developed is measured across the inner two probes. The semiconductor resistivity is then computed from

$$\rho = 2\pi s \frac{V}{I} \Gamma \tag{3.9}$$

where s is the probe-to-probe spacing and Γ is a well-documented "correction" factor. The correction factor typically depends on the thickness of the sample and on whether the bottom of the semiconductor is touching an insulator or a metal. Commercial instruments are



(a)



(b)

Figure 3.9 Resistivity measurement techniques. (a) Schematic drawing of the probe arrangement, placement, and biasing in the four-point probe measurement. (b) Schematic of a commercial eddy-current apparatus showing the RF coils and sonic components. [(b) From Schroder[6], © 1990 by John Wiley & Sons, Inc. Reprinted with permission.]

available that automatically compute the appropriate correction factor based on the sample thickness entered by the operator. Unlike the semiconductor-bar measurement, the four-point probe technique is obviously easy to implement, causes only slight surface damage in the vicinity of the probe contacts, and is ideally suited for working with wafers. The surface damage, although slight, does exclude the technique from being used on wafers intended for producing high-yield devices and ICs.

A second technique worthy of note utilizes a noncontacting eddy-current approach. The schematic of a commercial eddy-current apparatus is pictured in Fig. 3.9(b). There are fringing fields near a ferrite core excited with an RF coil. If a conducting material such as a semiconductor wafer is placed near the ferrite core, the fringing fields cause localized (eddy) currents to flow in the conducting material. The current flow in turn absorbs some of the RF power. The *sheet* resistivity of the conducting material is deduced by monitoring the power consumption in a calibrated system. The ultrasound generators and receivers built into the commercial apparatus are for the in-situ determination of the wafer thickness. The change in phase of ultrasound bounced off the top and bottom wafer surfaces permits the instrument to compute the intercore spacing occupied by the wafer. The sheet resistivity multiplied by the wafer thickness yields the wafer resistivity.

3.1.5 Band Bending

In previous encounters with the energy band diagram, we have consistently drawn E_c and E_v to be energies independent of the position coordinate x. When an electric field ($\mathscr{E}$) exists inside a material, the band energies become a function of position. The resulting variation of E_c and E_v with position on the energy band diagram, exemplified by Fig. 3.10(a), is popularly referred to as "band bending."

Seeking to establish the precise relationship between the electric field within a semiconductor and the induced band bending, let us carefully re-examine the energy band diagram. The diagram itself, as emphasized in Fig. 3.10(a), is a plot of the allowed electron energies within the semiconductor as a function of position, where E increasing upward is understood to be the *total* energy of the electrons. Furthermore, we know from previous discussions that if an energy of precisely E_G is added to break an atom-atom bond, the created electron and hole energies would be E_c and E_v, respectively, and the created carriers would be effectively motionless. Absorbing an energy in excess of E_G, on the other hand, would in all probability give rise to an electron energy greater than E_c and a hole energy less than E_v, with both carriers moving around rapidly within the lattice. We are led, therefore, to interpret $E - E_c$ to be the kinetic energy (K.E.) of the electrons and $E_v - E$ to be the kinetic energy of the holes [see Fig. 3.10(b)]. Moreover, since the total energy equals the sum of the kinetic energy and the potential energy (P. E.), E_c minus the energy reference level (E_{ref}) must equal the electron potential energy, as illustrated in Fig. 3.10(c). (Potential energy, it should be noted, is arbitrary to within a constant, and the position-independent reference energy, E_{ref}, may be chosen to be any convenient value.)

The potential energy is the key in relating the electric field within a semiconductor to positional variations in the energy bands. Specifically, assuming normal operational conditions, where magnetic field, temperature gradient, and stress-induced effects are negligible, only the force associated with an existing electric field can give rise to changes in the

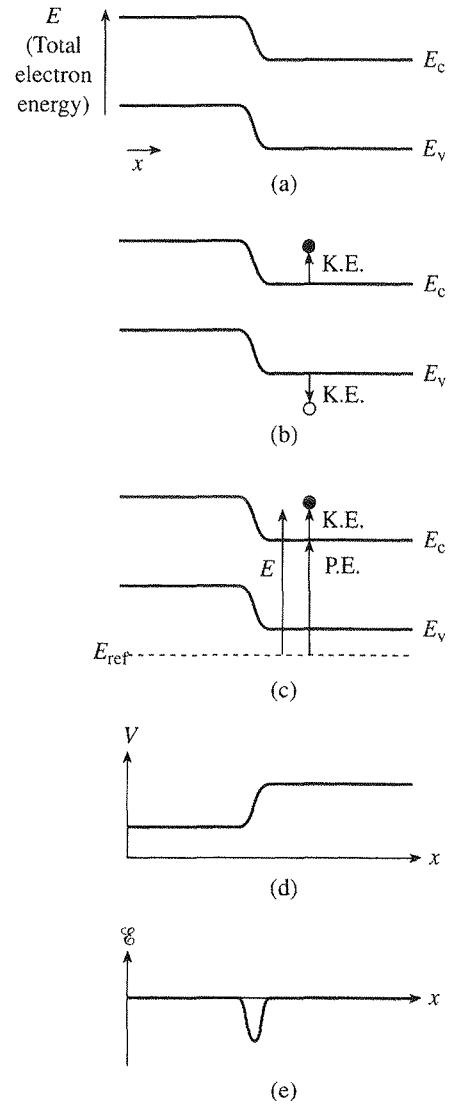


Figure 3.10 Relationship between band bending and the electrostatic variables inside a semiconductor: (a) sample energy band diagram exhibiting band bending; (b) identification of the carrier kinetic energies; (c) specification of the electron potential energy; (d) electrostatic potential and (e) electric field versus position dependence deduced from and associated with the part (a) energy band diagram.

potential energy of the carriers. Elementary physics, in fact, tells us that the potential energy of a $-q$ charged particle under such conditions is simply related to the electrostatic potential V at a given point by

$$\text{P.E.} = -qV \tag{3.10}$$

Having previously concluded that

$$\text{P.E.} = E_c - E_{ref} \tag{3.11}$$

we can state

$$\boxed{V = -\frac{1}{q}(E_c - E_{ref})} \tag{3.12}$$

By definition, moreover,

$$\mathscr{E} = -\nabla V \tag{3.13}$$

or, in one dimension,

$$\mathscr{E} = -\frac{dV}{dx} \tag{3.14}$$

Consequently,

$$\boxed{\mathscr{E} = \frac{1}{q}\frac{dE_c}{dx} = \frac{1}{q}\frac{dE_v}{dx} = \frac{1}{q}\frac{dE_i}{dx}} \tag{3.15}$$

The latter forms of Eq. (3.15) follow from the fact that E_c, E_v, and E_i differ by only an additive constant.

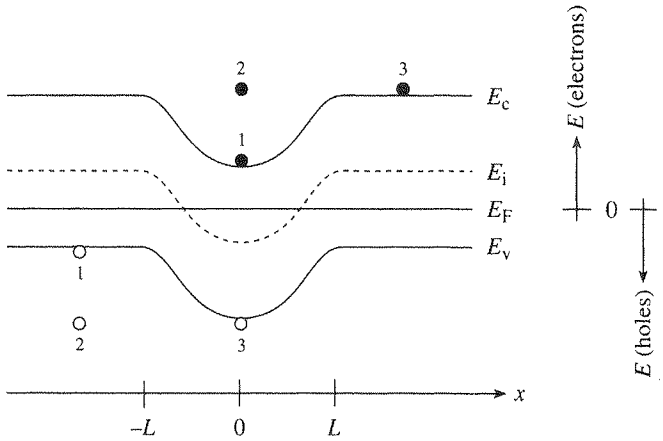
The preceding formulation provides a means of readily deducing the general form of the electrostatic variables associated with the "band bending" in Fig. 3.10(a) and other energy band diagrams. Making use of Eq. (3.12), or simply inverting $E_c(x)$ in Fig. 3.10(a), one obtains the V versus x dependence presented in Fig. 3.10(d). [Please note that the electrostatic potential, like the potential energy, is arbitrary to within a constant—the Fig. 3.10(d) plot can be translated upward or downward along the voltage axis without modifying the physical situation inside the semiconductor.] Finally, recording the slope of E_c versus position, as dictated by Eq. (3.15), produces the $\mathscr{E}$ versus x plot shown in Fig. 3.10(e).

In summary, the reader should be aware of the fact that the energy band diagram contains information relating to the electrostatic potential and electric field within the semiconductor. Moreover, the general form of the V and $\mathscr{E}$ dependencies within the semiconductor can be obtained almost by inspection. To deduce the form of the V versus x

relationship, merely sketch the "upside down" of E_c (or E_v or E_i) versus x; to determine the general $\mathscr{E}$ versus x dependence, simply note the slope of E_c (or E_v or E_i) as a function of position.

Exercise 3.2

P: Consider the following energy band diagram. Take the semiconductor represented to be Si maintained at 300 K with $E_i - E_F = E_G/4$ at $x = \pm L$ and $E_F - E_i = E_G/4$ at $x = 0$. Note the choice of E_F as the energy reference level and the identification of carriers at various points on the diagram.



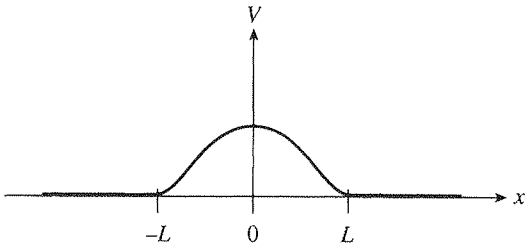
(a) Sketch the electrostatic potential (V) inside the semiconductor as a function of x.

(b) Sketch the electric field ($\mathscr{E}$) inside the semiconductor as a function of x.

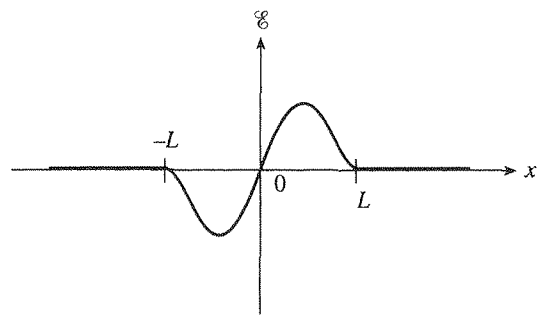
(c) Ascertain the K.E. and P.E. of the electrons and holes pictured on the diagram.

(d) Determine the resistivity of the $x > L$ portion of the semiconductor.

S: (a) The V versus x relationship must have the same functional form as the "upside down" of E_c (or E_i or E_v). If the arbitrary voltage reference point is taken to be $V = 0$ at $x = L$, then one concludes the V versus x dependence is as sketched here:



(b) The electric field, $\mathscr{E}$, is proportional to the slope of the bands:



(c) For the electrons, K.E. $= E - E_c$ and P.E. $= E_c - E_{ref} = E_c - E_F$. Since the total energy of the holes increases *downward* on the diagram, K.E. $= E_v - E$ and P.E. $= E_{ref} - E_v = E_F - E_v$ for the holes. The energy differences are of course evaluated at the same x. The deduced K.E. and P.E. values are summarized in the following table:

Carrier	*K.E. (eV)*	*P.E. (eV)*
Electron 1	0	0.28
Electron 2	0.56	0.28
Electron 3	0	0.84
Hole 1	0	0.28
Hole 2	0.56	0.28
Hole 3	0	0.84

(d) In the $x > L$ region, $E_i - E_F = E_G/4 = 0.28$ eV and

$$N_A = p = n_i e^{(E_i - E_F)/kT} = 10^{10} e^{0.28/0.0259} = 4.96 \times 10^{14}/\text{cm}^3$$

From Fig. 3.5(a) one deduces $\mu_p = 459$ cm^2/V-sec, and therefore,

$$\rho = \frac{1}{q\mu_p N_A} = \frac{1}{(1.6 \times 10^{-19})(459)(4.96 \times 10^{14})} = \boxed{27.5 \text{ ohm-cm}}$$

Alternatively making use of the Fig. 3.8(a) Si resistivity plot, one similarly concludes $\rho \cong 25$ ohm-cm.

3.2 DIFFUSION

3.2.1 Definition–Visualization

Diffusion is a process whereby particles tend to spread out or redistribute as a result of their random thermal motion, migrating on a macroscopic scale from regions of high particle concentration into regions of low particle concentration. If allowed to progress unabated, the diffusion process operates so as to produce a uniform distribution of particles. The diffusing entity, it should be noted, need not be charged; thermal motion, not interparticle repulsion, is the enabling action behind the diffusion process.

To cite an everyday example of diffusion, suppose an open bottle of perfume is placed in one corner of a room. Even in the absence of air currents, random thermal motion will spread the perfume molecules throughout the room in a relatively short period of time, with intermolecular collisions helping to uniformly redistribute the perfume to every nook and cranny within the room.

Seeking to obtain a more detailed understanding of the diffusion process, let us next conceptually "monitor" the process on a microscopic scale employing a simple hypothetical system. The system we propose to monitor is a one-dimensional box containing four compartments and 1024 mobile particles (see Fig. 3.11). The particles within the box obey certain stringent rules. Specifically, thermal motion causes all particles in a given compartment to "jump" into an adjacent compartment every τ seconds. In keeping with the random nature of the motion, each and every particle has an equal probability of jumping to the left and to the right. Hitting an "external wall" while attempting to jump to the left or right reflects the particle back to its pre-jump position. Finally, at time $t = 0$ it is assumed that all of the particles are confined in the left-most compartment.

Figure 3.11 records the evolution of our 1024 particle system as a function of time. At time $t = \tau$, 512 of the 1024 particles originally in compartment 1 jump to the right and come to rest in compartment 2. The remaining 512 particles jump to the left and are reflected back into the left-most compartment. The end result is 512 particles each in compartments 1 and 2 after τ seconds. At time $t = 2\tau$, 256 of the particles in compartment 2 jump into compartment 3 and the remainder jump back into compartment 1. In the meantime, 256 of the particles from compartment 1 jump into compartment 2 and 256 undergo a reflection at the left-hand wall. The net result after 2τ seconds is 512 particles in compartment 1 and 256 particles each in compartments 2 and 3. The state of the system after 3τ and 6τ seconds, also shown in Fig. 3.11, can be deduced in a similar manner. By $t = 6\tau$, the particles, once confined to the left-most compartment, have become almost uniformly distributed throughout the box, and it is unnecessary to consider later states. Indeed, the fundamental nature of the diffusion process is clearly self-evident from an examination of the existing states.

In semiconductors the diffusion process on a microscopic scale is similar to that occurring in the hypothetical system except, of course, the random motion of the diffusing particles is three-dimensional and not "compartmentalized." On a *macroscopic* scale the net effect of diffusion is precisely the same within both the hypothetical system and semi-

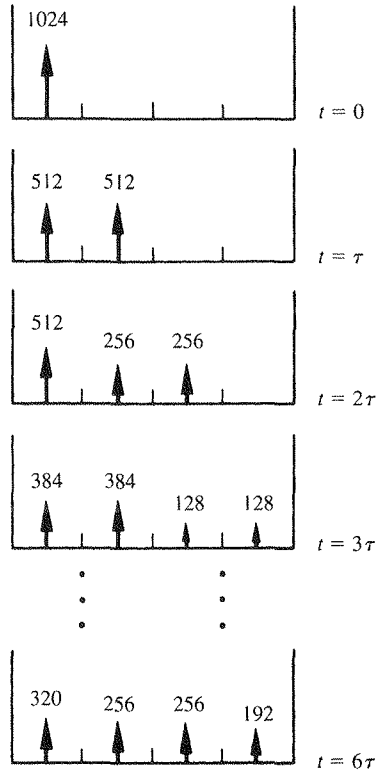


Figure 3.11 Diffusion on a microscopic scale in a hypothetical one-dimensional system. The numbers over the arrows indicate the number of particles in a given compartment; observation times are listed to the extreme right.

conductors; there is an overall migration of particles from regions of high particle concentration to regions of low particle concentration. Within semiconductors the mobile particles—the electrons and holes—are charged, and diffusion-related carrier transport therefore gives rise to particle currents as pictured in Fig. 3.12.

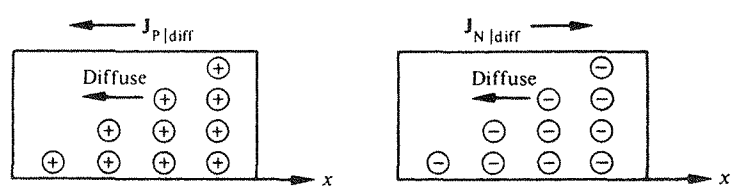


Figure 3.12 Visualization of electron and hole diffusion on a macroscopic scale.

(C) Exercise 3.3

The "DiffDemo" MATLAB file listed below was written to help the user visualize the diffusion process. The program provides a pseudo-animation of a one-dimensional particle system similar to that described in the text. The `y= [ ]` statement in the program controls the initial condition, `N` specifies the maximum number of monitored "jumps," and the number in the `pause` statements controls the time between jumps.

```
%Simulation of Diffusion (DiffDemo)
%One-dimensional system, right and left jumps equally probable

%Initialization
close
x=[0.5 1.5 2.5 3.5 4.5];
y=[1.0e6 0 0 0 0];    %NOTE: initial position can be changed
[xp,yp]=bar(x,y);
plot(xp,yp); text (0.5,1.1e6,'t = 0');
axis([0,5,0,1.2e6]);
pause (0.5)
N=15;                 %NOTE: increase N for extended run

%Computations and Plotting
for ii=1:N,
  %Diffusion step calculation
  bin(1)=round(y(1)/2 + y(2)/2);
  bin(2)=round(y(1)/2 + y(3)/2);
  bin(3)=round(y(2)/2 + y(4)/2);
  bin(4)=round(y(3)/2 + y(5)/2);
  bin(5)=round(y(4)/2 + y(5)/2);
  y=bin;
  %Plotting the result
  [xp,yp]=bar(x,y);
  axis(axis);
  plot(xp,yp); text(0.5,1.1e6,['t = ',num2str(ii)]);
  axis([0,5,0,1.2e6]);
  pause (0.5)
end
```

P: (a) Enter the program into your computer or locate the copy supplied on disk. Run the program. (The *command-period* on a Macintosh computer or *break* on an IBM-compatible computer can be used to prematurely terminate the program.)

(b) Rerun the program after changing the initial conditions so that all of the carriers are in the middle box at $t = 0$.

(c) Experiment with initially placing particles in more than one box and changing the number in the `pause` statements. You might also try increasing the number of particle boxes.

S: Sample output is reproduced in Fig. E3.3.

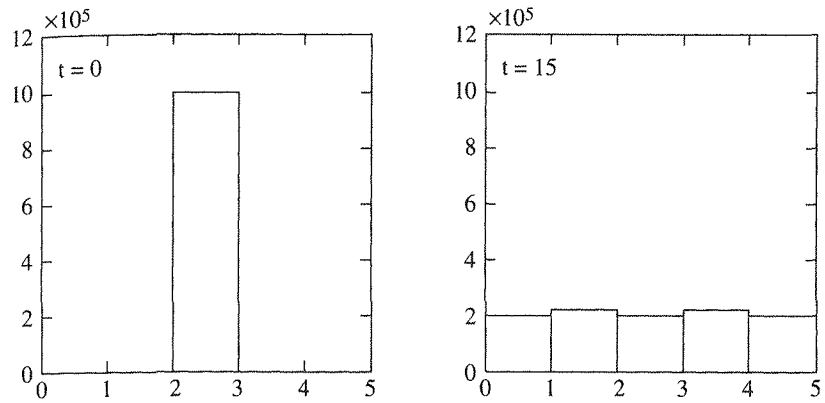


Figure E3.3

3.2.2 Hot-Point Probe Measurement

Before establishing the current associated with diffusion, let us digress somewhat and briefly consider the hot-point probe measurement. The *hot-point probe measurement* is a common technique for rapidly determining whether a semiconductor is *n*- or *p*-type. From a practical standpoint, knowledge of the semiconductor type is essential in device processing and must be known even before determining the doping concentration from resistivity measurements (refer back to Fig. 3.8). The hot-point probe "typing" experiment is considered here because it simultaneously provides an informative example of the diffusion process.

Examining Fig. 3.13(a), we find the only equipment required for performing the hot-point probe measurement is a hot-point probe, a cold-point probe, and a center-zero milliammeter. The hot-point probe is sometimes a refugee from a wood burning set or simply a soldering iron; the cold-point probe typically assumes the form of an electrical probe like that used with hand-held multimeters. No special requirements are imposed on the center-zero milliammeter connected between the probes. The measurement procedure itself is also extremely simple: After allowing the hot probe to heat up, one brings the two probes into contact with the semiconductor sample, and the ammeter deflects to the right or left, thereby indicating the semiconductor type. Observe that the spacing between the probes is set arbitrarily and can be reduced to enhance the ammeter deflection.

A simplified explanation of how the hot-point probe measurement works is presented in Fig. 3.13(b). In the vicinity of the probe contact the heated probe creates an increased

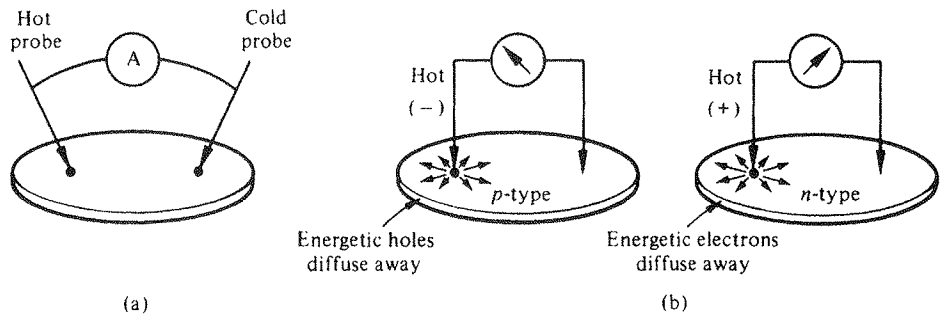


Figure 3.13 The hot-point probe measurement: (a) required equipment; (b) simplified explanation of how the measurement works.

number of higher-energy carriers. These energetic carriers will be predominantly holes in the case of a p-type material and electrons in an n-type material. With more energetic carriers near the heated probe than elsewhere, diffusion acts so as to spread the higher-energy carriers throughout the semiconductor wafer. The net effect is a deficit of holes or a net negative charge in the vicinity of the hot-point probe for a p-type material, and a positive charge buildup near the heated probe within an n-type material. Accordingly, the center-zero meter deflects in a different direction for p- and n-type materials.

3.2.3 Diffusion and Total Currents

Diffusion Currents

In defining the diffusion process and in citing diffusion examples, we have sought to emphasize the direct correlation between diffusion and a spatial variation in particle numbers. For diffusion to occur, more of the diffusing particles must exist at one point than at other points or, in mathematical terms, there must be a nonzero concentration gradient ($\nabla p \neq 0$ for holes, $\nabla n \neq 0$ for electrons). Logically, moreover, the greater the concentration gradient, the larger the expected flux of particles. Quantitative analysis of the diffusion process indeed confirms the foregoing and leads to what is known as Fick's law:

$$\mathcal{F} = -D\nabla\eta \tag{3.16}$$

where $\mathcal{F}$ is the flux or particles/cm^2-sec crossing a plane perpendicular to the particle flow, η is the particle concentration, and D is the *diffusion coefficient,* a positive proportionality factor. The diffusion current density due to electrons and holes is obtained by simply multiplying the carrier flux by the carrier charge:

$$\boxed{\mathbf{J}_{\mathrm{P|diff}} = -qD_{\mathrm{P}}\nabla p} \tag{3.17a}$$

$$\boxed{\mathbf{J}_{\mathrm{N|diff}} = qD_{\mathrm{N}}\nabla n} \tag{3.17b}$$

The constants of proportionality, D_P and D_N, have units of cm^2/sec and are referred to as the hole and electron diffusion coefficients, respectively.

Upon examining Eqs. (3.17), note that the current directions deduced from the equations are consistent with the Fig. 3.12 visualization of the macroscopic diffusion currents. For the positive concentration gradient shown in Fig. 3.12 ($dp/dx > 0$ and $dn/dx > 0$ for the pictured one-dimensional situation), both holes and electrons will diffuse in the $-x$ direction. $J_{P|diff}$ will therefore be negative or directed in the $-x$ direction, while $J_{N|diff}$ will be oriented in the $+x$ direction.

Total Currents

The total or net carrier currents in a semiconductor arise as the combined result of drift and diffusion. Summing the respective n and p segments of Eqs. (3.4) and Eqs. (3.17), one obtains

$$\boxed{\begin{aligned} \mathbf{J}_P &= \mathbf{J}_{P|drift} + \mathbf{J}_{P|diff} = q\mu_p p\mathscr{E} - qD_P\nabla p \\ &\qquad\qquad\qquad\qquad \updownarrow \text{drift} \qquad \updownarrow \text{diffusion} \\ \mathbf{J}_N &= \mathbf{J}_{N|drift} + \mathbf{J}_{N|diff} = q\mu_n n\mathscr{E} + qD_N\nabla n \end{aligned}} \qquad \begin{aligned}(3.18a)\\ \\ (3.18b)\end{aligned}$$

The total particle current flowing in a semiconductor is in turn computed from

$$\boxed{\mathbf{J} = \mathbf{J}_N + \mathbf{J}_P} \qquad (3.19)$$

The double boxes emphasize the importance of the total-current relationships; they are used directly or indirectly in the vast majority of device analyses.

3.2.4 Relating Diffusion Coefficients/Mobilities

The diffusion coefficients, the constants of the motion associated with diffusion, are obviously central parameters in characterizing carrier transport due to diffusion. Given the importance of the diffusion coefficients, one might anticipate an extended examination of relevant properties paralleling the mobility presentation in Subsection 3.1.3. Fortunately, an extended examination is unnecessary because the D's and the μ's are all interrelated. It is only necessary to establish the connecting formula known as the Einstein relationship.

In deriving the Einstein relationship, we consider a *nonuniformly* doped semiconductor maintained under equilibrium conditions. Special facts related to the nonuniform doping and the equilibrium state are invoked in the derivation. These facts, important in themselves, are reviewed prior to presenting the derivation proper.

Constancy of the Fermi Level

Consider a nondegenerate, nonuniformly doped n-type semiconductor sample, a sample where the doping concentration varies as a function of position. A concrete example of what we have in mind is shown in Fig. 3.14(a). Assuming equilibrium conditions prevail,

the energy band diagram characterizing the sample will have the form sketched in Fig. 3.14(b). In Chapter 2 the Fermi level in *uniformly* doped n-type semiconductors was found to move closer and closer to E_c when the donor doping was systematically increased (see Fig. 2.21). Consistent with this fact, E_c was drawn closer to E_F in going from $x = 0$ to $x = L$ on the Fig. 3.14(b) diagram. Construction of the diagram, however, could not have been completed without making use of a critical new piece of information. Specifically,

under equilibrium conditions, $dE_F/dx = dE_F/dy = dE_F/dz = 0$; i.e., *the Fermi level inside a material or a group of materials in intimate contact is invariant as a function of position.*

The constancy of the Fermi level means that E_F is to appear as a horizontal line on equilibrium energy band diagrams.

The position independence of the Fermi energy is established by examining the trans-

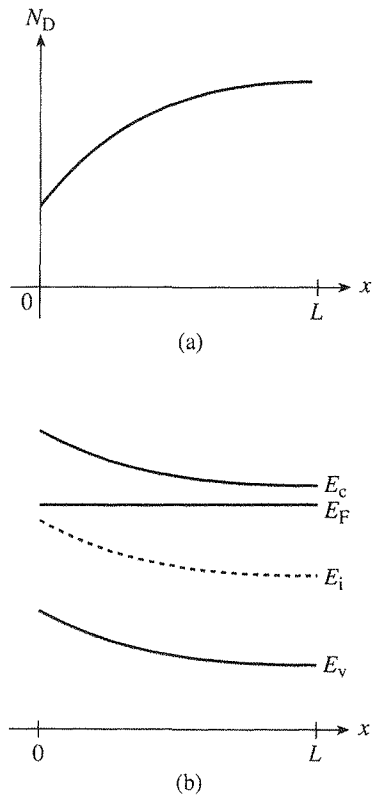


Figure 3.14 Nonuniformly doped semiconductor: (a) assumed doping variation with position; (b) corresponding equilibrium energy band diagram.

fer of carriers between allowed states with the same energy but at adjacent positions in an energy band. It is concluded the probability of filling the states at a given energy, $f(E)$, must be the same everywhere in the sample under equilibrium conditions. If this were not the case, carriers would preferentially transfer between states and thereby give rise to a net current. The existence of a net current is inconsistent with the specified equilibrium conditions. Referring to the Eq. (2.7) expression for $f(E)$, we find that the constancy of the Fermi function in turn requires the Fermi level to be independent of position.

Current Flow Under Equilibrium Conditions

Under equilibrium conditions the total current is of course identically zero. Because electron and hole activity is totally decoupled under equilibrium conditions, the electron and hole current densities, $\mathbf{J}_N$ and $\mathbf{J}_P$, must also independently vanish. Setting $\mathbf{J}_N = \mathbf{J}_P = 0$ in Eq. (3.18), however, leads to the conclusion that the drift and diffusion currents associated with a given carrier are merely required to be of equal magnitude and opposite polarity. In fact, the drift and diffusion components will vanish under equilibrium conditions only if $\mathscr{E} = 0$ and $\nabla n = \nabla p = 0$.

To provide a concrete illustration of non-vanishing equilibrium current components, let us return to an examination of the nonuniformly doped semiconductor sample characterized by Fig. 3.14. The dopant variation pictured in Fig. 3.14(a) clearly gives rise to a significant electron concentration gradient. An electron diffusion current flowing in the $+x$ direction must exist inside the sample. The band bending shown in the band diagram of Fig. 3.14(b), moreover, implies the existence of a "built-in" electric field oriented in the $-x$ direction. This gives rise to a drift current also in the $-x$ direction. Note that the drift and diffusion components are of opposite polarity. Since equilibrium conditions are assumed to prevail, the components must also be of equal magnitude. In point of fact, the electric field arises inside the semiconductor to counteract the diffusive tendencies of the carriers caused by the nonuniform doping. The major point here is that, even under equilibrium condition, nonuniform doping will give rise to carrier concentration gradients, a built-in electric field, and non-zero current components.

Einstein Relationship

Having laid the proper foundation, we can now proceed to derive the connecting formula between the D's and the μ's known as the Einstein relationship. To simplify the development, we work in only one dimension. The sample under analysis is taken to be a nondegenerate, nonuniformly doped semiconductor maintained under equilibrium conditions. Citing the fact that the net carrier currents must be identically zero under equilibrium conditions, and focusing on the electrons, we can state

$$J_{\mathrm{N|drift}} + J_{\mathrm{N|diff}} = q\mu_n n\mathscr{E} + qD_N \frac{dn}{dx} = 0 \tag{3.20}$$

However,

$$\mathscr{E} = \frac{1}{q}\frac{dE_i}{dx} \tag{3.21}$$

(same as 3.15)

and

$$n = n_i e^{(E_F - E_i)/kT} \tag{3.22}$$

Moreover, with $dE_F/dx = 0$ (due to the positional invariance of the Fermi level under equilibrium conditions),

$$\frac{dn}{dx} = -\frac{n_i}{kT} e^{(E_F - E_i)/kT} \frac{dE_i}{dx} = -\frac{q}{kT} n\mathscr{E} \tag{3.23}$$

Substituting dn/dx from Eq. (3.23) into Eq. (3.20), and rearranging the result slightly, one obtains

$$(qn\mathscr{E})\mu_n - (qn\mathscr{E})\frac{q}{kT} D_N = 0 \tag{3.24}$$

Since $\mathscr{E} \neq 0$ (a consequence of the nonuniform doping), it follows from Eq. (3.24) that

$$\boxed{\frac{D_N}{\mu_n} = \frac{kT}{q}} \quad \text{Einstein relationship for electrons} \tag{3.25a}$$

A similar argument for holes yields

$$\boxed{\frac{D_P}{\mu_p} = \frac{kT}{q}} \quad \text{Einstein relationship for holes} \tag{3.25b}$$

Although it was established while assuming equilibrium conditions, we can present more elaborate arguments that show the Einstein relationship to be valid even under nonequilibrium conditions. The nondegenerate restriction, however, still applies; slightly modified forms of Eqs. (3.25) result when the argument is extended to degenerate mate-

rials. We should emphasize again that the preliminary results, such as the positional invariance of the equilibrium Fermi level and the situation inside a nonuniformly doped semiconductor under equilibrium conditions, are important in themselves. Relative to numerical values, note that kT/q is a voltage, and at room temperature (300 K) is equal to 0.0259V.† Hence, for an $N_D = 10^{14}/\text{cm}^3$ Si sample maintained at room temperature, $D_N = (kT/q)\mu_n = (0.0259)(1358\ \text{cm}^2/\text{V-sec}) = 35.2\ \text{cm}^2/\text{sec}$. Finally, the Einstein relationship is one of the easiest equations to remember because it rhymes internally—D over μ equals kT over q. The rhyme holds even if the equation is inverted—μ over D equals q over kT.

Exercise 3.4

P: In this exercise we continue the examination of the semiconductor sample represented by the energy band diagram presented in Exercise 3.2.

(a) The semiconductor is in equilibrium. How does one deduce this fact from the given energy band diagram?

(b) What is the electron current density (J_N) and hole current density (J_P) at $x = \pm L/2$?

(c) Roughly sketch n and p versus x inside the sample.

(d) Is there an electron diffusion current at $x = \pm L/2$? If there is a diffusion current at a given point, indicate the direction of *current* flow.

(e) Noting the solution for the electric field presented in Exercise 3.2, is there an electron drift current at $x = \pm L/2$? If there is a drift current at a given point, indicate the direction of current flow.

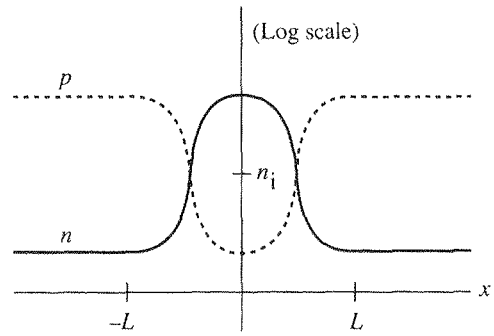
(f) What is the hole diffusion coefficient (D_P) in the $x > L$ region of the semiconductor?

S: (a) The semiconductor is concluded to be in equilibrium because the Fermi level is invariant as a function of position.

(b) $\boxed{J_N = 0}$ and $\boxed{J_P = 0}$ at both $x = -L/2$ and $x = +L/2$. The net electron and hole currents are always identically zero everywhere under equilibrium conditions.

† At room temperature $kT = 0.0259$ eV $= (1.6 \times 10^{-19})(0.0259)$ joule. Thus $kT/q = (1.6 \times 10^{-19})(0.0259)/(1.6 \times 10^{-19})$ joule/coul $= 0.0259$ V.

(c) Since $n = n_i \exp[(E_F - E_i)/kT]$ and $p = n_i \exp[(E_i - E_F)/kT]$, we conclude



(d) There is an electron diffusion current at both $x = -L/2$ and $x = +L/2$. From the answer to the preceding question we note that $dn/dx \neq 0$ at the cited points. With $dn/dx > 0$, $J_{N|diff}$ at $x = -L/2$ flows in the $+x$ direction. Conversely, $dn/dx < 0$ at $x = L/2$ and $J_{N|diff}$ at this point flows in the $-x$ direction.

(e) There is an electron drift current at both $x = -L/2$ and $x = +L/2$. Since both n and $\mathscr{E}$ are non-zero at $x = \pm L/2$, it follows that $J_{N|drift} = q\mu_n n\mathscr{E} \neq 0$. The drift component of the current always has the same direction as the electric field; $J_{N|drift}$ is in the $-x$ direction at $x = -L/2$ and in the $+x$ direction at $x = +L/2$. The drift component of the current must of course cancel the diffusion component of the current so that $J_N = J_{N|drift} + J_{N|diff} = 0$ under equilibrium conditions. The direction-related answers to parts (d) and (e), summarized next, are consistent with this requirement:

	$x = -L/2$	$x = +L/2$
$J_{N\|diff}$	$\rightarrow$	$\leftarrow$
$J_{N\|drift}$	$\leftarrow$	$\rightarrow$

(f) In working part (d) of Exercise 3.2, we concluded the $x > L$ region of the semiconductor (Si, 300 K) had a doping of $N_A \cong 5 \times 10^{14}/\text{cm}^3$ and a corresponding hole mobility of $\mu_p = 459$ cm^2/V-sec. Thus, employing the Einstein relationship, $D_P = (kT/q)\mu_p = (0.0259)(459) = \boxed{11.9 \text{ cm}^2/\text{sec}}$.

3.3 RECOMBINATION–GENERATION

3.3.1 Definition–Visualization

When a semiconductor is perturbed from the equilibrium state, an excess or deficit in the carrier concentrations relative to their equilibrium values is invariably created inside the semiconductor. Recombination–generation (R–G) is nature's order-restoring mechanism, the means whereby the carrier excess or deficit inside the semiconductor is stabilized (if the perturbation is maintained) or eliminated (if the perturbation is removed). Since non-equilibrium conditions prevail during device operation, recombination–generation more often than not plays a major role in shaping the characteristics exhibited by a device. Formally, recombination and generation can be defined as follows:

Recombination—a process whereby electrons and holes (carriers) are annihilated or destroyed.

Generation—a process whereby electrons and holes are created.

Unlike drift and diffusion, the terms *recombination* and *generation* do not refer to a single process. Rather, they are collective names for a group of similar processes; carriers can be created and destroyed within a semiconductor in a number of ways. The most common R–G processes are visualized in Fig. 3.15. The individual processes are described and discussed next.

Band-to-Band Recombination

Band-to-band recombination is conceptually the simplest of all recombination processes. As pictured in Fig. 3.15(a), it merely involves the direct annihilation of a conduction band electron and a valence band hole. An electron and hole moving in the semiconductor lattice stray into the same spatial vicinity and zap!—the electron and hole annihilate each other. The excess energy released during the process typically goes into the production of a photon (light).

R–G Center Recombination

The process pictured in Fig. 3.15(b) involves a "third party," or intermediary, and takes place only at special locations within the semiconductor known as R–G centers. Physically, R–G centers are lattice defects or special impurity atoms such as gold in Si. Lattice defects and the special atoms in the form of unintentional impurities are present even in semiconductors of the highest available purity. The R–G center concentration, however, is normally very low compared to the acceptor and donor concentrations in device-quality materials. The most important property of the R–G centers is the introduction of allowed electronic levels near the center of the band gap. These levels are identified by the E_T energy label in Fig. 3.15(b). The near midgap positioning of the levels is all-important because it distinguishes R–G centers from donors and acceptors. The midgap levels introduced by select R–G center impurities in Si are summarized in Fig. 3.16.

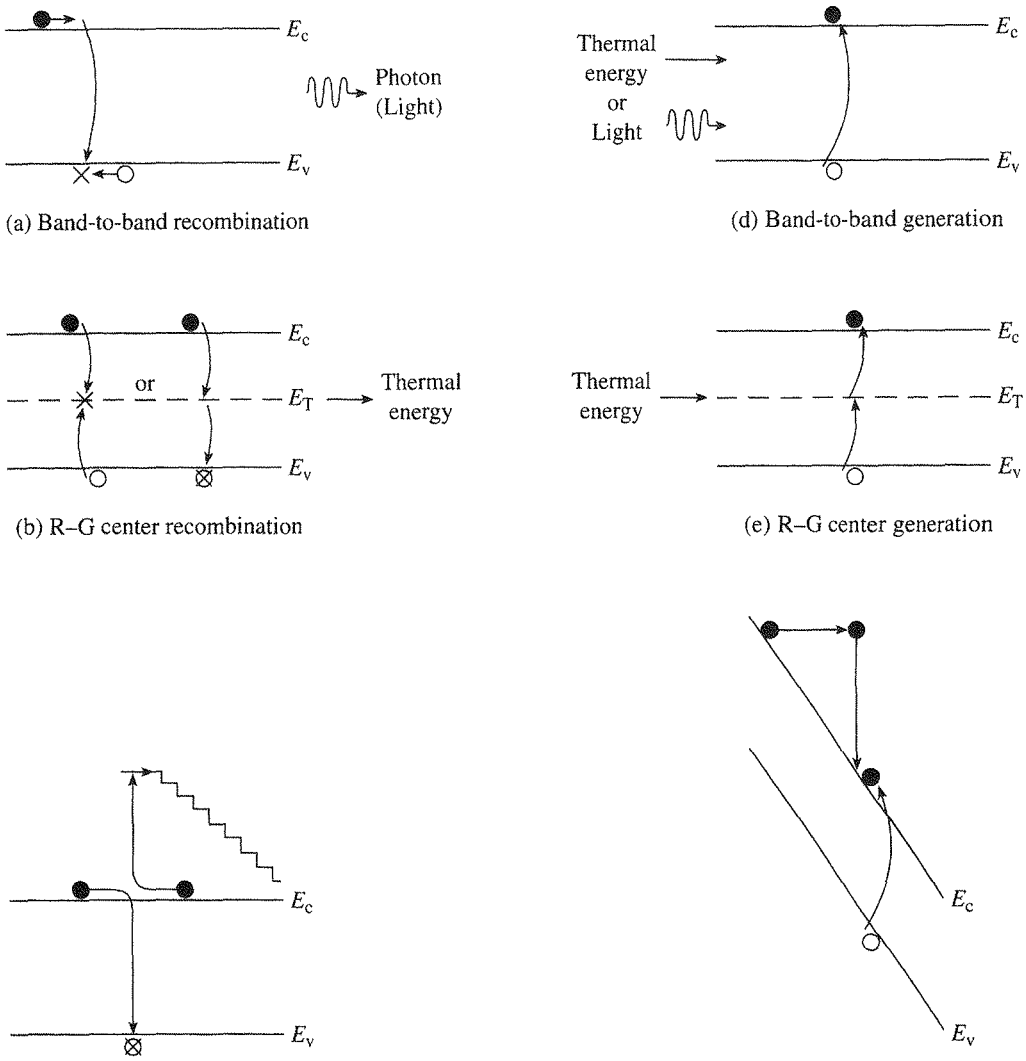


Figure 3.15 Energy band visualization of recombination and generation processes.

As noted in Fig. 3.15(b), recombination at an R–G center is a two-step process. First, one type of carrier, say an electron, strays into the vicinity of an R–G center, is caught by the potential well associated with the center, loses energy, and is trapped. Subsequently, a hole comes along, is attracted to the trapped electron, loses energy, and annihilates the electron within the center. Alternatively, as also shown in Fig. 3.15(b), one can think of the electron as losing energy a second time and annihilating the hole in the valence band. R–G center recombination, also called *indirect recombination,* typically releases thermal energy (heat) during the process or, equivalently, produces lattice vibrations.

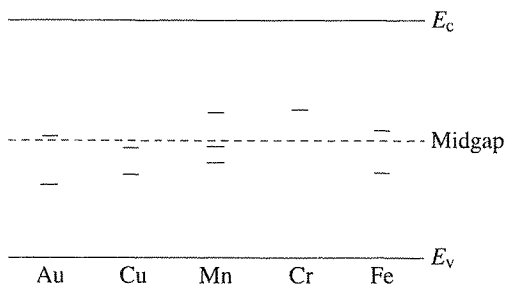


Figure 3.16 Near-midgap energy levels introduced by some common impurities in Si. When an impurity introduces multiple levels, one of the levels tends to dominate in a given semiconductor sample.

Auger Recombination

In the Auger (pronounced Oh-jay) process pictured in Fig. 3.15(c), band-to-band recombination occurs simultaneously with the collision between two like carriers. The energy released by the recombination is transferred during the collision to the surviving carrier. This highly energetic carrier subsequently "thermalizes"—loses energy in small steps through heat-producing collisions with the semiconductor lattice. The "staircase" in Fig. 3.15(c) represents the envisioned stepwise loss of energy.

Generation Processes

Any of the foregoing recombination processes can be reversed to generate carriers. Band-to-band generation, where an electron is excited directly from the valence band into the conduction band, is pictured in Fig. 3.15(d). Note that either thermal energy or light can provide the energy required for the band-to-band transition. If thermal energy is absorbed, the process is alternatively referred to as *direct thermal generation;* if externally introduced light is absorbed, the process is called *photogeneration.* The thermally assisted generation of carriers with R–G centers acting as intermediaries is envisioned in Fig. 3.15(e). Finally, impact ionization, the inverse of Auger recombination, is visualized in Fig. 3.15(f). In this process an electron–hole pair is produced as a result of energy released when a highly energetic carrier collides with the crystal lattice. The generation of carriers through impact ionization routinely occurs in the high $\mathscr{E}$-field regions of devices. We will have more to say about this process when treating breakdown in *pn* junctions.

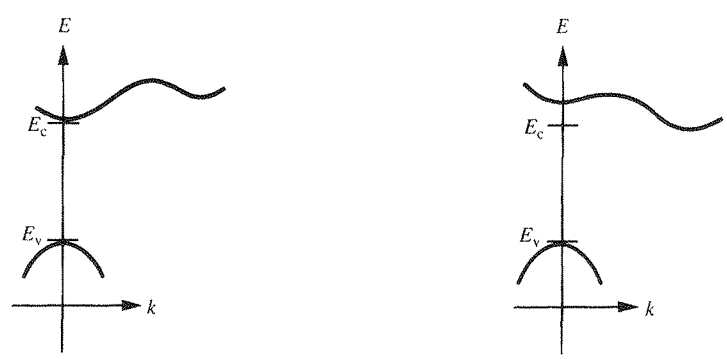
3.3.2 Momentum Considerations

All of the various recombination–generation processes occur at all times in all semiconductors—even under equilibrium conditions. The critical issue is not whether the processes are occurring, but rather, the rates at which the various processes are occurring. Typically, one need be concerned only with the dominant process, the process proceeding at the fastest rate. A number of the processes are important only under special conditions or are highly improbable, occurring at a much slower rate than competing processes. Auger recombina-

tion provides a case in point. Because the number of carrier-carrier collisions increases with increased carrier concentration, the frequency of Auger recombination likewise increases with carrier concentration. Auger recombination must be considered, for example, in treating highly doped regions of a device structure, but it is usually negligible otherwise. Similarly, significant impact ionization occurs only in the high $\mathscr{E}$-field regions of a device. In a somewhat different vein, photogeneration at room temperature becomes a significant process only when the semiconductor is exposed to external illumination. What then is the R–G process pair deserving of special attention, the process pair that typically dominates in the low $\mathscr{E}$-field regions of a nondegenerately doped semiconductor maintained at room temperature? From the information provided, the choice is obviously between thermal band-to-band R–G and recombination–generation via R–G centers.

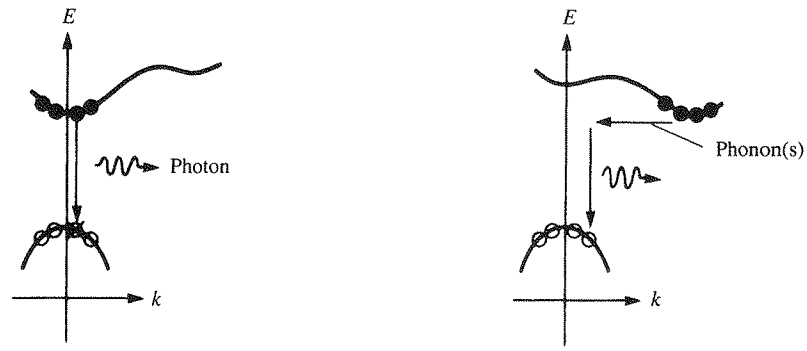
From the energy band explanation alone one might speculate that band-to-band recombination–generation would dominate under the cited "standard" conditions. However, visualization of R–G processes using the energy band diagram can be misleading. The E–x plot examines only changes in energy, whereas crystal momentum in addition to energy must be conserved in any R–G process. Momentum conservation requirements, as it turns out, play an important role in setting the process rate.

As first noted in Subsection 2.2.3, the momentum of an electron in an energy band, like the electron energy, can assume only certain quantized values. Momentum-related aspects of R–G processes are conveniently discussed with the aid of plots where the allowed electron energies in the conduction and valence bands are plotted versus the allowed momentum. These are referred to as E–$\mathbf{k}$ plots, where $\mathbf{k}$ is a parameter proportional to the electron crystal momentum. The E–$\mathbf{k}$ plots associated with semiconductors fall into two general categories. In the first, sketched in Fig. 3.17(a), both the minimum conduction band energy and maximum valence band energy occur at $\mathbf{k} = 0$. In the second, pictured in Fig. 3.17(b), the conduction band minimum is displaced to a $\mathbf{k} \neq 0$. Semiconductors exhibiting the former type of plot are referred to as *direct semiconductors,* and semiconductors exhibiting the latter type of plot, as *indirect semiconductors.* GaAs is a notable member of the direct semiconductor family; Ge and Si are both indirect semiconductors.



(a) Direct semiconductor (b) Indirect semiconductor

Figure 3.17 General forms of E–$\mathbf{k}$ plots for direct and indirect semiconductors.



(a) Direct semiconductor (b) Indirect semiconductor

Figure 3.18 E–**k** plot visualizations of recombination in direct and indirect semiconductors.

To make use of the E–**k** plots in visualizing an R–G process, one needs to know the nature of transitions associated with the absorption or emission of photons (light) and lattice vibration quanta called *phonons*. Photons, being massless entities, carry very little momentum, and a photon-assisted transition is essentially vertical on the E–**k** plot. Conversely, the thermal energy associated with lattice vibrations (phonons) is very small (in the 10–50 meV range), whereas the phonon momentum is comparatively large. Thus on an E–**k** plot a phonon-assisted transition is essentially horizontal. It is also important to point out that, as pictured in Fig. 3.18, electrons and holes normally only occupy states very close to the E_c minimum and E_v maximum, respectively. This is of course consistent with our earlier discussion of the distribution of carriers in the bands.

Let us now specifically re-examine the band-to-band recombination process. In a direct semiconductor where the **k**-values of electrons and holes are all bunched near $\mathbf{k} = 0$, little change in momentum is required for the recombination process to proceed. The conservation of both energy and momentum is readily met simply by the emission of a photon (see Fig. 3.18a). In an indirect semiconductor, on the other hand, there is a large change in momentum associated with the recombination process. The emission of a photon will conserve energy but cannot simultaneously conserve momentum. Thus for band-to-band recombination to proceed in an indirect semiconductor, the emission of a photon must be accompanied by the emission or absorption of a phonon (see Fig. 3.18b).

The rather involved nature of the band-to-band process in indirect semiconductors understandably leads to a vastly diminished recombination rate. Band-to-band recombination is in fact totally negligible compared to R–G center recombination in Si and other indirect semiconductors. Band-to-band recombination does proceed at a much faster rate in direct semiconductors and is the process producing the light observed from LEDs and junction lasers. Even in direct materials, however, the R–G center mechanism is often the dominant process. Given the universal importance of the R–G center process, and the fact that most devices are made from Si, which is an indirect material, the quantitative considerations in the next subsection are devoted primarily to the R–G center mechanism. Other processes are considered as the need arises.

3.3.3 R–G Statistics

R–G statistics is the technical name given to the mathematical characterization of recombination–generation processes. "Mathematical characterization" in the case of recombination–generation does not mean the development of a current-density relationship. An R–G event occurs at localized positions in the crystal, at a given x-value, and therefore the R–G action does not lead *per se* to charge transport. Rather, recombination–generation acts to change the carrier concentrations, thereby indirectly affecting the current flow. It is the time rate of change in the carrier concentrations ($\partial n/\partial t$, $\partial p/\partial t$) that must be specified to achieve the desired mathematical characterization. Because its characterization is relatively simple and its involvement in device operation quite common, we begin by considering photogeneration. The major portion of the subsection is of course devoted to recombination–generation via R–G centers, alternatively referred to as indirect thermal recombination–generation.

Photogeneration

Light striking the surface of a semiconductor as pictured in Fig. 3.19 will be partially reflected and partially transmitted into the material. Assume the light to be monochromatic with wavelength λ and frequency ν. If the photon energy ($h\nu$) is greater than the band gap energy, then the light will be absorbed and electron–hole pairs will be created as the light passes through the semiconductor. The decrease in intensity as monochromatic light passes through a material is given by

$$I = I_0 e^{-\alpha x} \tag{3.26}$$

where I_0 is the light intensity just inside the material at $x = 0^+$ and α is the *absorption coefficient.* Note from Fig. 3.20 that α is material dependent and a strong function of λ. Since there is a one-to-one correspondence between the absorption of photons and the creation of electron–hole pairs, the carrier creation rate should likewise exhibit an $\exp(-\alpha x)$ dependence. Moreover, referring to Fig. 3.15(d), the photogeneration process always acts so as to create an *equal* number of electrons and holes. We therefore conclude

$$\left.\frac{\partial n}{\partial t}\right|_{\text{light}} = \left.\frac{\partial p}{\partial t}\right|_{\text{light}} = G_{\text{L}}(x,\lambda) \tag{3.27}$$

where, for the situation analyzed,

$$G_{\text{L}}(x,\lambda) = G_{\text{L}0} e^{-\alpha x} \tag{3.28}$$

G_{L} is the simplified symbol for the photogeneration rate (a number/cm^3-sec), while $G_{\text{L}0}$ is the photogeneration rate at $x = 0$. The $|_{\text{light}}$ designation means "due to light" and is necessary because, as we have seen, n and p can be affected by a number of processes.

In working problems involving the illumination of a sample or device, we will often invoke one of two simplifying assumptions: (i) The illumination is assumed to be uniform everywhere inside the sample and (ii) the light is assumed to be absorbed in an

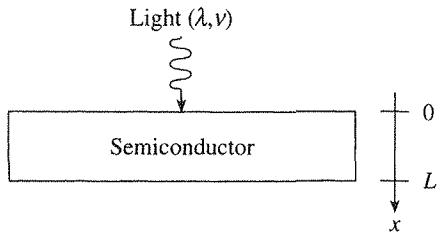


Figure 3.19 Semiconductor, light propagation, and coordinate orientation in the photogeneration analysis. The semiconductor is assumed to be sufficiently thick ($L \gg 1/\alpha$) so that bottom surface reflections may be neglected.

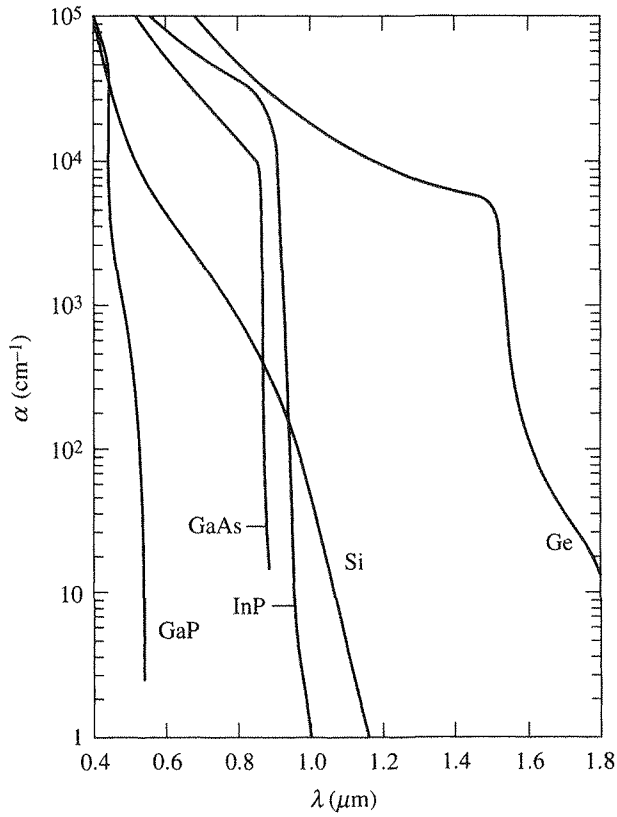


Figure 3.20 The absorption coefficient as a function of wavelength in Si and other select semiconductors. (From Schroder[6], © 1990 by John Wiley & Sons, Inc. Reprinted with permission.)

infinitesimally thin layer near the semiconductor surface. Seeking to identify when the assumed conditions are approached in practice, we note first that $1/\alpha$, which has the dimensions of a length, represents the average depth of penetration of the light into a material. At wavelengths where $h\nu \approx E_G$, the absorption coefficient is very small and $1/\alpha$ is very large, about 1 cm in Si. This makes $\exp(-\alpha x) \cong 1$ for all practical depths. To a very good approximation, $G_L \cong$ constant and the light is being uniformly absorbed. On the other hand, if $\lambda = 0.4\ \mu$m, an $\alpha(\text{Si}) \cong 10^5\ \text{cm}^{-1}$ is deduced from Fig. 3.20. At this wavelength the light penetration depth is only a few thousand angstroms and the assumption of surface absorption is closely approximated.

Indirect Thermal Recombination–Generation

Although the general case development of R–G center statistics is beyond the scope of this text, it is possible to present arguments leading to the special-case expression employed in many practical problems. Let us begin by carefully defining the various carrier and center concentrations involved in the analysis.

n_0, p_0	...	carrier concentrations in the material under analysis when equilibrium conditions prevail.
n, p	...	carrier concentrations in the material under arbitrary conditions.
$\Delta n \equiv n - n_0$ ... $\Delta p \equiv p - p_0$		deviations in the carrier concentrations from their equilibrium values. Δn and Δp can be both positive and negative, where a positive deviation corresponds to a carrier excess and a negative deviation corresponds to a carrier deficit.
N_T	...	number of R–G centers/cm^3.

In the special-case development, perturbation of the carrier concentrations from their equilibrium values must correspond to *low-level injection.* This requirement is just an alternative way of saying the perturbation must be relatively small. To be more precise,

low-level injection implies		
$\Delta p \ll n_0$,	$n \simeq n_0$	in an n-type material
$\Delta n \ll p_0$,	$p \simeq p_0$	in a p-type material

Consider a specific example of $N_D = 10^{14}/\text{cm}^3$ doped Si at room temperature subject to a perturbation where $\Delta p = \Delta n = 10^9/\text{cm}^3$. For the given material, $n_0 \simeq N_D = 10^{14}/\text{cm}^3$ and $p_0 \simeq n_i^2/N_D \simeq 10^6/\text{cm}^3$. Consequently, $n = n_0 + \Delta n \simeq n_0$ and $\Delta p = 10^9/\text{cm}^3 \ll n_0 = 10^{14}/\text{cm}^3$. The situation is clearly one of low-level injection. Observe, however, that $\Delta p \gg p_0$. Although the majority carrier concentration remains essentially unperturbed under low-level injection, the minority carrier concentration can, and routinely does, increase by many orders of magnitude.

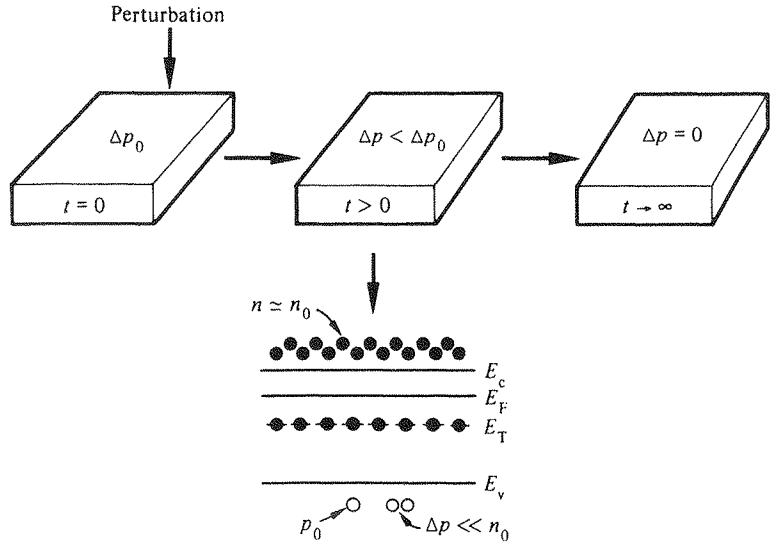


Figure 3.21 Situation inside an n-type semiconductor after a perturbation causing the low-level injection of excess holes.

We are now ready to analyze the specific situation envisioned in Fig. 3.21. A perturbation has caused a $\Delta p \ll n_0$ excess of holes in an n-type semiconductor sample. Assuming the system is in the midst of relaxing back to the equilibrium state from the pictured perturbed state via the R–G center interaction, what factors would you expect to have the greatest effect on $\partial p/\partial t|_R$, the rate of hole recombination? Well, to eliminate a hole, the hole must make the transition from the valence band to an electron-filled R–G center. Logically, the greater the number of filled R–G centers, the greater the probability of a hole annihilating transition and the faster the rate of recombination. Under equilibrium conditions, essentially all of the R–G centers are filled with electrons because the equilibrium Fermi level is comfortably above E_T. With $\Delta p \ll n_0$, electrons always vastly outnumber holes and rapidly fill R–G levels that become vacant. This keeps the number of filled R–G centers during the relaxation process $\approx N_T$. Thus, we expect $\partial p/\partial t|_R$ to be approximately proportional to N_T. Moreover, it also seems logical that the number of hole-annihilating transitions should increase almost linearly with the number of holes. All other factors being equal, the more holes available for annihilation, the greater the number of holes removed from the valence band per second. Consequently, we also expect $\partial p/\partial t|_R$ to be approximately proportional to p. Considering additional factors leads to no additional dependencies. Hence, introducing a positive proportionality constant, c_p, and realizing $\partial p/\partial t|_R$ is negative because p is decreasing, we conclude

$$\left.\frac{\partial p}{\partial t}\right|_R = -c_p N_T p \tag{3.29}$$

If the hole generation process is analyzed in a similar manner, $\partial p/\partial t|_G$ is deduced to depend only on the number of empty R–G centers. This number, which is small for the situation under consideration, is held approximately constant at its equilibrium value over the range of perturbations where hole generation is significant. This means that $\partial p/\partial t|_G$ can be replaced by $\partial p/\partial t|_{\text{G-equilibrium}}$. Moreover, the recombination and generation rates must precisely balance under equilibrium conditions,† or $\partial p/\partial t|_G = \partial p/\partial t|_{\text{G-equilibrium}} = -\partial p/\partial t|_{\text{R-equilibrium}}$. Thus, making use of Eq. (3.29), we obtain

$$\left.\frac{\partial p}{\partial t}\right|_G = c_p N_T p_0 \tag{3.30}$$

Finally, the net rate of change in the hole concentration due to the R–G center interaction is given by

$$\left.\frac{\partial p}{\partial t}\right|_{\substack{\text{i-thermal}\\ \text{R–G}}} = \left.\frac{\partial p}{\partial t}\right|_R + \left.\frac{\partial p}{\partial t}\right|_G = -c_p N_T(p - p_0) \tag{3.31}$$

or

$$\left.\frac{\partial p}{\partial t}\right|_{\substack{\text{i-thermal}\\ \text{R–G}}} = -c_p N_T \Delta p \qquad \text{for holes in an } n\text{-type material} \tag{3.32a}$$

An analogous set of arguments yields

$$\left.\frac{\partial n}{\partial t}\right|_{\substack{\text{i-thermal}\\ \text{R–G}}} = -c_n N_T \Delta n \qquad \text{for electrons in a } p\text{-type material} \tag{3.32b}$$

c_n like c_p is a positive proportionality constant. c_n and c_p are referred to as the electron and hole *capture coefficients,* respectively.

Although completely functional as is, it is desirable to manipulate Eqs. (3.32) into a slightly more compact and meaningful form. An examination of the left-hand sides of Eqs. (3.32) reveals the dimensional units are those of a concentration divided by time. Since Δp and Δn on the right-hand sides of the same equations are also concentrations, the constants $c_p N_T$ and $c_n N_T$ must have units of 1/time. It is therefore reasonable to introduce the time constants

† Under equilibrium conditions each fundamental process and its inverse must self-balance independent of any other process that may be occurring inside the material. This is known as the *Principle of Detailed Balance.*

$$\boxed{\begin{aligned} \tau_p &= \frac{1}{c_p N_T} \qquad (3.33a) \\ \tau_n &= \frac{1}{c_n N_T} \qquad (3.33b) \end{aligned}}$$

which, when substituted into Eqs. (3.32), yield

$$\boxed{\left.\frac{\partial p}{\partial t}\right|_{\substack{\text{i-thermal}\\ \text{R–G}}} = -\frac{\Delta p}{\tau_p}} \quad \text{for holes in an } n\text{-type material} \qquad (3.34a)$$

$$\boxed{\left.\frac{\partial n}{\partial t}\right|_{\substack{\text{i-thermal}\\ \text{R–G}}} = -\frac{\Delta n}{\tau_n}} \quad \text{for electrons in a } p\text{-type material} \qquad (3.34b)$$

Equations (3.34) are the desired end result, the special-case characterization of R–G center (indirect thermal) recombination–generation. Although steady state or slowly varying conditions are implicitly assumed in the development of Eqs. (3.34), the relationships can be applied with little error to most transient problems of interest. It should be restated that a $\Delta p < 0$ is possible and will give rise to a $\partial p/\partial t|_{\text{i-thermal R–G}} > 0$. A positive $\partial p/\partial t|_{\text{i-thermal R–G}}$ simply indicates that a carrier deficit exists inside the semiconductor and generation is occurring at a more rapid rate than recombination. $\partial p/\partial t|_{\text{i-thermal R–G}}$ and $\partial n/\partial t|_{\text{i-thermal R–G}}$, it must be remembered, characterize the *net* effect of the thermal recombination and thermal generation processes.

As duly emphasized, the Eq. (3.34) relationships apply only to minority carriers and to situations meeting the low-level injection requirement. The more general steady state result[7], valid for arbitrary injection levels and both carrier types in a nondegenerate semiconductor, is noted for completeness and future reference to be

$$\left.\frac{\partial p}{\partial t}\right|_{\substack{\text{i-thermal}\\ \text{R–G}}} = \left.\frac{\partial n}{\partial t}\right|_{\substack{\text{i-thermal}\\ \text{R–G}}} = \frac{n_i^2 - np}{\tau_p(n + n_1) + \tau_n(p + p_1)} \qquad (3.35)$$

where

$$n_1 \equiv n_i e^{(E_T - E_i)/kT} \qquad (3.36a)$$

$$p_1 \equiv n_i e^{(E_i - E_T)/kT} \qquad (3.36b)$$

It is left as an exercise to show that the general-case relationship reduces to Eqs. (3.34) under the assumed special-case conditions.

3.3.4 Minority Carrier Lifetimes

General Information

The time constants τ_n and τ_p, which were introduced without comment in writing down Eqs. (3.34), are obviously the "constants of the action" associated with recombination–generation. Seeking to provide insight relative to the standard interpretation and naming of the τ's, let us consider once again the situation pictured in Fig. 3.21. Examining the change in the hole concentration with time, it goes almost without saying that the excess holes do not all disappear at the same time. Rather, the hole excess present at $t = 0$ is systematically eliminated, with some of the holes existing for only a short period and others "living" for comparatively long periods of time. If thermal recombination–generation is the sole process acting to relax the semiconductor, the average excess hole lifetime, $\langle t \rangle$, can be computed in a relatively straightforward manner. Without going into details, the computation yields $\langle t \rangle = \tau_n$ (or τ_p). Physically, therefore, τ_n and τ_p have come to be interpreted as *the average time an excess minority carrier will live in a sea of majority carriers.* For identification purposes, τ_n and τ_p are simply referred to as the *minority carrier lifetimes.*

Like the μ's and the D's, the τ's are important material parameters invariably required in the modeling of devices. Unlike the μ's and the D's, there are few plots indicating the τ-values to be employed in a given device structure. In fact, subsidiary experimental measurements (see the following subsection) must be performed to determine the minority carrier lifetime in a given semiconductor sample. The reason for the lack of catalogued information can be traced to the extreme variability of the τ_n and τ_p parameters. Referring to Eqs. (3.33), we see that the carrier lifetimes depend on the often poorly controlled R–G center concentration (N_T), *not* on the carefully controlled doping parameters (N_A and N_D). Moreover, the physical nature of the dominant R–G center concentration changes even within a given sample during device fabrication. A fabrication procedure called *gettering* can reduce the R–G center concentration to a very low level and give rise to a $\tau_n(\tau_p) \sim$ 1 msec in Si. The intentional introduction of gold into Si, on the other hand, can controllably increase the R–G center concentration and give rise to a $\tau_n(\tau_p) \sim 1$ nsec. Minority carrier lifetimes in completed Si devices tend to lie about midway between the cited extremes.

A Lifetime Measurement

The carrier lifetimes are measured using a variety of different methods. The majority of measurements employ a device structure of some sort and deduce τ_n and/or τ_p by matching the theoretical and experimentally observed device characteristics. The measurement to be described employs a bar-like piece of semiconductor with ohmic contacts at its two ends. In a commercial implementation this structure is called a *photoconductor,* a type of photodetector. A situation very similar to that envisioned in Fig. 3.21 is created inside the semi-

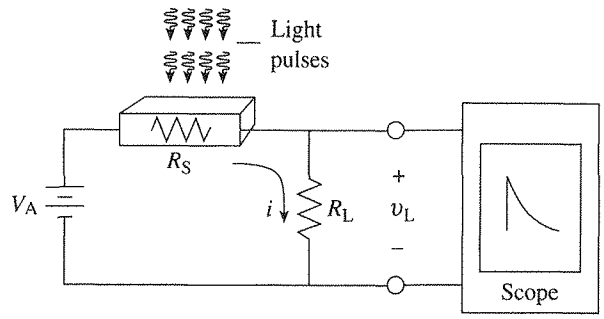


Figure 3.22 Schematic illustration of the photoconductivity decay measurement.

conductor during the measurement. The perturbation is in the form of a light pulse that gives rise to a carrier excess inside the semiconductor and a detectable change in conductivity. Once the light is removed the excess carriers are eliminated via recombination and the conductivity decays back to its value in the dark. The carrier lifetime is determined by monitoring the conductivity decay.

The experiment is schematically illustrated in Fig. 3.22. R_S is the sample resistance, R_L a load resistor, V_A the applied d.c. voltage, and v_L the measured load or output voltage. Obviously, R_S and therefore v_L are functions of time, with v_L reflecting the changes in conductivity inside the semiconductor sample.

We are specifically interested in how v_L changes with time after the light is turned off. v_L is related to the conductivity, which in turn is related to the excess carrier concentrations. Given an n-type sample and assuming uniform photogeneration throughout the semiconductor, we will verify later in the chapter (Section 3.5) that the decay of the excess minority carrier hole concentration (Δp) is described by

$$\Delta p(t) = \Delta p_0 e^{-t/\tau_p} \tag{3.37}$$

Δp_0 is the excess hole concentration at $t = 0$ when the light is turned off. Next considering the conductivity, we note

$$\sigma = 1/\rho = q(\mu_n n + \mu_p p) \tag{3.38a}$$

$$= q[\mu_n(n_0 + \Delta n) + \mu_p(p_0 + \Delta p)] \tag{3.38b}$$

$$= \underbrace{q\mu_n N_D}_{\sigma_{dark}} + \underbrace{q(\mu_n + \mu_p)\Delta p}_{\Delta\sigma(t)} \tag{3.39}$$

In establishing Eq. (3.39), we made use of the fact that $n_0 \cong N_D \gg p_0$ in an n-type semiconductor. Also Δn was assumed equal to Δp. σ_{dark} is the unperturbed sample conductivity in the dark and $\Delta\sigma(t) \propto \Delta p$ represents the time decaying component of the conductivity.

The last step is to relate v_L and $\Delta\sigma$. Inspecting Fig. 3.22, one concludes

$$v_L = iR_L = V_A \frac{R_L}{R_L + R_S} \tag{3.40}$$

but

$$R_S = \frac{\rho l}{A} = \frac{l}{\sigma A} = \frac{l}{(\sigma_{dark} + \Delta\sigma)A} = R_{Sd}\left(\frac{\sigma_{dark}}{\sigma_{dark} + \Delta\sigma}\right) \tag{3.41}$$

where l is the length, A the cross-sectional area, and $R_{Sd} = l/\sigma_{dark}A$ the dark resistance of the sample, respectively. Substituting Eq. (3.41) into Eq. (3.40) and rearranging yields

$$v_L = \frac{V_A}{1 + \dfrac{R_{Sd}/R_L}{1 + \Delta\sigma/\sigma_{dark}}} \tag{3.42}$$

or, if R_L is matched to R_{Sd} and the light intensity restricted so $\Delta\sigma/\sigma_{dark} \ll 1$ (equivalent to specifying low-level injection conditions), we obtain

$$v_L \cong \frac{V_A}{2}\left(1 + \frac{\Delta\sigma}{\sigma_{dark}}\right) = \underbrace{V_L}_{\text{d.c.}} + \underbrace{v_0 e^{-t/\tau_p}}_{\text{decay transient}} \tag{3.43}$$

An exponential decay is predicted with τ_p being the decay constant.

An actual implementation of the measurement is shown in Fig. 3.23.† The required V_A is supplied by the Tektronix PS5004 power supply (just about any power supply would do), while the Tektronix 11401 Digitizing Oscilloscope is used to capture the photoconductive decay. The bar-shaped Si test structures are housed in a special measurements box along with the $R_L \cong R_{Sd}$ load resistor. The box permits convenient handling of the samples and facilitates switching between samples. The light pulses required in the measurement are derived from the General Radio 1531-AB Stroboscope. The GR Stroboscope puts out high-intensity light flashes of $\sim 1\ \mu$sec duration spaced at controllable time intervals ranging from 2.5 msec to 0.5 sec. Fairly uniform photogeneration can be achieved in a sample

† This being the first text reference, it is perhaps worthwhile reiterating the General Introduction statement that measurement particulars, included at various points in the text, are derived from experiments performed in an undergraduate EE laboratory administered by the author. A somewhat expanded discussion of background theory, the measurement system, and measurement procedures can be found in R. F. Pierret, *Semiconductor Measurements Laboratory Operations Manual,* Supplement A, in the Modular Series on Solid State Devices, Addison-Wesley Publishing Co., Reading MA, © 1991.

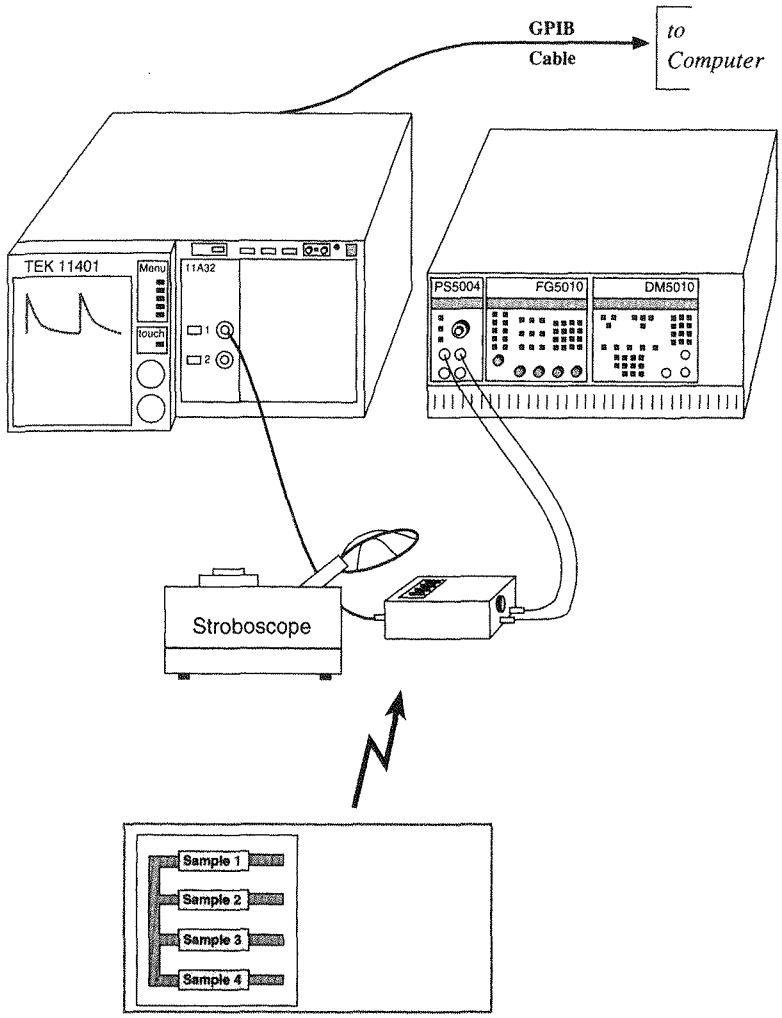


Figure 3.23 Photoconductive decay measurement system.

by first passing the stroboscope light through a small-diameter Si wafer. The Si wafer acts as a filter allowing only near-bandgap radiation to reach the sample under test. Although the 11401 Oscilloscope is capable of in situ data analysis, a computer is connected to the oscilloscope to add data manipulation and display flexibility.

An example of the photoconductive decay as displayed on the computer screen is reproduced in Fig. 3.24. The transient component of v_L in millivolts is plotted on the y-axis versus time in milliseconds on the x-axis. The part (a) linear-scale plot nicely exhibits the generally expected dependence. The part (b) semilog plot verifies the simple exponential nature of the decay (at least for times 0.1 msec $\leq t \leq$ 0.6 msec) and is more useful for

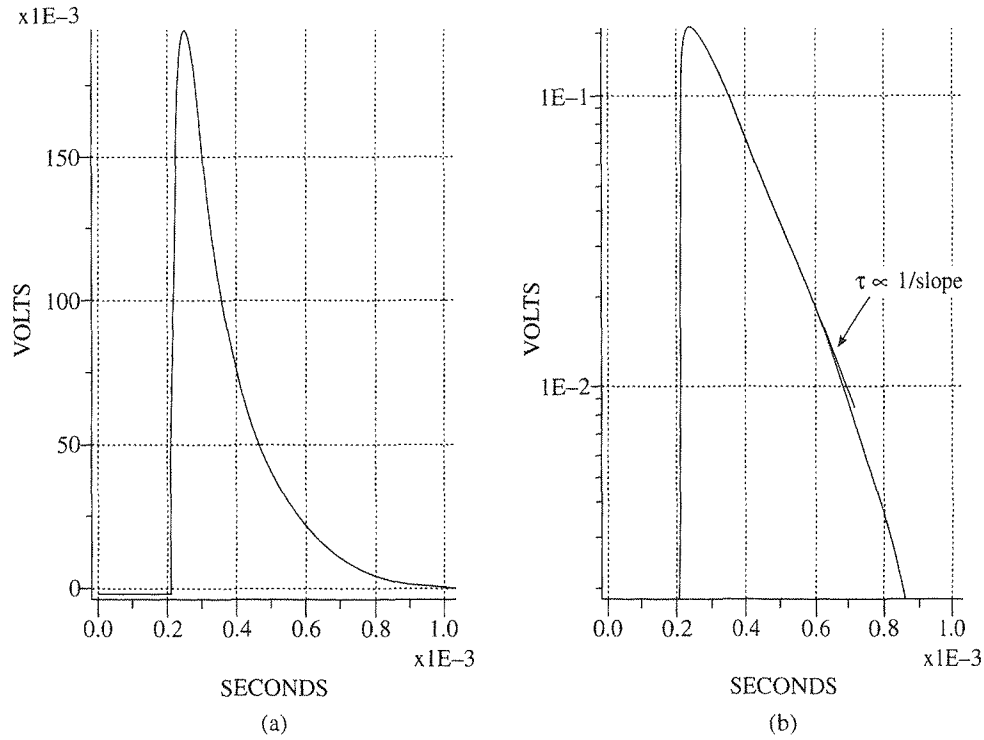


Figure 3.24 Photoconductivity transient response. The time-varying component of v_L in millivolts is plotted on the y-axis versus time in milliseconds on the x-axis. (a) is a linear and (b) is a semilog plot.

extracting the minority carrier lifetime. Specifically, from the slope of the straight-line region in the latter plot, the minority carrier lifetime in the probed sample is readily deduced to be $\tau_p \cong 150\ \mu$sec.

3.4 EQUATIONS OF STATE

In the first three sections of this chapter we examined and separately modeled the primary types of carrier action taking place inside semiconductors. Within actual semiconductors all the various types of carrier action occur at the same time, and the state of any semiconductor can be determined only by taking into account the combined effect of the individual types of carrier action. "Putting it all together," so to speak, leads to the basic set of starting equations employed in solving device problems, herein referred to as the *equations of state.* The first portion of this section is devoted to developing the equations of state; the latter portion includes a summary of common simplifications, a listing of special-case solutions, and problem-solving examples.

3.4.1 Continuity Equations

Each and every type of carrier action—whether it be drift, diffusion, indirect or direct thermal recombination, indirect or direct generation, or some other type of carrier action—gives rise to a change in the carrier concentrations with time. The combined effect of all types of carrier action can therefore be taken into account by equating the overall change in the carrier concentrations per unit time ($\partial n/\partial t$ or $\partial p/\partial t$) to the sum of the $\partial n/\partial t$'s or $\partial p/\partial t$'s due to the individual processes; that is,

$$\frac{\partial n}{\partial t} = \left.\frac{\partial n}{\partial t}\right|_{\text{drift}} + \left.\frac{\partial n}{\partial t}\right|_{\text{diff}} + \left.\frac{\partial n}{\partial t}\right|_{\substack{\text{thermal}\\ \text{R–G}}} + \left.\frac{\partial n}{\partial t}\right|_{\substack{\text{other processes}\\ \text{(light, etc.)}}} \tag{3.44a}$$

$$\frac{\partial p}{\partial t} = \left.\frac{\partial p}{\partial t}\right|_{\text{drift}} + \left.\frac{\partial p}{\partial t}\right|_{\text{diff}} + \left.\frac{\partial p}{\partial t}\right|_{\substack{\text{thermal}\\ \text{R–G}}} + \left.\frac{\partial p}{\partial t}\right|_{\substack{\text{other processes}\\ \text{(light, etc.)}}} \tag{3.44b}$$

The overall effect of the individual processes is established, in essence, by invoking the requirement of conservation of carriers. Electrons and holes cannot mysteriously appear and disappear at a given point, but must be transported to or created at the given point via some type of ongoing carrier action. There must be a spatial and time continuity in the carrier concentrations. For this reason, Eqs. (3.44) are known as the *continuity equations.*

The continuity equations can be written in a somewhat more compact form by noting

$$\left.\frac{\partial n}{\partial t}\right|_{\text{drift}} + \left.\frac{\partial n}{\partial t}\right|_{\text{diff}} = \frac{1}{q}\left(\frac{\partial J_{\text{Nx}}}{\partial x} + \frac{\partial J_{\text{Ny}}}{\partial y} + \frac{\partial J_{\text{Nz}}}{\partial z}\right) = \frac{1}{q}\nabla \cdot \mathbf{J}_{\text{N}} \tag{3.45a}$$

$$\left.\frac{\partial p}{\partial t}\right|_{\text{drift}} + \left.\frac{\partial p}{\partial t}\right|_{\text{diff}} = -\frac{1}{q}\left(\frac{\partial J_{\text{Px}}}{\partial x} + \frac{\partial J_{\text{Py}}}{\partial y} + \frac{\partial J_{\text{Pz}}}{\partial z}\right) = -\frac{1}{q}\nabla \cdot \mathbf{J}_{\text{P}} \tag{3.45b}$$

Equations (3.45), which can be established by a straightforward mathematical manipulation, merely state that there will be a change in the carrier concentrations within a given small region of the semiconductor if an imbalance exists between the total carrier currents into and out of the region. Utilizing Eqs. (3.45), we obtain

$$\boxed{\begin{aligned} \frac{\partial n}{\partial t} &= \frac{1}{q}\nabla \cdot \mathbf{J}_{\text{N}} + \left.\frac{\partial n}{\partial t}\right|_{\substack{\text{thermal}\\ \text{R–G}}} + \left.\frac{\partial n}{\partial t}\right|_{\substack{\text{other}\\ \text{processes}}} && (3.46a)\\ \frac{\partial p}{\partial t} &= -\frac{1}{q}\nabla \cdot \mathbf{J}_{\text{P}} + \left.\frac{\partial p}{\partial t}\right|_{\substack{\text{thermal}\\ \text{R–G}}} + \left.\frac{\partial p}{\partial t}\right|_{\substack{\text{other}\\ \text{processes}}} && (3.46b)\end{aligned}}$$

The (3.46) continuity equations are completely general and directly or indirectly constitute the starting point in most device analyses. In computer simulations the continuity

equations are often employed directly. The appropriate relationships for $\partial n/\partial t|_{\text{thermal R-G}}$, $\partial p/\partial t|_{\text{thermal R-G}}$ [which may or may not be the special-case relationships given by Eqs. (3.34)], along with the concentration changes due to "other processes," are substituted into Eqs. (3.46), and numerical solutions are sought for $n(x, y, z, t)$ and $p(x, y, z, t)$. In problems where a closed-form type of solution is desired, the continuity equations are typically utilized only in an indirect fashion. The actual starting point in such analyses is a simplified version of the continuity equations to be established in the next subsection.

3.4.2 Minority Carrier Diffusion Equations

The "workhorse" minority carrier diffusion equations are derived from the continuity equations by invoking the following set of simplifying assumptions:

(1) The particular system under analysis is *one-dimensional;* i.e., all variables are at most a function of just one coordinate (say the x-coordinate).

(2) The analysis is limited or restricted to *minority carriers.*

(3) $\mathscr{E} \simeq 0$ in the semiconductor or regions of the semiconductor subject to analysis.

(4) The equilibrium minority carrier concentrations are not a function of position. In other words, $n_0 \neq n_0(x)$, $p_0 \neq p_0(x)$.

(5) *Low-level injection* conditions prevail.

(6) *Indirect* thermal recombination–generation is the dominant thermal R–G mechanism.

(7) There are *no "other processes,"* except possibly photogeneration, taking place within the system.

Working on the continuity equation for electrons, we note that

$$\frac{1}{q}\nabla \cdot \mathbf{J}_{\mathrm{N}} \rightarrow \frac{1}{q}\frac{\partial J_{\mathrm{N}}}{\partial x} \tag{3.47}$$

if the system is one-dimensional. Moreover,

$$J_{\mathrm{N}} = q\mu_{\mathrm{n}} n\mathscr{E} + qD_{\mathrm{N}}\frac{\partial n}{\partial x} \simeq qD_{\mathrm{N}}\frac{\partial n}{\partial x} \tag{3.48}$$

when $\mathscr{E} \simeq 0$ and one is concerned only with minority carriers [simplifications (2) and (3)]. By way of explanation, the drift component can be neglected in the current density expression because $\mathscr{E}$ is small by assumption and minority carrier concentrations are also small, making the $n\mathscr{E}$ product extremely small. (Note that the same argument cannot be applied to majority carriers.) Since by assumption $n_0 \neq n_0(x)$, and by definition $n = n_0 + \Delta n$, we can also write

$$\frac{\partial n}{\partial x} = \frac{\partial n_0}{\partial x} + \frac{\partial \Delta n}{\partial x} = \frac{\partial \Delta n}{\partial x} \tag{3.49}$$

Combining Eqs. (3.47) through (3.49) yields

$$\frac{1}{q}\nabla \cdot \mathbf{J}_{\mathrm{N}} \rightarrow D_{\mathrm{N}}\frac{\partial^2 \Delta n}{\partial x^2} \tag{3.50}$$

Turning to the remaining terms in the continuity equation for electrons, the assumed dominance of recombination–generation via R–G centers, combined with the low-level injection and minority carrier restrictions, allows us to replace the thermal R–G term with the Eq. (3.34) special-case expression.

$$\left.\frac{\partial n}{\partial t}\right|_{\substack{\text{thermal}\\ \text{R–G}}} = -\frac{\Delta n}{\tau_{\mathrm{n}}} \tag{3.51}$$

In addition, applying simplification (7) yields

$$\left.\frac{\partial n}{\partial t}\right|_{\substack{\text{other}\\ \text{processes}}} = G_{\mathrm{L}} \tag{3.52}$$

where it is understood that $G_{\mathrm{L}} = 0$ if the semiconductor is not subject to illumination. Finally, the equilibrium electron concentration is never a function of time, $n_0 \neq n_0(t)$, and we can therefore write

$$\frac{\partial n}{\partial t} = \frac{\partial n_0}{\partial t} + \frac{\partial \Delta n}{\partial t} = \frac{\partial \Delta n}{\partial t} \tag{3.53}$$

Substituting Eqs. (3.50) through (3.53) into the (3.46a) continuity equation, and simultaneously recording the analogous result for holes, one obtains

$$\boxed{\begin{aligned} \frac{\partial \Delta n_{\mathrm{p}}}{\partial t} &= D_{\mathrm{N}}\frac{\partial^2 \Delta n_{\mathrm{p}}}{\partial x^2} - \frac{\Delta n_{\mathrm{p}}}{\tau_{\mathrm{n}}} + G_{\mathrm{L}} \\ \frac{\partial \Delta p_{\mathrm{n}}}{\partial t} &= D_{\mathrm{P}}\frac{\partial^2 \Delta p_{\mathrm{n}}}{\partial x^2} - \frac{\Delta p_{\mathrm{n}}}{\tau_{\mathrm{p}}} + G_{\mathrm{L}} \end{aligned}} \quad \begin{matrix} \text{Minority carrier} \\ \text{diffusion equations} \end{matrix} \quad \begin{matrix} (3.54\text{a}) \\ (3.54\text{b}) \end{matrix}$$

Subscripts are added to the carrier concentrations in Eqs. (3.54) to remind the user that the equations are valid only for minority carriers, applying to electrons in p-type materials and to holes in n-type materials.

Table 3.1 Common Diffusion Equation Simplifications.

Simplification	*Effect*
Steady state	$\frac{\partial \Delta n_p}{\partial t} \to 0 \quad \left(\frac{\partial \Delta p_n}{\partial t} \to 0\right)$
No concentration gradient or no diffusion current	$D_N \frac{\partial^2 \Delta n_p}{\partial x^2} \to 0 \quad \left(D_P \frac{\partial^2 \Delta p_n}{\partial x^2} \to 0\right)$
No drift current or $\mathscr{E} = 0$	No further simplification. ($\mathscr{E} \simeq 0$ is assumed in the derivation.)
No thermal R–G	$\frac{\Delta n_p}{\tau_n} \to 0 \quad \left(\frac{\Delta p_n}{\tau_p} \to 0\right)$
No light	$G_L \to 0$

3.4.3 Simplifications and Solutions

In the course of performing device analyses, the conditions of a problem often permit additional simplifications that drastically reduce the complexity of the minority carrier diffusion equations. Common simplifications and their effect on the minority carrier diffusion equations are summarized in Table 3.1. Extensively utilized solutions to simplified forms of the minority carrier diffusion equations are collected for future reference in Table 3.2.

3.4.4 Problem Solving

Although we have presented all necessary information, it may not be completely obvious how one works with the diffusion equations to obtain the carrier concentration solutions. Device analyses of course provide examples of "problem" solutions, and a number of device analyses are presented later in the text. Nevertheless, it is worthwhile to examine a few well-defined simple problems to illustrate problem-solving procedures. The chosen sample problems also provide a basis for the two supplemental concepts introduced in the next section.

Sample Problem No. 1

P: A uniformly donor-doped silicon wafer maintained at room temperature is suddenly illuminated with light at time $t = 0$. Assuming $N_D = 10^{15}/\text{cm}^3$, $\tau_p = 10^{-6}$ sec, and a light-induced creation of 10^{17} electrons and holes per cm^3-sec throughout the semiconductor, determine $\Delta p_n(t)$ for $t > 0$.

S: *Step 1*—Review precisely what information is given or implied in the statement of the problem.

Table 3.2 Common Special-Case Diffusion Equation Solutions.

Solution no. 1

GIVEN: Steady state, no light.

SIMPLIFIED DIFF. EQN:
$$0 = D_N \frac{d^2\Delta n_p}{dx^2} - \frac{\Delta n_p}{\tau_n}$$

SOLUTION:
$$\Delta n_p(x) = Ae^{-x/L_N} + Be^{x/L_N}$$
where $L_N \equiv \sqrt{D_n \tau_n}$
and A, B are solution constants.

Solution no. 2

GIVEN: No concentration gradient, no light.

SIMPLIFIED DIFF. EQN:
$$\frac{d\Delta n_p}{dt} = -\frac{\Delta n_p}{\tau_n}$$

SOLUTION:
$$\Delta n_p(t) = \Delta n_p(0)e^{-t/\tau_n}$$

Solution no. 3

GIVEN: Steady state, no concentration gradient.

SIMPLIFIED DIFF. EQN:
$$0 = -\frac{\Delta n_p}{\tau_n} + G_L$$

SOLUTION:
$$\Delta n_p = G_L \tau_n$$

Solution no. 4

GIVEN: Steady state, no R–G, no light.

SIMPLIFIED DIFF. EQN:
$$0 = D_N \frac{d^2\Delta n_p}{dx^2} \quad \text{or} \quad 0 = \frac{d^2\Delta n_p}{dx^2}$$

SOLUTION:
$$\Delta n_p(x) = A + Bx$$

The semiconductor is silicon, $T = 300$ K, the donor doping is the same everywhere with $N_D = 10^{15}/\text{cm}^3$, and $G_L = 10^{17}/\text{cm}^3\text{-sec}$ at all points inside the semiconductor. Also, the statement of the problem *implies* equilibrium conditions exist for $t < 0$.

Step 2—Characterize the system under equilibrium conditions.

In Si at room temperature $n_i = 10^{10}/\text{cm}^3$. Since $N_D \gg n_i$, $n_0 = N_D = 10^{15}/\text{cm}^3$ and $p_0 = n_i^2/N_D = 10^5/\text{cm}^3$. With the doping being uniform, the equilibrium n_0 and p_0 values are the same everywhere throughout the semiconductor.

Step 3—Analyze the problem qualitatively.

Prior to $t = 0$, equilibrium conditions prevail and $\Delta p_n = 0$. Starting at $t = 0$ the light creates added electrons and holes and Δp_n will begin to increase. The growing excess

carrier numbers, however, in turn lead to an increased indirect thermal recombination rate which is proportional to Δp_n. Consequently, as Δp_n grows as a result of photogeneration, more and more of the excess holes are eliminated per second by recombination through R–G centers. Eventually, a point is reached where the carriers annihilated per second by indirect thermal recombination balance the carriers created per second by the light, and a steady state condition is attained.

Summarizing, we expect $\Delta p_n(t)$ to start from zero at $t = 0$, to build up at a decreasing rate, and to ultimately become constant. Since the light generation and thermal recombination rates must balance under steady state conditions, we can even state $G_L = \Delta p_n(t \to \infty)/\tau_p$ or $\Delta p_n(t \to \infty) = \Delta p_{n|max} = G_L \tau_p$ if low-level injection prevails.

Step 4—Perform a quantitative analysis.

The minority carrier diffusion equation is the starting point for most first-order quantitative analyses. After examining the problem for obvious conditions that would invalidate the use of the diffusion equation, the appropriate minority carrier diffusion equation is written down, the equation is simplified, and a solution is sought subject to boundary conditions stated or implied in the problem.

For the problem under consideration a cursory inspection reveals that all simplifying assumptions involved in deriving the diffusion equations are readily satisfied. Specifically, only the minority carrier concentration is of interest; the equilibrium carrier concentrations are not a function of position, indirect thermal R–G is dominant in Si, and there are no "other processes" except for photogeneration. Because the photogeneration is uniform throughout the semiconductor, the perturbed carrier concentrations are also position-independent and the electric field $\mathscr{E}$ must clearly be zero in the perturbed system. Finally, a $\Delta p_{n|max} = G_L \tau_p = 10^{11}/\text{cm}^3 \ll n_0 = 10^{15}/\text{cm}^3$ is consistent with low-level injection prevailing at all times.

With no obstacles to utilizing the diffusion equation, the desired quantitative solution can now be obtained by solving

$$\frac{\partial \Delta p_n}{\partial t} = D_P \frac{\partial^2 \Delta p_n}{\partial x^2} - \frac{\Delta p_n}{\tau_p} + G_L \tag{3.55}$$

subject to the boundary condition

$$\Delta p_n(t)|_{t=0} = 0 \tag{3.56}$$

Since Δp_n is not a function of position, the diffusion equation becomes an ordinary differential equation and simplifies to

$$\frac{d\Delta p_n}{dt} + \frac{\Delta p_n}{\tau_p} = G_L \tag{3.57}$$

The general solution of Eq. (3.57) is

$$\Delta p_n(t) = G_L\tau_p + Ae^{-t/\tau_p} \tag{3.58}$$

Applying the boundary condition yields

$$A = -G_L\tau_p \tag{3.59}$$

and

$$\Delta p_n(t) = G_L\tau_p(1 - e^{-t/\tau_p}) \quad \Leftarrow \text{solution} \tag{3.60}$$

Step 5—Examine the solution.

Failing to examine the mathematical solution to a problem is like growing vegetables and then failing to eat the produce. Relative to the Eq. (3.60) result, $G_L\tau_p$ has the dimensions of a concentration (number/cm^3) and the solution is at least dimensionally correct. A plot of the Eq. (3.60) result is shown in Fig. 3.25. Note that, in agreement with qualitative predictions, $\Delta p_n(t)$ starts from zero at $t = 0$ and eventually saturates at $G_L\tau_p$ after a few τ_p.

Epilogue—We would be remiss if we did not point out the connection between the hypothetical problem just completed and the photoconductive decay measurement described in Subsection 3.3.4. Light output from the stroboscope used in the measurement can be modeled to first order by the pulse train pictured in Fig. 3.26(a). The Eq. (3.60) solution therefore approximately describes the carrier build-up during a light pulse. It should be noted, however, that the stroboscope light pulses have a duration of $t_{on} \cong 1$ μsec compared to a minority carrier lifetime of $\tau_p \cong 150$ μsec. With $t_{on}/\tau_p \ll 1$, one sees only the very

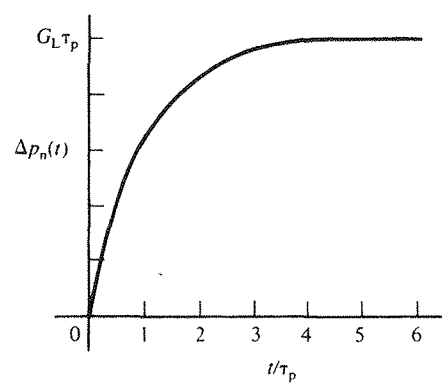


Figure 3.25 Solution to Sample Problem No. 1. Photogeneration-induced increase in the excess hole concentration as a function of time.

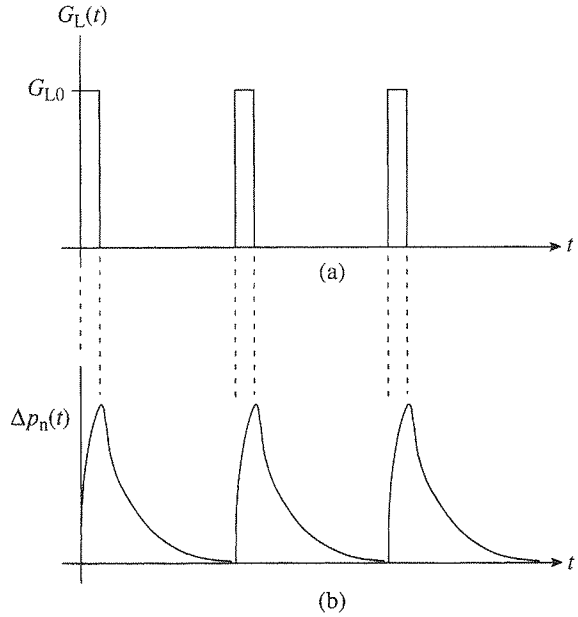


Figure 3.26 (a) Approximate model for the light output from the stroboscope in the photoconductivity decay measurement. (b) Sketch of the combined light-on/light-off solution for the excess minority-carrier concentration.

first portion of the Fig. 3.25 transient before the light is turned off and the semiconductor is allowed to decay back to equilibrium.

We are also interested in verifying the correctness of the light-off $\Delta p(t)$ expression used in the photoconductivity decay analysis. Derivation of this expression closely parallels the light-on sample problem solution. Exceptions are Δp_n at the beginning of the light-off transient equals Δp_n at the end of the light-on transient and G_L is set equal to zero in Eqs. (3.55) and (3.57). The solution to Eq. (3.57) with $G_L = 0$ is the same as Solution No. 2 in Table 3.2; i.e.,

$$\Delta p_n(t) = \Delta p_n(0)e^{-t/\tau_p} \tag{3.61}$$

where $t = 0$ has been reset to correspond to the beginning of the light-off transient. Equation (3.37) in the measurement analysis and Eq. (3.61) are of course identical. For completeness, the combined light-on and light-off $\Delta p_n(t)$ solution is sketched in Fig. 3.26(b).

Sample Problem No. 2

P: As pictured in Fig. 3.27(a), the $x = 0$ end of a uniformly doped semi-infinite bar of silicon with $N_D = 10^{15}/\text{cm}^3$ is illuminated so as to create $\Delta p_{n0} = 10^{10}/\text{cm}^3$ excess holes at

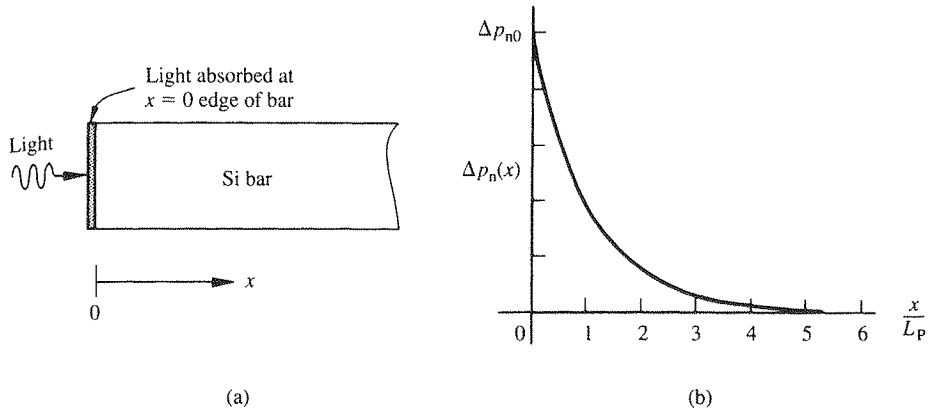


Figure 3.27 (a) Pictorial definition of Sample Problem No. 2. (b) Solution to Sample Problem No. 2 showing the excess hole concentration inside the Si bar as a function of position.

$x = 0$. The wavelength of the illumination is such that no light penetrates into the interior ($x > 0$) of the bar. Determine $\Delta p_n(x)$.

S: The semiconductor is again silicon uniformly doped with an $N_D = 10^{15}/\text{cm}^3$. Steady state conditions are inferred from the statement of the problem, since we are asked for $\Delta p_n(x)$ and not $\Delta p_n(x, t)$. Moreover, at $x = 0$, $\Delta p_n(0) = \Delta p_{n0} = 10^{10}/\text{cm}^3$, and $\Delta p_n \to 0$ as $x \to \infty$. The latter boundary condition follows from the semi-infinite nature of the bar. The perturbation in Δp_n due to the nonpenetrating light can't possibly extend out to $x = \infty$. The nonpenetrating nature of the light allows us to set $G_L = 0$ for $x > 0$. Finally, note that the problem statement fails to mention the temperature of operation. When this happens, it is reasonable to assume an intended $T = 300$ K.

If the light were removed, the silicon bar in Sample Problem 2 maintained at 300 K would revert to an equilibrium condition identical to that described in Sample Problem 1. Under equilibrium conditions, then, $n_0 = 10^{15}/\text{cm}^3$, $p_0 = 10^5/\text{cm}^3$, and the carrier concentrations are uniform throughout the semiconductor bar.

Qualitatively it is a simple matter to predict the expected effect of the nonpenetrating light on the silicon bar. The light first creates excess carriers right at $x = 0$. With more carriers at one point than elsewhere in the bar, the diffusion process next comes into play and the carrier excess spreads into the semiconductor proper. At the same time, however, the appearance of an excess hole concentration inside the bar enhances the thermal recombination rate. Thus, as the diffusing holes move into the bar their numbers are reduced by recombination. In addition, since the minority carrier holes live for only a limited period, a time τ_p on the average, fewer and fewer excess holes will survive as the depth of penetration into the bar becomes larger and larger. Under steady state conditions it is therefore reasonable to expect an excess distribution of holes near $x = 0$, with $\Delta p_n(x)$ monotonically decreasing from Δp_{n0} at $x = 0$ to $\Delta p_n = 0$ as $x \to \infty$.

In preparation for obtaining a quantitative solution, we observe that the system under consideration is one-dimensional, the analysis is restricted to the minority carrier holes, the equilibrium carrier concentrations are position-independent, indirect thermal R–G dominates, there are no "other processes" for $x > 0$, and low-level injection conditions clearly prevail ($\Delta p_{n|max} = \Delta p_{n0} = 10^{10}/\text{cm}^3 \ll n_0 = 10^{15}/\text{cm}^3$). The only question that might be raised concerning the use of the diffusion equation as the starting point for the quantitative analysis is whether $\mathscr{E} \simeq 0$. With the light on, a nonuniform distribution of holes and associated distribution of positive charge will appear near the $x = 0$ surface. The excess hole pile-up, however, is very small ($\Delta p_{n|max} \simeq n_i$) and the associated electric field is therefore expected to be correspondingly small. Moreover, in problems of this type it is found that the majority carriers, negatively charged electrons in the given problem, redistribute in such a way as to partly cancel the minority carrier charge. Thus experience indicates the $\mathscr{E} \simeq 0$ assumption to be reasonable, and use of the minority carrier diffusion equation to be justified.

Under steady state conditions with $G_L = 0$ for $x > 0$ the hole diffusion equation reduces to the form

$$D_P \frac{d^2 \Delta p_n}{dx^2} - \frac{\Delta p_n}{\tau_p} = 0 \qquad \text{for } x > 0 \tag{3.62}$$

which is to be solved subject to the boundary conditions

$$\Delta p_{n|x=0^+} = \Delta p_{n|x=0} = \Delta p_{n0} \tag{3.63}$$

and

$$\Delta p_{n|x\to\infty} = 0 \tag{3.64}$$

Equation (3.62) should be recognized as one of the simplified diffusion equations cited in Table 3.2, with the general solution

$$\Delta p_n(x) = Ae^{-x/L_P} + Be^{x/L_P} \tag{3.65}$$

where

$$L_P \equiv \sqrt{D_P \tau_p} \tag{3.66}$$

Because $\exp(x/L_P) \to \infty$ as $x \to \infty$, the only way that the Eq. (3.64) boundary condition can be satisfied is for B to be identically zero. With $B = 0$, application of the Eq. (3.63) boundary condition yields

$$A = \Delta p_{n0} \tag{3.67}$$

and

$$\Delta p_n(x) = \Delta p_{n0} e^{-x/L_P} \quad \Leftarrow \text{solution} \tag{3.68}$$

The Eq. (3.68) result is plotted in Fig. 3.27(b). In agreement with qualitative arguments, the nonpenetrating light merely gives rise to a monotonically decreasing $\Delta p_n(x)$ starting from Δp_{n0} at $x = 0$ and decreasing to $\Delta p_n = 0$ as $x \to \infty$. Note that the precise functional form of the falloff in the excess carrier concentration is exponential with a characteristic decay length equal to L_P.

3.5 SUPPLEMENTAL CONCEPTS

3.5.1 Diffusion Lengths

The situation just encountered in Sample Problem No. 2—the creation (or appearance) of an excess of minority carriers along a given plane in a semiconductor, the subsequent diffusion of the excess from the point of injection, and the exponential falloff in the excess carrier concentration characterized by a decay length (L_P)—occurs often enough in semiconductor analyses that the characteristic length has been given a special name. Specifically,

$$L_P \equiv \sqrt{D_P \tau_p} \quad \text{associated with the minority carrier holes in an } n\text{-type material} \tag{3.69a}$$

and

$$L_N \equiv \sqrt{D_N \tau_n} \quad \text{associated with the minority carrier electrons in a } p\text{-type material} \tag{3.69b}$$

are referred to as *minority carrier diffusion lengths.*

Physically, L_P and L_N represent the average distance minority carriers can diffuse into a sea of majority carriers before being annihilated. This interpretation is clearly consistent with Sample Problem No. 2, where the average position of the excess minority carriers inside the semiconductor bar is

$$\langle x \rangle = \int_0^\infty x \Delta p_n(x) dx \Big/ \int_0^\infty \Delta p_n(x) dx = L_P \tag{3.70}$$

For memory purposes, minority carrier diffusion into a sea of majority carriers might be likened to a small group of animals attempting to cross a piranha-infested stretch of the Amazon River. In the analogy, L_P and L_N correspond to the average distance the animals advance into the river before being eaten.

Seeking an idea as to the size of diffusion lengths, let us assume $T = 300$ K, the semiconductor is $N_D = 10^{15}/\text{cm}^3$ doped Si, and $\tau_p = 10^{-6}$ sec. In this sample

$$L_P = \sqrt{D_P \tau_p} = \sqrt{(kT/q)\mu_p \tau_p} = [(0.0259)(458)(10^{-6})]^{1/2}$$
$$= 3.44 \times 10^{-3} \text{ cm}$$

Although the computed value is fairly representative, it should be understood that diffusion lengths can vary over several orders of magnitude because of wide variations in the carrier lifetimes.

3.5.2 Quasi-Fermi Levels

Quasi-Fermi levels are energy levels used to specify the carrier concentrations inside a semiconductor under *non*equilibrium conditions.

To understand the need for introducing quasi-Fermi levels, let us first refer to Sample Problem No. 1. In this problem, equilibrium conditions prevailed prior to $t = 0$, with $n_0 = N_D = 10^{15}/\text{cm}^3$ and $p_0 = 10^5/\text{cm}^3$. The energy band diagram describing the equilibrium situation is shown in Fig. 3.28(a). A simple inspection of the energy band diagram and the Fermi level positioning also conveys the equilibrium carrier concentrations, since

$$n_0 = n_i e^{(E_F - E_i)/kT} \tag{3.71a}$$

$$p_0 = n_i e^{(E_i - E_F)/kT} \tag{3.71b}$$

The point we wish to emphasize is that under equilibrium conditions there is a one-to-one correspondence between the Fermi level and the carrier concentrations. Knowledge of E_F completely specifies n_0 and p_0 and vice versa.

Let us turn next to the nonequilibrium (steady state) situation inside the Problem 1 semiconductor at times $t \gg \tau_p$. For times $t \gg \tau_p$, $\Delta p_n = G_L \tau_p = 10^{11}/\text{cm}^3$, $p = p_0 + \Delta p \cong 10^{11}/\text{cm}^3$, and $n \cong n_0 = 10^{15}/\text{cm}^3$. Although n remains essentially unperturbed, p has

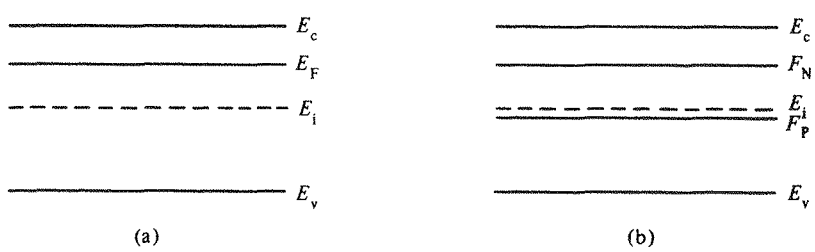


Figure 3.28 Sample use of quasi-Fermi levels. Energy band description of the situation inside the semiconductor of Sample Problem No. 1 under (a) equilibrium conditions and (b) nonequilibrium conditions ($t \gg \tau_p$).

increased by many orders of magnitude; and it is clear that the Fig. 3.28(a) diagram no longer describes the state of the system. In fact, the Fermi level is defined only for a system under equilibrium conditions and cannot be used to deduce the carrier concentrations inside a system in a nonequilibrium state.

The convenience of being able to deduce the carrier concentrations by inspection from the energy band diagram is extended to nonequilibrium conditions through the use of quasi-Fermi levels. This is accomplished by introducing two energies, F_N (the quasi-Fermi level for electrons) and F_P (the quasi-Fermi level for holes), which are *by definition* related to the nonequilibrium carrier concentrations in the same way E_F is related to the equilibrium carrier concentrations. To be specific, under nonequilibrium conditions and assuming the semiconductor to be nondegenerate,

$$n \equiv n_i e^{(F_N - E_i)/kT} \quad \text{or} \quad F_N \equiv E_i + kT \ln\left(\frac{n}{n_i}\right) \tag{3.72a}$$

and

$$p \equiv n_i e^{(E_i - F_P)/kT} \quad \text{or} \quad F_P \equiv E_i - kT \ln\left(\frac{p}{n_i}\right) \tag{3.72b}$$

Note that F_N and F_P are conceptual constructs, with the values of F_N and F_P being totally determined from a prior knowledge of n and p. Moreover, the quasi-Fermi level definitions have been carefully chosen so that when a perturbed system relaxes back toward equilibrium, $F_N \to E_F$, $F_P \to E_F$, and Eqs. (3.72) $\to$ Eqs. (3.71).

To provide a straightforward application of the quasi-Fermi level formalism, let us again consider the $t \gg \tau_p$ state of the semiconductor in Sample Problem No. 1. First, since $n \simeq n_0$, $F_N \simeq E_F$. Next, substitution of $p = 10^{11}/\text{cm}^3$ ($n_i = 10^{10}/\text{cm}^3$ and $kT = 0.0259$ eV) into Eq. (3.72b) yields $F_P = E_i - 0.06$ eV. By eliminating E_F from the energy band diagram and drawing lines at the appropriate energies to represent F_N and F_P, one obtains the diagram displayed in Fig. 3.28(b). Figure 3.28(b) clearly conveys to any observer that the system under analysis is in a nonequilibrium state. When referenced to part (a), part (b) of Fig. 3.28 further indicates at a glance that low-level injection is taking place inside the semiconductor, creating a concentration of minority carrier holes in excess of n_i. A second example use of the quasi-Fermi level formalism, a use based on Sample Problem No. 2 and involving position dependent quasi-Fermi levels, is presented in Exercise 3.5 at the end of the section.

As a final point, it should be mentioned that the quasi-Fermi level formalism can be used to recast some of the carrier-action relationships in a more compact form. For example, the standard form of the equation for the total hole current reads

$$\mathbf{J}_P = q\mu_p p\mathscr{E} - qD_P \nabla p \tag{3.73}$$

(same as 3.18a)

Differentiating both sides of Eq. (3.72b) with respect to position, one obtains

$$\nabla p = \left(\frac{n_i}{kT}\right) e^{(E_i - F_P)/kT}(\nabla E_i - \nabla F_P) \tag{3.74a}$$

$$= \left(\frac{qp}{kT}\right)\mathscr{E} - \left(\frac{p}{kT}\right)\nabla F_P \tag{3.74b}$$

The identity $\mathscr{E} = \nabla E_i/q$ (the three-dimensional version of Eq. 3.15) is employed in progressing from Eq. (3.74a) to Eq. (3.74b). Next, eliminating ∇p in Eq. (3.73) using Eq. (3.74b) gives

$$\mathbf{J}_P = q\left(\mu_p - \frac{qD_P}{kT}\right)p\mathscr{E} + \left(\frac{qD_P}{kT}\right)p\nabla F_P \tag{3.75}$$

From the Einstein relationship, however, $qD_P/kT = \mu_p$. We therefore conclude

$$\mathbf{J}_P = \mu_p p \nabla F_P \tag{3.76a}$$

Similarly,

$$\mathbf{J}_N = \mu_n n \nabla F_N \tag{3.76b}$$

Since $\mathbf{J}_P \propto \nabla F_P$ and $\mathbf{J}_N \propto \nabla F_N$ in Eqs. (3.76), one is led to a very interesting general interpretation of energy band diagrams containing quasi-Fermi levels. Namely, a quasi-Fermi level that varies with position ($dF_P/dx \neq 0$ or $dF_N/dx \neq 0$) indicates at a glance that current is flowing inside the semiconductor.

Exercise 3.5

P: In Sample Problem No. 2, nonpenetrating illumination of a semiconductor bar was found to cause a steady state, excess-hole concentration of $\Delta p_n(x) = \Delta p_{n0} \exp(-x/L_P)$. Given the prevailing low-level injection conditions, and noting that $p = p_0 + \Delta p$, we can therefore state

$$n \simeq n_0$$

$$p = p_0 + \Delta p_{n0} e^{-x/L_P}$$

for the illuminated sample.

(a) Making use of Eqs. (3.72), establish relationships for F_N and F_P in the illuminated bar.

(b) Show that F_P is a linear function of x at points where $\Delta p_n(x) \gg p_0$.

(c) Using the results of parts (a) and (b), sketch the energy band diagrams describing the semiconductor bar of Sample Problem No. 2 under equilibrium and illuminated steady state conditions. (Assume $\mathscr{E} = 0$ in the illuminated bar.)

(d) Is there a hole current in the illuminated bar under steady state conditions? Explain.

(e) Is there an electron current in the illuminated bar under steady state conditions? Explain.

S: (a) Since $n \cong n_0$, it follows from Eq. (3.72a) that $F_N \cong E_F$. Likewise, substituting the preceding p-expression into Eq. (3.72b), we conclude

$$F_P = E_i - kT \ln(p/n_i) = E_i - kT \ln[p_0/n_i + (\Delta p_{n0}/n_i)e^{-x/L_P}]$$

(b) If $\Delta p_n(x) \gg p_0$, then $(\Delta p_{n0}/n_i)\exp(-x/L_P) \gg p_0/n_i$ and

$$F_P \simeq E_i - kT \ln[(\Delta p_{n0}/n_i)e^{-x/L_P}]$$

or

$$F_P = E_i - kT \ln(\Delta p_{n0}/n_i) + kT(x/L_P)$$

(c) We know from Sample Problem No. 2 that $\Delta p_{n0} = 10^{10}/\text{cm}^3$, $n_i = 10^{10}/\text{cm}^3$, and $p_0 = n_i^2/N_D = 10^5/\text{cm}^3$. Thus

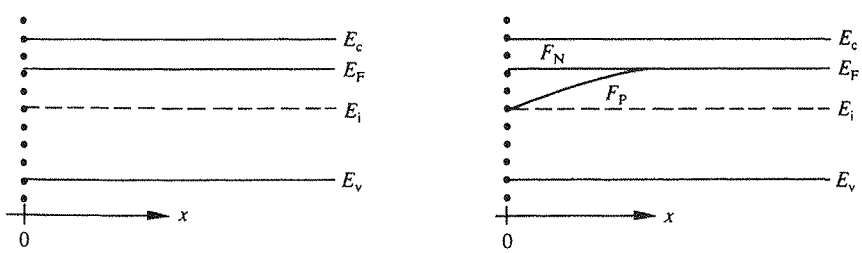
(i) Near $x = 0$, $\Delta p_n(x) \gg p_0$ and F_P is a linear function of x.

(ii) At $x = 0$, $\Delta p_{n0} = n_i$ and we deduce from the part (b) result that $F_P = E_i$.

(iii) For large x, F_P eventually approaches $F_N = E_F$.

(iv) $F_N - E_i \simeq E_F - E_i = kT \ln(N_D/n_i) = 0.30$ eV.

Utilizing the preceding information, one concludes



Equilibrium Steady state illuminated

(d) Assuming $p \neq 0$, it follows from Eq. (3.76a) that there will be a hole current whenever $dF_P/dx \neq 0$. There is obviously a hole current in the illuminated bar near $x = 0$.

(e) Appearances can sometimes be deceiving; $J_N \neq 0$ near $x = 0$! One might conclude from the part (c) result that $dF_N/dx = 0$ and therefore $J_N = 0$. Under steady state conditions, however, one must have J(total current) $= J_N + J_P =$ constant at all points in the bar. Since no current leaves the bar at $x = 0$, J moreover must vanish at $x = 0$. Thus $J_N(x) = -J_P(x)$ and we know $J_P \neq 0$ near $x = 0$. The apparent discrepancy here stems from the fact that J_N is proportional to both n and ∇F_N. Because the majority carrier electron concentration is much larger than the minority carrier hole concentration, dF_N/dx must be correspondingly smaller than dF_P/dx. The slope in F_N simply cannot be detected by inspecting the energy band diagram.

3.6 SUMMARY AND CONCLUDING COMMENTS

Most of the chapter was devoted to examining the three primary types of carrier action occurring inside of semiconductors: drift, diffusion, and recombination–generation. In each case the carrier action was first defined and visualized. Drift is charged particle motion in response to an applied electric field. Diffusion is migration from regions of high particle concentration to regions of low particle concentration due to the random thermal motion of the carriers. Recombination and generation are respectively the annihilation and creation of carriers. Next the quantitative effect of each type of carrier action was analyzed. Drift and diffusion give rise to particle currents (Eqs. 3.4 and 3.17–3.19); recombination–generation acts to change the carrier concentrations as a function of time (Eqs. 3.34). Associated with the quantitative analysis of each carrier action there arises a "constant of the motion," an important material-based parameter that specifies the magnitude of the carrier action in a given semiconductor sample. The carrier mobilities, the diffusion coefficients, and the minority carrier lifetimes are the material parameters associated with drift, diffusion, and recombination–generation, respectively. In the major semiconductors the carrier mobilities are very well characterized as a function of doping and temperature. Select mobility data are graphed in Figs. (3.5) and (3.7); an empirical fit relationship for the carrier mobilities in Si is presented in Exercise 3.1. The diffusion coefficients in nondegenerate semiconductors are computed from the carrier mobilities using the Einstein relationship (Eqs. 3.25). Conversely, an experimental measurement is normally performed to determine the carrier lifetimes in a given semiconductor sample.

Although introduced and examined individually, the various types of carrier action occur simultaneously inside semiconductors. Mathematically combining the overall effect of the various carrier activities leads to the continuity equations (Eqs. 3.46). The continuity equations in turn simplify to the minority carrier diffusion equations (Eqs. 3.54) under conditions encountered in many practical problems. Additional simplifications and their effect on the minority carrier diffusion equations are noted in Table 3.1; extensively utilized solutions to simplified forms of the equations are collected in Table 3.2. The continuity or minority carrier diffusion equations combined with the other relationships established in the chapter allow one to model the state of a semiconductor subject to an external perturbation. The more important working relationships are organized and repeated in Table 3.3.

Table 3.3 Carrier Action Equation Summary.

Equations of State

$$\frac{\partial n}{\partial t} = \frac{1}{q}\nabla \cdot \mathbf{J}_N + \left.\frac{\partial n}{\partial t}\right|_{\substack{\text{thermal}\\ \text{R–G}}} + \left.\frac{\partial n}{\partial t}\right|_{\substack{\text{other}\\ \text{processes}}} \qquad \frac{\partial \Delta n_p}{\partial t} = D_N \frac{\partial^2 \Delta n_p}{\partial x^2} - \frac{\Delta n_p}{\tau_n} + G_L$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q}\nabla \cdot \mathbf{J}_P + \left.\frac{\partial p}{\partial t}\right|_{\substack{\text{thermal}\\ \text{R–G}}} + \left.\frac{\partial p}{\partial t}\right|_{\substack{\text{other}\\ \text{processes}}} \qquad \frac{\partial \Delta p_n}{\partial t} = D_P \frac{\partial^2 \Delta p_n}{\partial x^2} - \frac{\Delta p_n}{\tau_p} + G_L$$

Current and R–G Relationships

$$\mathbf{J}_N = \mathbf{J}_{N|\text{drift}} + \mathbf{J}_{N|\text{diff}} = q\mu_n n\mathscr{E} + qD_N \nabla n \qquad \left.\frac{\partial n}{\partial t}\right|_{\substack{\text{i-thermal}\\ \text{R–G}}} = -\frac{\Delta n}{\tau_n}$$

$$\Updownarrow \text{ drift} \qquad \Updownarrow \text{ diffusion}$$

$$\mathbf{J}_P = \mathbf{J}_{P|\text{drift}} + \mathbf{J}_{P|\text{diff}} = q\mu_p p\mathscr{E} - qD_P \nabla p \qquad \left.\frac{\partial p}{\partial t}\right|_{\substack{\text{i-thermal}\\ \text{R–G}}} = -\frac{\Delta p}{\tau_p}$$

$$\mathbf{J} = \mathbf{J}_N + \mathbf{J}_P$$

Key Parametric Relationships

$$L_N \equiv \sqrt{D_N \tau_n} \qquad \frac{D_N}{\mu_n} = \frac{kT}{q} \qquad \tau_n = \frac{1}{c_n N_T}$$

$$L_P \equiv \sqrt{D_P \tau_p} \qquad \frac{D_P}{\mu_p} = \frac{kT}{q} \qquad \tau_p = \frac{1}{c_p N_T}$$

Resistivity and Electrostatic Relationships

$$\rho = \frac{1}{q(\mu_n n + \mu_p p)} \qquad \rho = \frac{1}{q\mu_n N_D} \quad \ldots\ n\text{-type semiconductor}$$

$$\rho = \frac{1}{q\mu_p N_A} \quad \ldots\ p\text{-type semiconductor}$$

$$\mathscr{E} = \frac{1}{q}\frac{dE_c}{dx} = \frac{1}{q}\frac{dE_v}{dx} = \frac{1}{q}\frac{dE_i}{dx} \qquad V = -\frac{1}{q}(E_c - E_{ref})$$

Quasi-Fermi Level Relationships

$$F_N \equiv E_i + kT \ln\left(\frac{n}{n_i}\right) \qquad \mathbf{J}_N = \mu_n n \nabla F_N$$

$$F_P \equiv E_i - kT \ln\left(\frac{p}{n_i}\right) \qquad \mathbf{J}_P = \mu_p p \nabla F_P$$

A number of topics related to carrier action were also addressed in the chapter, including resistivity and resistivity measurements, the hot-point probe measurement, constancy of the equilibrium Fermi level, nonuniform doping and the associated built-in electric field, E–$\mathbf{k}$ diagrams, measurement of the minority carrier lifetimes, and diffusion lengths. Moreover, although not identified as such, the energy band diagram was subject to further development. Specifically, it was pointed out that the existence of an electric field inside the semiconductor causes band bending or a variation of the energy bands with position. By simply inspecting an energy band diagram, it is possible to ascertain the general functional dependence of the electrostatic potential and electric field present in the material. In the discussion of recombination–generation, another level was added near midgap. This level, arising from R–G centers, plays a dominant role in the thermal communication between the energy bands. Lastly, quasi-Fermi levels were introduced to describe nonequilibrium conditions.

PROBLEMS

CHAPTER 3 PROBLEM INFORMATION TABLE

Problem	*Complete After*	*Difficulty Level*	*Suggested Point Weighting*	*Short Description*
3.1	3.3.1	1	16 (2 each part)	Energy band visualization
3.2	3.1.4 (a-d) 3.3.4 (e-h)	1–2	16 (2 each part)	Short answer
● 3.3	3.1.3	3	25	μ vs. T plots
● 3.4	"	3	18 (b-15, c-3)	μ_n vs. T student data
3.5	3.1.4	a-1, b-3	12 (a-6, b-6)	Intrinsic/maximum ρ
3.6	"	2	15 (3 each part)	Resistivity questions
● 3.7	"	2	15	ρ vs. N_A, N_D plot
● 3.8	"	3–4	20 (10 each part)	$N_D(x)$ variation in resistor
● 3.9	"	3	18 (a-12, b-6)	ρ vs. T plots
3.10	"	4	20 (5 each part)	Temperature sensor
3.11	3.1.5	2	5	Compute thermal speed
3.12	3.1.5 (a-d) 3.2.4 (e, f)	2, e-3	16/diagram (a::c-2, d-4, e-2, f-4)	Interpret E-band diagrams
3.13	3.2.4	3	15 (a-10, b-3, c-2)	Built-in electric field
● 3.14	"	2	15 (a-3, b-10, c-2)	D vs. N_A, N_D computation
3.15	3.3.3	3	10	Eq. (3.35) → Eqs. (3.34)
3.16	3.4.2	1	6 (2 each part)	Simple Diff. Eq. questions

3.17	3.4.4	1	6	Diff. Eq., no R–G region
3.18	"	2	12 (a::c-2, d-6)	Diff. Eq., mysterious ray
3.19	"	2	8	Diff. Eq., $G_L \rightarrow G_L/2$
3.20	"	2–3	10	Diff. Eq., light + edge R
3.21	"	3	15 (a::d-2, e-7)	Diff. Eq., double-ended
3.22	"	3–4	15 (a::c-2, d-9)	Diff. Eq., half-illuminated
3.23	"	2	8 (a-2, b-4, c-2)	CdS Photoconductor
3.24	3.5.2	2	12 (a-2, b-4, c::e-2)	Quasi-Fermi levels
3.25	"	3–4	15	Quasi-Fermi levels
● 3.26	"	3–4	15 (a-13, b-2)	Plot quasi-Fermi levels

3.1 Using the energy band diagram, indicate how one visualizes

(a) The existence of an electric field inside a semiconductor.

(b) An electron with a K.E. = 0.

(c) A hole with a K.E. = $E_G/4$.

(d) Photogeneration.

(e) Direct thermal generation.

(f) Band-to-band recombination.

(g) Recombination via R–G centers.

(h) Generation via R–G centers.

3.2 Short Answer

(a) An average hole drift velocity of 10^3 cm/sec results when 2 V is applied across a 1-cm-long semiconductor bar. What is the hole mobility inside the bar?

(b) Name the two dominant carrier *scattering* mechanisms in nondegenerately doped semiconductors of device quality.

(c) For a given semiconductor the carrier mobilities in intrinsic material are (choose one: higher than, lower than, the same as) those in heavily doped material. Briefly explain why the mobilities in intrinsic material are (chosen answer) those in heavily doped material.

(d) Two GaAs wafers, one n-type and one p-type, are uniformly doped such that N_D(wafer 1) = N_A(wafer 2) $\gg n_i$. Which wafer will exhibit the larger resistivity? Explain.

(e) The electron mobility in a silicon sample is determined to be 1300 cm^2/V-sec at room temperature. What is the electron diffusion coefficient?

(f) What is the algebraic statement of low-level injection?

(g) Light is used to create excess carriers in silicon. These excess carriers will predominantly recombine via (choose one: band-to-band, R–G center, or photo) recombination.

(h) Prior to processing, a silicon sample contains $N_D = 10^{14}/cm^3$ donors and $N_T = 10^{11}/cm^3$ R–G centers. After processing (say in the fabrication of a device), the sample contains $N_D = 10^{17}/cm^3$ donors and $N_T = 10^{10}/cm^3$ R–G centers. Did processing increase or decrease the minority carrier lifetime? Explain.

● **3.3** Complete part (b) of Exercise 3.1.

● **3.4** (a) (Optional) Read the Experiment No. 7 Introduction and Measurement System description in R. F. Pierret, *Semiconductor Measurement Laboratory Operations Manual,* Supplement A, in the Modular Series on Solid State Devices, Addison-Wesley Publishing Co., Reading, MA, © 1991.

(b) The measurement described in the part (a) reference is used to determine the mobility in an $N_D < 10^{14}/cm^3$ Si sample as a function of temperature from roughly room temperature to $T = 150$ K. Representative μ_n versus T (K) data derived by undergraduate students is tabulated in the following table. Assuming $\mu_n \propto T^{-\mathbf{b}}$ [or $\ln(\mu_n)$ = constant $-$ **b** $\ln(T)$] can be fitted to the data, use the MATLAB `polyfit` function or an equivalent least squares fit† to determine the best fit value of the power factor **b.** Draw the fit line and note the **b**-fit value on a log-log plot of the data.

(c) Compare your part (b) result with Fig. 3.7(a). Is it significant that the experimental data was derived from a lowly doped sample? Is the experimentally derived **b**-value consistent with the **b**-value noted on the text plot? Are the measured mobility values of the proper magnitude?

---

† Given N data points, (x_i, y_i) with $i = 1$ to N, the straight line $y = a + bx$, which is the "best fit" to the data, will have a y-axis intercept and slope respectively equal to

$$a = \frac{\Sigma y_i \Sigma x_i^2 - \Sigma x_i \Sigma x_i y_i}{N \Sigma x_i^2 - (\Sigma x_i)^2} \quad \text{and} \quad b = \frac{N \Sigma x_i y_i - \Sigma x_i \Sigma y_i}{N \Sigma x_i^2 - (\Sigma x_i)^2}$$

All summations are from $i = 1$ to $i = N$. The "best fit" criterion leading to the expressions for a and b is that the square of the difference between the data points and the fitted line is minimized; hence the name "least squares fit."

T (K)	μ_n (cm^2/V-sec)
290	1501
280	1646
270	1805
260	1985
250	2185
240	2415
230	2675
220	2978
210	3306
200	3743
190	4209
180	4619
170	5216
160	5910
150	6757

3.5 Intrinsic and Maximum Resistivity

(a) Determine the resistivity of intrinsic Ge, Si, and GaAs at 300 K.

(b) Determine the maximum possible resistivity of Ge, Si, and GaAs at 300 K.

3.6 More Resistivity Questions

(a) A silicon sample maintained at room temperature is uniformly doped with $N_D = 10^{16}/\text{cm}^3$ donors. Calculate the resistivity of the sample using Eq. (3.8a). Compare your calculated result with the ρ deduced from Fig. 3.8(a).

(b) The silicon sample of part (a) is "compensated" by adding $N_A = 10^{16}/\text{cm}^3$ acceptors. Calculate the resistivity of the compensated sample. (Exercise caution in choosing the mobility values to be employed in this part of the problem.)

(c) Compute the resistivity of intrinsic ($N_A = 0$, $N_D = 0$) silicon at room temperature. How does your result here compare with that for part (b)?

(d) A 500-ohm resistor is to be made from a bar-shaped piece of n-type Si. The bar has a cross-sectional area of 10^{-2} cm^2 and a current-carrying length of 1 cm. Determine the doping required.

(e) A lightly doped ($N_D < 10^{14}/\text{cm}^3$) Si sample is heated up from room temperature to 100°C. $N_D \gg n_i$ at both room temperature and 100°C. Is the resistivity of the sample expected to increase or decrease? Explain.

● **3.7** Making use of the mobility fit relationships and parameters quoted in Exercise 3.1, construct a plot of the silicon resistivity versus impurity concentration at $T = 300$ K. Include curves for both n- and p-type silicon over the range $10^{13}/\text{cm}^3 \leq N_A$ or $N_D \leq 10^{20}/\text{cm}^3$. Compare your result with Fig. 3.8(a).

3.8 Resistors in ICs are sometimes thin semiconductor layers near the surface of a wafer. However, formation of the layer by diffusion or ion implantation (discussed in Chapter 4) is likely to give rise to a doping concentration that varies with depth into the layer. Let us examine how the resistance is computed when the doping varies with depth.

(a) Given a bar-shaped layer of width W, length L, and depth d, and assuming an arbitrary $N_D(x)$ variation with the depth x from the wafer surface, show that the resistance of the layer is to be computed from

$$R = \frac{L}{W}\left[\frac{1}{q\int_0^d \mu_n(x)N_D(x)\,dx}\right]$$

● (b) Taking $N_D(x) = N_{D0}\exp(-ax) + N_{DB}$, compute and plot R versus N_{D0} for $10^{14}/\text{cm}^3 \leq N_{D0} \leq 10^{18}/\text{cm}^3$ when $L = W$, $N_{DB} = 10^{14}/\text{cm}^3$, $d = 5\ \mu$m, and $1/a = 1\ \mu$m.

● **3.9** (a) Modify your μ versus T program from Problem 3.3 to compute and construct *semi-log* plots of ρ (log scale) versus T for 200 K $\leq T \leq$ 500 K and N_D or N_A stepped in decade values from $10^{14}/\text{cm}^3$ to $10^{18}/\text{cm}^3$. [NOTE: The lower doping curves are in error at the higher temperatures if Eqs. (3.8) are used to compute ρ. Can you explain why? Assuming Eqs. (3.8) were used in the computation, how must the ρ versus T program be modified to correct the error?]

(b) Compare your n-type plot with Fig. 7 in Li and Thurber, Solid-State Electronics, **20,** 609 (1977). Compare your p-type plot with Fig. 7 in Li, Solid-State Electronics, **21,** 1109 (1978). Comment on the comparisons.

3.10 Your boss asks you to construct a temperature sensor for measuring the ambient outside temperature in W. Lafayette IN ($-30°\text{C} \leq T \leq 40°\text{C}$). You decide to base operation of the temperature sensor on the change in resistance associated with a bar-shaped piece of Si.

(a) Restricting yourself to nondegenerate dopings and reasonable device dimensions, and also requiring the resistance to be readily measured with a hand-held multimeter (say $1\ \Omega \leq R \leq 1000\ \Omega$), indicate the doping and dimensions of your sensor.

(b) Derive an expression for the sensitivity (dR/dT in $\Omega/°C$) of your sensor over the operating temperature range. Relative to sensitivity, is it preferable to use low or high dopings? Explain.

(c) What are the upper and lower temperature limits of operation of your sensor (approximately)? Explain.

• (d) Plot R versus T of your sensor over the required temperature range of operation.

3.11 Thermal energy alone was noted to contribute to relatively high carrier velocities inside of a semiconductor. Considering a nondegenerate semiconductor maintained at room temperature, compute the thermal velocity of electrons having a kinetic energy corresponding to the peak of the electron distribution in the conduction band. Set $m^* = m_0$ in performing your calculation. (This problem assumes the prior working of either Problem 2.7 or Problem 2.8.)

3.12 Interpretation of Energy Band Diagrams

Six different silicon samples maintained at 300 K are characterized by the energy band diagrams in Fig. P3.12. Answer the questions that follow after choosing a specific diagram for analysis. Possibly repeat using other energy band diagrams. (Excessive repetitions have been known to lead to the onset of insanity.)

(a) Do equilibrium conditions prevail? How do you know?

(b) Sketch the electrostatic potential (V) inside the semiconductor as a function of x.

(c) Sketch the electric field ($\mathscr{E}$) inside the semiconductor as a function of x.

(d) The carrier pictured on the diagram moves back and forth between $x = 0$ and $x = L$ without changing its total energy. Sketch the K.E. and P.E. of the carrier as a function of position inside the semiconductor. Let E_F be the energy reference level.

(e) Roughly sketch n and p versus x.

(f) On the same set of coordinates, make a rough sketch of the electron drift-current density ($J_{N|drift}$) and the electron diffusion-current density ($J_{N|diff}$) inside the Si sample as a function of position. Be sure to graph the proper polarity of the current densities at all points and clearly identify your two current components. Also briefly explain how you arrived at your sketch.

3.13 The nonuniform doping in the central region of bipolar junction transistors (BJTs) creates a built-in field that assists minority carriers across the region and increases the maximum operating speed of the device. Suppose the BJT is a Si device maintained under equilibrium conditions at room temperature with a central region of length L. Moreover, the nonuniform acceptor doping is such that

$$p(x) \cong N_A(x) = n_i e^{(a-x)/b} \qquad \ldots 0 \leq x \leq L$$

where $a = 1.8\ \mu m$, $b = 0.1\ \mu m$, and $L = 0.8\ \mu m$.

(a) Draw the energy band diagram for the $0 \leq x \leq L$ region specifically showing E_c, E_F, E_i, and E_v on your diagram. Explain how you arrived at your diagram.

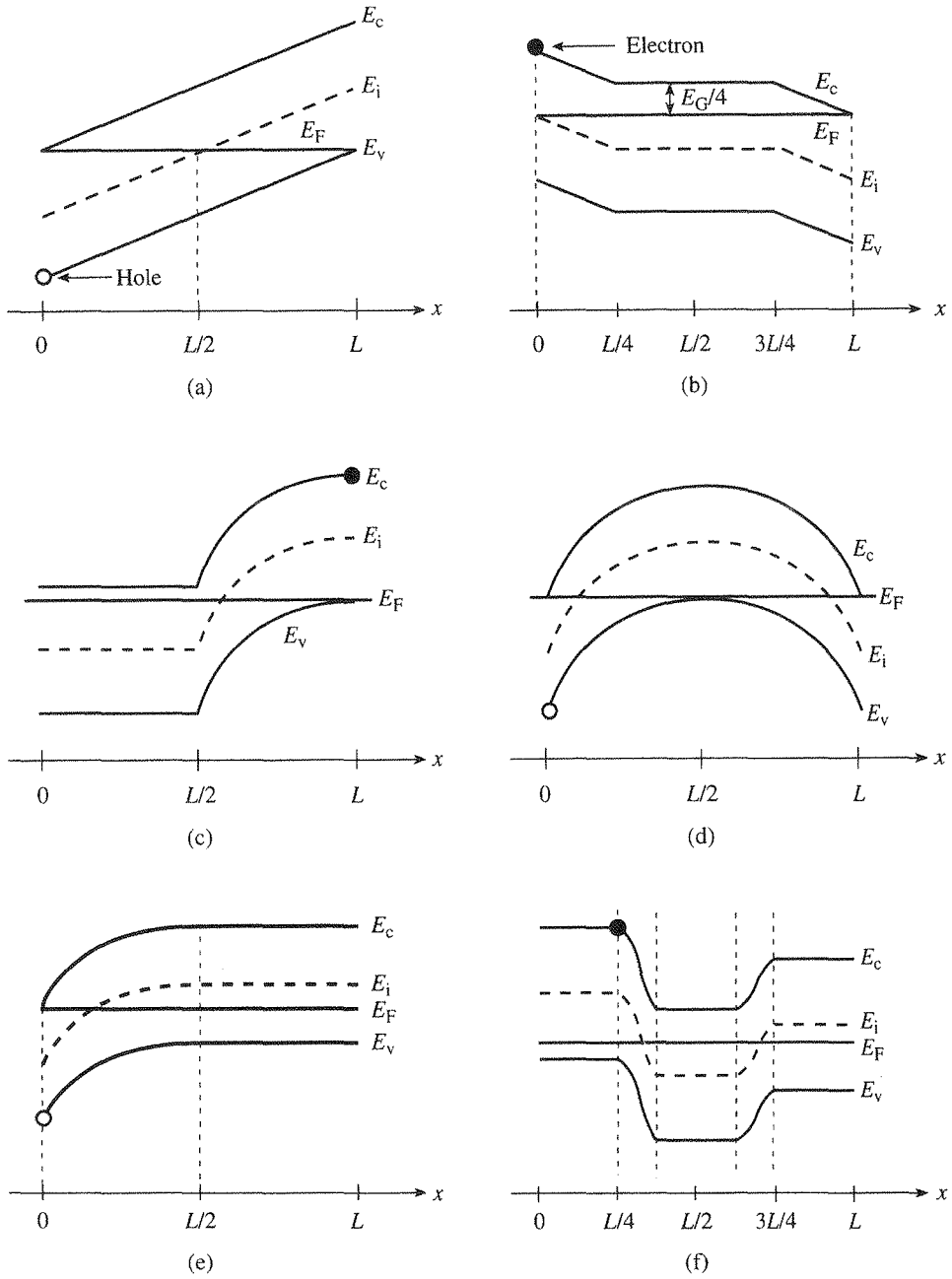

E_c
E_i
E_F
E_v
Hole
x
0
$L/2$
L
(a)
Electron
$E_G/4$
E_c
E_F
E_i
E_v
x
0
$L/4$
$L/2$
$3L/4$
L
(b)
E_c
E_i
E_F
E_v
x
0
$L/2$
L
(c)
E_c
E_F
E_i
E_v
x
0
$L/2$
L
(d)
E_c
E_i
E_F
E_v
x
0
$L/2$
L
(e)
E_c
E_i
E_F
E_v
x
0
$L/4$
$L/2$
$3L/4$
L
(f)

Figure P3.12

(b) Make a sketch of the $\mathscr{E}$-field inside the region as a function of position, and compute the value of $\mathscr{E}$ at $x = L/2$.

(c) Is the built-in electric field such as to aid the motion of minority carrier electrons in going from $x = 0$ to $x = L$? Explain.

3.14 (a) Based on information found in the text proper, *roughly sketch* the expected variation of D_N and D_P versus doping appropriate for $10^{14}/\text{cm}^3 \leq N_A$ or $N_D \leq 10^{18}/\text{cm}^3$ doped Si maintained at $T = 300$ K. Explain how you arrived at the form of your sketch.

(b) Making use of the fit relationships and parameters quoted in Exercise 3.1, construct a plot of D_N and D_P versus N_A or N_D for $10^{14}/\text{cm}^3 \leq N_A$ or $N_D \leq 10^{18}/\text{cm}^3$ doped Si maintained at $T = 300$ K.

(c) Why was the upper doping limit taken to be $10^{18}/\text{cm}^3$ in the part (b) computation?

3.15 Taking $E_T \cong E_i$ so that $n_1 \cong p_1 \cong n_i$, setting $\Delta n = \Delta p$, and assuming τ_n is comparable to τ_p, show that the general-case R–G relationship of Eq. (3.35) reduces to the special-case relationships of Eqs. (3.34) when one has low-level injection in a specific material type.

3.16 $\partial \Delta n_p/\partial t = D_N \partial^2 \Delta n_p/\partial x^2 - \Delta n_p/\tau_n + G_L$ is known as the minority carrier diffusion equation for electrons.

(a) Why is it called a *diffusion equation?*

(b) Why is it referred to as a *minority carrier equation?*

(c) The equation is valid only under low-level injection conditions. Why?

3.17 Show that, under steady state conditions

$$\Delta p_n(x) = \Delta p_{n0}(1 - x/L) \qquad \ldots 0 \leq x \leq L$$

is the special-case solution of the minority carrier diffusion equation that will result if (1) one assumes *all* recombination–generation processes are negligible within an n-type semiconductor of length L, and (2) one employs the boundary conditions $\Delta p_n(0) = \Delta p_{n0}$ and $\Delta p_n(L) = 0$. (Neglecting recombination–generation is an excellent approximation when L is much less than a minority carrier diffusion length. A $\Delta p(x)$ solution of the above type is frequently encountered in practical problems.)

3.18 The earth is hit by a mysterious ray that momentarily wipes out all minority carriers. Majority carriers are unaffected. Initially in equilibrium and not affected by room light, a uniformly doped silicon wafer sitting on your desk is struck by the ray at time $t = 0$. The wafer doping is $N_A = 10^{16}/\text{cm}^3$, $\tau_n = 10^{-6}$ sec, and $T = 300$ K.

(a) What is Δn at $t = 0^+$? ($t = 0^+$ is an imperceptible fraction of a second after $t = 0$.)

(b) Does generation or recombination dominate at $t = 0^+$? Explain.

(c) Do low-level injection conditions exist inside the wafer at $t = 0^+$? Explain.

(d) Starting from the appropriate differential equation, derive $\Delta n_p(t)$ for $t > 0$.

3.19 A silicon wafer ($N_A = 10^{14}/\text{cm}^3$, $\tau_n = 1\ \mu\text{sec}$, T = room temperature) is first illuminated for a time $t \gg \tau_n$ with light which generates $G_{L0} = 10^{16}$ electron-hole pairs per cm^3-sec uniformly throughout the volume of the silicon. At time $t = 0$ the light intensity is reduced, making $G_L = G_{L0}/2$ for $t \geq 0$. Determine $\Delta n_p(t)$ for $t \geq 0$.

3.20 A semi-infinite p-type bar (see Fig. P3.20) is illuminated with light which generates G_L electron-hole pairs per cm^3-sec uniformly throughout the volume of the semiconductor. Simultaneously, carriers are extracted at $x = 0$ making $\Delta n_p = 0$ at $x = 0$. Assuming a steady state condition has been established and $\Delta n_p(x) \ll p_0$ for all x, determine $\Delta n_p(x)$.

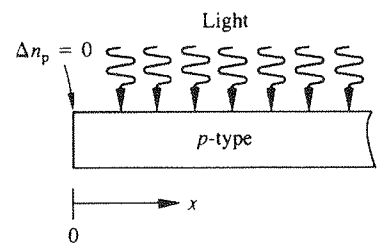


Figure P3.20

3.21 The two ends of a uniformly doped n-type Si bar of length L are simultaneously illuminated so as to create γN_D excess holes at both $x = 0$ and $x = L$ (see Fig. P3.21). The wavelength and intensity of the illumination are such that no light penetrates into the interior ($0 < x < L$) of the bar and $\gamma = 10^{-3}$. Also, steady state conditions prevail, $T = 300$ K, and $N_D \gg n_i$.

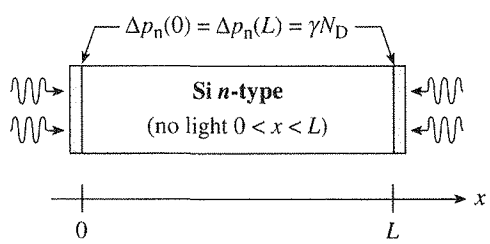


Figure P3.21

(a) Based on *qualitative reasoning*, sketch the expected general form of the $\Delta p_n(x)$ solution.

(b) Do low-level injection conditions prevail inside the illuminated bar? Explain.

(c) Write down the differential equation (simplest form possible) you must solve to determine $\Delta p_n(x)$ inside the bar.

(d) Write down the general form of the $\Delta p_n(x)$ solution and the boundary condition(s) appropriate for this particular problem.

(e) Establish an expression for the hole current (J_P) flowing in the illuminated bar at $x = 0$. [Your answer may be left in terms of the arbitrary constant(s) appearing in the general form of the $\Delta p_n(x)$ solution.]

3.22 As pictured in Fig. P3.22, the $x > 0$ portion of an infinite semiconductor is illuminated with light. The light generates $G_L = 10^{15}$ electron-hole pairs/cm^3-sec uniformly throughout the $x > 0$ region of the bar. $G_L = 0$ for $x < 0$, steady state conditions prevail, the semiconductor is silicon, the entire bar is uniformly doped with $N_D = 10^{18}$/cm^3, $\tau_p = 10^{-6}$ sec, and $T = 300$ K.

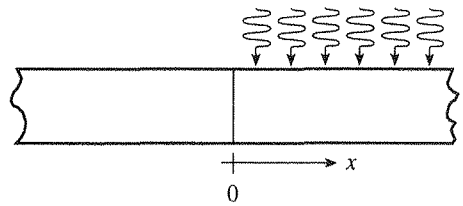


Figure P3.22

(a) What is the hole concentration at $x = -\infty$? Explain.

(b) What is the hole concentration at $x = +\infty$? Explain.

(c) Do low-level injection conditions prevail? Explain.

(d) Determine $\Delta p_n(x)$ for all x.
NOTE: (1) Separate $\Delta p_n(x)$ expressions apply for $x > 0$ and $x < 0$.
(2) Both Δp_n and $d\Delta p_n/dx$ must be continous at $x = 0$.

3.23 CdS is the most widely used material for constructing commercial photoconductors operating in the visible portion of the spectrum. The CdS photoconductor has a high sensitivity and its spectral response closely matches that of the human eye. A model VT333 CdS Photoconductor is pictured in Fig. P3.23.

Figure P3.23

(a) Speculate why the conducting film has a snake-like pattern.

(b) The VT333 resistor pattern is approximately 0.3 mm wide and 3 cm long. Estimating the deposited CdS film to be 5 μm thick, $N_D = 10^{13}$/cm$^3 \gg n_i$, and $\mu_n = 100$ cm^2/V-sec, compute the dark resistance of the device.

(c) The VT333 exhibited a 250 Ω resistance when illuminated with a microscope light. Can the usual relationships be used to determine the G_L required to produce the cited resistance? Explain.

3.24 The equilibrium and steady state conditions before and after illumination of a semiconductor are characterized by the energy band diagrams shown in Fig. P3.24. $T = 300$ K, $n_i = 10^{10}/\text{cm}^3$, $\mu_n = 1345$ cm^2/V-sec, and $\mu_p = 458$ cm^2/V-sec. From the information provided, determine

(a) n_0 and p_0, the equilibrium carrier concentrations.

(b) n and p under steady state conditions.

(c) N_D.

(d) Do we have "low-level injection" when the semiconductor is illuminated? Explain.

(e) What is the resistivity of the semiconductor before and after illumination?

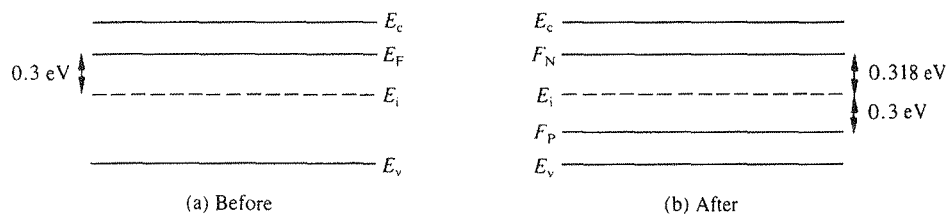


Figure P3.24

3.25 A portion ($0 \le x \le L$) of a Si sample, uniformly doped with $N_D = 10^{15}/\text{cm}^3$ donors and maintained at room temperature, is subject to a *steady state* perturbation such that

$$\begin{aligned} n &\simeq N_D \\ p &= n_i(1 - x/L) + n_i^2/N_D \end{aligned} \qquad \ldots 0 \le x \le L$$

Since $n \simeq N_D$, it is reasonable to assume $\mathscr{E} \simeq 0$ in the $0 \le x \le L$ region. Given $\mathscr{E} \simeq 0$, sketch the energy band diagram for the perturbed region specifically including E_c, E_i, E_v, F_N, and F_P on your diagram.

● **3.26** (a) Write a MATLAB (computer) program that automatically plots F_N and F_P versus x for the experimental situation first considered in Sample Problem No. 2 and further analyzed in Exercise 3.5. Assume a Si sample maintained at 300 K with $\tau_p = 10^{-6}$ sec. In writing the program, take N_D and Δp_{n0} to be input variables. Scale the x-axis in L_P units with $(x/L_P)_{max} = \ln[100\Delta p_{n0}/p_0]$. Scale the y-axis in kT units plotting $(F_N - E_v)/kT$ and $(F_P - E_v)/kT$. Set $y_{min} = -5$ and $y_{max} = 45$. Identify the positions of E_v and E_c on your plot.

(b) Run your program with $N_D = 10^{15}/\text{cm}^3$ and $\Delta p_{n0} = 10^{10}/\text{cm}^3$. Compare the resultant plot with the corresponding diagram sketched in Exercise 3.5.

(c) Remembering the assumed low-level injection requires $\Delta p_{n0} \ll N_D$, try running your program using different values of N_D and Δp_{n0}.

4 Basics of Device Fabrication

Read Only

This chapter provides a very brief overview of silicon device fabrication. The fabrication overview constitutes the final preparatory step before the consideration and analysis of specific devices. The goal is to develop a general feel for the physical nature of device structures. Moreover, the way a device is made affects key device parameters, which in turn affect the simplifications, assumptions, and so forth possible in the device analysis. Even in purely theoretical endeavors, some familiarity with fabrication processes and procedures is obviously necessary.

In what follows we first examine common "building block" processes. The processes are used in combination and are routinely repeated several times in producing a modern device or IC. Subsequently we describe the sequencing of processes to produce a *pn* junction diode, the first device to be analyzed in Part II of the text. To better illustrate IC fabrication, the chapter concludes with a subsection on the making of a CPU (central processing unit). For additional information the reader is referred to one of the excellent device-fabrication books cited in the Part I reading list.

Finally, it should be noted that this is the first "read only" chapter. As explained in the General Introduction, chapters with the read-only designation are primarily intended to convey qualitative information of general interest. They contain a small number of equations, no exercises, and few, if any, end-of-chapter problems. Also, there is no reference to the contents of read-only chapters in the test-like Review Problem Sets that are included at the end of the three text segments. If the material in this chapter were being transmitted verbally, you would be advised to put down your note-taking pencils and to just sit back and listen.

4.1 FABRICATION PROCESSES

4.1.1 Oxidation

The existence of a readily produced high-quality oxide helped establish the dominance of Si in the production of commercial devices. SiO_2 functions as both an insulator in a number of device structures and as a barrier to diffusion (see Subsection 4.1.2) during device fabrication. Silicon is highly reactive and a thin native oxide layer is rapidly produced by simply exposing virgin Si to the atmosphere at room temperature. SiO_2 layers of controllable thickness are produced during device fabrication by reacting Si with either oxygen gas or water vapor at an elevated temperature. In either case the oxidizing species diffuses

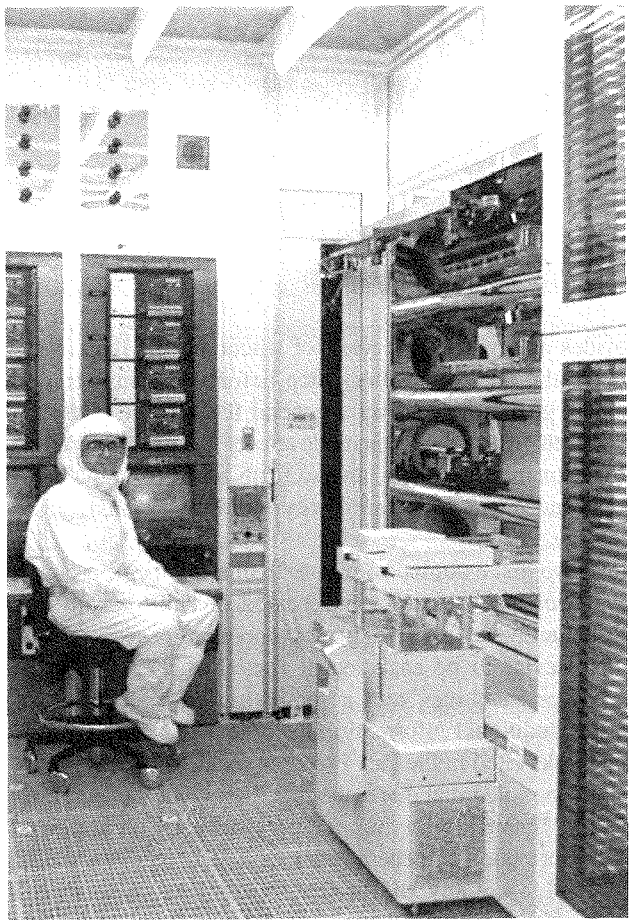


Figure 4.1 Production-line (Intel Fab-9 fabrication facility) bank of furnaces of the type used for oxidation and diffusion. The ends of three horizontal furnace tubes are clearly visible in the center-right of the photograph. An operator in standard cleanroom garb is seated next to the furnace controls. (Photograph courtesy of Intel Corporation.)

through the existing oxide and reacts at the Si–SiO_2 interface to form more SiO_2. The relevant overall reactions are

$$Si + O_2 \rightarrow SiO_2 \tag{4.1a}$$

$$Si + 2H_2O \rightarrow SiO_2 + 2H_2 \tag{4.1b}$$

Growth of SiO_2 using O_2 and water vapor are referred to as *dry* and *wet oxidation*, respectively. Dry oxidation is used to form the most critical insulator regions in a device structure, such as the gate oxide in a MOSFET, primarily because it yields superior Si–SiO_2 interface

properties. Wet oxidation, on the other hand, proceeds at a faster rate and is therefore preferred in forming the thicker barrier oxides.

A photograph of a production-line oxidation/diffusion furnace is shown in Fig. 4.1. A simplified sketch of the furnace and input system is presented in Fig. 4.2 along with a short pictorial description of the oxidation process. Standard oxidation temperatures of 800°C to 1200°C are produced in the furnace by resistance heating. The coil-like heating elements are commonly arranged in three separately controlled zones. The outer zones operate at a higher power to compensate for heat losses out the ends of the furnace. A relatively long center zone can thereby be maintained at a uniform temperature controllable to ± 1°C. The tube down the center of the furnace is usually made of clear fused quartz, although SiC and polycrystalline Si tubes are also used. The Si wafers to be oxidized are loaded vertically onto a slotted quartz boat and pushed into the center of the furnace. During dry oxidation the oxygen gas is simply fed into the back of the tube. Wet oxidation is performed by bubbling a carrier gas (Ar or N_2) through water in a heated flask or burning O_2 and H_2 at the input to the tube. The latter approach is called *pyrogenic wet oxidation.* The required time in the furnace depends on the furnace temperature, the desired oxidation thickness, and the Si surface orientation. (Other factors can enter under certain conditions.) Representative dry and wet oxidation growth curves are graphed in Fig. 4.3. It should be noted that in a commercial system the wafer loading, insertion into the furnace, ramping of the furnace temperature, and gas control are all automated.

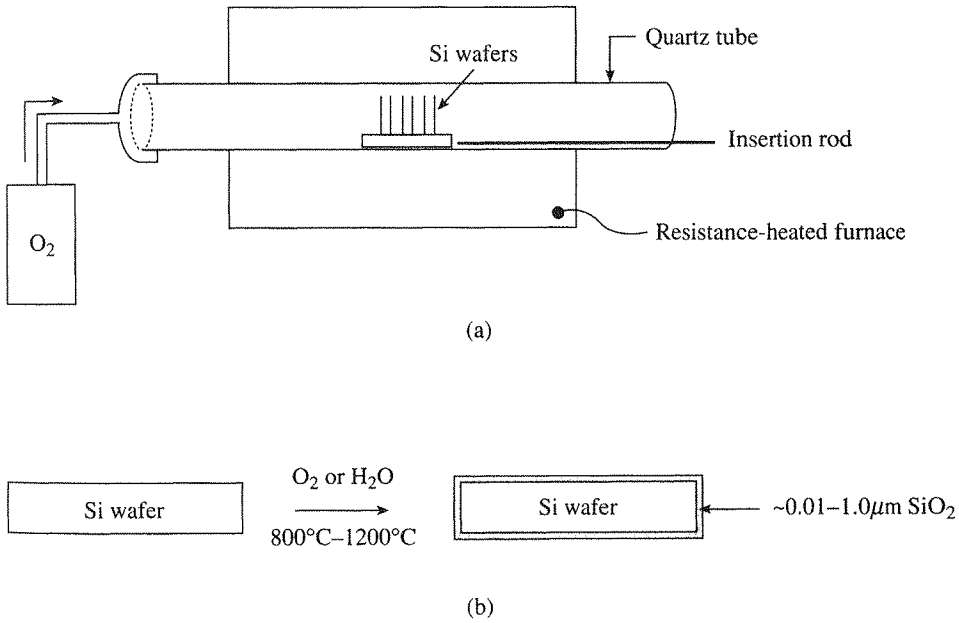
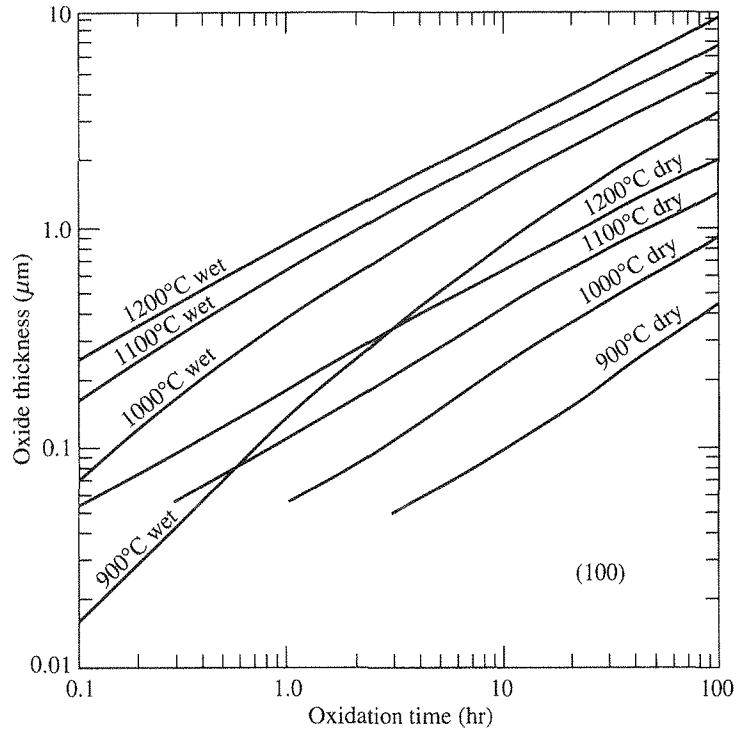


Figure 4.2 (a) Simplified schematic illustration of an oxidation system. (b) Short pictorial description of the oxidation process.

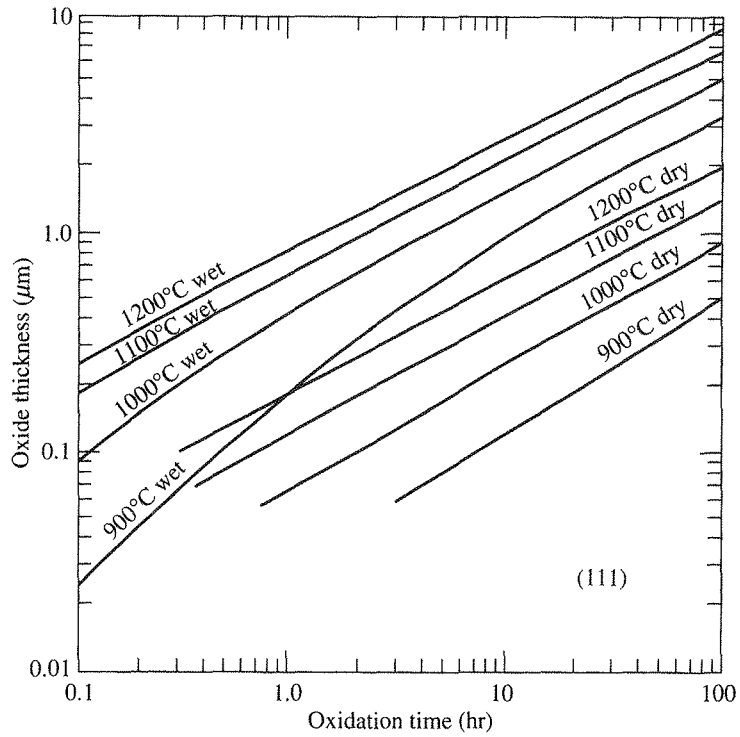


(a)

Figure 4.3 Dry and wet oxidation growth curves. The SiO_2 thickness formed on (a) (100) and (b) (111) silicon surfaces as a function of time. (From Jaeger[8], © 1988 by Addison-Wesley Publishing Co., Inc. Reprinted by permission of the publisher.)

4.1.2 Diffusion

Solid state diffusion is a historically important and still widely used method for introducing dopant atoms into a semiconductor lattice. The basic process is sketched in Fig. 4.4. The semiconductor crystal, a Si wafer in Fig. 4.4, is exposed to a solid, liquid, or gaseous source containing the desired impurity. A reaction at the wafer surface establishes a supply of dopant atoms immediately adjacent to the semiconductor crystal. At elevated temperatures the atoms slowly diffuse into regions of the semiconductor not protected by the oxide. Diffusion takes place of course because the concentration of impurity atoms is greater outside the crystal than inside the crystal. The surface doping concentrations produced by this method are very high (up to $10^{21}/cm^3$), and thus the surface region takes on the *n*- or *p*-type character associated with the diffusing atoms. It should be noted that diffusion also takes place into the SiO_2, albeit at a relatively slow rate for the common dopants. This means the SiO_2 can protect the underlying Si for only a limited period of time. For a given dopant the masking time depends on the oxide thickness, the diffusion temperature, and the background doping.



(b)

Figure 4.3 *Continued.*

Diffusion is commonly performed in an open tube system identical in construction to that used for oxidation. Diffusion temperatures are also similar, ranging from roughly 900°C to 1200°C. The only major difference is the introduction of a dopant source in place of the oxidizing ambient. The specific example of phosphorus diffusion employing a liquid source is schematically illustrated in Fig. 4.5. The pictured N_2 carrier gas is passed through a bubbler containing phosphorus oxychloride ($POCl_3$) that is a liquid at room temperature. The N_2 picks up vapors from the liquid source and carries them into the furnace tube. The small percentage of oxygen simultaneously admitted into the furnace reacts with the $POCl_3$ to deposit P_2O_5 on the surface of the Si wafer, and Si in turn reacts with P_2O_5 liberating phosphorus atoms that diffuse into the silicon. Because undesirable surface compounds tend to form at the higher diffusion temperatures, the foregoing is often combined with a second step where the dopant source is cut off, the furnace temperature increased, and the already-introduced impurities driven deeper into the semiconductor at a more rapid rate. When a two-step process is used, the portion with the source present is called the *predeposition,* the latter portion with the source removed, the *drive-in.*

Neglecting second-order effects, the impurity concentration versus position inside the semiconductor after the predeposition and drive-in steps can be computed from

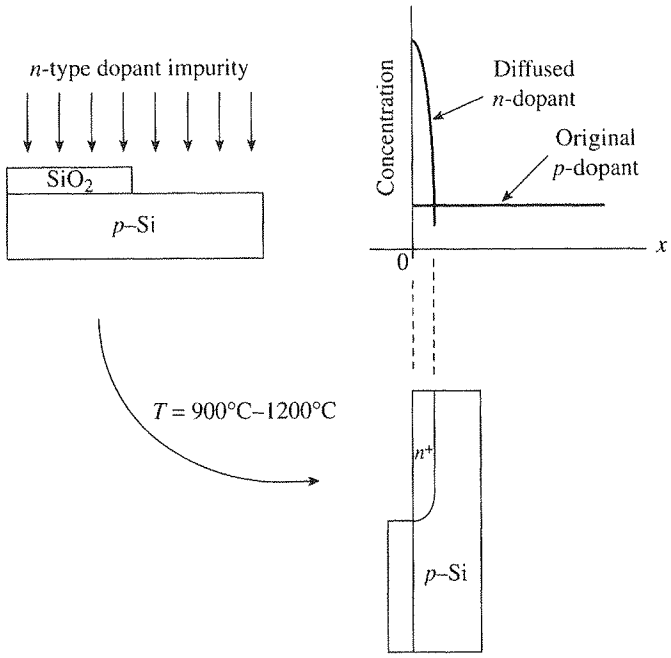


Figure 4.4 The basic diffusion process.

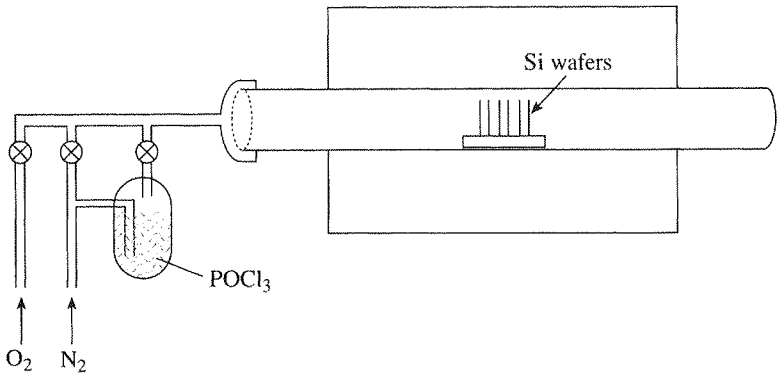


Figure 4.5 Schematic illustration of phosphorus diffusion using a liquid source.

$$N_1(x,t_1) = N_0 \mathrm{erfc}(x/2\sqrt{D_1 t_1}) \qquad \ldots \text{ after predep} \tag{4.2a}$$

$$N_2(x,t_2) = N_0\left(\frac{2}{\pi}\sqrt{D_1 t_1/D_2 t_2}\right)e^{-(x/2\sqrt{D_2 t_2})^2} \qquad \ldots \text{ after drive-in } D_2 t_2 \gg D_1 t_1 \tag{4.2b}$$

The subscripts 1 and 2 refer to the predeposition and drive-in, respectively, x is the distance into the semiconductor as measured from the semiconductor surface, $N(x, t)$ the impurity

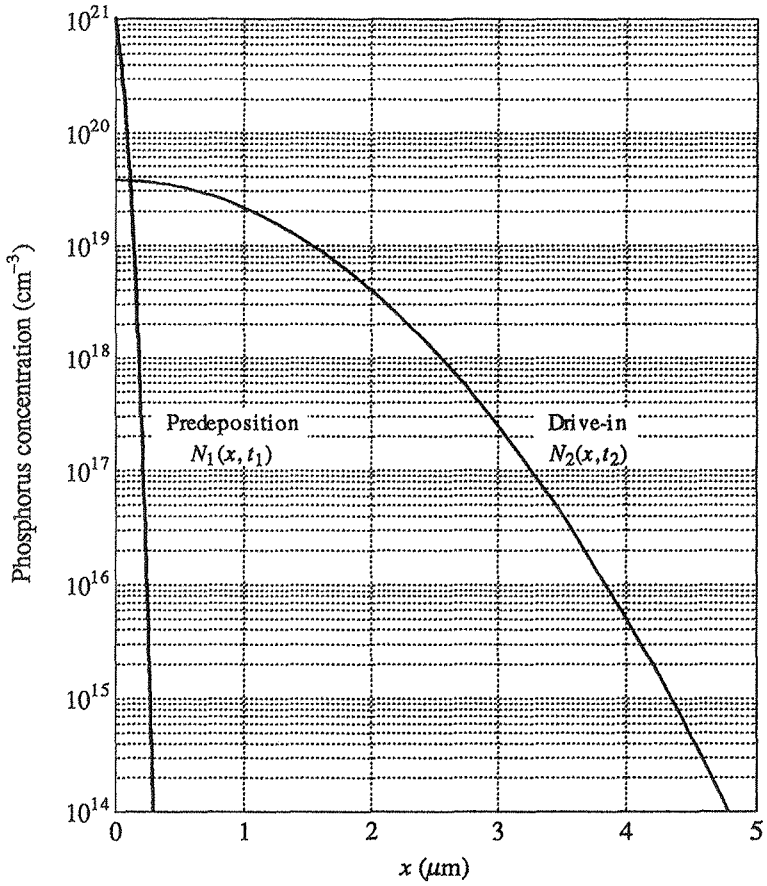


Figure 4.6 Computed phosphorus diffusion profiles. ($N_0 = 10^{21}/\text{cm}^3$, $D_1 = 2.58 \times 10^{-14}$ cm^2/sec, $t_1 = 600$ sec, $D_2 = 2.49 \times 10^{-12}$ cm^2/sec, $t_2 = 1800$ sec.)

concentration at a depth x after a given time t, N_0 the concentration at $x = 0$ during the predeposition, D the diffusion coefficient for the given impurity and furnace temperature, and t the time duration of the diffusion step. The complementary error function, erfc, is a function cited in most mathematical handbooks. The results of a sample computation corresponding to a phosphorus predeposition at 1000°C for 10 minutes followed by a drive-in at 1200°C for 30 minutes are presented in Fig. 4.6.

(4.2a)

4.1.3 Ion Implantation

(4.2b)

Ion implantation affords an alternative means of introducing dopants and other atoms into the near-surface region of a semiconductor. With the need for shallower junctions, lower-temperature processing, and more precise control, ion implantation has become a workhorse manufacturing process. In ion implantation an impurity is introduced into the

listance
npurity

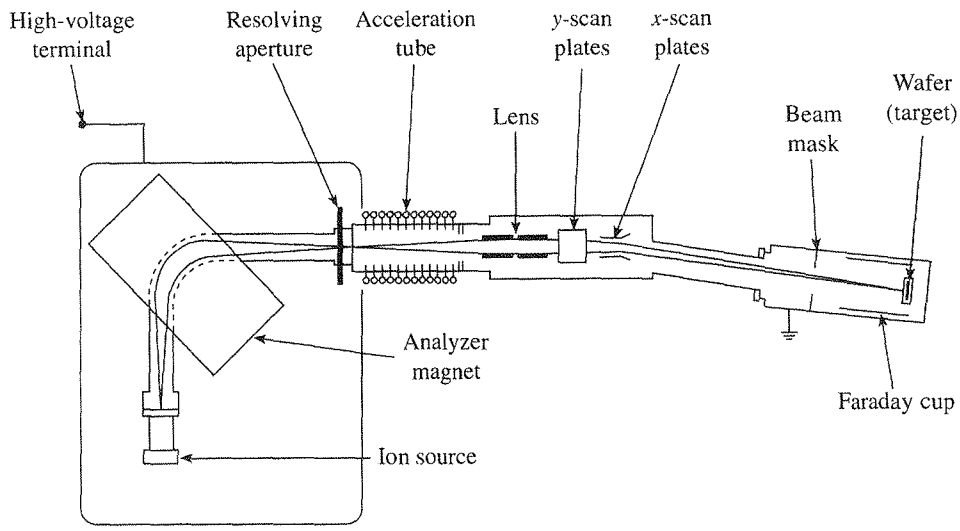


Figure 4.7 Simplified schematic of an ion implantation system. (From Runyan and Bean[9], © 1990 by Addison-Wesley Publishing Co., Inc. Reprinted by permission of the publisher.)

semiconductor by creating ions of the impurity, accelerating the ions to high energies ranging from 5 keV to 1 MeV, and then literally shooting the ions into the semiconductor. As one might suspect, the implanted ions displace semiconductor atoms along their path into the crystal. Moreover, the ions themselves do not necessarily come to rest on lattice sites. A follow-up anneal (heating of the semiconductor), an integral part of the overall process, is therefore necessary to remove the crystal damage and "activate" implanted dopants.

A simplified schematic of an ion implantation system and photographs of a commercial implantation end-station are presented in Figs. 4.7 and 4.8, respectively. Ions of the desired impurity are produced in the ion source shown at the extreme left in Fig. 4.7. The ions are next accelerated into the mass analyzer where unwanted ions, also produced in the ion source, are removed. The resulting ion beam is then accelerated to the preset operating potential, focused, and finally scanned over the surface of the wafer. Scanning is accomplished electrostatically, by mechanically moving the wafer, or by a combination of the two methods. An electrical contact to the wafer allows a flow of electrons to neutralize the implanted ions. A very precise determination of the total number of implanted ions/cm^2, called the *dose* and given the symbol ϕ, is obtained by integrating the target current over the time of the implant.

The concentration profile produced by ion implantation has the general form of a Gaussian distribution function and is mathematically described by

$$N(x) = \frac{\phi}{\sqrt{2\pi}(\Delta R_p)} e^{-(1/2)[(x - R_p)/\Delta R_p]^2} \tag{4.3}$$

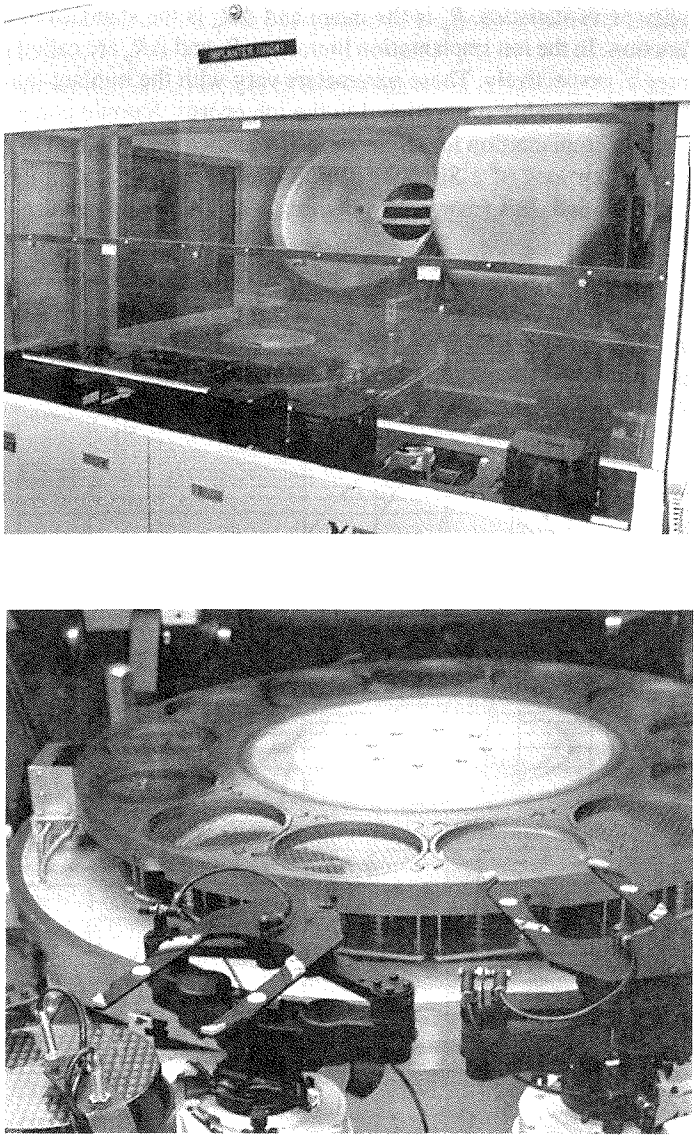


Figure 4.8 Photographs of an ion implantation end-station at the Intel Fab-9 fabrication facility. The top photograph shows the two ports associated with the station. The left port is open and awaiting the loading of wafers onto the wafer carousel, and ion implantation is taking place in the closed right port. The bottom photograph is a close-up of the wafer carousel and the robot arms that automatically load and unload the carousel. (Photographs courtesy of Intel Corporation.)

For those familiar with statistics, R_p is the mean and ΔR_p is the standard deviation of the distribution function. In the ion implantation literature, R_p and ΔR_p are called the *projected range* and *straggle,* respectively. These parameters vary with the implant ion and the substrate material and are roughly proportional to the ion energy. Sample computed distributions for phosphorus implantation into Si at various energies are shown in Fig. 4.9.

Like diffusion, portions of a Si wafer can be masked during ion implantation. Thin films of SiO_2, Si_3N_4, and, less frequently, photoresist and Al have been used for this pur-

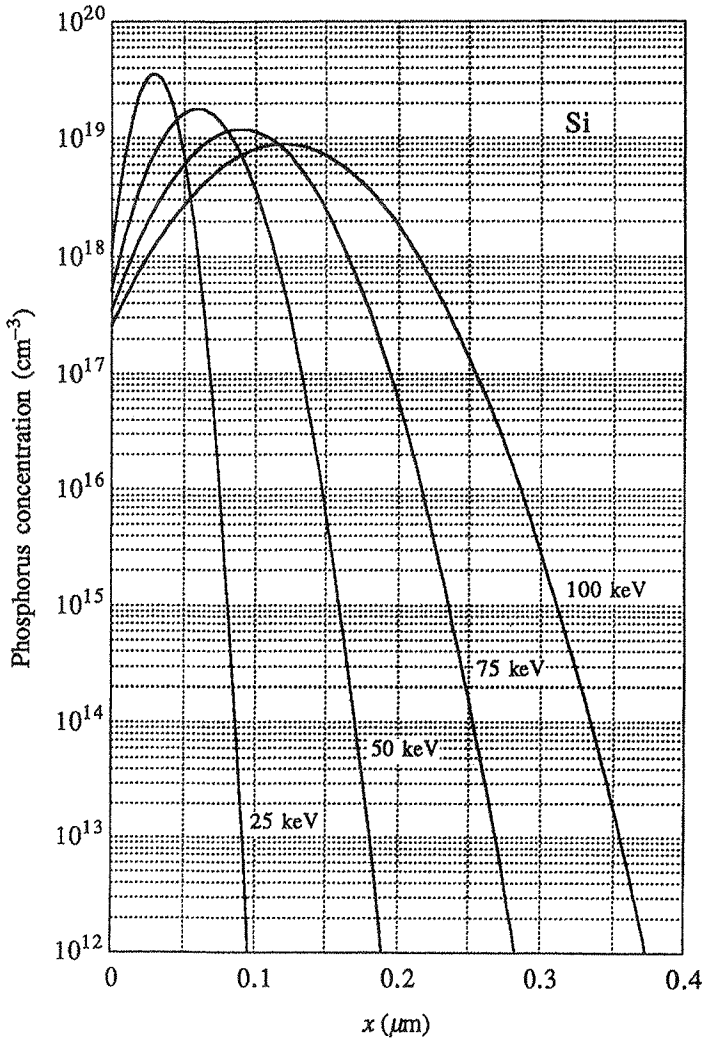


Figure 4.9 Computed phosphorus implantation profiles assuming a constant dose of $10^{14}/cm^2$.

pose. Basically the implanted ions are stopped in the masking material before reaching the underlying silicon.

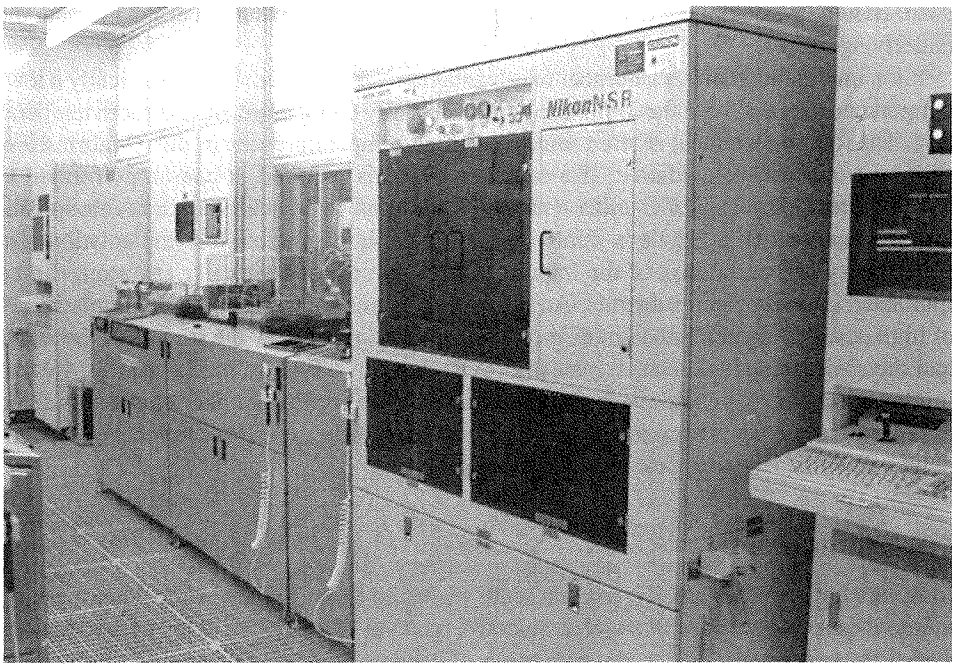
Ion implantation offers a number of advantages over diffusion. It is intrinsically a low-temperature process, with the implantation itself usually being performed at room temperature. The follow-up anneal is often some other high-temperature step in the IC fabrication, such as a subsequent oxidation. If a separate anneal is required, it can be performed at temperatures as low as 600°C under optimum conditions. In any event, undesirable spreading of a concentration profile is minimized using ion implantation. Ion implantation also affords very precise control of the impurity concentration, and virtually any atom can be implanted into any given substrate. An interesting recent development is the implantation of high doses of oxygen into silicon, producing a SiO_2 layer below the surface of the wafer and thereby leading to silicon on insulator (SOI) structures. Finally, note from Fig. 4.9 that implantation typically yields very shallow concentration profiles. Implants have thus come to replace the predeposition step in diffusions and are ideally suited for a number of modern device structures requiring extemely shallow junctions.

4.1.4 Lithography

In discussing diffusion and ion implantation, we talked about SiO_2 or some other masking material covering portions of a wafer. Selective removal of a thin film at prescribed regions across the surface of a wafer, the patterning of insulators and metals required in forming ICs, is achieved through a process called lithography. Photographs of a production-line lithography station are reproduced in Fig. 4.10.

Major steps in the lithography process are depicted in Fig. 4.11 using the patterning of an SiO_2 film as an example. The top surface of the SiO_2-covered wafer is first coated with an ultraviolet-light sensitive material called *photoresist,* which is supplied as a liquid. A few drops of the liquid are placed on the wafer, and the wafer is spun at high speeds to produce a thin uniform coating. After spinning, a short "prebake" at relatively low temperatures (80°C–100°C) is performed to drive solvent out of the resist and to improve surface adhesion. The hardened photoresist is similar to a photographic emulsion.

The next step is to expose the resist through a "mask" using UV-light as illustrated in Fig. 4.11(b). The mask here is a carefully prepared glass or quartz photoplate containing a copy of the pattern to be transferred to the SiO_2 film. Darkened regions in the emulsion on the mask block the UV-light. Regions of the photoresist exposed to the light undergo a chemical reaction which varies with the type of resist being employed. In negative resists the areas where the light strikes become polymerized and more difficult to remove. When placed in a developer the polymerized regions remain, while the unexposed regions dissolve and wash away. The net result after development is pictured on the right-hand side of Fig. 4.11(c). Positive resists contain large amounts of a sensitizer that dramatically slows down the dissolution rate of the resist in an alkaline developer. This sensitizer breaks down when exposed to light leading to the preferential removal of the exposed regions as shown on the left-hand side of Fig. 4.11(c). As an informational aside, negative resist was widely used in early IC processing. Positive resist is now the main type in use because it affords better small-geometry control.



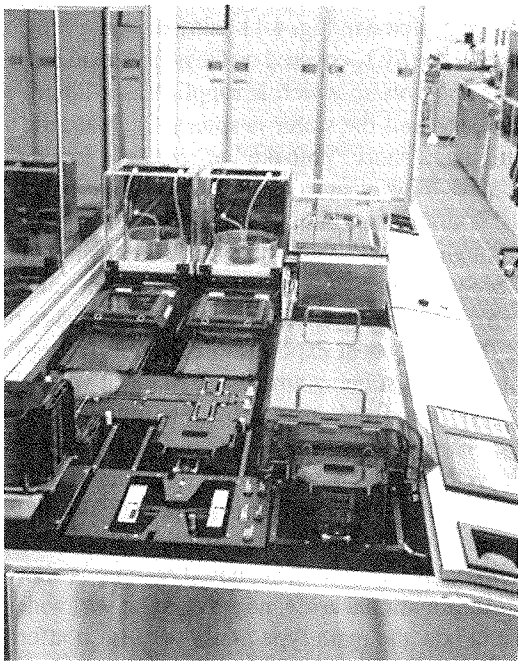


Figure 4.10 Production-line lithography station consisting of an SSI 150 coat and develop system (center-left in upper photograph) and a Nikon stepper (center-right in the upper photograph). The top of the SSI 150 system is shown more clearly in the lower photograph. The SSI 150 automatically performs the coat, prebake, develop, and harden operations described in the text. The intervening pattern exposure is performed in the Nikon stepper. Without human intervention, the wafer moves along the input track of the SSI 150, enters the Nikon stepper, exits the Nikon stepper, and moves back along one of the two output tracks of the SSI 150. (Photographs courtesy of Intel Corp.)

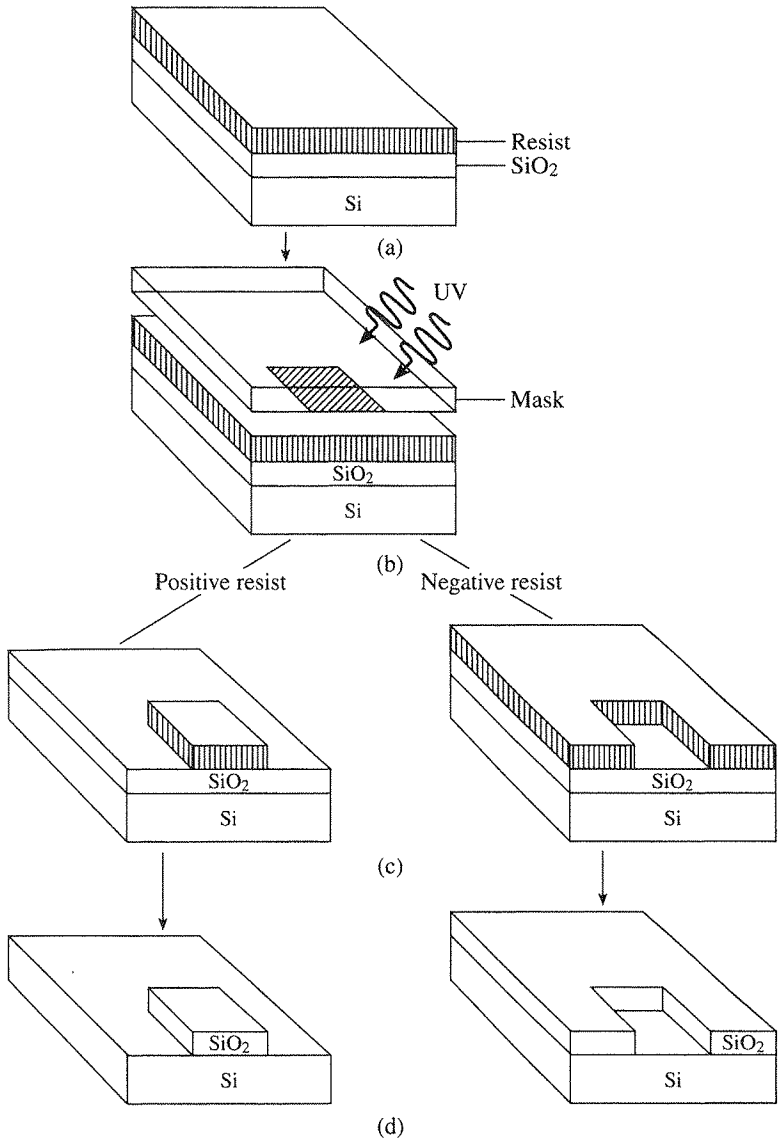


Figure 4.11 Major steps in the lithography process: (a) Application of resist; (b) resist exposure through a mask; (c) after development; (d) after oxide etching and resist removal. (From Jaeger[8], © 1988 by Addison-Wesley Publishing Co., Inc. Reprinted by permission of the publisher.)

The final set of steps involves transferring the pattern to the SiO_2 film. Immediately following exposure and development the photoresist is given a hard bake, heating for 20 to 30 minutes at 120°C–180°C, to further increase adhesion to the wafer and to improve resistance to the subsequent etch. An acid etch, buffered HF in the case of SiO_2, is then used to dissolve unprotected regions of the underlying film. Lastly, the photoresist, having completed its job, is stripped away. This is accomplished by using a chemical solution that swells and lifts the resist or by oxidizing (burning) the resist in an oxygen plasma system called an asher.

The limits of the UV-based lithography process just described are being pushed by the ever shrinking dimensions of modern devices. UV sources with shorter wavelengths and special compensating techniques have expanded the small dimension limit. Nevertheless, lithographic systems employing x-rays are being developed and are likely to come on-line in the near future.

4.1.5 Thin-Film Deposition

To connect device structures to the "outside world" requires the deposition and patterning of a metal layer. In fact, complex ICs have three and sometimes four electrically isolated metallization layers. Electrical isolation of the metal layers in turn requires the deposition of intervening dielectric layers. Thin films are also deposited to prevent the interdiffusion of materials and to protect the device or circuit from contamination. In-use methods for depositing the required films are reviewed next.

Evaporation

Evaporation is one of the older and more straightforward methods of thin-film deposition. As envisioned in Fig. 4.12, the material to be evaporated is placed in or on a resistance-heated source holder inside a vacuum chamber. To evaporate Al, for example, a short piece of Al wire would be placed on a tungsten filament or boat. The substrate on which the film is to be deposited is also positioned inside the chamber facing the source. The chamber is then evacuated, power supplied to the holder, and the source vaporized. Because of the reduced pressure, the source material travels unimpeded to the substrate and deposits as a thin film.

Generally speaking, hot-filament evaporation is subject to moderately high levels of contamination. Electron-beam evaporation, a variation of the process where the source is heated by an electron beam, eliminates contamination but generates device-degrading x-rays. Consequently, evaporation is seldom used in the production-line fabrication of modern ICs, although it is still extensively used in making simple devices where the cited problems are of minimal concern.

Sputtering

Sputtering, like evaporation, is performed in a vacuum chamber. The source material and the substrate (wafer) are placed on opposing parallel plates connected to a high-voltage power supply as pictured in Fig. 4.13. During a deposition the chamber is first evacuated

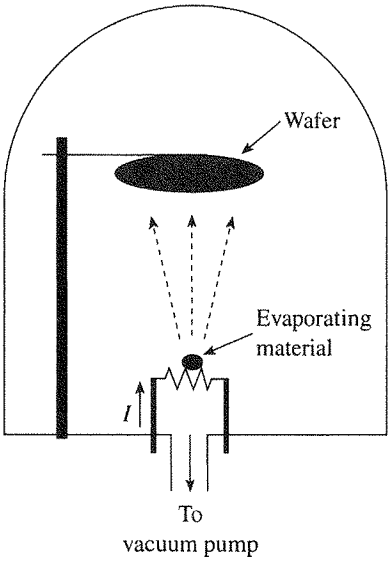


Figure 4.12 Hot-filament evaporation.

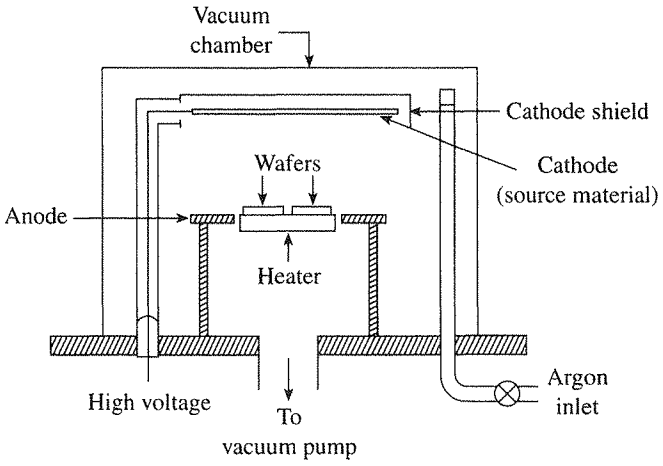


Figure 4.13 Schematic of a d.c. sputtering system. The source material covers the cathode while the wafer is mounted on the system anode. (From Jaeger[8], © 1988 by Addison-Wesley Publishing Co., Inc. Reprinted by permission of the publisher.)

of air and then a low-pressure amount of sputtering gas, typically Ar, is admitted into the chamber. Applying an interelectrode voltage ionizes the Ar gas and creates a plasma between the plates. Since the plate covered with source material is maintained at a negative potential relative to the substrate, Ar^+ ions are accelerated toward and into the source covered plate. The impacting Ar^+ ions in turn cause source atoms or molecules to be ejected from this plate. Being neutral, the ejected atoms or molecules readily travel to the substrate where they deposit to form the desired thin film. A d.c. power supply can be used when depositing metals, and an RF supply is necessary when depositing insulating films. When sputtering compounds it is sometimes necessary to introduce a gas of one of the components to assure the formation of a near-stoichiometric film. Providing low-temperature, low-contamination films with an acceptable throughput, sputtering has become the chief commercial method of depositing Al and other metals.

Chemical Vapor Deposition (CVD)

In chemical vapor deposition the thin film is formed from one or more gas phase components. Either a compound decomposes to form the film or a reaction between gas components takes place to form the film. Invariably the CVD reactions are surface catalyzed, preferentially taking place on the surface of wafers inserted into the gas stream. In-use CVD processes fall into one of three general categories. They are atmospheric pressure (APCVD or simply CVD), low-pressure (LPCVD), and plasma-enhanced (PECVD) processes. Atmospheric pressure depositions can be performed in relatively simple systems. Low-pressure often offers comparable kinetics with improved uniformity and less gas consumption. In plasma CVD the electrons in the plasma impart energy to the reaction gases, thereby enhancing the reactions and permitting very low deposition temperatures.

CVD reactors come in a variety of shapes and configurations; an example configuration employed in AP/LPCVD depositions is pictured in Fig. 4.14. CVD processing is routinely used to produce the masking and intermetallic dielectric films required in the formation of complex ICs. Polycrystalline Si, which functions as a pseudo-metal when heavily doped, is also deposited employing atmospheric and low-pressure CVD. APCVD, LPCVD, and PECVD are all typically used at some point in an IC process flow.

4.1.6 Epitaxy

Epitaxy is a special type of thin-layer deposition. Whereas depositions described in the preceding subsection yield either amorphous or polycrystalline films, epitaxy produces a crystalline layer that is an extension of the underlying semiconductor lattice. The word *epitaxy* is in fact derived from two Greek words meaning "upon-arranged." Commonly formed from the vapor-phase decomposition of silicon tetrachloride ($SiCl_4$) or a silane compound (SiH_4, SiH_2Cl_2, $SiHCl_3$) in a reactor very similar to those employed in CVD, additional Si literally grows following the lattice pattern of the pre-existing crystal. The doping of the epi-layer, intentionally distinct from that of the substrate, is controlled by introducing a dopant containing gas such as phosphine (PH_3), diborane (B_2H_6), or arsine

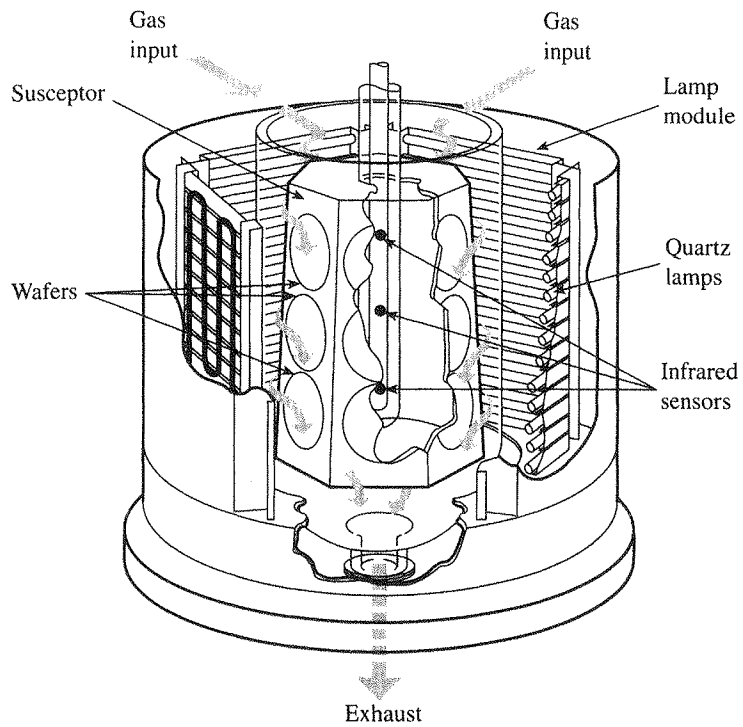


Figure 4.14 Cutaway view of a radiantly heated barrel-type reactor employed in both CVD processes and Si epitaxy. Wafers positioned vertically on the central susceptor are heated by the light from the surrounding quartz lamps. Process gases enter at the top, flow across the wafers, and exit coaxially at the bottom. (From Deacon[10], © 1984 by Lake Publishing. All rights reserved. Reprinted by permission of the publisher.)

(AsH_3) during the deposition. Standard whole-wafer epitaxy must of course be completed early in the process flow, prior to the growth or deposition of any permanent surface films. We should also note that starting Si wafers are sometimes purchased from the wafer supplier with the desired epi-layer already in place.

4.2 DEVICE FABRICATION EXAMPLES

The purpose of this section is to illustrate how the individual fabrication processes are combined and sequenced to produce solid state devices. Two examples are presented: The first describes the fabrication of a simple *pn* junction diode. The second outlines the making of a complex IC. Both presentations are intentionally qualitative in nature, including only major processing steps and a minimum of processing details.

4.2.1 *pn* Junction Diode Fabrication

Figure 4.15 graphically summarizes the major processing steps in the formation of a *pn* junction diode. The starting point is a flat, damage-free, single-crystal Si wafer. It is assumed that a preclean has removed all particulates, organic films, and adsorbed metal ions from the semiconductor surfaces. For this particular illustration we further assume the wafer is *p*-type, having been uniformly doped with boron during the formation of the crystal.

The initial steps in the process flow are in preparation for a subsequent phosphorus diffusion. First a thermal oxide is grown that will serve as a diffusion barrier. The oxide thickness must be comfortably greater than the projected masking thickness. Step 2 is a lithography process performed to open "diffusion" holes in the oxide that will eventually become the positions of the *pn* junction diodes. Specifically note that the Step 2 illustration in Fig. 4.15 assumes the use of a positive photoresist.

After a proper clean-up the wafer is next inserted into a phosphorus predeposition furnace and subsequently into a phosphorus drive furnace. The net result, as pictured in Step 3 in Fig. 4.15, is the formation of n^+-p junctions in surface regions not protected by the oxide. (The + in n^+ indicates a very high doping.) By way of clarification, some oxygen is required during the phosphorus predeposition as noted in the diffusion discussion. Also, the drive may be performed in an oxidizing atmosphere to minimize out-diffusion. Thus, to reopen the oxide holes, a subsidiary lithography step, not shown in the simplified process flow of Fig. 4.15, is usually necessary after diffusion.

The final steps facilitate connecting the device to the "outside world." Sputtering or possibly evaporation of Al yields a thin metal film over the entire surface of the wafer as pictured in Step 4 of Fig. 4.15. A lithography process, Step 5 in Fig. 4.15, is then performed to remove excess metal external to the area of the diffused junction. Normally a low-temperature ($\leq$500°C) anneal would also be performed to promote a low-resistance contact between the metal and Si.

With the completion of the metallization contact, the diodes become functional across the wafer. To produce commercial diodes, a diamond-edged saw would be used to cut the wafer into pieces containing a single device. (A wafer piece containing a single device or IC is called a *die*.) A large area metal contact is then made to the back of the die, a lead attached to the top surface contact, and the device encapsulated in protective plastic or hermetically sealed in a metallic package.

4.2.2 Computer CPU Process Flow

In this subsection we examine a fabrication process flow that has been used by Intel Corporation to produce computer CPUs and other ICs. The process flow description and associated figures are a direct excerpt, reprinted with permission, from *Components Quality and Reliability 1991/1992*, Intel Corporation, © 1990. No changes have been made to the excerpted material except for a renumbering of the figures and a modification of the figure captions. Figure 4.16, reproduced from the cited reference, displays a simplified block dia-

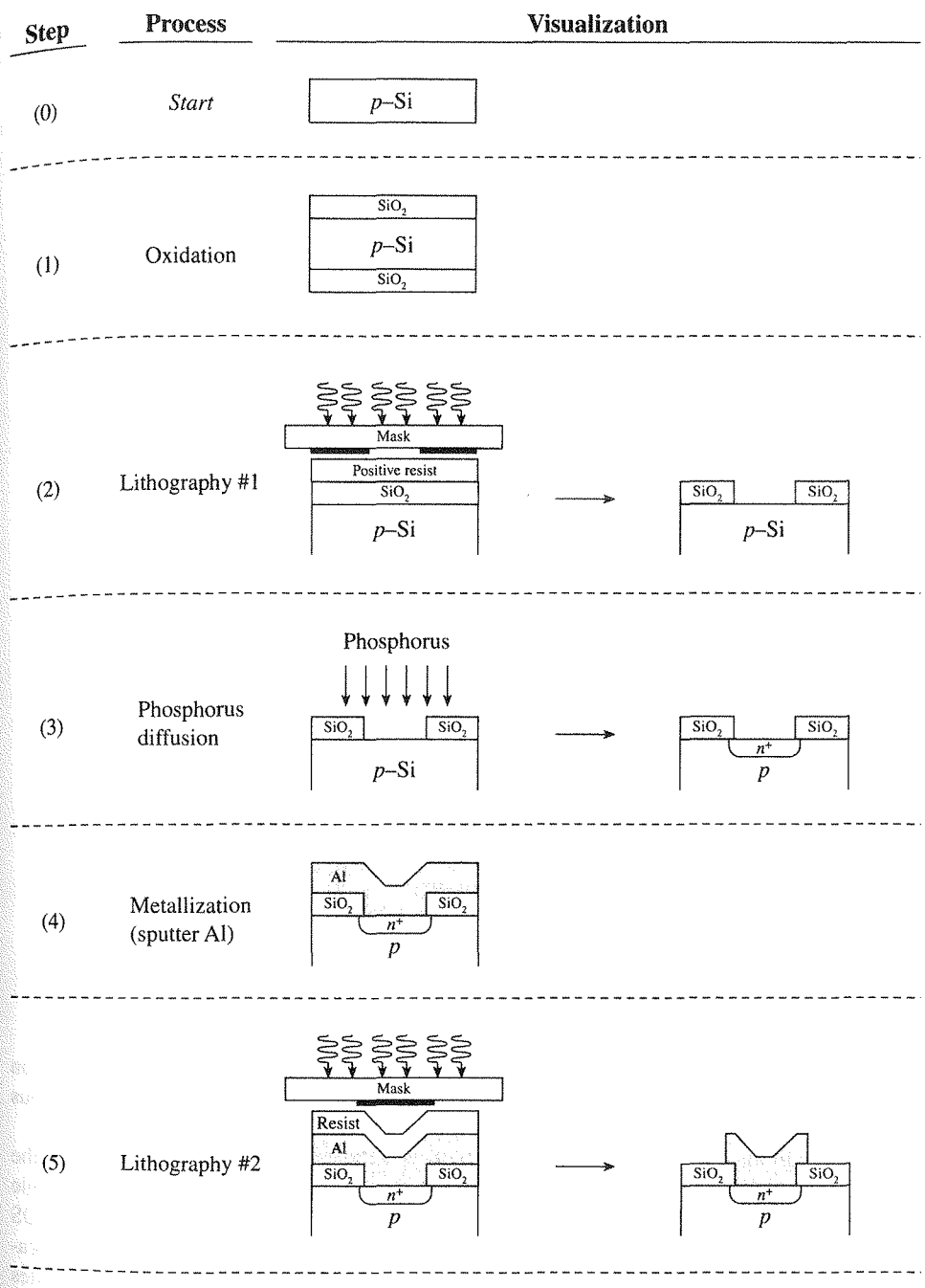


Figure 4.15 Graphical summary of the major processing steps in the formation of a *pn* junction diode.

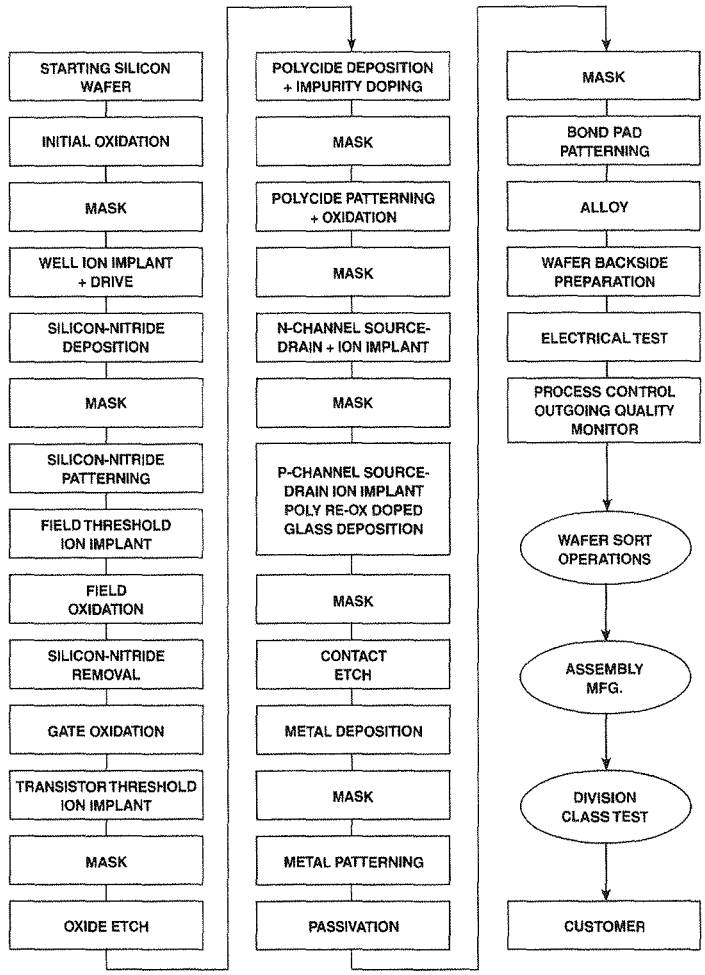


Figure 4.16 Simplified block diagram of the single poly, single metal, CHMOS process flow.

gram of the overall process flow. Figure 4.17, which is correlated with the text description of the component processes, shows cross-sectional sketches of the IC structure at various stages of construction.

Although most of the terms employed in the process description will be familiar to the reader, a few need to be clarified. For one, the fabrication sequence is referred to as a single poly, single metal, CHMOS process flow. "Poly" is short for polycrystalline Si. CHMOS stands for Complementary High-density Metal-Oxide-Semiconductor—a type of fabrication technology used to produce MOS transistors. "Single poly, single metal" indicates that one level of heavily doped polycrystalline Si and one level of metal are used to contact

and interconnect the MOS devices. Text references are also made to "wet" and "dry" etches. Wet etching is the familiar dissolution of a material, such as SiO_2, in a liquid chemical bath. Dry etching is the removal of a material using a plasma-enhanced gas-phase reaction.

Finally, the following process flow description is intended to be "looked over," not digested. At this point in the development, the reader is not equipped and not expected to understand the intricacies of either the processing or the IC structure.

---

1. **Starting Silicon Material and Well Definition** [Fig. 4.17(a)] To start CHMOS device production 150 mm (6″), high-resistivity, ⟨100⟩ orientation, single crystal, p-type (Boron doped) silicon wafers are used. P-type silicon is required to create n-channel transistors. To create p-channel transistors, necessary for CHMOS devices n-type (arsenic or phosphorous doped) silicon regions (n-wells) are implanted. The wafer is masked, then implanted to create p-type and n-type silicon regions on the same wafer. The n-well provides the background doping for the p-channel transistors while the p-type, start material (protected from the implant by the unexposed photoresist) serves as the background doping for n-channel transistors. A high temperature drive cycle completes the formation of the well by thermal dopant transition.

2. **Field Threshold Implant and Field Oxidation** [Fig. 4.17(b)]. Nitride (Si_3N_4) is deposited, masked, then etched. The etched nitride regions define the location of the field threshold ion plant [Fig. 4.17(b)] and the locations where oxide SiO_2 is permitted to grow during field oxidation [Fig. 4.17(c)]. Areas where nitride remains mask (prevent) oxide growth. These regions are where transistors will be built. The thick (approximately 6000 Å) field oxide isolates adjacent transistors in order to prevent electrical interactions. After field oxidation, the nitride mask is removed.

3. **Gate Oxidation** [Fig. 4.17(d)]. A thin thermal oxide is grown across the wafer. The portions of this oxide remaining after subsequent processing will provide the required gate oxide for the MOS transistors. Device performance is closely related to the growth of a dense, high-quality gate oxide.

4. **Transistor Threshold Ion Implant** [Fig. 4.17(d)]. The boron implant adjusts the threshold voltage (Vt) of p-channel and n-channel devices to the desired level. Thick field oxide prevents the boron from penetrating into the isolation regions.

5. **Polysilicon-to-Diffusion Contacts** [Fig. 4.17(e)]. Openings are defined in the thin gate oxide region where contact between poly conductors and diffused regions in the silicon substrate (buried contacts) is required.

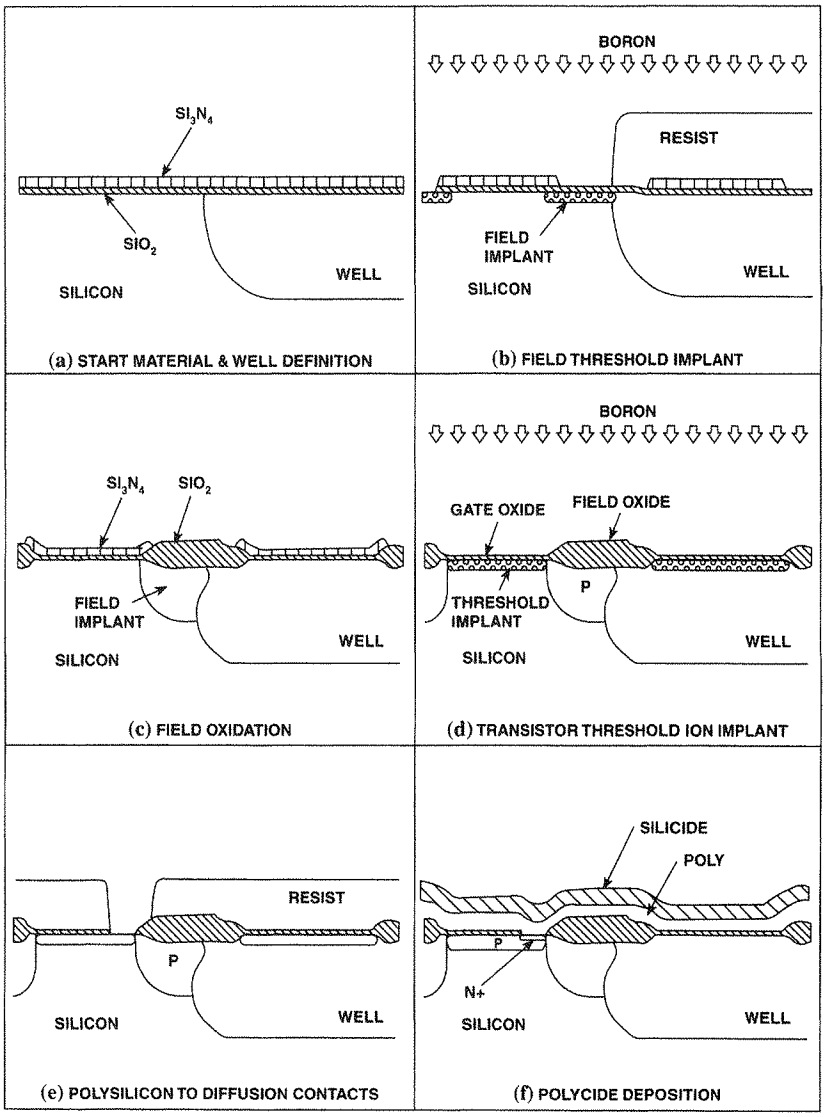


Figure 4.17 Cross-sections through the CHMOS structure at select points in the process flow.

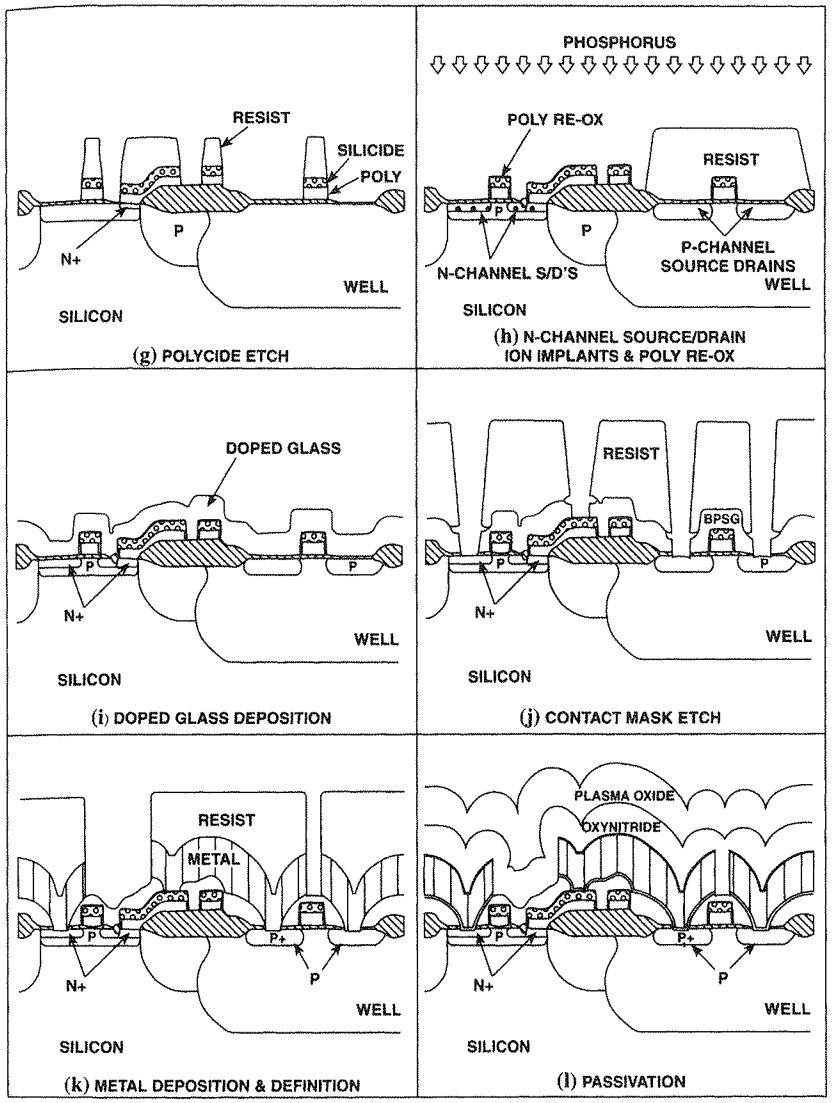


Figure 4.17 *Continued.*

ow.

6. **Polycide Deposition, Impurity Doping, Patterning and Oxidation** [Fig. 4.17(f)]. A polycrystalline silicon layer, about 1500 Å thick, is deposited over the entire wafer by CVD techniques. The polysilicon is subsequently doped with phosphorous via gas-source diffusion techniques for conductivity. The phosphorous is diffused through the poly and into the substrate, minimizing the resistance of the buried contact. About 1500 Å of silicide is conformally deposited on the doped polysilicon. Silicide is used in small geometry transistors to increase device performance speeds. The polycide stack is defined by plasma dry etch techniques with resist acting as the mask [Fig. 4.17(g)]. The polycide stack is oxidized to protect the transistor gates.

7. **Source and Drain Implant and Poly Re-Ox** [Fig. 4.17(h)]. The remaining polycide stack, thin oxide over the poly, field oxide and resist provide a mask for p-channel and n-channel source-drain implants which occurs at this point. (This technique is commonly known as a "self-aligned source/drain" process, since the source and drain are aligned directly to the gate, which defines their location relative to the channel area.) All remaining poly oxide is then removed and the poly re-ox step performed. The re-ox step is a high-temperature thermal oxidation step which drives the source-drain implant into the silicon, provides a high-integrity dielectric on the polysilicon (essential in a double poly process) and grows an oxide on the exposed source-drain regions to prevent dopant out-diffusion during subsequent phosphosilicate glass deposition and processing.

8. **Doped Glass Deposition** [Fig. 4.17(i)]. A doped (boron, phosphorous or both) silicate glass layer, 5000 Å to 10,000 Å thick, is deposited via low-temperature CVD techniques to provide electrical isolation between the subsequent metal conductor lines and the underlying polycide gates and active device structures. The glass is then prepared for subsequent masking steps.

9. **Contact Mask and Etch** [Fig. 4.17(j)]. Windows in the resist are exposed and opened to define the location of metal-to-polycide and metal-to-silicon contact holes. Depending on the process, contact holes are wet, dry, or wet and dry etched. In a wet/dry combination, the wet etch gives a shallow slope near the top of the contact. This improves metal step coverage. The remaining doped silicon glass is then removed by an isotropic dry etch which leaves the contact sidewalls vertical.

10. **Metal Deposition and Definition** [Fig. 4.17(k)]. Metal conductor layer is sputter-deposited into the wafer and defined. The metal electrically connects the transistors to the outside world. The voltage applied by these metal lines turns the transistors "on" and "off."

11. **Passivation** [Fig. 4.17(l)]. Plasma enhanced CVD techniques are used to deposit a dual film passivation layer over the entire wafer surface. The dual layer consist of an

oxynitride underneath a plasma oxide. The oxynitride provides long-term field reliability, by being designed to be a contamination and moisture barrier. The latter is particularly critical if the dice are to be assembled in plastic packages, plastic itself being a poor moisture barrier. The plasma oxide layer provides handling protection for the wafer and individual devices.

Most of the remaining production-line and testing steps cited in the Fig. 4.16 block diagram are in preparation for eventual IC die separation and packaging. The photograph of a sample CPU die, along with an identification of its functional units, are displayed in Fig. 4.18. The pictured Pentium™ Processor die, fabricated following an upgraded version of the described processing, is approximately 1.5 cm square and contains 3.1 million transistors.

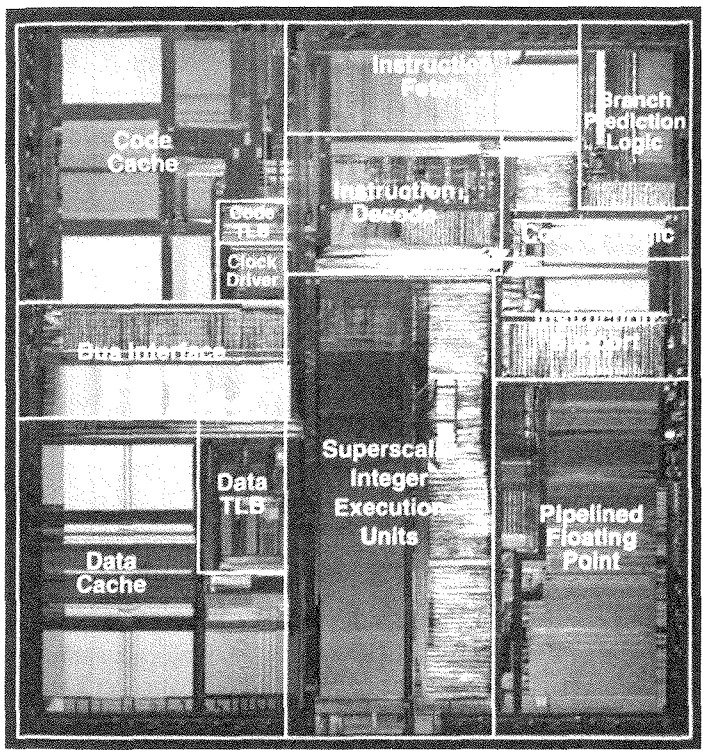


Figure 4.18 The Intel Pentium™ Processor and its functional units. (Photograph and functional unit identification courtesy of Intel Corporation.)

4.3 SUMMARY

The overview of silicon device fabrication hopefully provided insight into the construction and physical nature of device structures. First, common component processes were examined, including oxidation, diffusion, ion implantation, lithography, evaporation, sputtering, chemical vapor deposition, and epitaxy. The fabrication of a simple *pn* junction diode and a moderately complex IC process flow were then considered to illustrate how the individual processes are combined and sequenced in producing actual structures.

uction
exam-
ering,
le and
vidual

R1 Part I
SUPPLEMENT AND REVIEW

R1

ALTERNATIVE / SUPPLEMENTAL READING LIST

Recommended Readings

Author(s)	*Type (A = Alt., S = Supp.)*	*Level*	*Relevant Chapters*
For Semiconductor Physics			
Ferendeci	A/S	Advanced Undergraduate/ Introductory Graduate	1–6
Neamen	A	Undergraduate	1–6
Pierret	S	Introductory Graduate	1–6
Streetman	A	Undergraduate	1–4
Tyagi	A/S	Advanced Undergraduate/ Introductory Graduate	1–5
For Device Fabrication			
Jaeger	S	Undergraduate	All
Runyan and Bean	S	Professional	All
Sze (editor)	S	Professional	All

(1) A. M. Ferendeci, *Physical Foundations of Solid State and Electron Devices,* McGraw-Hill, New York, © 1991.

(2) R. C. Jaeger, *Introduction to Microelectronic Fabrication,* Vol. V in the Modular Series on Solid State Devices edited by G. W. Neudeck and R. F. Pierret, Addison-Wesley, Reading, MA, © 1988.

(3) D. A. Neamen, *Semiconductor Physics and Devices, Basic Principles,* Irwin, Homewood, IL, © 1992.

(4) R. F. Pierret, *Advanced Semiconductor Fundamentals,* Vol. VI in the Modular Series on Solid State Devices edited by G. W. Neudeck and R. F. Pierret, Addison-Wesley, Reading, MA, © 1987.

R1

(5) W. R. Runyan and K. E. Bean, *Semiconductor Integrated Circuit Processing Technology,* Addison-Wesley, Reading, MA, © 1990.

(6) B. G. Streetman, *Solid State Electronic Devices,* 4th edition, Prentice Hall, Englewood Cliffs, NJ, © 1995.

(7) S. M. Sze (editor), *VSLI Technology,* 2nd edition, McGraw Hill, New York, © 1988.

(8) M. S. Tyagi, *Introduction to Semiconductor Materials and Devices,* John Wiley & Sons, New York, © 1991.

Other Recent Texts

(1) A. Bar-Lev, *Semiconductors and Electronic Devices,* 3rd edition, Prentice Hall, Inc., New York, © 1993.

(2) D. H. Navon, *Semiconductor Microdevices and Materials,* Holt, Rinehart and Winston, New York, © 1986.

(3) D. L. Pulfrey and N. G. Tarr, *Introduction to Microelectronic Devices,* Prentice Hall, Englewood Cliffs, NJ, © 1989.

(4) C. T. Sah, *Fundamentals of Solid-State Electronics,* World Scientific, Singapore, © 1991.

(5) J. Singh, *Semiconductor Devices, an Introduction,* McGraw-Hill, New York, © 1994.

(6) S. M. Sze, *Semiconductor Devices, Physics and Technology,* John Wiley & Sons, New York, © 1985.

(7) E. S. Yang, *Microelectronic Devices,* McGraw Hill, New York, © 1988.

(8) M. Zambuto, *Semiconductor Devices,* McGraw Hill, New York, © 1989.

(↓ Reader-added texts)

FIGURE SOURCES / CITED REFERENCES

(1) W. Shockley, *Electrons and Holes in Semiconductors,* Litton Educational Publishing, Inc., © 1950.

(2) S. M. Sze, *Physics of Semiconductor Devices,* 2nd edition, John Wiley & Sons, New York, © 1981.

(3) W. Zuhlehner and D. Huber, "Czochralski Grown Silicon," *Crystals* **8,** Springer-Verlag, Berlin, © 1982.

(4) C. Jacoboni, C. Canali, G. Ottaviani, and A. A. Quaranta, "A Review of Some Charge Transport Properties of Silicon," Solid-State Electronics, **20,** 77 (1977).

(5) P. M. Smith, J. Frey, and P. Chatterjee, "High-Field Transport of Holes in Silicon," Applied Physics Letters, **39,** 332 (Aug. 1981).

(6) D. K. Schroder, *Semiconductor Material and Device Characterization,* John Wiley & Sons, New York, © 1990.

(7) R. F. Pierret, *Advanced Semiconductor Fundamentals,* Vol. VI in the Modular Series on Solid State Devices edited by G. W. Neudeck and R. F. Pierret, Addison-Wesley, Reading, MA, © 1987.

(8) R. C. Jaeger, *Introduction to Microelectronic Fabrication,* Vol. V in the Modular Series on Solid State Devices edited by G. W. Neudeck and R. F. Pierret, Addison-Wesley, Reading, MA, © 1988.

(9) W. R. Runyan and K. E. Bean, *Semiconductor Integrated Circuit Processing Technology,* Addison-Wesley, Reading, MA, © 1990.

(10) T. Deacon, "Silicon Epitaxy: An Overview," Microelectronic Manufacturing and Testing, September 1984.

(↓ Reader-added references)

R1

REVIEW LIST OF TERMS

Defining the following terms using your own words provides a rapid review of the Part I material.

(1) amorphous
(2) polycrystalline
(3) crystalline
(4) lattice
(5) unit cell
(6) ingot
(7) carrier
(8) electron
(9) hole
(10) conduction band
(11) valence band
(12) band gap
(13) effective mass
(14) intrinsic semiconductor
(15) extrinsic semiconductor
(16) dopant
(17) donor
(18) acceptor
(19) n-type material
(20) p-type material
(21) n^+ (or p^+) material
(22) majority carrier
(23) minority carrier
(24) density of states
(25) Fermi function
(26) Fermi energy (or level)
(27) nondegenerate semiconductor
(28) degenerate semiconductor
(29) charge neutrality
(30) extrinsic temperature region
(31) intrinsic temperature region
(32) freeze-out
(33) drift
(34) scattering
(35) thermal motion
(36) drift velocity
(37) saturation velocity
(38) current density
(39) mobility
(40) resistivity
(41) conductivity
(42) band bending
(43) diffusion
(44) diffusion coefficient
(45) recombination
(46) generation
(47) band-to-band R–G
(48) R–G centers
(49) E–$\mathbf{k}$ plots
(50) direct semiconductor
(51) indirect semiconductor
(52) photogeneration
(53) absorption coefficient
(54) low-level injection
(55) equilibrium
(56) perturbation
(57) steady state
(58) minority carrier lifetime
(59) minority carrier diffusion length
(60) quasi-Fermi level
(61) dry oxidation
(62) wet oxidation
(63) predeposition
(64) drive-in
(65) diffusion mask
(66) ion implantation
(67) dose
(68) projected range
(69) photoresist
(70) negative resist
(71) positive resist
(72) sputtering
(73) CVD
(74) LPCVD
(75) PECVD
(76) epitaxy
(77) die
(78) poly
(79) wet etch
(80) dry etch

PART I—REVIEW PROBLEM SETS AND ANSWERS

The following problem sets were designed assuming a knowledge—at times an integrated knowledge—of the subject matter in Chapters 1–3 of Part I. The sets could serve as a review or as a means of evaluating the reader's mastery of the subject. Problem Set A was adapted from a one-hour "open-book" examination; Problem Set B was adapted from a one-hour "closed-book" examination. An answer key is included at the end of the problem sets.

Problem Set A

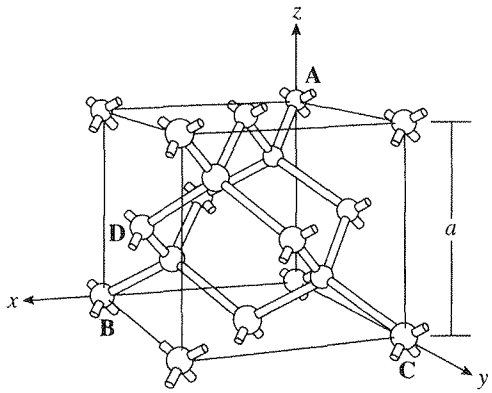
Special Instructions:

(1) You may employ an OPEN book in answering the Problem Set A questions.

(2) Unless instructed otherwise, employ parameters accurate to three-places in performing numerical computations.

Problem A1

Answer the questions that follow making use of the unit cell for the Si crystal lattice reproduced here:



(a) If, as pictured, the origin of coordinates is located at the lower back corner of the unit cell, what are the Miller indices of the plane passing through the points ABC?

(b) What are the Miller indices of the direction vector running from the origin of coordinates to the point D?

(c) In terms of the lattice constant a, what is the distance between nearest neighbors in the Si lattice?

(d) Modifying the pictured Si unit cell, indicate how one would visualize a *donor.*

R1

Problem A2

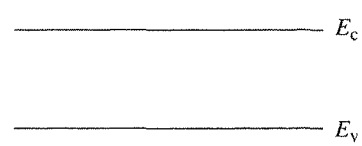
A Si wafer is uniformly doped with $N_D = 10^{17}/\text{cm}^3$ donors and maintained under equilibrium conditions at $T = 300$ K. $\tau_p = 10^{-6}$ sec. Complete the following table characterizing the wafer.

Parameter	*Value*	*Units*	*Explanation*
E_G			—
n_i	10^{10}	cm^{-3}	—
n_0			
p_0			
$E_F - E_i$			
ρ(charge density)			
μ_n			
ρ(resistivity)			
L_P			
D_P			

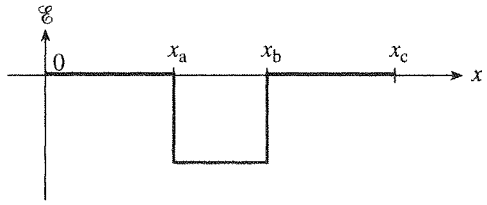
Problem A3

(a) On the energy band diagram provided below, indicate the usual positioning of the following energy levels:
   (i) E_i . . . instrintsic Fermi level
   (ii) E_D . . . donor energy level
   (iii) E_A . . . acceptor energy level
   (iv) E_T . . . trap or R–G center energy level
   (v) E_F for a *degenerately* doped p-type material

   Add comments as necessary to forestall any misinterpretation of your graphical answer.



(b) A Si device is maintained under equilibrium conditions at $T = 300$ K. Given the electric field inside the device as is pictured next, $N_A = 10^{16}/\text{cm}^3$ for $0 \le x \le x_a$, and $N_D = 10^{16}/\text{cm}^3$ for $x_b \le x \le x_c$, draw the energy band diagram for the device. Include E_c, E_v, E_i, and a carefully positioned E_F on your diagram. Also indicate how you arrived at your answer.

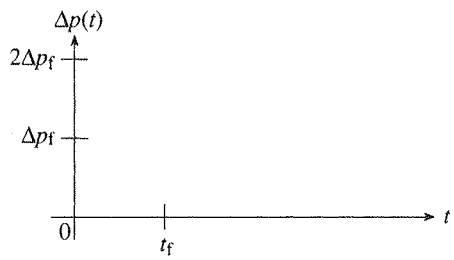


(c) What is the electrostatic potential drop across the device described in part (b); i.e., what is $V(x = x_c) - V(x = 0)$? Record your work.

Problem A4

An $N_D = 10^{15}/\text{cm}^3$ Si sample maintained at room temperature is subject to two flashes of light from a stroboscope. Each flash creates $\Delta p_f = 10^{10}/\text{cm}^3$ holes *uniformly* throughout the Si sample. The first flash occurs at $t = 0$ and the second occurs at $t = t_f$, where t_f is comparable to τ_p. The flashes of light themselves are of infinitesimally short duration compared to τ_p.

(a) Sketch (based on qualitative reasoning) the expected general variation of $\Delta p(t)$ as a function of time.



(b) Do low-level injection conditions prevail at all times inside the Si sample? Explain.

(c) Starting from the appropriate differential equation, solve for $\Delta p(t)$ for all $t \geq 0$.

(d) Assuming $\Delta n(t) = \Delta p(t)$ inside the bar, determine the fractional change in the Si sample conductivity immediately after the first flash of light. Give both a symbolic and numerical value for the desired $[\sigma(0^+) - \sigma(0^-)]/\sigma(0^-)$. 0^+ and 0^- refer to times just before and just after the first flash, respectively.

Problem Set B

Special Instructions:

(1) Problem Set B questions are to be answered with the book CLOSED.

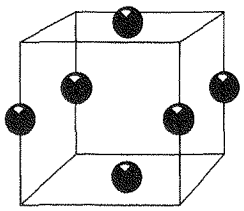
R1

(2) In answering multiple-choice questions, choose the *best available answer.* Select only *one answer* per question.

(3) Employ $n_i = 10^{10}/\text{cm}^3$ and $kT = 0.0259$ eV in numerical computations involving Si samples maintained at 300 K.

I. General Information

(1) As shown in the accompanying figure, a cubic unit cell has one atom centered halfway down each of the vertical edges of the cell *and* one atom positioned in the middle of the top and bottom faces of the cell. How many atoms are there within the cell?

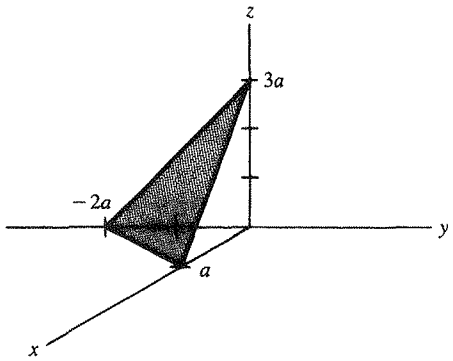


(a) 4
(b) 2
(c) 1
(d) 1/2

(2) What is the name of the *lattice* characterized by the unit cell in Question (1)?

(a) Simple cubic
(b) bcc
(c) fcc
(d) Diamond

(3) Determine the Miller indices of the plane pictured below.



(a) $(1\bar{2}3)$
(b) (123)
(c) $(3\bar{2}1)$
(d) (321)
(e) $(6\bar{3}2)$

(4) Which of the following directions is *not* perpendicular to the [100] direction?

(a) $[0\bar{1}1]$

(b) [032]

(c) [001]

(d) $[\bar{1}00]$

(5) In a particular semiconductor the probability of electrons occupying states at an energy kT above the bottom of the conduction band is e^{-10}. Determine the position of the Fermi level in the given material.

(a) $E_F = E_c$

(b) $E_c - E_F \simeq 9kT$

(c) $E_c - E_F = 10kT$

(d) $E_F = E_c + kT$

(6) Which of the following is the dominant carrier scattering mechanism in lowly doped (high-purity) silicon at room temperature?

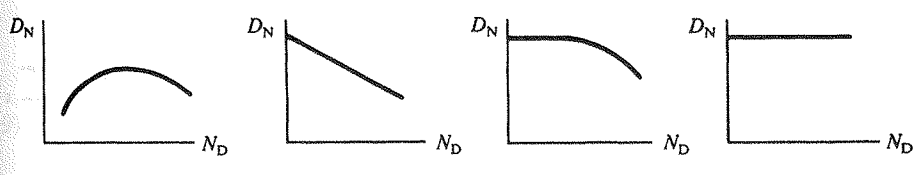
(a) Carrier-carrier scattering

(b) Lattice scattering

(c) Impurity ion scattering

(d) Piezoelectric scattering

(7) Which of the following sketches best describes the D_N versus N_D dependence of electrons in silicon at room temperature?



(a) (b) (c) (d)

(8) Using the *bonding* model, indicate how one visualizes a donor.

(9) Using the energy band diagram, indicate how one would visualize the freeze-out of electrons on donor sites as $T \rightarrow 0$ K.

(10) The resistivity of n-type material is typically smaller than the resistivity of comparably doped p-type material. Explain why.

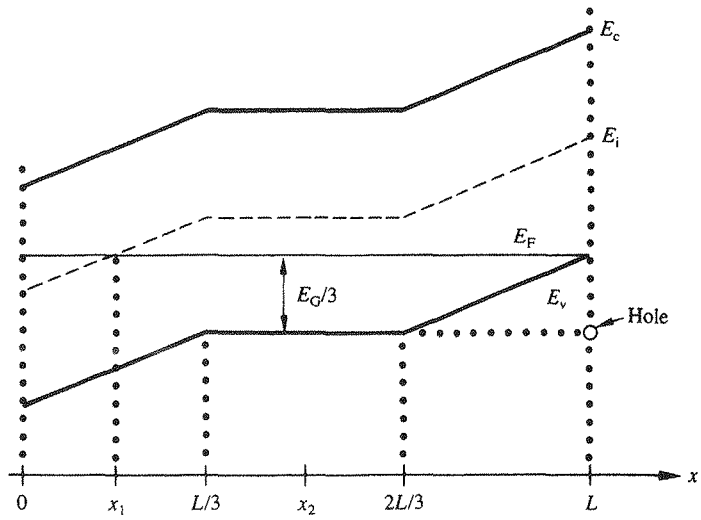
(11) Using the energy band diagram, indicate how one visualizes *recombination* via R–G centers.

(12) Define *drift velocity.*

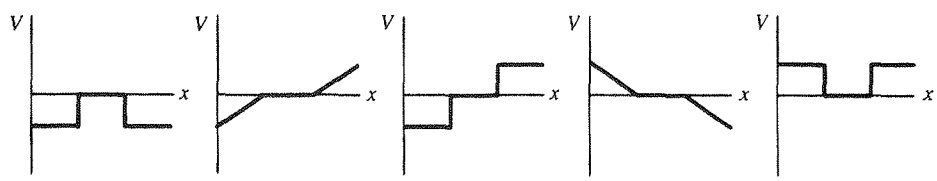
R1

II. Interpretation of Energy Band Diagrams

A silicon device maintained at 300 K is characterized by the following energy band diagram. Use the cited energy band diagram in answering Questions (13) to (20).

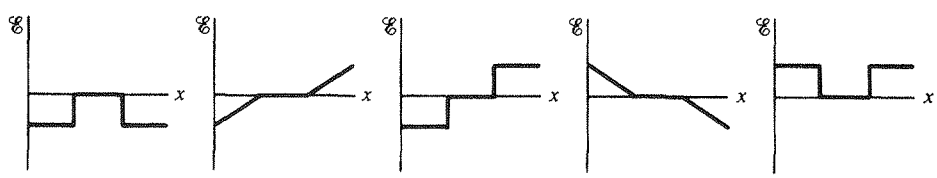


(13) The electrostatic potential (V) inside the semiconductor is as sketched below.



(a) (b) (c) (d) (e)

(14) The electric field ($\mathscr{E}$) is as sketched below.



(a) (b) (c) (d) (e)

(15) Do equilibrium conditions prevail?

(a) Yes

(b) No

(c) Can't be determined

(16) The semiconductor is degenerate

(a) Near $x = 0$

(b) For $L/3 \leq x \leq 2L/3$

(c) Near $x = L$

(d) Nowhere

(17) At $x = x_2$, $p = ?$

(a) $7.63 \times 10^6/\text{cm}^3$

(b) $1.35 \times 10^{13}/\text{cm}^3$

(c) $10^{10}/\text{cm}^3$

(d) $1.72 \times 10^{16}/\text{cm}^3$

(18) The electron current density (J_N) flowing at $x = x_1$ is

(a) Zero

(b) $\mu_n n_i E_G/L$

(c) $-\mu_n n_i E_G/L$

(d) $D_N[n(x_2) - n(0)]/L$

(19) The hole drift current density ($J_{P|drift}$) flowing at $x = x_1$ is

(a) Zero

(b) $\mu_p n_i E_G/L$

(c) $-\mu_p n_i E_G/L$

(d) $q\mu_p N_D(kT/q)/L$

(20) The kinetic energy of the hole shown on the diagram is

(a) E_v

(b) $-E_G/3$

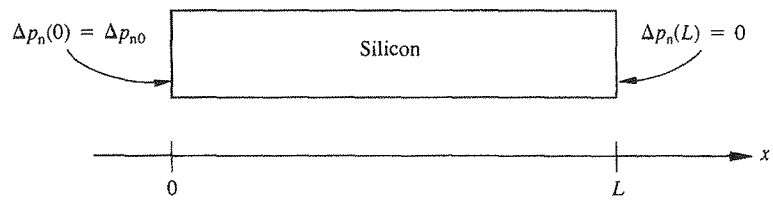
(c) $E_G/3$

(d) Zero

R1

III. Problem Solving

A uniformly doped n-type silicon bar of length L is maintained at room temperature under steady state conditions such that $\Delta p_n(0) = \Delta p_{n0} > 0$ and $\Delta p_n(L) = 0$. $N_D = 10^{15}$ / cm^3, $\Delta p_{n0} \ll n_0$, and there are no "other processes" (including photogeneration) occurring in side the bar.



(21) Determine p_0, the equilibrium hole concentration.

(a) $p_0 \approx 10^{10}/\text{cm}^3$

(b) $p_0 \approx 10^{15}/\text{cm}^3$

(c) $p_0 \approx 10^{5}/\text{cm}^3$

(d) Not given

(22) Does low-level injection exist under steady state conditions?

(a) Yes

(b) No

(c) Can't tell

(23) The equation (simplest form possible) one must solve to determine $\Delta p_n(x)$ is

(a) $0 = D_P \dfrac{d^2 \Delta p_n}{dx^2}$

(b) $\dfrac{d\Delta p_n}{dt} = \dfrac{\Delta p_n}{\tau_p}$

(c) $\dfrac{\partial \Delta p_n}{\partial t} = D_P \dfrac{\partial^2 \Delta p_n}{\partial x^2} - \dfrac{\Delta p_n}{\tau_p}$

(d) $\dfrac{\partial p_n}{\partial t} = D_P \dfrac{\partial^2 \Delta p_n}{\partial x^2}$

(e) $0 = D_P \dfrac{d^2 \Delta p_n}{dx^2} - \dfrac{\Delta p_n}{\tau_p}$

(24) The general solution for $\Delta p_n(x)$ will be ($L_P^2 \equiv D_P \tau_p$)

(a) $A \exp(-x/L_P) + B \exp(x/L_P)$

(b) $A \exp(-t/\tau_p)$

(c) $A + Bx$

(d) $A \exp(-x/L_P)$

(25) The boundary conditions to be employed in solving for $\Delta p_n(x)$ are

(a) $\Delta p_n(0) = \Delta p_{n0}$ and $\Delta p_n(\infty) = 0$

(b) $\Delta p_n(0) = 10^{15}/\text{cm}^3$ and $\Delta p_n(\infty) = 0$

(c) $\Delta p_n(0) = \Delta p_{n0}$ and $\Delta p_n(L) = 0$

(d) $\Delta p_n(0) = 10^{15}/\text{cm}^3$ and $\Delta p_n(L) = 0$

Answers—Set A

Problem A1

(a) (111)

(b) [211]

(c) $(\sqrt{3}/4)a$

(d) One of the Si atoms in the unit cell is to be replaced with a non-Si atom exhibiting an extra bond (representing a weakly bound 5th valence electron).

Problem A2

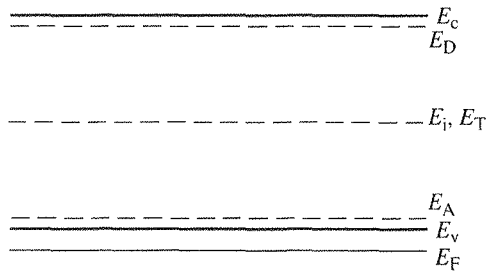
Parameter	*Value*	*Units*	*Explanation*
E_G	1.12	eV	—
n_i	10^{10}	cm^{-3}	—
n_0	10^{17}	cm^{-3}	$n_0 = N_D \quad (N_D \gg n_i)$
p_0	10^3	cm^{-3}	$p_0 = n_i^2/N_D$
$E_F - E_i$	0.417	eV	$E_F - E_i = kT\ln(N_D/n_i) = (0.0259)\ln(10^{17}/10^{10})$
ρ(charge density)	0	coul/cm^3	Charge neutrality prevails under the specified conditions.
μ_n	801	cm^2/V-sec	From Fig. 3.5

R1

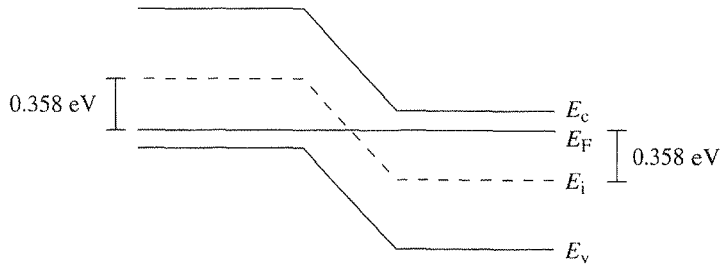
ρ (resistivity)	0.078	ohm-cm	$\rho = \dfrac{1}{q\mu_n N_D} = \dfrac{1}{(1.6 \times 10^{-19})(801)(10^{17})}$
L_P	2.93×10^{-3}	cm	$L_P = \sqrt{D_P \tau_p} = \sqrt{(8.57)(10^{-6})}$
D_P	8.57	cm²/sec	$D_P = (kT/q)\mu_p = (0.0259)(331)$

Problem A3

(a) E_T is located near midgap, while E_F must be positioned below $E_v + 3kT$.



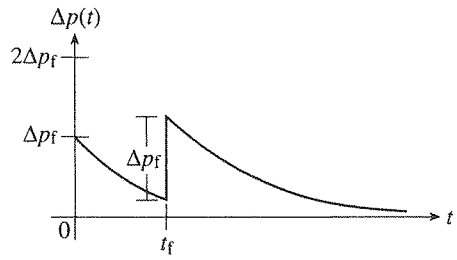
(b) Under equilibrium conditions, we know that E_F is position-independent. Since $\mathscr{E} = (1/q)(dE_c/dx) = (1/q)(dE_i/dx) = (1/q)(dE_v/dx)$, from the $\mathscr{E}$ versus x plot, we deduce that the bands are flat for $0 \le x \le x_a$ and $x_b \le x \le x_c$. In the $x_a \le x \le x_b$ region the bands must exhibit a constant negative slope. Also, for $0 \le x \le x_a$, $E_i - E_F = kT\ln(N_A/n_i) = (0.0259)\ln(10^{16}/10^{10}) = 0.358$ eV; for $x_b \le x \le x_c$, $E_F - E_i = kT\ln(N_D/n_i) = 0.358$ eV. Combining the cited information, one concludes the energy band diagram is as sketched below.



(c) $\Delta V = -\dfrac{1}{q}\Delta E_c = \dfrac{1}{q}[E_c(0) - E_c(x_c)] = \dfrac{1}{q}[E_i(0) - E_i(x_c)] = 0.716$ V

Problem A4

(a)



(b) Yes. $\Delta p_{max} \leq 2\Delta p_f = 2 \times 10^{10}/\text{cm}^3 \ll n_0 = N_D = 10^{15}/\text{cm}^3$. Thus $\Delta p \ll n_0$ as required for low-level injection.

(c) In the given problem, the minority carrier diffusion equation simplifies to

$$\frac{d\Delta p_n}{dt} + \frac{\Delta p_n}{\tau_p} = 0$$

For $0 \leq t \leq t_f \ldots$

$$\Delta p(0) = \Delta p_f \qquad \ldots \text{ boundary condition}$$

$$\Delta p(t) = Ae^{-t/\tau_p} \qquad \ldots \text{ general solution}$$

$$A = \Delta p_f$$

$$\Delta p(t) = \Delta p_f e^{-t/\tau_p} \qquad \Leftarrow \text{ answer}$$

For $t \geq t_f \ldots$

$$\Delta p(t_f) = \Delta p_f + \Delta p_f e^{-t_f/\tau_p} = \Delta p_f(1 + e^{-t_f/\tau_p}) \qquad \ldots \text{ boundary condition}$$

$$\Delta p(t) = Ae^{-(t-t_f)/\tau_p} \qquad \ldots \text{ general solution}$$

$$A = \Delta p_f(1 + e^{-t_f/\tau_p})$$

$$\Delta p(t) = \Delta p_f(1 + e^{-t_f/\tau_p})e^{-(t-t_f)/\tau_p} \qquad \Leftarrow \text{ answer}$$

(d)

$$\sigma = 1/\rho = q(\mu_n n + \mu_p p)$$

$$\sigma(0^-) \cong q\mu_n N_D \qquad \ldots n_0 \cong N_D \text{ and } n_0 \gg p_0$$

$$\sigma(0^+) = q[\mu_n(n_0 + \Delta n) + \mu_p(p_0 + \Delta p)] \cong q\mu_n N_D + q(\mu_n + \mu_p)\Delta p_f$$

$$\frac{\Delta\sigma}{\sigma(0^-)} = \frac{q(\mu_n + \mu_p)\Delta p_f}{q\mu_n N_D} = \left(1 + \frac{\mu_p}{\mu_n}\right)\frac{\Delta p_f}{N_D} \qquad \Leftarrow \text{ answer symbolic}$$

$$= \left(1 + \frac{458}{1345}\right)\frac{10^{10}}{10^{15}} = 1.34 \times 10^{-5} \qquad \Leftarrow \text{ answer numeric}$$

R1

Answers—Set B

(1) b . . . 1/2 × (2 face atoms) + 1/4 × (4 edge atoms) = 2

(2) b

(3) e . . . 1, −2, 3 — intercepts
1, −1/2, 1/3 — 1/intercepts
$(6\bar{3}2)$ — indices

(4) d . . . Any direction with a nonzero first integer will *not* be perpendicular to the [100] direction.

(5) b . . . From the statement of the problem, we conclude

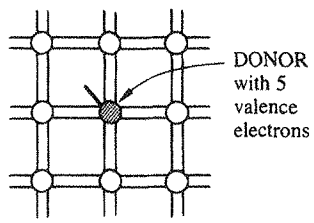
$$f(E_c + kT) = \frac{1}{1 + e^{(E_c + kT - E_F)/kT}} = e^{-10}$$

$$E_c - E_F = kT[\ln(e^{10} - 1) - 1] \simeq 9kT$$

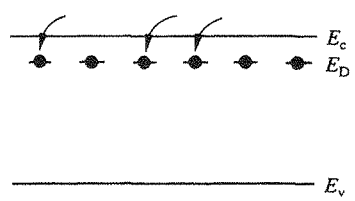
(6) b

(7) c . . . For nondegenerate dopings $D_N = (kT/q)\mu_n$ and μ_n versus N_D has the form of sketch (c). (See Fig. 3.5.)

(8)



(9)



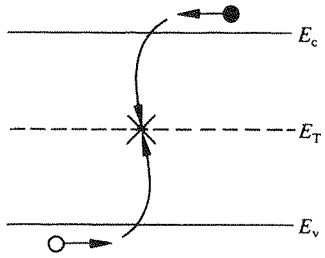
(10)

$$\rho = \frac{1}{q\mu_n N_D} \quad \ldots n\text{-type}$$

$$\rho = \frac{1}{q\mu_p N_A} \quad \ldots p\text{-type}$$

Typically, $\mu_p < \mu_n$ for a given semiconductor and temperature. Thus $\rho(p\text{-type})$ is greater than $\rho(n\text{-type})$ if $N_D = N_A$.

(11)



(12) *Drift velocity* is the average electron or hole velocity resulting from an applied electric field. (One might also mention that at low fields $v_d = \mu\mathscr{E}$, where μ is the carrier mobility.)

(13) d . . . V has the same shape as the "upside-down" of the bands.

(14) e . . . $\mathscr{E}$ is proportional to the slope of the bands.

(15) a . . . E_F is invariant with position.

(16) c . . . E_F approaches E_v near $x = L$.

(17) b . . .

$$E_i - E_F = E_G/2 - E_G/3 = E_G/6$$

$$p = n_i e^{(E_i - E_F)/kT} = 10^{10} e^{1.12/(6\times 0.0259)} = 1.35 \times 10^{13}/\text{cm}^3$$

(18) a . . . Under equilibrium conditions $J_N = 0$.

(19) b . . .

$$\begin{aligned} J_{P|drift} &= q\mu_p p\mathscr{E} \\ \text{At } x &= x_1 \\ p &= n_i \qquad (E_i = E_F) \\ \mathscr{E} &= \frac{1}{q}\frac{dE_i}{dx} = \frac{1}{q}\frac{\Delta E_i}{\Delta x} = \frac{1}{q}\frac{(E_G/3)}{(L/3)} = \frac{E_G}{qL} \\ J_{P|drift} &= \mu_p n_i E_G/L \end{aligned}$$

(20) c . . . K. E. $= E_v(L) - E_{hole} = E_G/3$

(21) c . . . $n_0 \simeq N_D = 10^{15}/\text{cm}^3$

$$p_0 \simeq n_i^2/n_0 \simeq 10^{20}/10^{15} = 10^5/\text{cm}^3$$

(22) a . . . $\Delta p_{n|max} = \Delta p_{n0} \ll n_0$

R1

(23) e . . . Steady state conditions and the absence of photogeneration allow one to eliminate the $\partial/\partial t$ and G_L terms in the minority carrier diffusion equation (Eq. 3.54b).

(24) a . . . This is Solution No. 1 cited in Subsection 3.4.3.

(25) c

Electrical Engineering/Semiconductor Devices

Semiconductor Device Fundamentals

Robert F. Pierret

Although roughly a half-century old, the field of study associated with semiconductor devices continues to be dynamic and exciting. New and improved devices are being developed at an almost frantic pace. While the number of devices in complex integrated circuits increases through the millions and the side-length of chips is measured in centimeters, the individual devices are literally being shrunk to atomic dimensions. Moreover, semiconductor properties desired for a given device structure but not available in nature are being produced artificially; in essence, the semiconductor properties themselves are now being engineered to fit device specifications. This book should be viewed as a gateway to what the reader hopefully will agree is the fascinating realm of semiconductor devices.

Highlights

- Introduces and explains the basic terminology, models, properties, and concepts associated with semiconductors and semiconductor devices.
- Provides detailed insight into the internal workings of the "building-block" device structures such as the *pn* junction diode, Schottky diode, BJT, and MOSFET.
- Presents information about a wide variety of additional devices, including solar cells, LEDs, HBTs, and modern field effect devices.
- Systematically develops the analytical tools needed to solve practical device problems.

Special Features

- The *computer-based exercises and homework* [illegible] permit the replacement of plug-and-chug boredom with more challenging a[illegible]
- The *supplement-and-review mini-chapters* c[illegible] review problem sets with answers, and deta[illegible]
- Providing a change of pace with a minimum [illegible] *read-only chapters* are intended to be fun-re[illegible]
- Other special features include tables that s[illegible] difficulty level, and suggested weighting of [illegible] summaries in the critical beginning chapte[illegible] computer-generated plots.

X0043AS939
Semiconductor Device Fundamentals
New

Robert F. Pierret is a Professor of Electrical and Computer [illegible] [illegible]yette, Indiana. He has been a member of the Purdue faculty since 1970 and administers the undergraduate Semiconductor Measurements Laboratory. Promoting an innovative total-quality approach to curriculum development, he has served in recent years as the Chairman of the ECE Curriculum Committee. Robert F. Pierret conceived and co-edits the popular Modular Series on Solid State Devices published by Addison-Wesley. He is author of four volumes in the series. He is also a consulting editor for Addison-Wesley and on the Editorial Advisory Board of the journal of *Solid-State Electronics*.

90000
9 780201 543933
ISBN 0-201-54393-1

Addison-Wesley Publishing Company