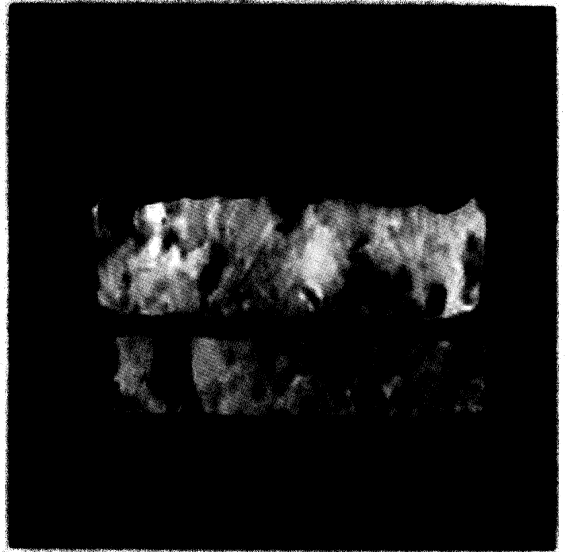


SEMICONDUCTOR DEVICES

**Physics
and
Technology**



2nd Edition

S.M. Sze

CIRRUS EX. 1061 - 1/72

2ND EDITION

Semiconductor Devices

Physics and Technology

S. M. SZE

UMC Chair Professor

National Chiao Tung University

National Nano Device Laboratories

Hsinchu, Taiwan



JOHN WILEY & SONS, INC.

Acquisitions Editor *William Zobrist*
Marketing Manager *Katherine Hepburn*
Production Services Manager *Jeanine Furino*
Production Editor *Sandra Russell*
Designer *Harold Nolan*
Production Management Services *Argosy Publishing Services*

Cover Photography: A transmission-electron micrograph of a floating-gate nonvolatile semiconductor memory with a magnification of 100,000 times. (Photography courtesy of George T. T. Sheng.) For a discussion of the device, see Chapters 1, 6, and 14.

This book was typeset in *New Caledonia* by *Argosy Publishing* and printed and bound by *R. R. Donnelley and Sons, Inc. (Willard)*. The cover was printed by *The Lehigh Press*.

The paper in this book was manufactured by a mill whose forest management programs include sustained yield harvesting of its timberlands. Sustained yield harvesting principles ensure that the number of trees cut each year does not exceed the amount of new growth.

The book is printed on acid-free paper. 

Copyright © 1985, 2002 by John Wiley & Sons, Inc. 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, scanning, or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (508) 750-8400, fax (508) 750-4470. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc. 605 Third Avenue, New York, NY 10158-0012, (212) 850-6008, E-mail: PERMREQ@WILEY.COM. To order books or for customer service call 1-800-CALL-WILEY (225-5945).

Library of Congress Cataloging in Publication Data:

Sze, S. M., 1936-

Semiconductor devices, physics and technology/S.M. Sze.—2nd ed.

p. cm.

Includes bibliographical references and index.

ISBN 0-471-33372-7 (cloth: alk. paper)

1. Semiconductors. I. Title.

TK7871.85 .S9883 2001

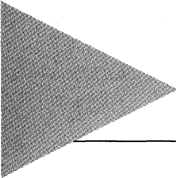
621.3815'2—dc21

2001026003

ISBN 0-471-33372-7

Printed in the United States of America

10 9 8 7 6 5 4 3



Contents

Preface iv

► CHAPTER 1

Introduction 1

- 1.1 Semiconductor Devices 1
- 1.2 Semiconductor Technology 7
- Summary 13

PART I

SEMICONDUCTOR PHYSICS

► CHAPTER 2

Energy Bands and Carrier Concentration in Thermal Equilibrium 17

- 2.1 Semiconductor Materials 17
- 2.2 Basic Crystal Structure 19
- 2.3 Basic Crystal Growth Technique 24
- 2.4 Valence Bonds 27
- 2.5 Energy Bands 28
- 2.6 Intrinsic Carrier Concentration 34
- 2.7 Donors and Acceptors 37
- Summary 44

► CHAPTER 3

Carrier Transport Phenomena 47

- 3.1 Carrier Drift 48
- 3.2 Carrier Diffusion 57
- 3.3 Generation and Recombination Processes 60
- 3.4 Continuity Equation 66
- 3.5 Thermionic Emission Process 72
- 3.6 Tunneling Process 73
- 3.7 High-Field Effects 75
- Summary 80

PART II

SEMICONDUCTOR DEVICES

► CHAPTER 4

p - n Junction 84

- 4.1 Basic Fabrication Steps 85
- 4.2 Thermal Equilibrium Condition 88
- 4.3 Depletion Region 93
- 4.4 Depletion Capacitance 100
- 4.5 Current-Voltage Characteristics 104

- 4.6 Charge Storage and Transient Behavior 114
- 4.7 Junction Breakdown 117
- 4.8 Heterojunction 124
- Summary 127

► CHAPTER 5

Bipolar Transistor and Related Devices 130

- 5.1 The Transistor Action 131
- 5.2 Static Characteristics of Bipolar Transistor 137
- 5.3 Frequency Response and Switching of Bipolar Transistor 146
- 5.4 The Heterojunction Bipolar Transistor 151
- 5.5 The Thyristor and Related Power Devices 156
- Summary 165

► CHAPTER 6

MOSFET and Related Devices 169

- 6.1 The MOS Diode 170
- 6.2 MOSFET Fundamentals 186
- 6.3 MOSFET Scaling 199
- 6.4 CMOS and BiCMOS 205
- 6.5 MOSFET on Insulator 209
- 6.6 MOS Memory Structures 214
- 6.7 The Power MOSFET 218
- Summary 220

► CHAPTER 7

MESFET and Related Devices 224

- 7.1 Metal-Semiconductor Contacts 225
- 7.2 MESFET 237
- 7.3 MODFET 246
- Summary 251

► CHAPTER 8

Microwave Diodes, Quantum-Effect, and Hot-Electron Devices 254

- 8.1 Basic Microwave Technology 255
- 8.2 Tunnel Diode 259
- 8.3 IMPATT Diode 261
- 8.4 Transferred-Electron Devices 264
- 8.5 Quantum-Effect Devices 270
- 8.6 Hot-Electron Devices 275
- Summary 278

► **CHAPTER 9**

Photonic Devices 282

- 9.1 Radiative Transitions and Optical Absorption 282
- 9.2 Light-Emitting Diodes 288
- 9.3 Semiconductor Laser 298
- 9.4 Photodetector 311
- 9.5 Solar Cell 318
- Summary 328

PART III

SEMICONDUCTOR TECHNOLOGY

► **CHAPTER 10**

Crystal Growth and Epitaxy 332

- 10.1 Silicon Crystal Growth from the Melt 333
- 10.2 Silicon Float-Zone Process 339
- 10.3 GaAs Crystal-Growth Techniques 343
- 10.4 Material Characterization 347
- 10.5 Epitaxial-Growth Techniques 354
- 10.6 Structures and Defects in Epitaxial Layers 361
- Summary 365

► **CHAPTER 11**

Film Formation 369

- 11.1 Thermal Oxidation 370
- 11.2 Dielectric Deposition 378
- 11.3 Polysilicon Deposition 388
- 11.4 Metallization 390
- Summary 400

► **CHAPTER 12**

Lithography and Etching 404

- 12.1 Optical Lithography 404
- 12.2 Next-Generation Lithographic Methods 418
- 12.3 Wet Chemical Etching 426
- 12.4 Dry Etching 431
- 12.5 Microelectromechanical Systems 443
- Summary 447

► **CHAPTER 13**

Impurity Doping 452

- 13.1 Basic Diffusion Process 453
- 13.2 Extrinsic Diffusion 462
- 13.3 Diffusion-Related Processes 466
- 13.4 Range of Implanted Ions 469
- 13.5 Implant Damage and Annealing 477
- 13.6 Implantation-Related Processes 481
- Summary 485

► **CHAPTER 14**

Integrated Devices 489

- 14.1 Passive Components 491
- 14.2 Bipolar Technology 496
- 14.3 MOSFET Technology 503
- 14.4 MESFET Technology 519
- 14.5 Challenges for Microelectronics 522
- Summary 527

► **APPENDIX A**

List of Symbols 531

► **APPENDIX B**

International Systems of Units (SI Units) 533

► **APPENDIX C**

Unit Prefixes 534

► **APPENDIX D**

Greek Alphabet 535

► **APPENDIX E**

Physical Constants 536

► **APPENDIX F**

Properties of Important Element and Binary Compound Semiconductors at 300 K 537

► **APPENDIX G**

Properties of Si and GaAs at 300 K 538

► **APPENDIX H**

Derivation of the Density of States in a Semiconductor 539

► **APPENDIX I**

Derivation of Recombination Rate for Indirect Recombination 541

► **APPENDIX J**

Calculation of the Transmission Coefficient for a Symmetric Resonant-Tunneling Diode 543

► **APPENDIX K**

Basic Kinetic Theory of Gases 545

► **APPENDIX L**

Answers to Selected Problems 547

Index 551

Energy Bands and Carrier Concentration in Thermal Equilibrium

- ▶ 2.1 SEMICONDUCTOR MATERIALS
- ▶ 2.2 BASIC CRYSTAL STRUCTURE
- ▶ 2.3 BASIC CRYSTAL GROWTH TECHNIQUE
- ▶ 2.4 VALENCE BONDS
- ▶ 2.5 ENERGY BANDS
- ▶ 2.6 INTRINSIC CARRIER CONCENTRATION
- ▶ 2.7 DONORS AND ACCEPTORS
- ▶ SUMMARY

In this chapter, we consider some basic properties of semiconductors. We begin with a discussion of crystal structure, which is the arrangement of atoms in a semiconductor. This is followed by a brief description of the crystal growth technique. We then present the concepts of valence bonds and energy bands, which relate to conduction in semiconductors. Finally, we discuss the concept of carrier concentration in thermal equilibrium. These concepts are used throughout this book.

Specifically, we cover the following topics:

- Element and compound semiconductors and their basic properties.
- The diamond structure and its related crystal planes.
- The bandgap and its impact on electrical conductivity.
- The intrinsic carrier concentration and its dependence on temperature.
- The Fermi level and its dependence on carrier concentration.

▶ 2.1 SEMICONDUCTOR MATERIALS

Solid-state materials can be grouped into three classes—insulators, semiconductors, and conductors. Figure 1 shows the range of electrical conductivities σ (and the corresponding resistivities $\rho = 1/\sigma$)^{*} associated with some important materials in each of the three classes. Insulators such as fused quartz and glass have very low conductivities, on the order of

^{*} A list of symbols is given in Appendix A.

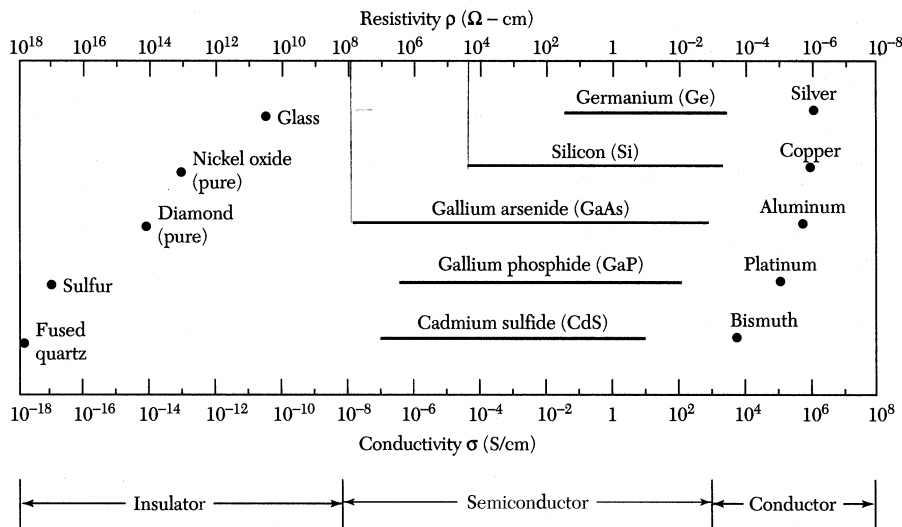


Fig. 1 Typical range of conductivities for insulators, semiconductors, and conductors.

$10^{-18} - 10^{-8}$ S/cm; and conductors such as aluminum and silver have high conductivities, typically from 10^4 to 10^6 S/cm.* Semiconductors have conductivities between those of insulators and those of conductors. The conductivity of a semiconductor is generally sensitive to temperature, illumination, magnetic field, and minute amounts of impurity atoms (typically, about 1 μ g to 1 g of impurity atoms in 1 kg of semiconductor materials). This sensitivity in conductivity makes the semiconductor one of the most important materials for electronic applications.

2.1.1 Element Semiconductors

The study of semiconductor materials began in the early nineteenth century.¹ Over the years many semiconductors have been investigated. Table 1 shows a portion of the periodic table related to semiconductors. The element semiconductors, those composed of single

TABLE 1 Portion of the Periodic Table Related to Semiconductors

Period	Column II	III	IV	V	VI
2		B	C	N	O
		Boron	Carbon	Nitrogen	Oxygen
3	Mg	Al	Si	P	S
	Magnesium	Aluminum	Silicon	Phosphorus	Sulfur
4	Zn	Ga	Ge	As	Se
	Zinc	Gallium	Germanium	Arsenic	Selenium
5	Cd	In	Sn	Sb	Te
	Cadmium	Indium	Tin	Antimony	Tellurium
6	Hg		Pb		
	Mercury		Lead		

* The international system of units is presented in Appendix B.

species of atoms, such as silicon (Si) and germanium (Ge), can be found in Column IV. In the early 1950s, germanium was the major semiconductor material. Since the early 1960s silicon has become a practical substitute and has now virtually supplanted germanium as a material for semiconductor fabrication. The main reasons we now use silicon are that silicon devices exhibit better properties at room temperature, and high-quality silicon dioxide can be grown thermally. There is also an economic consideration. Device-grade silicon costs much less than any other semiconductor material. Silicon in the form of silica and silicates comprises 25% of the Earth's crust, and silicon is second only to oxygen in abundance. Currently, silicon is one of the most studied elements in the periodic table; and silicon technology is by far the most advanced among all semiconductor technologies.

2.1.2 Compound Semiconductors

In recent years a number of compound semiconductors have found applications for various devices. The important compound semiconductors as well as the two element semiconductors are listed² in Table 2. A binary compound semiconductor is a combination of two elements from the periodic table. For example, gallium arsenide (GaAs) is a III-V compound that is a combination of gallium (Ga) from Column III and arsenic (As) from Column V.

In addition to binary compounds, ternary compounds and quaternary compounds are made for special applications. The alloy semiconductor $\text{Al}_x\text{Ga}_{1-x}\text{As}$, which has Al and Ga from Column III and As from Column V is an example of a ternary compound, whereas quaternary compounds of the form $\text{A}_x\text{B}_{1-x}\text{C}_y\text{D}_{1-y}$ can be obtained from combination of many binary and ternary compound semiconductors. For example, GaP, InP, InAs, plus GaAs can be combined to yield the alloy semiconductor $\text{Ga}_x\text{In}_{1-x}\text{As}_y\text{P}_{1-y}$. Compared with the element semiconductors, the preparation of compound semiconductors in single-crystal form usually involves much more complex processes.

Many of the compound semiconductors have electrical and optical properties that are different from silicon. These semiconductors, especially GaAs, are used mainly for high-speed electronic and photonic applications. Although we do not know as much about the technology of compound semiconductors as we do about that of silicon, advances in silicon technology have also helped progress in compound semiconductor technology. In this book we are concerned mainly with device physics and processing technology of silicon and gallium arsenide.

2.2 BASIC CRYSTAL STRUCTURE

The semiconductor materials we will be studying are single crystals, that is, the atoms are arranged in a three-dimensional periodic fashion. The periodic arrangement of atoms in a crystal is called a *lattice*. In a crystal, an atom never strays far from a single, fixed position. The thermal vibrations associated with the atom are centered about this position. For a given semiconductor, there is a *unit cell* that is representative of the entire lattice; by repeating the unit cell throughout the crystal, one can generate the entire lattice.

2.2.1 Unit Cell

A generalized primitive, three-dimensional unit cell is shown in Fig. 2. The relationship between this cell and the lattice is characterized by three vectors **a**, **b**, and **c**, which need

TABLE 2 Semiconductor Materials²

General Classification	Semiconductor	
	Symbol	Name
Element	Si	Silicon
	Ge	Germanium
Binary compound		
IV-IV -----	SiC	Silicon carbide
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
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
IV-VI -----	PbS	Lead sulfide
	PbSe	Lead selenide
	PbTe	Lead telluride
Ternary compound	$\text{Al}_x\text{Ga}_{1-x}\text{As}$	Aluminum gallium arsenide
	$\text{Al}_x\text{In}_{1-x}\text{As}$	Aluminum indium arsenide
	$\text{GaAs}_{1-x}\text{P}_x$	Gallium arsenic phosphide
	$\text{Ga}_x\text{In}_{1-x}\text{As}$	Gallium indium arsenide
	$\text{Ga}_x\text{In}_{1-x}\text{P}$	Gallium indium phosphide
Quaternary compound	$\text{Al}_x\text{Ga}_{1-x}\text{As}_y\text{Sb}_{1-y}$	Aluminum gallium arsenic antimonide
	$\text{Ga}_x\text{In}_{1-x}\text{As}_{1-y}\text{P}_y$	Gallium indium arsenic phosphide

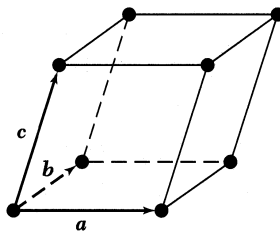


Fig. 2 A generalized primitive unit cell.

not be perpendicular to each other and which may or may not be equal in length. Every equivalent lattice point in the three-dimensional crystal can be found using the set

$$\mathbf{R} = m\mathbf{a} + n\mathbf{b} + p\mathbf{c} \quad (1)$$

where m , n , and p are integers.

Figure 3 shows some basic cubic-crystal unit cells. Figure 3a shows a simple cubic (sc) crystal; it has an atom at each corner of the cubic lattice, and each atom has six equidistant nearest-neighbor atoms. The dimension a is called the lattice constant. In the periodic table, only polonium is crystallized in the simple cubic lattice. Figure 3b is a body-centered cubic (bcc) crystal where, in addition to the eight corner atoms, an atom is located at the center of the cube. In a bcc lattice, each atom has eight nearest-neighbor atoms. Crystals exhibiting bcc lattices include those of sodium and tungsten. Figure 3c shows the face-centered cubic (fcc) crystal that has one atom at each of the six cubic faces in addition to the eight corner atoms. In this case, each atom has 12 nearest-neighbor atoms. A large number of elements exhibit the fcc lattice form, including aluminum, copper, gold, and platinum.

▶ EXAMPLE 1

If we pack hard spheres in a bcc lattice such that the atom in the center just touches the atoms at the corners of the cube, find the fraction of the bcc unit cell volume filled with hard spheres.

SOLUTION Each corner sphere in a bcc unit cell is shared with eight neighboring cells; thus, each unit cell contains one-eighth of a sphere at each of the eight corners for a total of one sphere. In addition, each unit cell contains one central sphere. We have the following:

Spheres (atoms) per unit cell = $(1/8) \times 8$ (corner) + 1 (center) = 2;

Nearest-neighbor distance (along the diagonal AE in Fig. 3b) = $a\sqrt{3}/2$;

Radius of each sphere = $a\sqrt{3}/4$;

Volume of each sphere = $4\pi/3 \times (a\sqrt{3}/4)^3 = \pi a^3\sqrt{3}/16$; and

Maximum fraction of unit cell filled = $\text{Number of spheres} \times \text{volume of each sphere} / \text{total volume of each unit cell} = 2\pi a^3\sqrt{3}/16 a^3 = \pi\sqrt{3}/8 \approx 0.68$.

Therefore, about 68% of the bcc unit cell volume is filled with hard spheres, and about 32% of the volume is empty. ◀

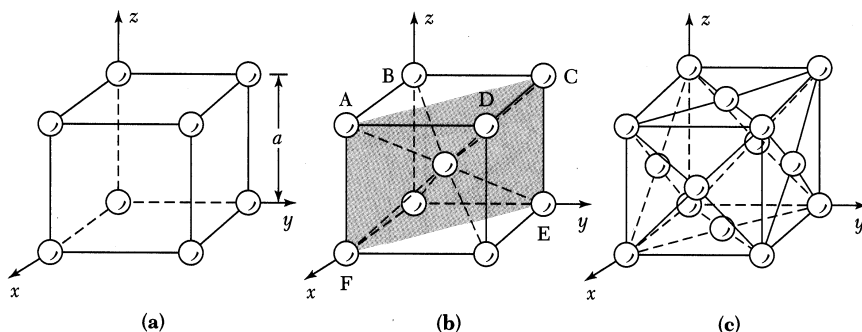


Fig. 3 Three cubic-crystal unit cells. (a) Simple cubic. (b) Body-centered cubic. (c) Face-centered cubic.

2.2.2 The Diamond Structure

The element semiconductors, silicon and germanium, have a diamond lattice structure as shown in Fig. 4a. This structure also belongs to the fcc crystal family and can be seen as two interpenetrating fcc sublattices with one sublattice displaced from the other by one-quarter of the distance along the body diagonal of the cube (i.e., a displacement of $a\sqrt{3}/4$). Although chemically identical, the two sets of atoms belonging to the two sublattices are different from the point view of the crystal structure. It can be seen in Fig. 4a that if a corner atom has one nearest neighbor in the body diagonal direction, then it has no nearest neighbor in the reverse direction. Consequently, it requires two such atoms in the unit cell. Alternatively, a unit cell of a diamond lattice consists of a tetrahedron in which each atom is surrounded by four equidistant nearest neighbors that lie at the corners (refer to the spheres connected by darkened bars in Fig. 4a).

Most of the III-V compound semiconductors (e.g., GaAs) have a *zincblende lattice*, shown in Fig. 4b, which is identical to a diamond lattice except that one fcc sublattice has Column III atoms (Ga) and the other has Column V atoms (As). Appendix F at the end of the book gives a summary of the lattice constants and other properties of important element and binary compound semiconductors.

► EXAMPLE 2

At 300 K the lattice constant for silicon is $5.43 \text{ \AA}$. Calculate the number of silicon atoms per cubic centimeter and the density of silicon at room temperature.

SOLUTION There are eight atoms per unit cell. Therefore,

$$8/a^3 = 8/(5.43 \times 10^{-8})^3 = 5 \times 10^{22} \text{ atoms/cm}^3; \text{ and}$$

$$\text{Density} = \text{no. of atoms/cm}^3 \times \text{atomic weight/Avogadro constant} = 5 \times 10^{22} \text{ (atoms/cm}^3) \times 28.09 \text{ (g/mol)} / 6.02 \times 10^{23} \text{ (atoms/mol)} = 2.33 \text{ g/cm}^3. \quad \blacktriangleleft$$

2.2.3 Crystal Planes and Miller Indices

In Fig. 3b we note that there are four atoms in the *ABCD* plane and five atoms in the *ACEF* plane (four atoms from the corners and one from the center) and that the atomic spacing is different for the two planes. Therefore, the crystal properties along different planes are different, and the electrical and other device characteristics can be dependent

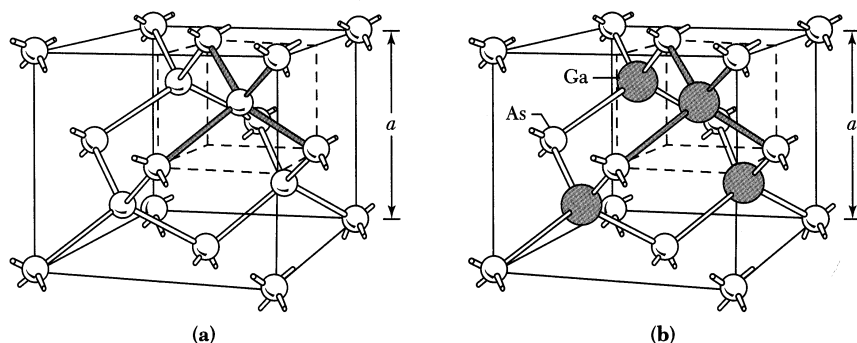


Fig. 4 (a) Diamond lattice. (b) Zincblende lattice.

on the crystal orientation. A convenient method of defining the various planes in a crystal is to use *Miller indices*.³ These indices are obtained using the following steps:

1. Find the intercepts of the plane on the three Cartesian coordinates in terms of the lattice constant.
2. Take the reciprocals of these numbers and reduce them to the smallest three integers having the same ratio.
3. Enclose the result in parentheses (hkl) as the Miller indices for a single plane.

EXAMPLE 3

As shown in Fig. 5, the plane has intercepts at a , $3a$, and $2a$ along the three coordinates. Taking the reciprocals of these intercepts, we get $1/a$, $1/3a$, and $1/2a$. The smallest three integers having the same ratio are 6, 2, and 3 (obtained by multiplying each fraction by $6a$). Thus, the plane is referred to as a (623)-plane. ◀

Figure 6 shows the Miller indices of important planes in a cubic crystal.^o Some other conventions are the following:

1. $(\bar{h}kl)$: For a plane that intercepts the x -axis on the negative side of the origin, such as $(\bar{1}00)$.
2. $\{hkl\}$: For planes of equivalent symmetry, such as $\{100\}$ for (100) , (010) , (001) , $(\bar{1}00)$, $(0\bar{1}0)$, and $(00\bar{1})$ in cubic symmetry.
3. $[hkl]$: For a crystal direction, such as $[100]$ for the x -axis. By definition, the $[100]$ -direction is perpendicular to (100) -plane, and the $[111]$ -direction is perpendicular to the (111) -plane.
4. $\langle hkl \rangle$: For a full set of equivalent directions, such as $\langle 100 \rangle$ for $[100]$, $[010]$, $[001]$, $[\bar{1}00]$, $[0\bar{1}0]$, and $[00\bar{1}]$.

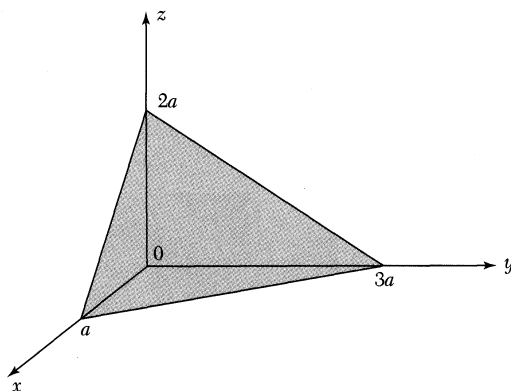


Fig. 5 A (623)-crystal plane.

^o In Chapter 6, we show that the $\langle 100 \rangle$ orientation is preferred for silicon metal-oxide-semiconductor field-effect transistors (MOSFETs).

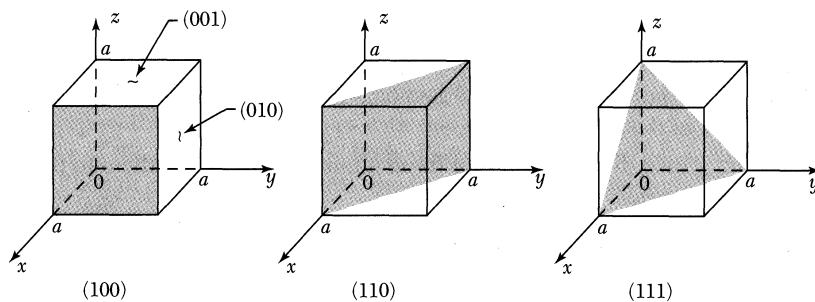
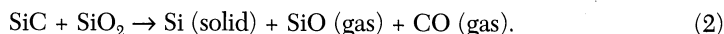


Fig. 6 Miller indices of some important planes in a cubic crystal.

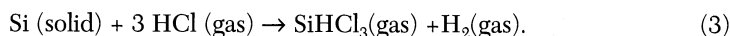
► 2.3 BASIC CRYSTAL GROWTH TECHNIQUE

In this section, we consider briefly how we grow semiconductor crystals, specifically the silicon crystal because 95% of the semiconductor materials used by the electronic industry is silicon.

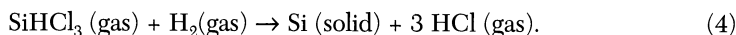
The starting material for silicon is a relatively pure form of sand (SiO_2) called quartzite. It is mixed with various forms of carbon and allowed to react to form silicon (98% pure):



The silicon product is reacted with hydrogen chloride to form trichlorosilane,



The trichlorosilane is decomposed using an electric current in a chamber with a controlled ambient, producing rods of ultrapure polycrystalline silicon (i.e., silicon material that contains many single-crystal regions with different size and orientation with respect to one another),



The polycrystalline silicon is now ready for the crystal-growing process. Figure 7 shows many pieces of polycrystalline silicon in a silica (SiO_2) crucible.

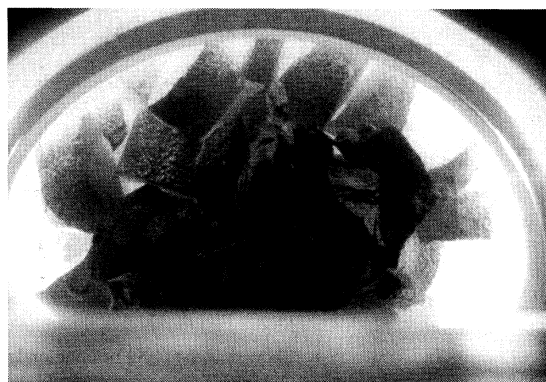


Fig. 7 Photograph of polycrystalline silicon in a silica crucible.

The most common crystal growth method is the Czochralski technique. Figure 8 shows a schematic drawing of Czochralski crystal puller. The crucible containing the polycrystalline silicon is heated either by radio-frequency induction or by a thermal resistance method to the melting point of silicon (1412°C). The crucible rotates during the growth to prevent the formation of local hot or cold regions.

The atmosphere around the crystal-growing apparatus or crystal puller is controlled to prevent contamination of the molten silicon. Argon is often used as the ambient gas. When the temperature of the silicon has stabilized, a piece of silicon with a suitable orientation (e.g., $\langle 111 \rangle$), which is called the seed crystal, is lowered to the melt and is the starting point for the subsequent growth of a much larger crystal. As the bottom of the seed crystal begins to melt in the molten silicon, the downward motion of the rod holding the seed is reversed. As the seed crystal is slowly withdrawn from the melt (Fig. 9), the molten silicon adhering to the crystal freezes or solidifies, using the crystal structure of the seed crystal as a template. The seed crystal is, therefore, used to initiate the growth of the ingot with the correct crystal orientation. The rod continues its upward movement, forming an even larger crystal. The crystal growth is terminated when the silicon in the crucible is depleted. By carefully controlling the temperature of the crucible and the rotation speeds of the crucible and the rod, precise control of the diameter of the crystal can be maintained. Figure 10 shows a 200 mm diameter silicon crystal ingot. The desired impurity concentration is obtained by adding impurities to the melt in the form of heavily doped silicon prior to crystal growth. A more detailed discussion on the crystal growth of silicon as well as other semiconductors can be founded in Chapter 10.

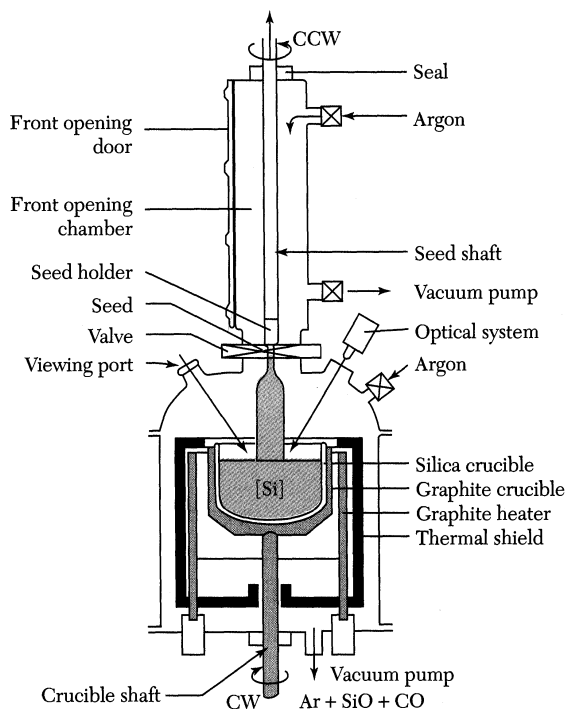


Fig. 8 Simplified schematic drawing of the Czochralski puller. Clockwise (CW), counterclockwise (CCW).

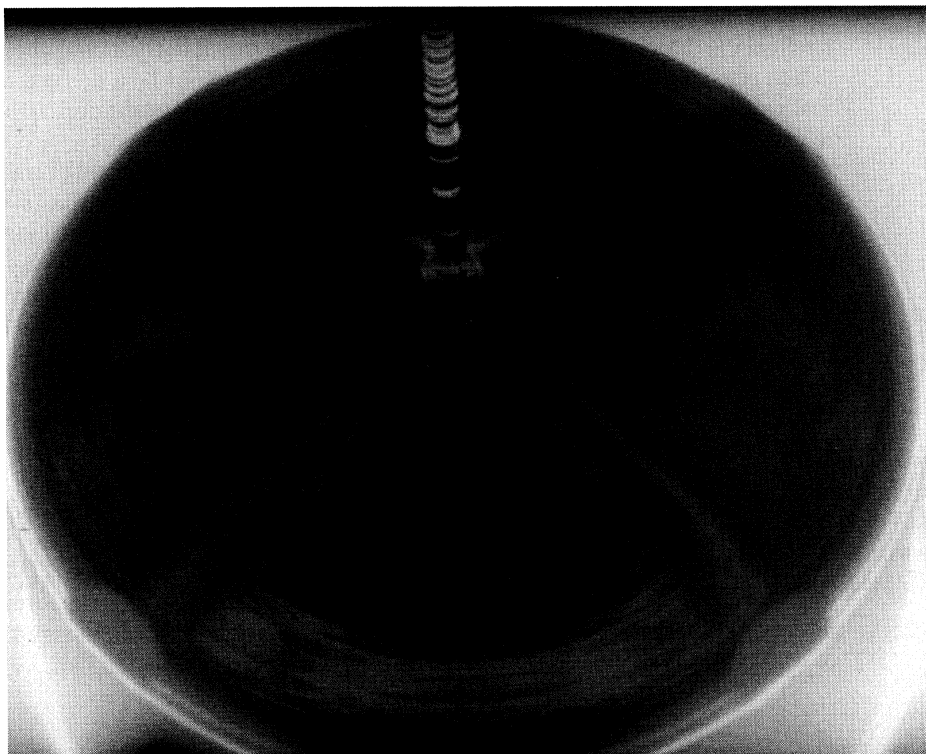


Fig. 9 Photograph of a 200 mm diameter, (100)-oriented Si crystal being pulled from the melt. (Photograph courtesy of Taisil Electronic Materials Corp., Taiwan.)

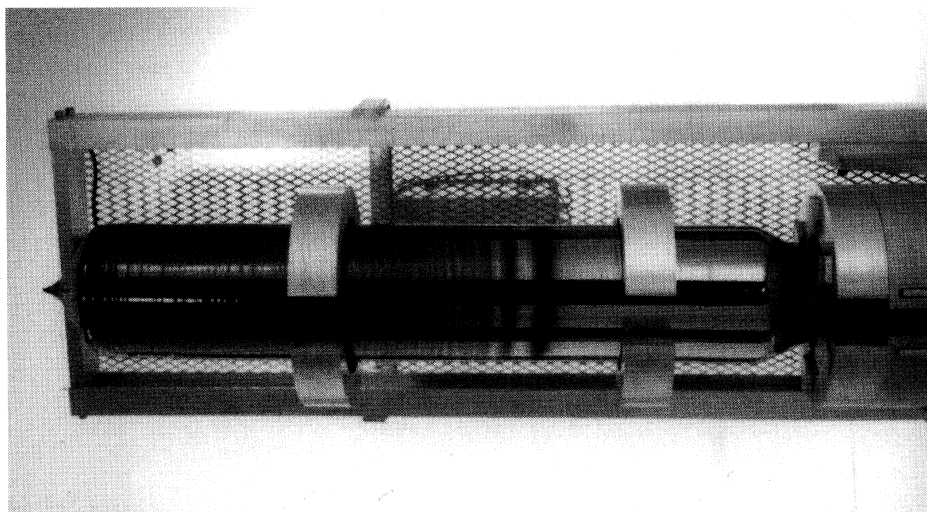


Fig. 10 A 200 mm diameter silicon crystal ingot grown by the Czochralski technique.

2.4 VALENCE BONDS

As discussed in Section 2.2, each atom in a diamond lattice is surrounded by four nearest neighbors. Figure 11a shows the tetrahedron bonds of a diamond lattice. A simplified two-dimensional bonding diagram for the tetrahedron is shown in Fig. 11b. Each atom has four electrons in the outer orbit, and each atom shares these valence electrons with its four neighbors. This sharing of electrons is known as *covalent bonding*; each *electron pair* constitutes a covalent bond. Covalent bonding occurs between atoms of the same element or between atoms of different elements that have similar outer-shell electron configurations. Each electron spends an equal amount of time with each nucleus. However, both electrons spend most of their time between the two nuclei. The force of attraction for the electrons by both nuclei holds the two atoms together.

Gallium arsenide crystallizes in a zincblende lattice, which also has tetrahedron bonds. The major bonding force in GaAs is also due to the covalent bond. However, gallium arsenide has a small ionic contribution that is an electrostatic attractive force between each Ga^+ ion and its four neighboring As^- ions, or between each As^- ion and its four neighboring Ga^+ ions. Electronically, this means that the paired bonding electrons spend slightly more time in the As atom than in the Ga atom.

At low temperatures, the electrons are bound in their respective tetrahedron lattice; consequently, they are not available for conduction. At higher temperatures, thermal vibrations may break the covalent bonds. When a bond is broken or partially broken, a free electron results that can participate in current conduction. Figure 12a shows the situation when a valence electron in silicon becomes a free electron. An electron deficiency is left in the covalent bond. This deficiency may be filled by one of the neighboring electrons, which results in a shift of the deficiency location, as from location A to location B in Fig. 12b. We may therefore consider this deficiency as a particle similar to an electron. This fictitious particle is called a *hole*. It carries a positive charge and moves, under the influence of an applied electric field, in the direction opposite to that of an electron. Therefore, both the electron and the hole contribute to the total electric current. The concept of a hole is analogous to that of a bubble in a liquid. Although it is actually the liquid that moves, it is much easier to talk about the motion of the bubble in the opposite direction.

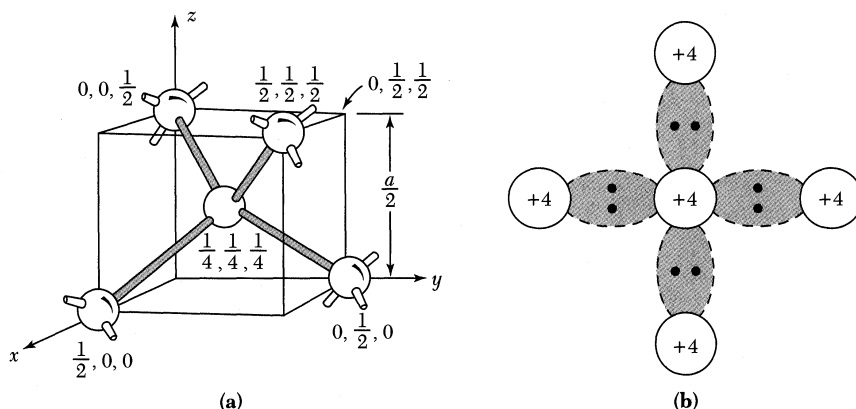


Fig. 11 (a) A tetrahedron bond. (b) Schematic two-dimensional representation of a tetrahedron bond.

► 2.5 ENERGY BANDS

2.5.1 Energy Levels of Isolated Atoms

For an isolated atom, the electrons can have discrete energy levels. For example, the energy levels for an isolated hydrogen atom are given by the Bohr model⁴:

$$E_H = -m_0 q^4 / 8 \epsilon_0^2 h^2 n^2 = -13.6 / n^2 \text{ eV}, \quad (5)$$

where m_0 is the free-electron mass, q is the electronic charge, ϵ_0 is the free-space permittivity, h is the Planck constant, and n is a positive integer called the principal quantum number. The quantity eV (electron volt) is an energy unit corresponding to the energy gained by an electron when its potential is increased by one volt. It is equal to the product of q (1.6×10^{-19} coulomb) and one volt, or 1.6×10^{-19} J. The discrete energies are -13.6 eV for the ground state energy level ($n = 1$), -3.4 eV for the first excited-state energy level ($n = 2$), etc. Detailed studies reveal that for higher principle quantum numbers ($n \geq 2$), energy levels are split according to their angular momentum quantum number ($\ell = 0, 1, 2, \dots, n - 1$).

We now consider two identical atoms. When they are far apart, the allowed energy levels for a given principal quantum number (e.g., $n = 1$) consist of one doubly degenerate level, that is, both atoms have exactly the same energy. When they are brought closer, the doubly degenerate energy levels will split into two levels by the interaction between the atoms. As N isolated atoms are brought together to form a solid, the orbits of the outer electrons of different atoms overlap and interact with each other. This interaction, including those forces of attraction and repulsion between atoms, causes a shift in the energy levels, as in the case of two interacting atoms. However, instead of two levels, N separate but closely spaced levels are formed. When N is large, the result is an essentially continuous band of energy. This band of N levels can extend over a few eV depending on the inter-atomic spacing for the crystal. Figure 13 shows the effect, where the parameter a represents the equilibrium interatomic distance of the crystal.

The actual band splitting in a semiconductor is much more complicated. Figure 14 shows an isolated silicon atom that has 14 electrons. Of the 14 electrons, 10 occupy deep-lying energy levels whose orbital radius is much smaller than the interatomic separation in the crystal. The four remaining valence electrons are relatively weakly bound and can be involved in chemical reactions. Therefore, we only need to consider the outer shell

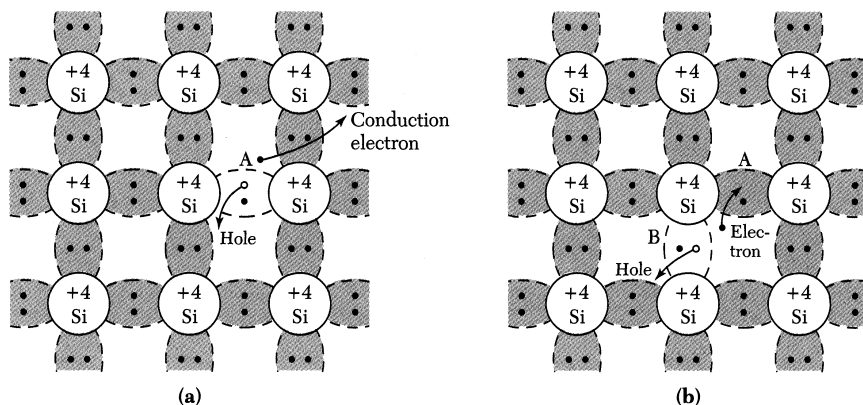


Fig. 12 The basic bond representation of intrinsic silicon. (a) A broken bond at position A, resulting in a conduction electron and a hole. (b) A broken bond at position B.

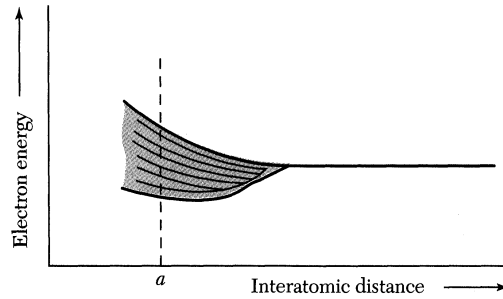


Fig. 13 The splitting of a degenerate state into a band of allowed energies.

(the $n = 3$ level) for the valence electrons, since the two inner shells are completely full and tightly bound to the nucleus. The 3s subshell (i.e., for $n = 3$ and $\ell = 0$) has two allowed quantum states per atom. This subshell will contain two valence electrons at $T = 0$ K. The 3p subshell (i.e., $n = 3$, and $\ell = 1$) has six allowed quantum states per atom. This subshell will contain the remaining two valence electrons of an individual silicon atom.

Figure 15 is a schematic diagram of the formation of a silicon crystal from N isolated silicon atoms. As the interatomic distance decreases, the 3s and 3p subshell of the N silicon atoms will interact and overlap. At the equilibrium interatomic distance, the bands will again split, with four quantum states per atom in the lower band and four quantum states per atom in the upper band. At a temperature of absolute zero, electrons occupy the lowest energy states, so that all states in the lower band (*the valence band*) will be full and all states in the upper band (*the conduction band*) will be empty. The bottom of the conduction band is called E_C , and the top of the valence band is called E_V . The *bandgap energy* E_g between the bottom of the conduction band and the top of the valence band ($E_C - E_V$) is the width of the forbidden energy gap, as shown on the far left of Fig. 15. Physically, E_g is the energy required to break a bond in the semiconductor to free an electron to the conduction band and leave a hole in the valence band.

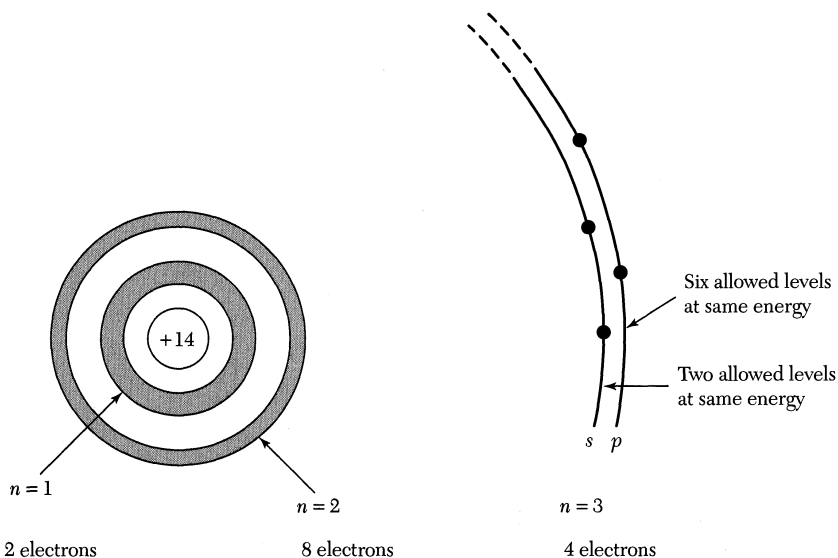


Fig. 14 Schematic representation of an isolated silicon atom.

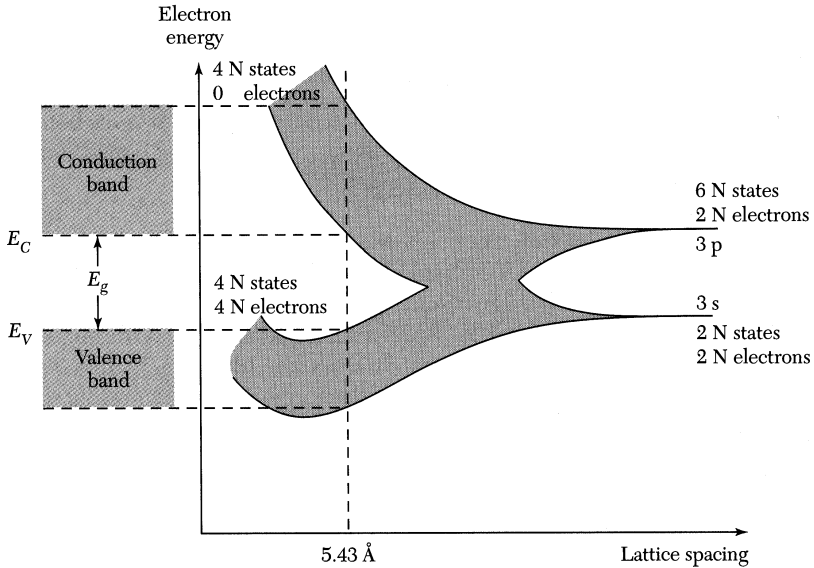


Fig. 15 Formation of energy bands as a diamond lattice crystal is formed by bringing isolated silicon atoms together.

2.5.2 The Energy-Momentum Diagram

The energy E of a free electron is given by

$$E = \frac{p^2}{2m_0}, \quad (6)$$

where p is the momentum and m_0 is the free-electron mass. If we plot E vs. p , we obtain a parabola as shown in Fig. 16. In a semiconductor crystal, an electron in the conduction band is similar to a free electron in that it is relatively free to move about in the crystal. However, because of the periodic potential of the nuclei, Eq. 6 can no longer be valid. However, it turns out that we can still use Eq. 6 if we replace the free-electron mass in Eq. 6 by an effective mass m_n (the subscript n refers to the negative charge on an electron), that is,

$$E = \frac{p^2}{2m_n}. \quad (7)$$

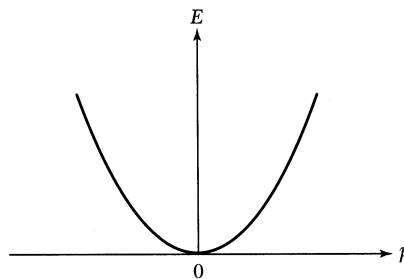


Fig. 16 The parabolic energy (E) vs. momentum (p) curve for a free electron.

The electron effective mass depends on the properties of the semiconductor. If we have an energy-momentum relationship described by Eq. 7, we can obtain the effective mass from the second derivative of E with respect to p :

$$m_n \equiv \left(\frac{d^2 E}{dp^2} \right)^{-1}. \quad (8)$$

Therefore, the narrower the parabola, corresponding to a larger second derivative, the smaller the effective mass. A similar expression can be written for holes (with effective mass m_p where the subscript p refers to the positive charge on a hole). The effective-mass concept is very useful because it enables us to treat electrons and holes essentially as classical charged particles.

Figure 17 shows a simplified energy-momentum relationship of a special semiconductor with an electron effective mass of $m_n = 0.25 m_0$ in the conduction band (the upper parabola) and a hole effective mass of $m_p = m_0$ in the valence band (the lower parabola). Note that the electron energy is measured upward and the hole energy is measured downward. The spacing at $p = 0$ between these two parabolas is the bandgap E_g , shown previously in Fig. 15.

The actual energy-momentum relationships (also called energy-band diagram) for silicon and gallium arsenide are much more complex. They are shown in Fig. 18 for two crystal directions. We note that the general features in Fig. 18 are similar to those in Fig. 17. First of all, there is a bandgap E_g between the bottom of the conduction band and the top of the valence band. Second, near the minimum of the conduction band or the maximum of the valence band, the E - p curves are essentially parabolic. For silicon, Fig 18a, the maximum in the valence band occurs at $p = 0$, but the minimum in the conduction band occurs along the $[100]$ direction at $p = p_c$. Therefore, in silicon, when an electron makes a transition from the maximum point in the valence band to the minimum point in the conduction, it requires not only an energy change ($\geq E_g$) but also some momentum change ($\geq p_c$).

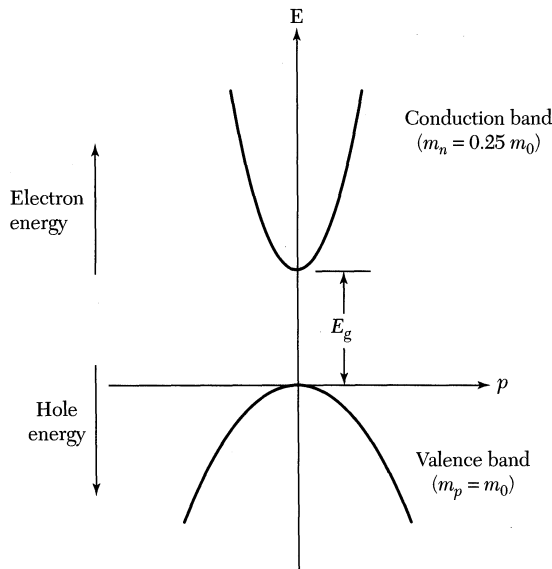


Fig. 17 A schematic energy-momentum diagram for a special semiconductor with $m_n = 0.25 m_0$ and $m_p = m_0$.

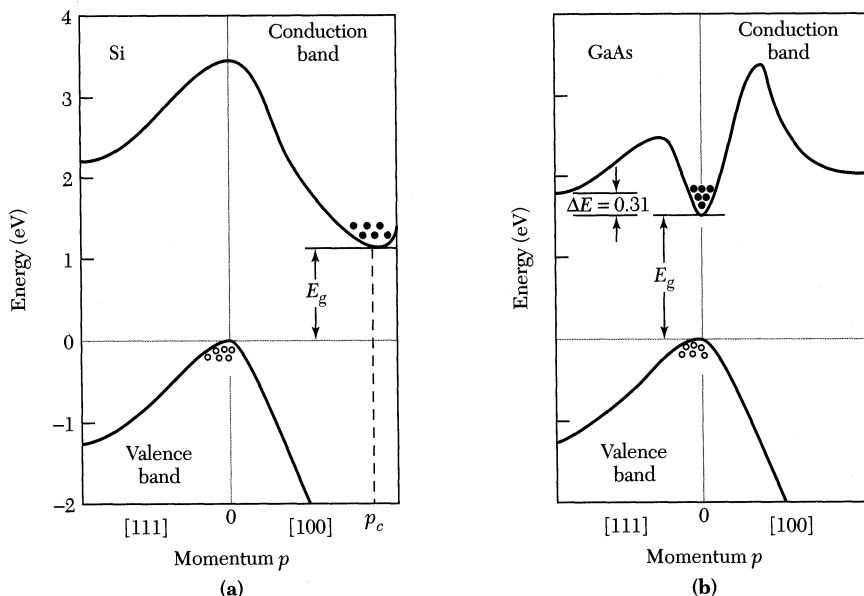


Fig. 18 Energy band structures of Si and GaAs. Circles (o) indicate holes in the valence bands and dots (•) indicate electrons in the conduction bands.

For gallium arsenide, Fig. 18b, the maximum in the valence band and the minimum in the conduction band occur at the same momentum ($p = 0$). Thus, an electron making a transition from the valence band to the conduction band can do so without a change in momentum.

Gallium arsenide is called a *direct semiconductor*, because it does not require a change in momentum for an electron transition from the valence band to the conduction band. Silicon is called an *indirect semiconductor*, because a change of momentum is required in a transition. This difference between direct and indirect band structures is very important for light-emitting diodes and semiconductor lasers. These devices require direct semiconductors for efficient generation of photons (refer to Chapter 9).

We can obtain the effective mass from Fig. 18 using Eq. 8. For example, for gallium arsenide with a very narrow conduction-band parabola, the electron effective mass is $0.063 m_0$, where for silicon, with a wider conduction-band parabola, the electron effective mass is $0.19 m_0$.

2.5.3 Conduction in Metals, Semiconductors, and Insulators

The enormous variation in electrical conductivity of metals, semiconductors, and insulators as shown in Fig. 1 may be explained qualitatively in terms of their energy bands. We shall see that the electron occupation of the highest band or of the highest two bands determines the conductivity of a solid. Figure 19 shows the energy band diagrams of three classes of solids—metals, semiconductors, and insulators.

Metals

The characteristics of a metal (also called a conductor) include a very low value of resistivity, and the conduction band either is partially filled (such as Cu) or overlaps the valence band (such as Zn or Pb) so that there is no bandgap, as shown in Fig. 19a. As a conse-

quence, the uppermost electrons in the partially filled band or electrons at the top of the valence band can move to the next-higher available energy level when they gain kinetic energy (e.g., from an applied electric field). Electrons are free to move with only a small applied field in a metal because there are many unoccupied states close to the occupied energy states. Therefore, current conduction can readily occur in conductors.

Insulators

In an insulator such as silicon dioxide (SiO_2), the valence electrons form strong bonds between neighboring atoms. These bonds are difficult to break, and consequently there are no free electrons to participate in current conduction at or near room temperature. As shown in the energy band diagram Fig. 19c, insulators are characterized by a large bandgap. Note that electrons occupy all energy levels in the valence band and all energy levels in the conduction band are empty. Thermal energy^o or the energy of an applied electric field is insufficient to raise the uppermost electron in the valence band to the conduction band. Thus, although an insulator has many vacant states in the conduction band that can accept electrons, so few electrons actually occupy conduction band states that the overall contribution to electrical conductivity is very small, resulting in a very high resistivity. Therefore, silicon dioxide is an insulator; it can not conduct current.

Semiconductors

Now, consider a material that has a much smaller energy gap, on the order of 1 eV (Fig. 19b). Such materials are called semiconductors. At $T = 0$ K, all electrons are in the valence band, and there are no electrons in the conduction band. Thus, semiconductors are poor conductors at low temperatures. At room temperature and under normal atmosphere, values of E_g are 1.12 eV for Si and 1.42 eV for GaAs. The thermal energy kT at room temperature is a good fraction of E_g , and an appreciable numbers of electrons are thermally excited from the valence band to the conduction band. Since there are many empty states in the conduction band, a small applied potential can easily move these electrons, resulting in a moderate current.

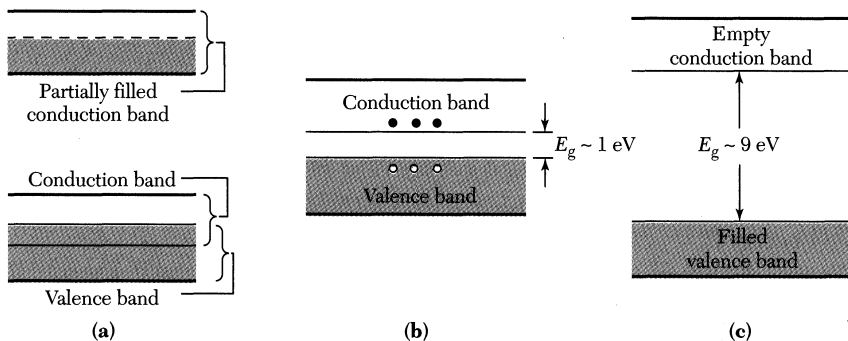


Fig. 19 Schematic energy band representations of (a) a conductor with two possibilities (either the partially filled conduction band shown at the upper portion or the overlapping bands shown at the lower portion), (b) a semiconductor, and (c) an insulator.

^o The thermal energy is of the order of kT . At room temperature, kT is 0.026 eV, which is much smaller than the bandgap of an insulator.

► 2.6 INTRINSIC CARRIER CONCENTRATION

We now derive the carrier concentration in thermal equilibrium condition, that is, the steady-state condition at a given temperature without any external excitations such as light, pressure, or electric field. At a given temperature, continuous thermal agitation results in the excitation of electrons from the valence band to the conduction band and leaves an equal number of holes in the valence band. An *intrinsic semiconductor* is one that contains relatively small amounts of impurities compared with the thermally generated electrons and holes.

To obtain the electron density (i.e., the number of electrons per unit volume) in an intrinsic semiconductor, we first evaluate the electron density in an incremental energy range dE . This density $n(E)$ is given by the product of the density of states $N(E)$, that is, the density of allowed energy states (including electron spin) per energy range per unit volume,* and by the probability of occupying that energy range $F(E)$. Thus, the electron density in the conduction band is given by integrating $N(E) F(E) dE$ from the bottom of the conduction band (E_C initially taken to be $E = 0$ for simplicity) to the top of the conduction band E_{top} :

$$n = \int_0^{E_{\text{top}}} n(E) dE = \int_0^{E_{\text{top}}} N(E) F(E) dE, \quad (9)$$

where n is in cm^{-3} , and $N(E)$ is in $(\text{cm}^3\text{-eV})^{-1}$.

The probability that an electron occupies an electronic state with energy E is given by the Fermi–Dirac distribution function, which is also called the Fermi distribution function

$$F(E) = \frac{1}{1 + e^{(E - E_F)/kT}}, \quad (10)$$

where k is the Boltzmann constant, T is the absolute temperature in degrees Kelvin, and E_F is the energy of the Fermi level. The Fermi energy is the energy at which the probability of occupation by an electron is exactly one-half. The Fermi distribution is illustrated in Fig. 20 for different temperatures. Note that $F(E)$ is symmetrical around the Fermi energy E_F . For energies that are $3kT$ above or below the Fermi energy, the exponential term in Eq. 10 becomes larger than 20 or smaller than 0.05, respectively. The Fermi distribution function can thus be approximated by simpler expressions:

$$F(E) \cong e^{-(E - E_F)/kT} \quad \text{for } (E - E_F) > 3kT, \quad (11a)$$

$$\text{and} \quad F(E) \cong 1 - e^{-(E - E_F)/kT} \quad \text{for } (E - E_F) < 3kT. \quad (11b)$$

Equation 11b can be regarded as the probability that a hole occupies a state located at energy E .

Figure 21 shows schematically from left to right the band diagram, the density of states $N(E)$, which varies as $\sqrt{E}$ for a given electron effective mass, the Fermi distribution function, and the carrier concentrations for an intrinsic semiconductor. The carrier concentration can be obtained graphically from Fig. 21 using Eq. 9; that is, the product of $N(E)$ in Fig. 21b and $F(E)$ in Fig. 21c gives the $n(E)$ -versus- E curve (upper curve) in Fig. 21d. The upper shaded area in Fig. 21d corresponds to the electron density.

* The density of states $N(E)$ is derived in Appendix H.

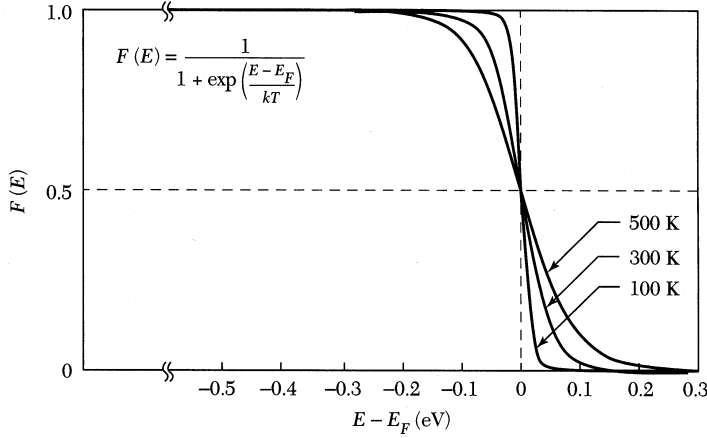


Fig. 20 Fermi distribution function $F(E)$ versus $(E - E_F)$ for various temperatures.

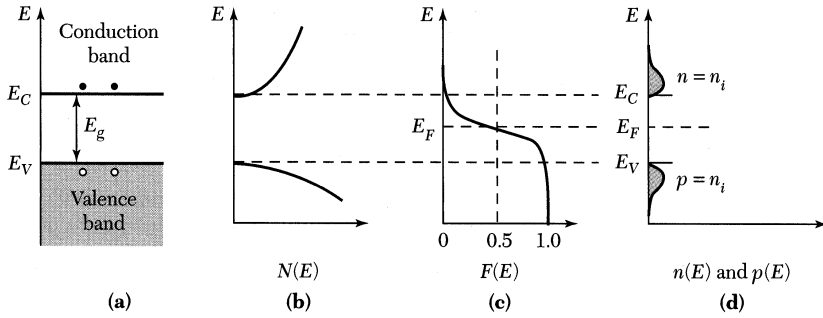


Fig. 21 Intrinsic semiconductor. (a) Schematic band diagram. (b) Density of states. (c) Fermi distribution function. (d) Carrier concentration.

There are a large number of allowed states in the conduction band. However, for an intrinsic semiconductor there will not be many electrons in the conduction band. Therefore, the probability of an electron occupying one of these states is small. There also are a large number of allowed states in the valence band. By contrast, most of these are occupied by electrons. Thus, the probability of an electron occupying one of these states in the valence band is nearly unity. There will be only a few unoccupied electron states, that is, holes, in the valence band. As can be seen, the Fermi level is located near the middle of the bandgap (i.e., E_F is many kT below E_C). Substituting the last equation in Appendix H and Eq. 11a into Eq. 9 yields*

$$n = \frac{2}{\sqrt{\pi}} N_C (kT)^{-3/2} \int_0^\infty E^{1/2} \exp [-(E - E_F) / kT] dE, \quad (12)$$

where

$$N_C \equiv 12(2\pi m_n kT/h^2)^{3/2} \text{ for Si} \quad (13a)$$

$$\equiv 2(2\pi m_n kT/h^2)^{3/2} \text{ for GaAs.} \quad (13b)$$

* We have taken E_{top} to be ∞ , because $F(E)$ becomes very small when $(E - E_C) \gg kT$.

If we let $x \equiv E/kT$, Eq. 12 becomes

$$n = \frac{2}{\sqrt{\pi}} N_C \exp(E_F/kT) \int_0^\infty x^{1/2} e^{-x} dx. \quad (14)$$

The integral in Eq. 14 is of the standard form and equals $\sqrt{\pi}/2$. Therefore, Eq. 14 becomes

$$n = N_C \exp(E_F/kT). \quad (15)$$

If we refer to the bottom of the conduction band as E_C instead of $E = 0$, we obtain for the electron density in the conduction band

$$n = N_C \exp [-(E_C - E_F)/kT], \quad (16)$$

where N_C defined in Eq. 13 is the *effective density of states* in the conduction band. At room temperature (300 K), N_C is $2.86 \times 10^{19} \text{ cm}^{-3}$ for silicon and $4.7 \times 10^{17} \text{ cm}^{-3}$ for gallium arsenide.

Similarly, we can obtain the hole density p in the valence band:

$$p = N_V \exp [-(E_F - E_V)/kT], \quad (17)$$

and

$$N_V \equiv 2 (2\pi m_p kT/h^2)^{3/2}, \quad (18)$$

where N_V is the *effective density of states in the valence band* for both Si and GaAs. At room temperature, N_V is $2.66 \times 10^{19} \text{ cm}^{-3}$ for silicon and $7.0 \times 10^{18} \text{ cm}^{-3}$ for gallium arsenide.

For an intrinsic semiconductor, the number of electrons per unit volume in the conduction band is equal to the number of holes per unit volume in the valence band, that is, $n = p = n_i$ where n_i is the *intrinsic carrier density*. This relationship of electrons and holes is depicted in Fig. 21d. Note that the shaded area in the conduction band is the same as that in the valence band.

The Fermi level for an intrinsic semiconductor is obtained by equating Eq. 16 and Eq. 17:

$$E_F = E_i = (E_C + E_V)/2 + (kT/2) \ln (N_V/N_C). \quad (19)$$

At room temperature, the second term is much smaller than the bandgap. Hence, the intrinsic Fermi level E_i of an intrinsic semiconductor generally lies very close to the middle of the bandgap.

The intrinsic carrier density is obtained from Eqs. 16, 17, and 19:

$$np = n_i^2, \quad (20)$$

$$n_i^2 = N_C N_V \exp (-E_g/kT), \quad (21)$$

and

$$n_i = \sqrt{N_C N_V} \exp (-E_g/2kT), \quad (22)$$

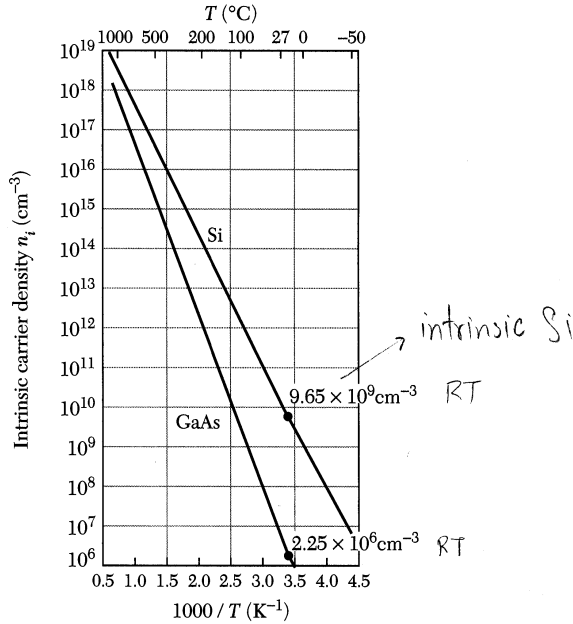


Fig. 22 Intrinsic carrier densities in Si and GaAs as a function of the reciprocal of temperature.⁵⁻⁷

where $E_g \equiv E_C - E_V$. Figure 22 shows the temperature dependence of n_i for silicon and gallium arsenide.⁵ At room temperature (300 K), n_i is $9.65 \times 10^9 \text{ cm}^{-3}$ for silicon⁶ and $2.25 \times 10^6 \text{ cm}^{-3}$ for gallium arsenide.⁷ As expected, the larger the bandgap, the smaller the intrinsic carrier density.

► 2.7 DONORS AND ACCEPTORS

When a semiconductor is doped with impurities, the semiconductor becomes *extrinsic* and impurity energy levels are introduced. Figure 23a shows schematically that a silicon atom is replaced (or substituted) by an arsenic atom with five valence electrons. The arsenic atom forms covalent bonds with its four neighboring silicon atoms. The fifth electron has a relatively small binding energy to its host arsenic atom and can be “ionized” to become a conduction electron at a moderate temperature. We say that this electron has been “donated” to the conduction band. The arsenic atom is called a *donor* and the silicon becomes *n*-type because of the addition of the negative charge carrier. Similarly, Fig. 23b shows that when a boron atom with three valence electrons substitutes for a silicon atom, an additional electron is “accepted” to form four covalent bonds around the boron, and a positively charged “hole” is created in the valence band. This is a *p*-type semiconductor, and the boron is an *acceptor*.

We can estimate the *ionization energy* for the donor E_D by replacing m_0 with the electron effective mass m_n and taking into account the semiconductor permittivity ϵ_s in the hydrogen atom model, Eq. 5:

$$E_D = \left(\frac{\epsilon_0}{\epsilon_s} \right)^2 \left(\frac{m_n}{m_0} \right) E_H. \quad (23)$$

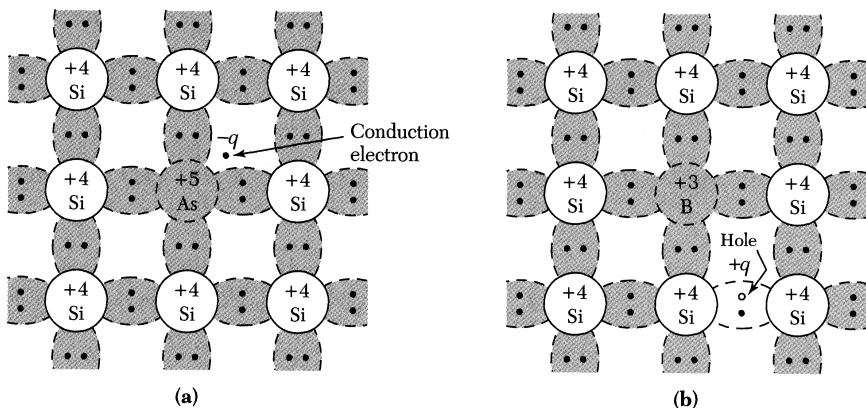


Fig. 23 Schematic bond pictures for (a) *n*-type Si with donor (arsenic) and (b) *p*-type Si with acceptor (boron).

The ionization energy for donors, measured from the conduction band edge, as calculated from Eq. 23 is 0.025 eV for silicon and 0.007 eV for gallium arsenide. The hydrogen atom calculation for the ionization level of acceptors is similar to that for donors. We consider the unfilled valence band as a filled band plus a hole in the central force field of a negative charged acceptor. The calculated ionization energy, measured from the valence band edge, is 0.05 eV for both silicon and gallium arsenide.

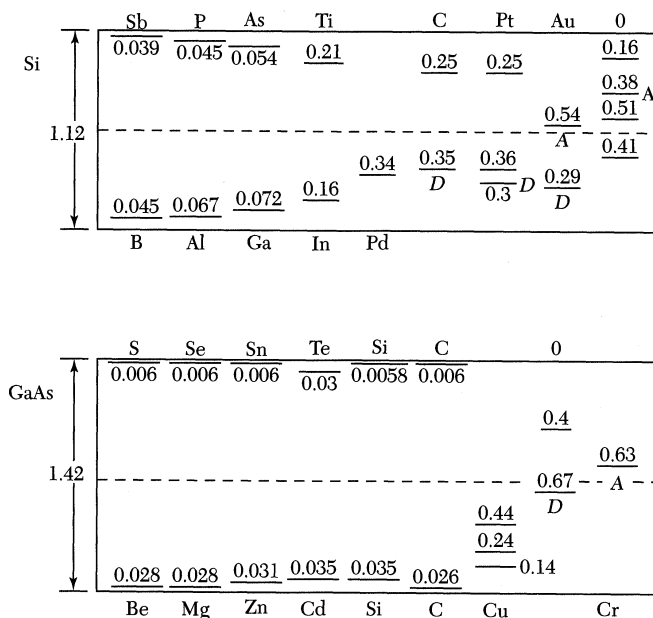


Fig. 24 Measured ionization energies (in eV) for various impurities in Si and GaAs. The levels below the gap center are measured from the top of the valence band and are acceptor levels unless indicated by *D* for donor level. The levels above the gap center are measured from the bottom of the conduction band and are donor levels unless indicated by *A* for acceptor level.⁸

This simple hydrogen atom model cannot account for the details of the ionization energy, particularly for the deep impurity levels in semiconductors (i.e., with ionization energies $\geq 3 kT$). However, the calculated values do predict the correct order of magnitude of the true ionization energies for shallow impurity levels. Figure 24 shows the measured ionization energies for various impurities in silicon and gallium arsenide.⁸ Note that it is possible for a single atom to have many levels; for example, oxygen in silicon has two donor levels and two acceptor levels in the forbidden energy gap.

2.7.1 Nondegenerate Semiconductor

In our previous discussion, we have assumed that the electron or hole concentration is much lower than the effective density of states in the conduction band or the valence band, respectively. In other words, the Fermi level E_F is at least $3kT$ above E_V or $3kT$ below E_C . For such cases, the semiconductor is referred to as a *nondegenerate* semiconductor.

For shallow donors in silicon and gallium arsenide, there usually is enough thermal energy to supply the energy E_D to ionize all donor impurities at room temperature and thus provide the same number of electrons in the conduction band. This condition is called complete ionization. Under a complete ionization condition, we can write the electron density as

$$n = N_D, \quad (24)$$

where N_D is the donor concentration. Figure 25a illustrates complete ionization where the donor level E_D is measured with respect to the bottom of the conduction band and equal concentrations of electrons (which are mobile) and donor ions (which are immobile) are shown. From Eqs. 16 and 24, we obtain the Fermi level in terms of the effective density of states N_C and the donor concentration N_D :

$$E_C - E_F = kT \ln (N_C / N_D). \quad (25)$$

Similarly, for shallow acceptors as shown in Fig. 25b, if there is complete ionization, the concentration of holes is

$$p = N_A, \quad (26)$$

Where N_A is the acceptor concentration. We can obtain the corresponding Fermi level from Eqs. 17 and 26:

$$E_F - E_V = kT \ln (N_V / N_A). \quad (27)$$

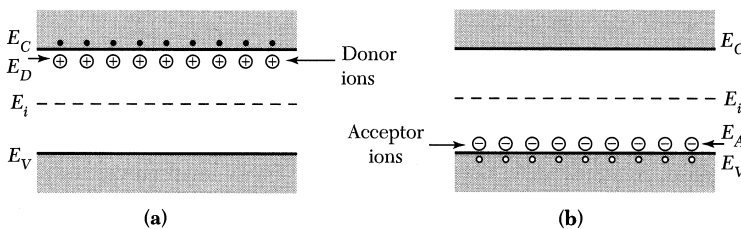


Fig. 25 Schematic energy band representation of extrinsic semiconductors with (a) donor ions and (b) acceptor ions.

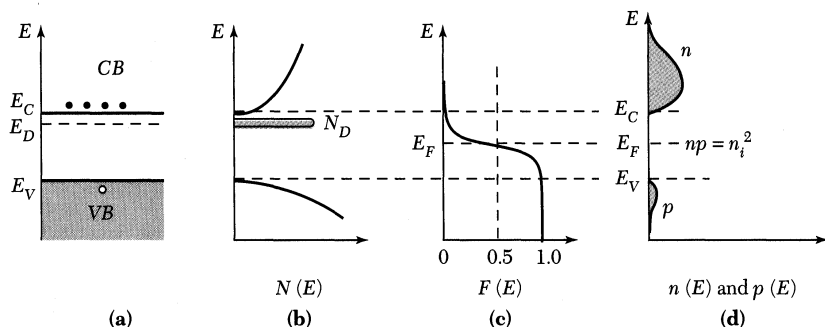


Fig. 26 *n*-Type semiconductor. (a) Schematic band diagram. (b) Density of states. (c) Fermi distribution function. (d) Carrier concentration. Note that $np = n_i^2$.

From Eq. 25 we can see that the higher the donor concentration, the smaller the energy difference $(E_C - E_F)$, that is, the Fermi level will move closer to the bottom of the conduction band. Similarly, for higher acceptor concentration, the Fermi level will move closer to the top of the valence band. Figure 26 illustrates the procedure for obtaining the carrier concentration. This figure is similar to that shown in Fig. 21. However, the Fermi level is closer to the bottom of the conduction band, and the electron concentration (upper shaded area) is much larger than the hole concentration (lower shaded area).

It is useful to express electron and hole densities in terms of the intrinsic carrier concentration n_i and the intrinsic Fermi level E_i since E_i is frequently used as a reference level when discussing extrinsic semiconductors. From Eq. 16 we obtain

$$\begin{aligned} n &= N_C \exp [-(E_C - E_F)/kT], \\ &= N_C \exp [-(E_C - E_i)/kT] \exp [(E_F - E_i)/kT], \end{aligned}$$

or

$$n = n_i \exp [(E_F - E_i)/kT], \quad (28)$$

and similarly,

$$p = n_i \exp [(E_i - E_F)/kT]. \quad (29)$$

Note that the product of n and p from Eqs. 28 and 29 equals n_i^2 . This result is identical to that for the intrinsic case, Eq. 20. Equation 20 is called the *mass action law*, which is valid for both intrinsic and extrinsic semiconductors under thermal equilibrium conduction. In an extrinsic semiconductor, the Fermi level moves toward either the bottom of the conduction band (*n*-type) or the top of the valence band (*p*-type). Either *n*- or *p*-type carriers will then dominate, but the product of the two types of carriers will remain constant at a given temperature.

► EXAMPLE 4

A silicon ingot is doped with 10^{16} arsenic atoms/cm³. Find the carrier concentrations and the Fermi level at room temperature (300 K).

SOLUTION At 300 K, we can assume complete ionization of impurity atoms. We have

$$n \approx N_D = 10^{16} \text{ cm}^{-3}.$$

From Eq. 20, $p \approx n_i^2/N_D = (9.65 \times 10^9)^2/10^{16} = 9.3 \times 10^3 \text{ cm}^{-3}$.

The Fermi level measured from the bottom of the conduction band is given by Eq. 25:

$$\begin{aligned} E_C - E_F &= kT \ln(N_C/N_D) \\ &= 0.0259 \ln(2.86 \times 10^{19}/10^{16}) = 0.205 \text{ eV}. \end{aligned}$$

The Fermi level measured from the intrinsic Fermi level is given by Eq. 28:

$$\begin{aligned} E_F - E_i &= kT \ln(N_D/n_i) \approx kT \ln(N_D/n_i) \\ &= 0.0259 \ln(10^{16}/9.65 \times 10^9) = 0.358 \text{ eV}. \end{aligned}$$

These results are shown graphically in Fig. 27. ◀

If both donor and acceptor impurities are present simultaneously, the impurity that is present in a greater concentration determines the type of conductivity in the semiconductor. The Fermi level must adjust itself to preserve charge neutrality, that is, the total negative charges (electrons and ionized acceptors) must equal the total positive charges (holes and ionized donors). Under complete ionization condition, we have

$$n + N_A = p + N_D. \quad (30)$$

Solving Eqs. 20 and 30 yields the equilibrium electron and hole concentrations in an n -type semiconductor:

$$n = \frac{1}{2} \left[N_D - N_A + \sqrt{(N_D - N_A)^2 + 4n_i^2} \right], \quad (31)$$

$$p_n = n_i^2/n_n. \quad (32)$$

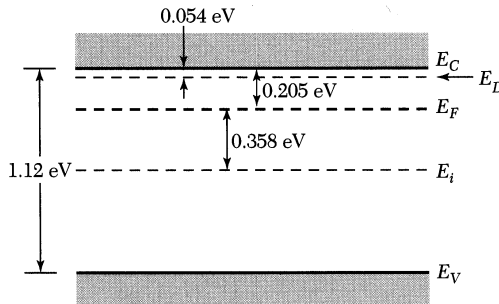


Fig. 27 Band diagram showing Fermi level E_F and intrinsic Fermi level E_i .

The subscript n refers to the n -type semiconductor. Because the electron is the dominant carrier, it is called the *majority carrier*. The hole in the n -type semiconductor is called the *minority carrier*. Similarly, we obtain the concentration of holes (majority carrier) and electrons (minority carrier) in a p -type semiconductor:

$$p_p = \frac{1}{2} \left[N_A - N_D + \sqrt{(N_D - N_A)^2 + 4n_i^2} \right], \quad (33)$$

$$n_p = n_i^2 / p_p \quad (34)$$

The subscript p refers to the p -type semiconductor.

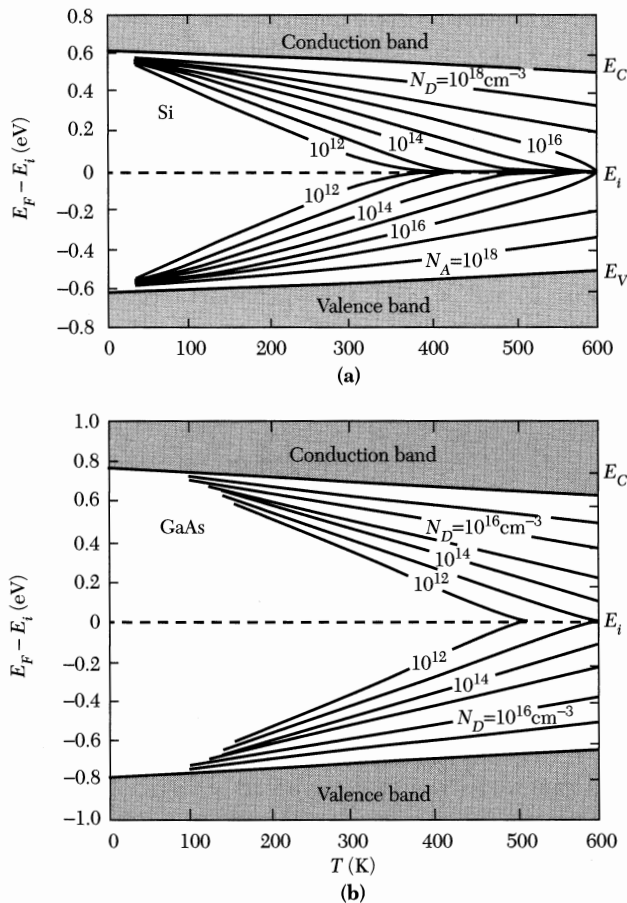


Fig. 28 Fermi level for Si and GaAs as a function of temperature and impurity concentration. The dependence of the bandgap on temperature is shown.⁹

Generally, the magnitude of the net impurity concentration $|N_D - N_A|$ is larger than the intrinsic carrier concentration n_i ; therefore, the above relationships can be simplified to

$$n_n \approx N_D - N_A \quad \text{if} \quad N_D > N_A, \quad (35)$$

$$p_p \approx N_A - N_D \quad \text{if} \quad N_A > N_D. \quad (36)$$

From Eqs. 31 to 34 together with Eqs. 16 and 17, we can calculate the position of the Fermi level as a function of temperature for a given acceptor or donor concentration. Figure 28 shows a plot of these calculations for silicon⁹ and gallium arsenide. We have incorporated in the figure the variation of the bandgap with temperature (see Problem 8). Note that as the temperature increases, the Fermi level approaches the intrinsic level, that is, the semiconductor becomes intrinsic.

Figure 29 shows electron density in Si as a function of temperature for a donor concentration of $N_D = 10^{15} \text{ cm}^{-3}$. At low temperatures, the thermal energy in the crystal is not sufficient to ionize all the donor impurities present. Some electrons are “frozen” at the donor level and the electron density is less than the donor concentration. As the temperature is increased, the condition of complete ionization is reached, (i.e., $n_n = N_D$). As the temperature is further increased, the electron concentration remains essentially the same over a wide temperature range. This is the extrinsic region. However, as the temperature is increased even further, we reach a point where the intrinsic carrier concentration becomes comparable to the donor concentration. Beyond this point, the semiconductor becomes intrinsic. The temperature at which the semiconductor becomes intrinsic depends on the impurity concentrations and the bandgap value and can be obtained from Fig. 22 by setting the impurity concentration equal to n_i .

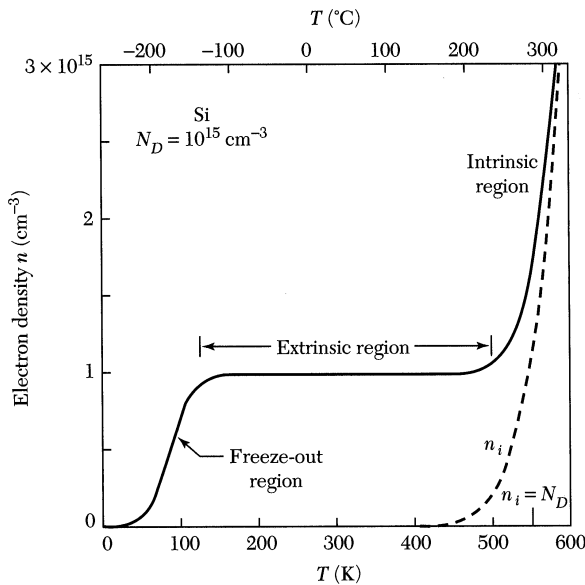


Fig. 29 Electron density as a function of temperature for a Si sample with a donor concentration of 10^{15} cm^{-3} .

2.7.2 Degenerate Semiconductor

When the doping concentration becomes equal or larger than the corresponding effective density of states, we can no longer use the approximation of Eq. 11 and the electron density, Eq. 9, has to be integrated numerically. For very heavily doped n -type or p -type semiconductor, E_F will be above E_C or below E_V . The semiconductor is referred to as a *degenerate* semiconductor.

An important aspect of high doping is the bandgap narrowing effect, that is, high impurity concentration causes a reduction of the bandgap. The bandgap reduction ΔE_g for silicon at room temperature is given by

$$\Delta E_g = 22 \left(\frac{N}{10^{18}} \right)^{1/2} \text{ meV}, \quad (37)$$

where the doping is in cm^{-3} . For example, for $N_D \leq 10^{18} \text{ cm}^{-3}$, $\Delta E_g \leq 0.022 \text{ eV}$, which is less than 2% of the original bandgap. However, for $N_D \geq N_C = 2.86 \times 10^{19} \text{ cm}^{-3}$, $\Delta E_g \geq 0.12 \text{ eV}$, which is a significant fraction of E_g .

► SUMMARY

At the beginning of the chapter we listed a few important semiconductor materials. The properties of semiconductors are determined to a large extent by the crystal structure. We have defined the Miller indices to describe the crystal surfaces and crystal orientations and briefly described how semiconductor crystals are grown. More detailed discussions can be found in Chapter 10.

The bonding of atoms and the electron energy-momentum relationship in a semiconductor were considered in connection with the electrical properties. The energy band diagram can be used to understand why some materials are good conductors of electric current whereas others are poor conductors. We have also shown that changing the temperature or the amount of impurities can drastically vary the conductivity of a semiconductor.

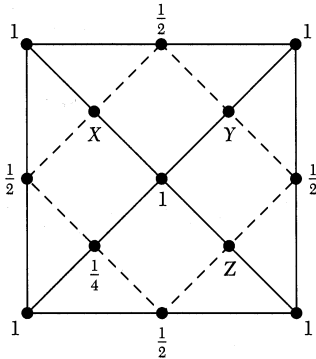
► REFERENCES

1. R. A. Smith, *Semiconductors*, 2nd ed., Cambridge Univ. Press, London, 1979.
2. R. F. Pierret, *Semiconductor Device Fundamentals*, Addison Wesley, Boston, MA, 1996.
3. C. Kittel, *Introduction to Solid State Physics*, 6th ed., Wiley, New York, 1986.
4. D. Halliday and R. Resnick, *Fundamentals of Physics*, 2nd ed., Wiley, New York, 1981.
5. C. D. Thurmond, "The Standard Thermodynamic Function of the Formation of Electrons and Holes in Ge, Si, GaAs, and GaP," *J. Electrochem. Soc.*, **122**, 1133 (1975).
6. P. P. Altermatt, et al., "The Influence of a New Bandgap Narrowing Model on Measurement of the Intrinsic Carrier Density in Crystalline Silicon," *Tech. Dig., 11th Int. Photovoltaic Sci. Eng. Conf.*, Sapporo, p. 719 (1999).
7. J. S. Blackmore, "Semiconducting and Other Major Properties of Gallium Arsenide," *J. Appl. Phys.*, **53**, 123–181 (1982).
8. S. M. Sze, *Physics of Semiconductors Devices*, 2nd ed., Wiley, New York, 1981.
9. A. S. Grove, *Physics and Technology of Semiconductor Devices*, Wiley, New York, 1967.

► PROBLEMS (* DENOTES DIFFICULT PROBLEMS)

FOR SECTION 2.2 BASIC CRYSTAL STRUCTURE

1. (a) What is the distance between nearest neighbors in silicon?
(b) Find the number of atoms per square centimeter in silicon in the (100), (110), and (111) planes.
2. If we project the atoms in a diamond lattice onto the bottom surface with the heights of the atoms in unit of the lattice constant shown in the figure below. Find the heights of the three atoms (X, Y, Z) on the figure.



3. Find the maximum fraction of the unit cell volume, which can be filled by identical hard spheres in the simple cubic, face-centered cubic, and diamond lattices.
- *4. Calculate the tetrahedral bond angle, the angle between any pair of the four bonds in a diamond lattice. (Hint: represent the four bonds as vectors of equal length. What must be the sum of the four vectors equal? Take components of this vector equation along the direction of one of these vectors.)
5. If a plane has intercepts at $2a$, $3a$, and $4a$ along the three Cartesian coordinates, where a is the lattice constant, find the Miller indices of the planes.
6. (a) Calculate the density of GaAs (the lattice constant of GaAs is $5.65 \text{ \AA}$, and the atomic weights of Ga and As are 69.72 and 74.92 g/mol, respectively).
(b) A gallium arsenide sample is doped with tin. If the tin displaces gallium atoms in the crystal lattice, are donors or acceptors formed? Why? Is the semiconductor n - or p -type?

FOR SECTION 2.3 BASIC CRYSTAL GROWTH TECHNIQUE

7. (a) Does silicon or silicon dioxide have a higher melting point? Why?
(b) Why is a seed crystal used for crystal growth?
(c) Why is the crystal orientation of a wafer important?
(d) What two variables are used to control the diameter of the silicon rod?

FOR SECTION 2.5 ENERGY BANDS

8. The variation of silicon and GaAs bandgaps with temperature can be expressed as $E_g(T) = E_g(0) - \alpha T^2/(T + \beta)$, where $E_g(0) = 1.17 \text{ eV}$, $\alpha = 4.73 \times 10^{-4} \text{ eV/K}$, and $\beta = 636 \text{ K}$ for

silicon; and $E_g(0) = 1.519$ eV, $\alpha = 5.405 \times 10^{-4}$ eV/K, and $\beta = 204$ K for GaAs. Find the bandgaps of Si and GaAs at 100 K and 600 K.

FOR SECTION 2.6 INTRINSIC CARRIER CONCENTRATION

- *9. Derive Eq. 17. (Hint: In the valence band, the probability of occupancy of a state by a hole is $[1 - F(E)]$.)
10. At room temperature (300 K) the effective density of states in the valence band is 2.66×10^{19} cm⁻³ for silicon and 7×10^{18} cm⁻³ for gallium arsenide. Find the corresponding effective masses of holes. Compare these masses with the free-electron mass.
11. Calculate the location of E_i in silicon at liquid nitrogen temperature (77 K), at room temperature (300 K), and at 100°C (let $m_p = 1.0 m_0$ and $m_n = 0.19 m_0$). Is it reasonable to assume that E_i is in the center of the forbidden gap?
12. Find the kinetic energy of electrons in the conduction band of a nondegenerate n -type semiconductor at 300 K.
13. (a) For a free electron with a velocity of 10^7 cm/s, what is its de Broglie wavelength.
(b) In GaAs, the effective mass of electrons in the conduction band is $0.063 m_0$. If they have the same velocity, find the corresponding de Broglie wavelength.
14. The intrinsic temperature of a semiconductor is the temperatures at which the intrinsic carrier concentration equals the impurity concentration. Find the intrinsic temperature for a silicon sample doped with 10^{15} phosphorus atoms/cm³.

FOR SECTION 2.7 DONORS AND ACCEPTORS

15. A silicon sample at $T = 300$ K contains an acceptor impurity concentration of $N_A = 10^{16}$ cm⁻³. Determine the concentration of donor impurity atoms that must be added so that the silicon is n -type and the Fermi energy is 0.20 eV below the conduction band edge.
16. Draw a simple flat energy band diagram for silicon doped with 10^{16} arsenic atoms/cm³ at 77 K, 300 K, and 600 K. Show the Fermi level and use the intrinsic Fermi level as the energy reference.
17. Find the electron and hole concentrations and Fermi level in silicon at 300 K (a) for 1×10^{15} boron atoms/cm³ and (b) for 3×10^{16} boron atoms/cm³ and 2.9×10^{16} arsenic atoms/cm³.
18. A Si sample is doped with 10^{17} As atoms/cm³. What is the equilibrium hole concentration p_0 at 300 K? Where is E_F relative to E_i ?
19. Calculate the Fermi level of silicon doped with 10^{15} , 10^{17} , and 10^{19} phosphorus atoms/cm³ at room temperature, assuming complete ionization. From the calculated Fermi level, check if the assumption of complete ionization is justified for each doping. Assume that the ionized donors is given by $n = N_D [1 - F(E_D)] = \frac{N_D}{1 + e^{(E_F - E_D)/kT}}$.
20. For an n -type silicon sample with 10^{16} cm⁻³ phosphorous donor impurities and a donor level at $E_D = 0.045$ eV, find the ratio of the neutral donor density to the ionized donor density at 77 K where the Fermi level is 0.0459 below the bottom of the conduction band. The expression for ionized donors is given in Prob. 19.

Carrier Transport Phenomena

-
- ▶ 3.1 CARRIER DRIFT
 - ▶ 3.2 CARRIER DIFFUSION
 - ▶ 3.3 GENERATION AND RECOMBINATION PROCESSES
 - ▶ 3.4 CONTINUITY EQUATION
 - ▶ 3.5 THERMIONIC EMISSION PROCESS
 - ▶ 3.6 TUNNELING PROCESS
 - ▶ 3.7 HIGH-FIELD EFFECTS
 - ▶ SUMMARY
-

In this chapter, we investigate various transport phenomena in semiconductor devices. The transport processes include drift, diffusion, recombination, generation, thermionic emission, tunneling, and impact ionization. We consider the motion of charge carriers (electrons and holes) in semiconductors under the influence of an electric field and a carrier concentration gradient. We also discuss the concept of the nonequilibrium condition where the carrier concentration product pn is different from its equilibrium value n_i^2 . Returning to an equilibrium condition through the generation-recombination processes is considered next. We then derive the basic governing equations for semiconductor device operation, which includes the current density equation and the continuity equation. This is followed by a discussion of thermionic emission and tunneling process. The chapter closes with a brief discussion of high-field effects, which include velocity saturation and impact ionization.

Specifically, we cover the following topics:

- The current density equation and its drift and diffusion components.
- The continuity equation and its generation and recombination components.
- Other transport phenomena, including the thermionic emission, tunneling, transferred-electron effect, and impact ionization.
- Methods to measure key semiconductor parameters such as resistivity, mobility, majority-carrier concentration, and minority-carrier lifetime.

► 3.1 CARRIER DRIFT

3.1.1 Mobility

Consider an n -type semiconductor sample with uniform donor concentration in thermal equilibrium. As discussed in Chapter 2, the conduction electrons in the semiconductor conduction band are essentially free particles, since they are not associated with any particular lattice or donor site. The influence of crystal lattices is incorporated in the effective mass of conduction electrons, which differs somewhat from the mass of free electrons. Under thermal equilibrium, the average thermal energy of a conduction electron can be obtained from the theorem for equipartition of energy, $1/2 kT$ units of energy per degree of freedom, where k is Boltzmann's constant and T is the absolute temperature. The electrons in a semiconductor have three degrees of freedom; they can move about in a three-dimensional space. Therefore, the kinetic energy of the electrons is given by

$$\frac{1}{2} m_n v_{th}^2 = \frac{3}{2} kT, \quad (1)$$

where m_n is the effective mass of electrons and v_{th} is the average thermal velocity. At room temperature (300 K) the thermal velocity of electrons in Eq. 1 is about 10^7 cm/s for silicon and gallium arsenide.

The electrons in the semiconductor are therefore moving rapidly in all directions. The thermal motion of an individual electron may be visualized as a succession of random scattering from collisions with lattice atoms, impurity atoms, and other scattering centers, as illustrated in Fig. 1a. The random motion of electrons leads to a zero net displacement of an electron over a sufficiently long period of time. The average distance between collisions is called the *mean free path*, and the average time between collisions is called the *mean free time* τ_c . For a typical value of 10^{-5} cm for the mean free path, τ_c is about 1 ps (ie., $10^{-5}/v_{th} \cong 10^{-12}$ s).

When a small electric field $\mathcal{E}$ is applied to the semiconductor sample, each electron will experience a force $-q\mathcal{E}$ from the field and will be accelerated along the field (in the opposite direction to the field) during the time between collisions. Therefore, an additional velocity component will be superimposed upon the thermal motion of electrons. This additional component is called the *drift velocity*. The combined displacement of an electron due to the random thermal motion and the drift component is illustrated in Fig. 1b. Note that there is a net displacement of the electron in the direction opposite to the applied field.

We can obtain the drift velocity v_n by equating the momentum (force $\times$ time) applied to an electron during the free flight between collisions to the momentum gained by the electron in the same period. The equality is valid because in a steady state all momen-

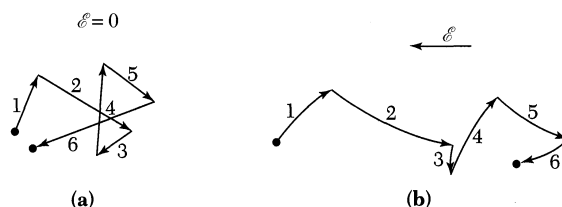


Fig. 1 Schematic path of an electron in a semiconductor. (a) Random thermal motion. (b) Combined motion due to random thermal motion and an applied electric field.

tum gained between collisions is lost to the lattice in the collision. The momentum applied to an electron is given by $-q\mathcal{E}\tau_c$, and the momentum gained is $m_n v_n$. We have

$$-q\mathcal{E}\tau_c = m_n v_n \quad (2)$$

or

$$v_n = -\left(q\tau_c / m_n\right)\mathcal{E} \quad (2a)$$

Equation 2a states that the electron drift velocity is proportional to the applied electric field. The proportionality factor depends on the mean free time and the effective mass. The proportionality factor is called the *electron mobility* μ_n in units of $\text{cm}^2/\text{V}\cdot\text{s}$, or

$$\mu_n \equiv \frac{q\tau_c}{m_n} \quad (3)$$

Thus,

$$\boxed{v_n = -\mu_n \mathcal{E}} \quad (4)$$

Mobility is an important parameter for carrier transport because it describes how strongly the motion of an electron is influenced by an applied electric field. A similar expression can be written for holes in the valence band:

$$\boxed{v_p = \mu_p \mathcal{E}}, \quad (5)$$

where v_p is the hole drift velocity and μ_p is the hole mobility. The negative sign is removed in Eq. 5 because holes drift in the same direction as the electric field.

In Eq. 3 the mobility is related directly to the mean free time between collisions, which in turn is determined by the various scattering mechanisms. The two most important mechanisms are lattice scattering and impurity scattering. Lattice scattering results from thermal vibrations of the lattice atoms at any temperature above absolute zero. These vibrations disturb the lattice periodic potential and allow energy to be transferred between the carriers and the lattice. Since lattice vibration increases with increasing temperature, lattice scattering becomes dominant at high temperatures; hence the mobility decreases with increasing temperature. Theoretical analysis¹ shows that the mobility due to lattice scattering μ_L will decrease in proportion to $T^{-3/2}$.

Impurity scattering results when a charge carrier travels past an ionized dopant impurity (donor or acceptor). The charge carrier path will be deflected owing to Coulomb force interaction. The probability of impurity scattering depends on the total concentration of ionized impurities, that is, the sum of the concentration of negatively and positively charged ions. However, unlike lattice scattering, impurity scattering becomes less significant at higher temperatures. At higher temperatures, the carriers move faster; they remain near the impurity atom for a shorter time and are therefore less effectively scattered. The mobility due to impurity scattering μ_i can theoretically be shown to vary as $T^{3/2}/N_T$, where N_T is the total impurity concentration.²

The probability of a collision taking place in unit time, $1/\tau_c$, is the sum of the probabilities of collisions due to the various scattering mechanisms:

$$\frac{1}{\tau_c} = \frac{1}{\tau_{c, \text{ lattice}}} + \frac{1}{\tau_{c, \text{ impurity}}}, \quad (6)$$

or

$$\frac{1}{\mu} = \frac{1}{\mu_L} + \frac{1}{\mu_I}. \tag{6a}$$

Figure 2 shows the measured electron mobility as a function of temperature for silicon with five different donor concentrations.³ The inset shows the theoretical temperature dependence of mobility due to both lattice and impurity scatterings. For lightly doped samples (e.g., the sample with doping of 10^{14} cm^{-3}), the lattice scattering dominates, and the mobility decreases as the temperature increases. For heavily doped samples, the effect of impurity scattering is most pronounced at low temperatures. The mobility increases as the temperature increases, as can be seen for the sample with doping of 10^{19} cm^{-3} . For a given temperature, the mobility decreases with increasing impurity concentration because of enhanced impurity scatterings.

Figure 3 shows the measured mobilities in silicon and gallium arsenide as a function of impurity concentration at room temperature.³ Mobility reaches a maximum value at low impurity concentrations; this corresponds to the lattice-scattering limitation. Both electron and hole mobilities decrease with increasing impurity concentration and eventually approach a minimum value at high concentrations. Note also that the mobility of electrons is greater than that of holes. Greater electron mobility is due mainly to the smaller effective mass of electrons.

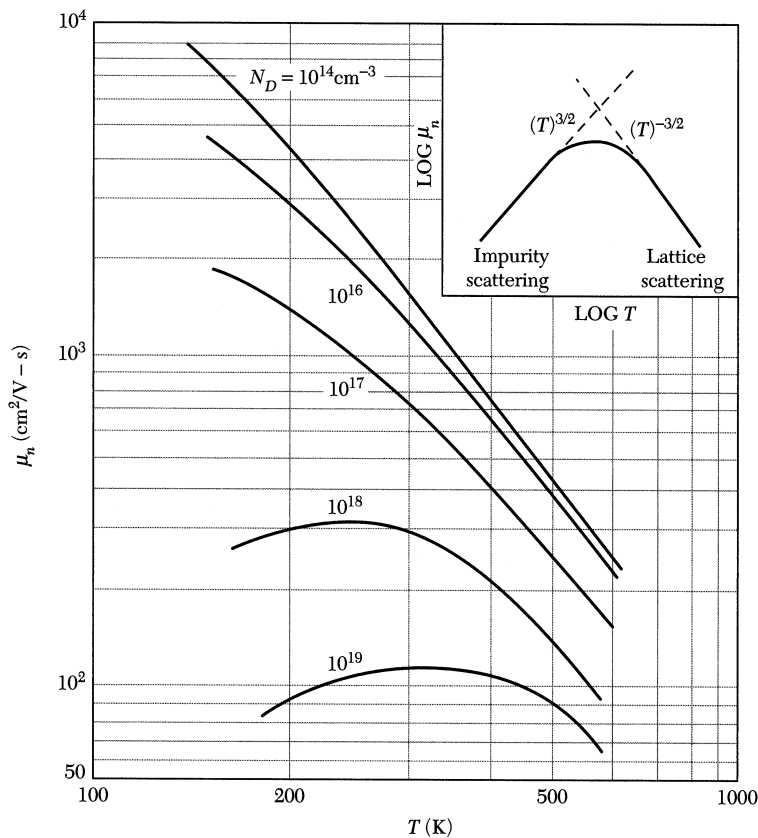


Fig. 2 Electron mobility in silicon versus temperature for various donor concentrations. Insert shows the theoretical temperature dependence of electron mobility.³

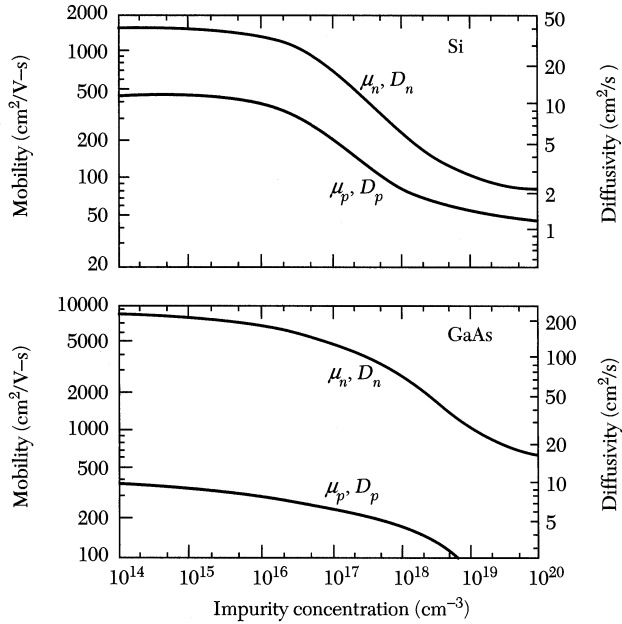


Fig. 3 Mobilities and diffusivities in Si and GaAs at 300 K as a function of impurity concentration.³

► EXAMPLE 1

Calculate the mean free time of an electron having a mobility of $1000 \text{ cm}^2/\text{V}\cdot\text{s}$ at 300 K; also calculate the mean free path. Assume $m_n = 0.26 m_0$ in these calculations.

SOLUTION From Eq. 3, the mean free time is given by

$$\begin{aligned}\tau_c &= \frac{m_n \mu_n}{q} = \frac{(0.26 \times 0.91 \times 10^{-30} \text{ kg}) \times (1000 \times 10^{-4} \text{ m}^2/\text{V}\cdot\text{s})}{1.6 \times 10^{-19} \text{ C}} \\ &= 1.48 \times 10^{-13} \text{ s} = 0.148 \text{ ps.}\end{aligned}$$

The mean free path is given by

$$\begin{aligned}l &= v_{th} \tau_c = (10^7 \text{ cm/s})(1.48 \times 10^{-13} \text{ s}) = 1.48 \times 10^{-6} \text{ cm} \\ &= 14.8 \text{ nm.}\end{aligned}$$

3.1.2 Resistivity

We now consider conduction in a homogeneous semiconductor material. Figure 4a shows an *n*-type semiconductor and its band diagram at thermal equilibrium. Figure 4b shows the corresponding band diagram when a biasing voltage is applied to the right-hand terminal. We assume that the contacts at the left-hand and right-hand terminals are ohmic, that is, there is negligible voltage drop at each of the contacts. The behavior of ohmic contacts are considered in Chapter 7. As mentioned previously, when an electric field $\mathcal{E}$

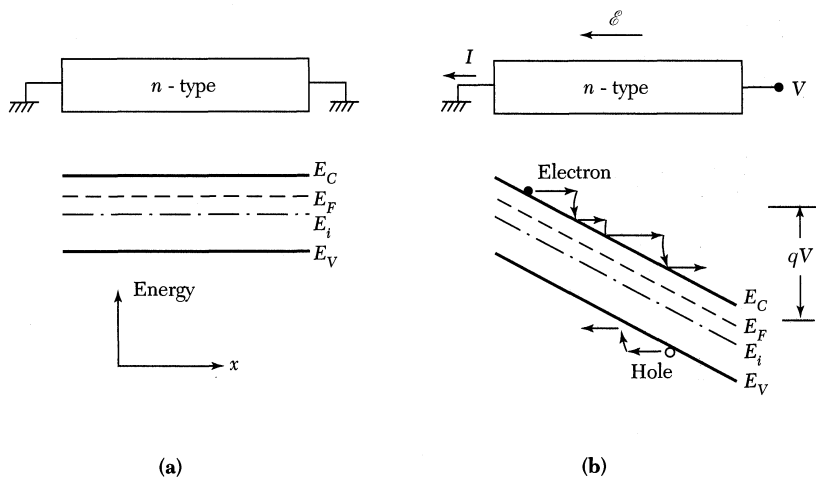


Fig. 4 Conduction process in an *n*-type semiconductor (a) at thermal equilibrium and (b) under a biasing condition.

is applied to a semiconductor, each electron will experience a force $-q\mathcal{E}$ from the field. The force is equal to the negative gradient of potential energy; that is,

$$-q\mathcal{E} = -(\text{gradient of electron potential energy}) = -\frac{dE_C}{dx}. \quad (7)$$

Recall that in Chapter 2 the bottom of the conduction band E_C corresponds to the potential energy of an electron. Since we are interested in the gradient of the potential energy, we can use any part of the band diagram that is parallel to E_C (e.g., E_F , E_i , or E_V , as shown in Fig. 4b). It is convenient to use the intrinsic Fermi level E_i because we shall use E_i when we consider *p-n* junctions in Chapter 4. Therefore, from Eq. 7 we have

$$\mathcal{E} = \frac{1}{q} \frac{dE_C}{dx} = \frac{1}{q} \frac{dE_i}{dx}. \quad (8)$$

We can define a related quantity ψ as the *electrostatic potential* whose negative gradient equals the electric field:

$$\mathcal{E} \equiv -\frac{d\psi}{dx}. \quad (9)$$

Comparison of Eqs. 8 and 9 gives

$$\psi = -\frac{E_i}{q}, \quad (10)$$

which provides a relationship between the electrostatic potential and the potential energy of an electron. For a homogeneous semiconductor shown in Fig. 4b, the potential energy and E_i decrease linearly with distance; thus, the electric field is a constant in the negative x -direction. Its magnitude equals the applied voltage divided by the sample length.

The electrons in the conduction band move to the right side as shown in Fig. 4b. The kinetic energy corresponds to the distance from the band edge (i.e., E_C for electrons). When an electron undergoes a collision, it loses some or all of its kinetic energy to the lattice and drops toward its thermal equilibrium position. After the electron has lost some

or all its kinetic energy, it will again begin to move toward the right, and the same process will be repeated many times. Conduction by holes can be visualized in a similar manner but in the opposite direction.

The transport of carriers under the influence of an applied electric field produces a current called the *drift current*. Consider a semiconductor sample shown in Fig. 5, which has a cross-sectional area A , a length L , and a carrier concentration of n electrons/cm³. Suppose we now apply an electric field $\mathcal{E}$ to the sample. The electron current density J_n flowing in the sample can be found by summing the product of the charge ($-q$) on each electron times the electron's velocity over all electrons per unit volume n :

$$J_n = \frac{I_n}{A} = \sum_{i=1}^n (-qv_i) = -qn v_n = qn\mu_n \mathcal{E}, \quad (11)$$

where I_n is the electron current. We have employed Eq. 4 for the relationship between v_n and $\mathcal{E}$.

A similar argument applies to holes. By taking the charge on the hole to be positive, we have

$$J_p = qp v_p = qp\mu_p \mathcal{E}. \quad (12)$$

The total current flowing in the semiconductor sample due to the applied field $\mathcal{E}$ can be written as the sum of the electron and hole current components:

$$J = J_n + J_p = (qn\mu_n + qp\mu_p)\mathcal{E}. \quad (13)$$

The quantity in parentheses is known as *conductivity*:

$$\sigma = q(n\mu_n + p\mu_p). \quad (14)$$

The electron and hole contributions to conductivity are simply additive.

The corresponding resistivity of the semiconductor, which is the reciprocal of σ , is given by

$$\rho \equiv \frac{1}{\sigma} = \frac{1}{q(n\mu_n + p\mu_p)}. \quad (15)$$

Generally, in extrinsic semiconductors, only one of the components in Eq. 13 or 14 is significant because of the many orders-of-magnitude difference between the two carrier densities. Therefore, Eq. 15 reduces to

$$\rho = \frac{1}{qn\mu_n} \quad (15a)$$

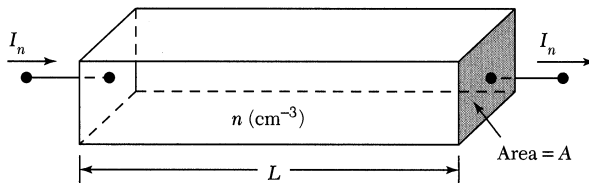


Fig. 5 Current conduction in a uniformly doped semiconductor bar with length L and cross-sectional area A .

for an n -type semiconductor (since $n \gg p$), and to

$$\rho = \frac{1}{qn\mu_n} \quad (15b)$$

for a p -type semiconductor (since $p \gg n$).

The most common method for measuring resistivity is the four-point probe method shown in Fig. 6. The probes are equally spaced. A small current I from a constant-current source is passed through the outer two probes and a voltage V is measured between the inner two probes. For a thin semiconductor sample with thickness W that is much smaller than the sample diameter d , the resistivity is given by

$$\rho = \frac{V}{I} \cdot W \cdot CF \quad \Omega \cdot \text{cm}, \quad (16)$$

where CF is a well-documented "correction factor." The correction factor depends on the ratio of d/s , where s is the probe spacing. When $d/s > 20$, the correction factor approaches 4.54.

Figure 7 shows the measured room-temperature resistivity as a function of the impurity concentration for silicon and gallium arsenide.³ At this temperature and for low impurity concentrations, all donor (e.g., P and As in Si) or acceptor (e.g., B in Si) impurities that have shallow energy levels will be ionized. Under these conditions, the carrier concentration is equal to the impurity concentration. From these curves we can obtain the impurity concentration of a semiconductor if the resistivity is known, or vice versa.

► EXAMPLE 2

Find the room-temperature resistivity of an n -type silicon doped with 10^{16} phosphorus atoms/cm³.

SOLUTION At room temperature we assume that all donors are ionized; thus,

$$n \approx N_D = 10^{16} \text{ cm}^{-3}.$$

From Fig. 7 we find $\rho \approx 0.5 \Omega \cdot \text{cm}$. We can also calculate the resistivity from Eq. 15a:

$$\rho = \frac{1}{qn\mu_n} = \frac{1}{1.6 \times 10^{-19} \times 10^{16} \times 1300} = 0.48 \quad \Omega \cdot \text{cm}.$$

The mobility μ_n is obtained from Fig. 3.

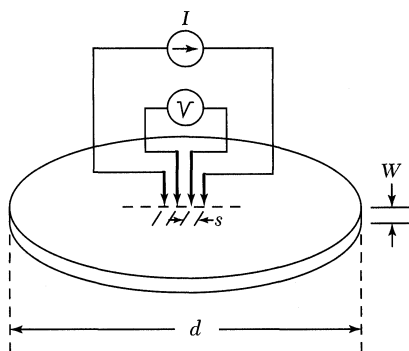


Fig. 6 Measurement of resistivity using a four-point probe.³

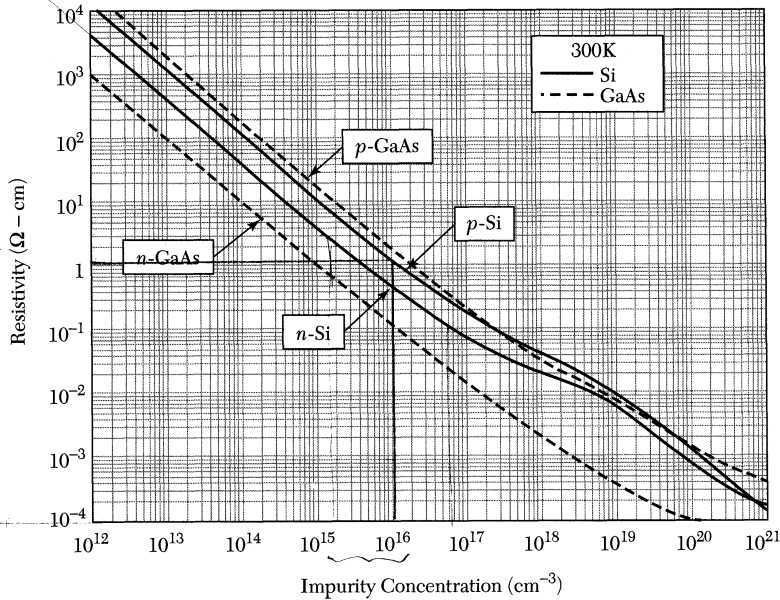


Fig. 7 Resistivity versus impurity concentration³ for Si and GaAs.

3.1.3 The Hall Effect

The carrier concentration in a semiconductor may be different from the impurity concentration, because the ionized impurity density depends on the temperature and the impurity energy level. To measure the carrier concentration directly, the most commonly used method is the Hall effect. Hall measurement is also one of the most convincing methods to show the existence of holes as charge carriers, because the measurement can give directly the carrier type. Figure 8 shows an electric field applied along the x -axis and a magnetic field applied along the z -axis. Consider a p -type semiconductor sample. The Lorentz force $q\mathbf{v} \times \mathbf{B}$ ($= qv_x B_z$) due to the magnetic field will exert an average upward force on the holes flowing in the x -direction. The upward directed current causes an

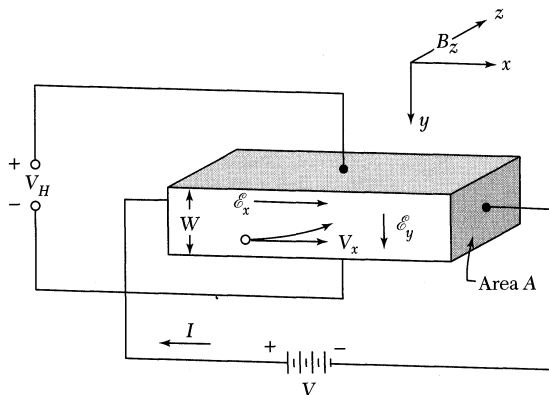


Fig. 8 Basic setup to measure carrier concentration using the Hall effect.

accumulation of holes at the top of the sample that gives rise to a downward-directed electric field $\mathcal{E}_y$. Since there is no net current flow along the y -direction in the steady state, the electric field along the y -axis exactly balances the Lorentz force; that is,

$$q\mathcal{E}_y = qv_x B_z, \quad (17)$$

or

$$\mathcal{E}_y = v_x B_z. \quad (18)$$

Once the electric field $\mathcal{E}_y$ becomes equal to $v_x B_z$, no net force along the y -direction is experienced by the holes as they drift in the x -direction.

The establishment of the electric field is known as the *Hall effect*. The electric field in Eq. 18 is called the *Hall field*, and the terminal voltage $V_H = \mathcal{E}_y W$ (Fig. 8) is called the *Hall voltage*. Using Eq. 12 for the hole drift velocity, the Hall field $\mathcal{E}_y$ in Eq. 18 becomes

$$\mathcal{E}_y = \left(\frac{J_p}{qp} \right) B_z = R_H J_p B_z, \quad (19)$$

where

$$R_H \equiv \frac{1}{qp}. \quad (20)$$

The Hall field $\mathcal{E}_y$ is proportional to the product of the current density and the magnetic field. The proportionality constant R_H is the *Hall coefficient*. A similar result can be obtained for an n -type semiconductor, except that the Hall coefficient is negative:

$$R_H = -\frac{1}{qn}. \quad (21)$$

A measurement of the Hall voltage for a known current and magnetic field yields

$$p = \frac{1}{qR_H} = \frac{J_p B_z}{q\mathcal{E}_y} = \frac{(I/A)B_z}{q(V_H/W)} = \frac{IB_z W}{qV_H A}, \quad (22)$$

where all the quantities in the right-hand side of the equation can be measured. Thus, the carrier concentration and carrier type can be obtained directly from the Hall measurement.

► EXAMPLE 3

A sample of Si is doped with 10^{16} phosphorus atoms/cm³. Find the Hall voltage in a sample with $W = 500 \mu\text{m}$, $A = 2.5 \times 10^{-3} \text{ cm}^2$, $I = 1 \text{ mA}$, and $B_z = 10^{-4} \text{ Wb/cm}^2$.

SOLUTION The Hall coefficient is

$$R_H = -\frac{1}{qn} = -\frac{1}{1.6 \times 10^{-19} \times 10^{16}} = -625 \text{ cm}^3/\text{C}.$$

The Hall voltage is

$$\begin{aligned} V_H &= \mathcal{E}_y W = \left(R_H \frac{I}{A} B_z \right) W \\ &= \left(-625 \cdot \frac{10^{-3}}{2.5 \times 10^{-3}} \cdot 10^{-4} \right) 500 \times 10^{-4} \\ &= -1.25 \text{ mV}. \end{aligned}$$

► 3.2 CARRIER DIFFUSION

3.2.1 Diffusion Process

In the preceding section, we considered the drift current, that is, the transport of carriers when an electric field is applied. Another important current component can exist if there is a spatial variation of carrier concentration in the semiconductor material. The carriers tend to move from a region of high concentration to a region of low concentration. This current component is called *diffusion current*.

To understand the diffusion process, let us assume an electron density that varies in the x -direction, as shown in Fig. 9. The semiconductor is at uniform temperature, so that the average thermal energy of electrons does not vary with x , only the density $n(x)$ varies. Consider the number of electrons crossing the plane at $x = 0$ per unit time and per unit area. Because of finite temperature, the electrons have random thermal motions with a thermal velocity v_{th} and a mean free path l (note that $l = v_{th} \tau_c$, where τ_c is the mean free time.) The electrons at $x = -l$, one mean free path away on the left side, have equal chances of moving left or right; and in a mean free time τ_c , one half of them will move across the plane $x = 0$. The average rate of electron flow per unit area F_1 of electrons crossing plane $x = 0$ from the left is then

$$F_1 = \frac{\frac{1}{2} n(-l) \cdot l}{\tau_c} = \frac{1}{2} n(-l) \cdot v_{th}. \quad (23)$$

Similarly, the average rate of electron flow per unit area F_2 of electrons at $x = l$ crossing plane $x = 0$ from the right is

$$F_2 = \frac{1}{2} n(l) \cdot v_{th}. \quad (24)$$

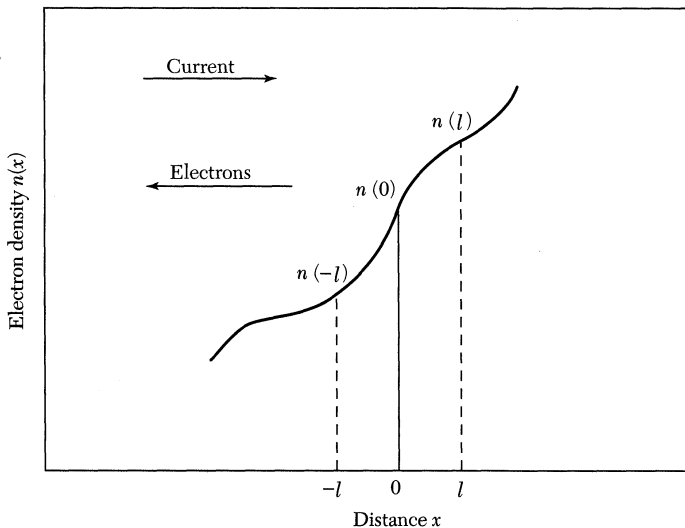


Fig. 9 Electron concentration versus distance; l is the mean free path. The directions of electron and current flows are indicated by arrows.

The net rate of carrier flow from left to right is

$$F = F_1 - F_2 = \frac{1}{2} v_{th} [n(-l) - n(l)]. \quad (25)$$

Approximating the densities at $x = \pm l$ by the first two terms of a Taylor series expansion, we obtain

$$\begin{aligned} F &= \frac{1}{2} v_{th} \left\{ n \left[(0) - l \frac{dn}{dx} \right] - \left[n(0) + l \frac{dn}{dx} \right] \right\} \\ &= -v_{th} l \frac{dn}{dx} \equiv -D_n \frac{dn}{dx}, \end{aligned} \quad (26)$$

where $D_n \equiv v_{th} l$ is called the *diffusion coefficient* also called the *diffusivity*. Because each electron carries a charge $-q$, the carrier flow gives rise to a current

$$J_n = -qF = qD_n \frac{dn}{dx}. \quad (27)$$

The diffusion current is proportional to the spatial derivative of the electron density. Diffusion current results from the random thermal motion of carriers in a concentration gradient. For an electron density that increases with x , the gradient is positive, and the electrons will diffuse toward the negative x -direction. The current is positive and flows in the direction opposite to that of the electrons as indicated in Fig. 9.

► EXAMPLE 4

Assume that, in an n -type semiconductor at $T = 300$ K, the electron concentration varies linearly from 1×10^{18} to $7 \times 10^{17} \text{ cm}^{-3}$ over a distance of 0.1 cm. Calculate the diffusion current density if the electron diffusion coefficient is $D_n = 22.5 \text{ cm}^2/\text{s}$.

SOLUTION The diffusion current density is given by

$$\begin{aligned} J_{n, \text{diff}} &= qD_n \frac{dn}{dx} \approx qD_n \frac{\Delta n}{\Delta x} \\ &= (1.6 \times 10^{-19})(22.5) \left(\frac{1 \times 10^{18} - 7 \times 10^{17}}{0.1} \right) = 10.8 \text{ A/cm}^2. \end{aligned}$$

3.2.2 Einstein Relation

Equation 27 can be written in a more useful form using the theorem for the equipartition of energy for this one-dimensional case. We can write

$$\frac{1}{2} m_n v_{th}^2 = \frac{1}{2} kT. \quad (28)$$

From Eqs. 3, 26, and 28 and using the relationship $l = v_{th} \tau_c$, we obtain

$$D_n = v_{th} l = v_{th} (v_{th} \tau_c) = v_{th}^2 \left(\frac{\mu_n m_n}{q} \right) = \left(\frac{kT}{m_n} \right) \left(\frac{\mu_n m_n}{q} \right), \quad (29)$$

or

$$D_n = \left(\frac{kT}{q} \right) \mu_n. \quad (30)$$

Equation 30 is known as the *Einstein relation*. It relates the two important constants (diffusivity and mobility) that characterize carrier transport by diffusion and by drift in a semiconductor. The Einstein relation also applies between D_p and μ_p . Values of diffusivities for silicon and gallium arsenide are shown in Fig. 3.

► EXAMPLE 5

Minority carriers (holes) are injected into a homogeneous n -type semiconductor sample at one point. An electric field of 50 V/cm is applied across the sample, and the field moves these minority carriers a distance of 1 cm in 100 μ s. Find the drift velocity and the diffusivity of the minority carriers.

SOLUTION

$$v_p = \frac{1 \text{ cm}}{100 \times 10^{-6} \text{ s}} = 10^4 \text{ cm/s};$$

$$\mu_p = \frac{v_p}{\mathcal{E}} = \frac{10^4}{50} = 200 \text{ cm}^2/\text{V}\cdot\text{s};$$

$$D_p = \frac{kT}{q} \mu_p = 0.0259 \times 200 = 5.18 \text{ cm}^2/\text{s}.$$

3.2.3 Current Density Equations

When an electric field is present in addition to a concentration gradient, both drift current and diffusion current will flow. The total current density at any point is the sum of the drift and diffusion components:

$$J_n = q\mu_n n \mathcal{E} + qD_n \frac{dn}{dx}, \quad (31)$$

where $\mathcal{E}$ is the electric field in the x -direction.

A similar expression can be obtained for the hole current:

$$J_p = q\mu_p p \mathcal{E} - qD_p \frac{dp}{dx}. \quad (32)$$

We use the negative sign in Eq. 32 because for a positive hole gradient the holes will diffuse in the negative x -direction. This diffusion results in a hole current that also flows in the negative x -direction.

The total conduction current density is given by the sum of Eqs. 31 and 32:

$$J_{\text{cond}} = J_n + J_p. \quad (33)$$

The three expressions (Eqs. 31–33) constitute the current density equations. These equations are important for analyzing device operations under low electric fields. However, at sufficiently high electric fields the terms $\mu_n \mathcal{E}$ and $\mu_p \mathcal{E}$ should be replaced by the saturation velocity v_s discussed in Section 3.7.

► 3.3 GENERATION AND RECOMBINATION PROCESSES

In thermal equilibrium the relationship $pn = n_i^2$ is valid. If excess carriers are introduced to a semiconductor so that $pn > n_i^2$, we have a *nonequilibrium situation*. The process of introducing excess carriers is called *carrier injection*. Most semiconductor devices operate by the creation of charge carriers in excess of the thermal equilibrium values. We can introduce excess carriers by optical excitation or forward-biasing a p – n junction (discussed in Chapter 4).

Whenever the thermal-equilibrium condition is disturbed (i.e., $pn \neq n_i^2$), processes exist to restore the system to equilibrium (i.e., $pn = n_i^2$). In the case of the injection of excess carriers, the mechanism that restores equilibrium is recombination of the injected minority carriers with the majority carriers. Depending on the nature of the recombination process, the released energy that results from the recombination process can be emitted as a photon or dissipated as heat to the lattice. When a photon is emitted, the process is called radiative recombination; otherwise, it is called nonradiative recombination.

Recombination phenomena can be classified as direct and indirect processes. Direct recombination, also called band-to-band recombination, usually dominates in direct-bandgap semiconductors, such as gallium arsenide, whereas indirect recombination via bandgap recombination centers dominates in indirect bandgap semiconductors, such as silicon.

3.3.1 Direct Recombination

Consider a direct-bandgap semiconductor in thermal equilibrium. The continuous thermal vibration of lattice atoms causes some bonds between neighboring atoms to be broken. When a bond is broken, an electron-hole pair is generated. In terms of the band diagram, the thermal energy enables a valence electron to make an upward transition to the conduction band, leaving a hole in the valence band. This process is called carrier generation and is represented by the generation rate G_{th} (number of electron-hole pairs generated per cm^3 per second) in Fig. 10a. When an electron makes a transition downward from the conduction band to the valence band, an electron-hole pair is annihilated. This reverse process is called recombination; it is represented by the recombination rate R_{th} in Fig. 10a. Under thermal equilibrium conditions, the generation rate G_{th} must equal

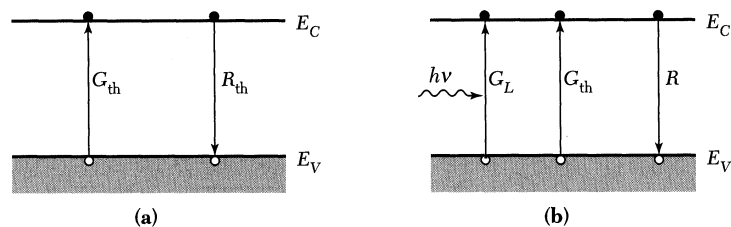


Fig. 10 Direct generation and recombination of electron-hole pairs: (a) at thermal equilibrium and (b) under illumination.

the recombination rate R_{th} , so that the carrier concentrations remain constant and the condition $pn = n_i^2$ is maintained.

When excess carriers are introduced to a direct-bandgap semiconductor, the probability is high that electrons and holes will recombine directly, because the bottom of the conduction band and the top of the valence band are lined up and no additional momentum is required for the transition across the bandgap. The rate of the direct recombination R is expected to be proportional to the number of electrons available in the conduction band and the number of holes available in the valence band; that is,

$$R = \beta np, \quad (34)$$

where β is the proportionality constant. As discussed previously, in thermal equilibrium the recombination rate must be balanced by the generation rate. Therefore, for an n -type semiconductor, we have

$$G_{th} = R_{th} = \beta n_{no} p_{no}. \quad (35)$$

In this notation for carrier concentrations the first subscript refers to the type of the semiconductor. The subscript o indicates an equilibrium quantity. The n_{no} and p_{no} represent electron and hole densities, respectively, in an n -type semiconductor at thermal equilibrium. When we shine a light on the semiconductor to produce electron-hole pairs at a rate G_L (Fig. 10b), the carrier concentrations are above their equilibrium values. The recombination and generation rate become

$$R = \beta n_n p_n = \beta(n_{no} + \Delta n)(p_{no} + \Delta p), \quad (36)$$

$$G = G_L + G_{th}, \quad (37)$$

where Δn and Δp are the excess carrier concentrations, given by

$$\Delta n = n_n - n_{no} \quad (38a)$$

$$\Delta p = p_n - p_{no}, \quad (38b)$$

and $\Delta n = \Delta p$ to maintain overall charge neutrality.

The net rate of change of hole concentration is given by

$$\frac{dp_n}{dt} = G - R = G_L + G_{th} - R. \quad (39)$$

In steady state, $dp_n / dt = 0$. From Eq. 39 we have

$$G_L = R - G_{th} \equiv U, \quad (40)$$

where U is the net recombination rate. Substituting Eqs. 35 and 36 into Eq. 40 yields

$$U = \beta(n_{no} + p_{no} + \Delta p)\Delta p. \quad (41)$$

For low-level injection $\Delta p, p_{no} \ll n_{no}$, Eq. 41 is simplified to

$$U \equiv \beta n_{no} \Delta p = \frac{p_n - p_{no}}{\frac{1}{\beta n_{no}}}. \quad (42)$$

Therefore, the net recombination rate is proportional to excess minority carrier concentration. Obviously, $U = 0$ in thermal equilibrium. The proportionality constant $1/\beta n_{no}$ is called the *lifetime* τ_p of the excess minority carriers, or

$$U \equiv \frac{p_n - p_{no}}{\tau_p}, \quad (43)$$

where

$$\tau_p \equiv \frac{1}{\beta n_{no}}. \quad (44)$$

The physical meaning of lifetime can best be illustrated by the transient response of a device after the sudden removal of the light source. Consider an n -type sample, as shown in Fig. 11a, that is illuminated with light and in which the electron-hole pairs are generated uniformly throughout the sample with a generation rate G_L . The time-dependent expression is given by Eq. 39. In steady state, from Eqs. 40 and 43

$$G_L = U = \frac{p_n - p_{no}}{\tau_p} \quad (45)$$

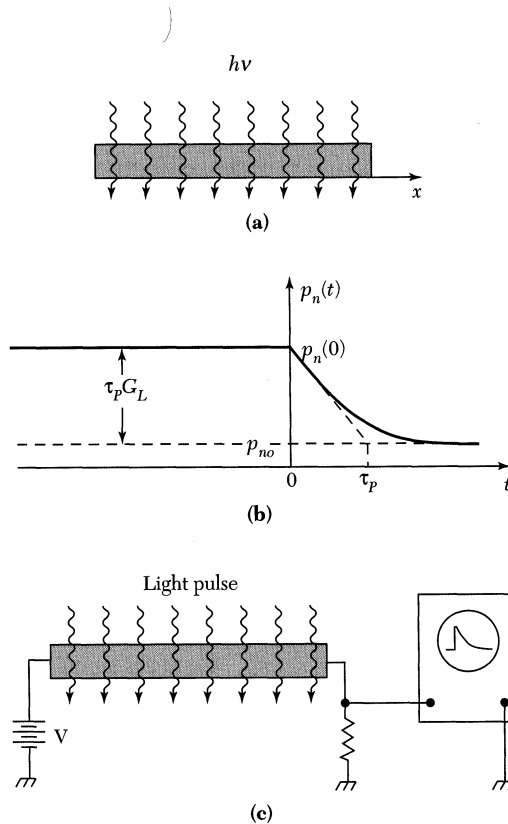


Fig. 11 Decay of photoexcited carriers. (a) n -type sample under constant illumination. (b) Decay of minority carriers (holes) with time. (c) Schematic setup to measure minority carrier lifetime.

or

$$p_n = p_{no} + \tau_p G_L. \quad (45a)$$

If at an arbitrary time, say $t = 0$, the light is suddenly turned off, the boundary conditions are $p_n(t = 0) = p_{no} + \tau_p G_L$, as given by Eq. 45a, and $p_n(t \rightarrow \infty) = p_{no}$. The time-dependent expression of Eq. 39 becomes

$$\frac{dp_n}{dt} = G_{th} - R = -U = -\frac{p_n - p_{no}}{\tau_p} \quad (46)$$

and the solution is

$$p_n(t) = p_{no} + \tau_p G_L \exp(-t/\tau_p). \quad (47)$$

Figure 11b shows the variation of p_n with time. The minority carriers recombine with majority carriers and decay exponentially with a time constant τ_p , which corresponds to the lifetime defined in Eq. 44.

This case illustrates the main idea of measuring the carrier lifetime using the photoconductivity method. Figure 11c shows a schematic setup. The excess carriers, generated uniformly throughout the sample by the light pulse, cause a momentary increase in the conductivity. The increase in conductivity manifests itself by a drop in voltage across the sample when a constant current is passed through it. The decay of the conductivity can be observed on an oscilloscope and is a measure of the lifetime of the excess minority carriers.

► EXAMPLE 6

A Si sample with $n_{no} = 10^{14} \text{ cm}^{-3}$ is illuminated with light and 10^{13} electron-hole pairs/cm³ are created every microsecond. If $\tau_n = \tau_p = 2 \text{ } \mu\text{s}$, find the change in the minority carrier concentration.

SOLUTION Before illumination

$$p_{no} = n_i^2 / n_{no} = (9.65 \times 10^9)^2 / 10^{14} \approx 9.31 \times 10^5 \text{ cm}^{-3}.$$

After illumination

$$p_n = p_{no} + \tau_p G_L = 9.31 \times 10^5 + 2 \times 10^{-6} \times \frac{10^{13}}{1 \times 10^{-6}} \approx 2 \times 10^{13} \text{ cm}^{-3}. \quad \blacktriangleleft$$

3.3.2 Indirect Recombination

For indirect-bandgap semiconductors, such as silicon, a direct recombination process is very unlikely, because the electrons at the bottom of the conduction band have nonzero momentum with respect to the holes at the top of the valence band (see Chapter 2). A direct transition that conserves both energy and momentum is not possible without a simultaneous lattice interaction. Therefore the dominant recombination process in such semiconductors is indirect transition via localized energy states in the forbidden energy gap.⁴ These states act as stepping stones between the conduction band and the valence band.

Figure 12 shows various transitions that occur in the recombination process through intermediate-level states (also called recombination centers). We illustrate the charging state of the center before and after each of the four basic transitions takes place. The arrows in the figure designate the transition of the electron in a particular process. The illustration is for the case of a recombination center with a single energy level that is neutral

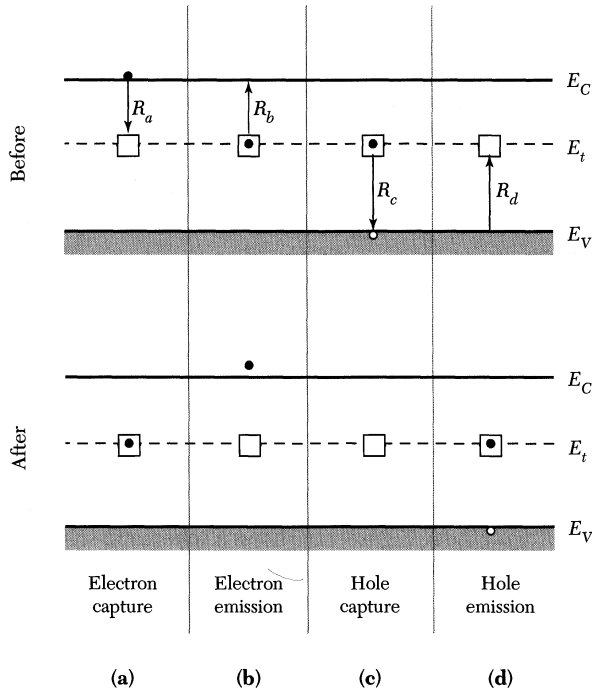


Fig. 12 Indirect generation-recombination processes at thermal equilibrium.

when not occupied by an electron or negative when it is occupied. In indirect recombination, the derivation of the recombination rate is more complicated. The detailed derivation is given in Appendix I, and the recombination rate is given by ⁴

$$U = \frac{v_{th} \sigma_n \sigma_p N_t (p_n n_n - n_i^2)}{\sigma_p [p_n + n_i e^{(E_t - E_i)/kT}] + \sigma_n [n_n + n_i e^{(E_t - E_i)/kT}]}, \quad (48)$$

where v_{th} is the thermal velocity of carriers given in Eq. 1, N_t is the concentration of the recombination center in the semiconductor, and σ_n is the capture cross section of electrons. The quantity σ_n describes the effectiveness of the center to capture an electron and is a measure of how close the electron has to come to the center to be captured. σ_p is the capture cross section of holes.

We can simplify the general expression for the dependence of U on E_t by assuming equal electron and hole capture cross sections, that is, $\sigma_n = \sigma_p = \sigma_o$. Equation 48 then becomes

$$U = v_{th} \sigma_o N_t \frac{(p_n n_n - n_i^2)}{p_n + n_n + 2n_i \cosh\left(\frac{E_t - E_i}{kT}\right)}. \quad (49)$$

Under a low-injection condition in an n -type semiconductor so that $n_n \gg p_n$, the recombination rate can be written as

$$U \approx v_{th} \sigma_o N_t \frac{p_n - p_{no}}{1 + \left(\frac{2n_i}{n_{no}} \right) \cosh \left(\frac{E_t - E_i}{kT} \right)} = \frac{p_n - p_{no}}{\tau_p}. \quad (50)$$

The recombination rate for indirect recombination is given by the same expression as Eq. 43; however, τ_p depends on the locations of the recombination centers.

3.3.3 Surface Recombination

Figure 13 shows schematically the bonds at a semiconductor surface.⁵ Because of the abrupt discontinuity of the lattice structure at the surface, a large number of localized energy states or generation-recombination centers may be introduced at the surface region. These energy states, called *surface states*, may greatly enhance the recombination rate at the surface region. The kinetics of surface recombination are similar to those considered before for bulk centers. The total number of carriers recombining at the surface *per unit area* and unit time can be expressed in a form analogous to Eq. 48. For a low-injection condition, and for the limiting case where electron concentration at the surface is essentially equal to the bulk majority carrier concentration, the total number of carriers recombining at the surface per unit area and unit time can be simplified to

$$U_s \equiv v_{th} \sigma_p N_{st} (p_s - p_{no}), \quad (51)$$

where p_s denotes the hole concentrations at the surface, and N_{st} is the recombination center density per unit area in the surface region. Since the product $v_{th} \sigma_p N_{st}$ has its dimension in centimeters per second, it is called the *low-injection surface recombination velocity* S_{lr} :

$$S_{lr} \equiv v_{th} \sigma_p N_{st}. \quad (52)$$

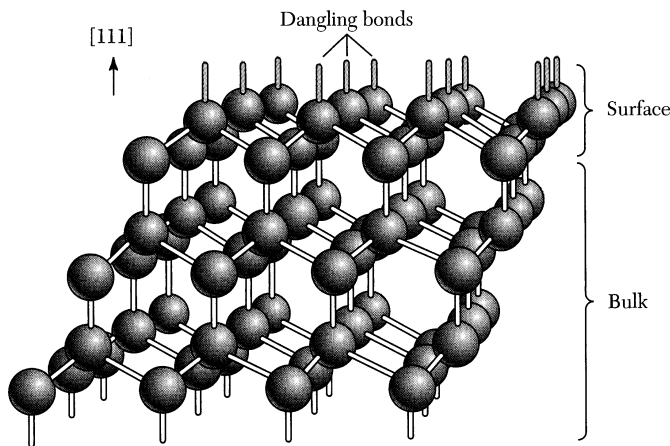


Fig. 13 Schematic diagram of bonds at a clean semiconductor surface. The bonds are anisotropic and differ from those in the bulk.⁵

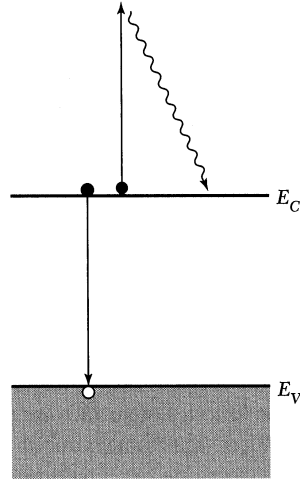


Fig. 14 Auger recombination.

3.3.4 Auger Recombination

The Auger recombination process occurs by the transfer of the energy and momentum released by the recombination of an electron-hole pair to a third particle that can be either an electron or a hole. The example of Auger recombination process is shown in Fig. 14. A second electron in the conduction band absorbs the energy released by the direct recombination. After the Auger process, the second electron becomes an energetic electron. It loses its energy to the lattice by scattering events. Usually, Auger recombination is important when the carrier concentration is very high as a result of either high doping or high injection level. Because the Auger process involves three particles, the rate of Auger recombination can be expressed as

$$R_{Aug} = Bn^2p \text{ or } Bnp^2. \quad (53)$$

The proportionality constant B has a strong temperature dependence.

► 3.4 CONTINUITY EQUATION

In the previous sections we considered individual effects such as drift due to an electric field, diffusion due to a concentration gradient, and recombination of carriers through intermediate-level recombination centers. We now consider the overall effect when drift, diffusion, and recombination occur simultaneously in a semiconductor material. The governing equation is called the *continuity equation*.

To derive the one-dimensional continuity equation for electrons, consider an infinitesimal slice with a thickness dx located at x shown in Fig. 15. The number of electrons in the slice may increase due to the *net* current flow into the slice and the *net* carrier generation in the slice. The overall rate of electron increase is the algebraic sum of four components: the number of electrons flowing into the slice at x , minus the number of electrons flowing out at $x + dx$, plus the rate at which electrons are generated, minus the rate at which they are recombined with holes in the slice.

The first two components are found by dividing the currents at each side of the slice by the charge of an electron. The generation and recombination rates are designated by

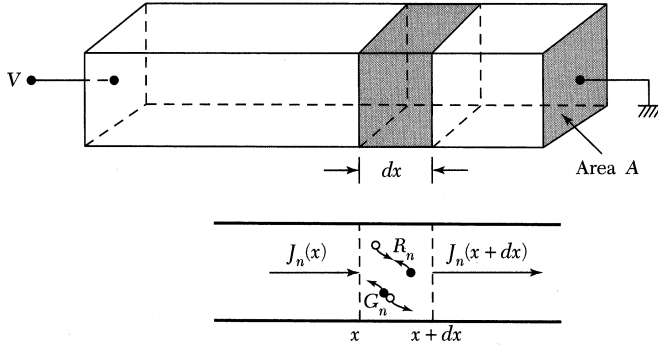


Fig. 15 Current flow and generation-recombination processes in an infinitesimal slice of thickness dx .

G_n and R_n , respectively. The overall rate of change in the number of electrons in the slice is then

$$\frac{\partial n}{\partial t} Adx = \left[\frac{J_n(x)A}{-q} - \frac{J_n(x+dx)A}{-q} \right] + (G_n - R_n)Adx, \quad (54)$$

where A is the cross-sectional area and Adx is the volume of the slice. Expanding the expression for the current at $x+dx$ in Taylor series yields

$$J_n(x+dx) = J_n(x) + \frac{\partial J_n}{\partial x} dx + \dots \quad (55)$$

We thus obtain the basic *continuity equation* for electrons:

$$\frac{\partial n}{\partial t} = \frac{1}{q} \frac{\partial J_n}{\partial x} + (G_n - R_n). \quad (56)$$

A similar continuity equation can be derived for holes, except that the sign of the first term on the right-hand side of Eq. 56 is changed because of the positive charge associated with a hole:

$$\frac{\partial p}{\partial t} = -\frac{1}{q} \frac{\partial J_p}{\partial x} + (G_p - R_p). \quad (57)$$

We can substitute the current expressions from Eqs. 31 and 32 and the recombination expressions from Eq. 43 into Eqs. 56 and 57. For the one-dimensional case under low-injection condition, the continuity equations for minority carriers (i.e., n_p in a p -type semiconductor or p_n in an n -type semiconductor) are

$$\frac{\partial n_p}{\partial t} = n_p \mu_n \frac{\partial \mathcal{E}}{\partial x} + \mu_n \mathcal{E} \frac{\partial n_p}{\partial x} + D_n \frac{\partial^2 n_p}{\partial x^2} + G_n - \frac{n_p - n_{p0}}{\tau_n}, \quad (58)$$

$$\frac{\partial p_n}{\partial t} = -p_n \mu_p \frac{\partial \mathcal{E}}{\partial x} - \mu_p \mathcal{E} \frac{\partial p_n}{\partial x} + D_p \frac{\partial^2 p_n}{\partial x^2} + G_p - \frac{p_n - p_{n0}}{\tau_p}. \quad (59)$$

In addition to the continuity equations, Poisson's equation

$$\frac{d\mathcal{E}}{dx} = \frac{\rho_s}{\epsilon_s} \quad (60)$$

must be satisfied, where ϵ_s is the semiconductor dielectric permittivity and ρ_s is the space charge density given by the algebraic sum of the charge carrier densities and the ionized impurity concentrations, $q(p - n + N_D^+ - N_A^-)$.

In principle, Eqs. 58 through 60 together with appropriate boundary conditions have a unique solution. Because of the algebraic complexity of this set of equations, in most cases the equations are simplified with physical approximations before a solution is attempted. We solve the continuity equations for three important cases.

3.4.1 Steady-State Injection from One Side

Figure 16a shows an *n*-type semiconductor where excess carriers are injected from one side as a result of illumination. It is assumed that light penetration is negligibly small (i.e., the assumptions of zero field and zero generation for $x > 0$). At steady state there is a

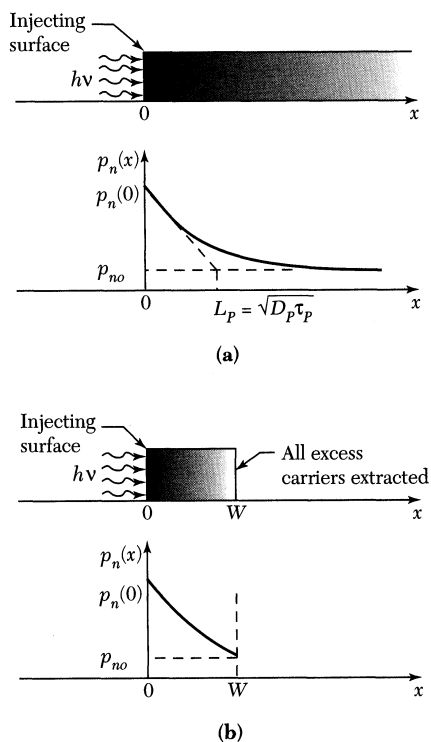


Fig. 16 Steady-state carrier injection from one side. (a) Semiinfinite sample. (b) Sample with thickness W .

concentration gradient near the surface. From Eq. 59 the differential equation for the minority carriers inside the semiconductor is

$$\frac{\partial p_n}{\partial t} = 0 = D_p \frac{\partial^2 p_n}{\partial x^2} - \frac{p_n - p_{no}}{\tau_p} \quad (61)$$

The boundary conditions are $p_n(x=0) = p_n(0) = \text{constant value}$ and $p_n(x \rightarrow \infty) = p_{no}$. The solution of $p_n(x)$ is

$$p_n(x) = p_{no} + [p_n(0) - p_{no}]e^{-x/L_p} \quad (62)$$

The length L_p is equal to $\sqrt{D_p \tau_p}$ and is called the *diffusion length*. Figure 16a shows the variation of the minority carrier density, which decays with a characteristic length given by L_p .

If we change the second boundary condition as shown in Fig. 16b so that all excess carriers at $x = W$ are extracted, that is, $p_n(W) = p_{no}$, then we obtain a new solution for Eq. 61:

$$p_n(x) = p_{no} + [p_n(0) - p_{no}] \left[\frac{\sinh\left(\frac{W-x}{L_p}\right)}{\sinh(W/L_p)} \right] \quad (63)$$

The current density at $x = W$ is given by the diffusion current expression, Eq. 32 with $\mathcal{E} = 0$:

$$J_p = -qD_p \left. \frac{\partial p_n}{\partial x} \right|_W = q[p_n(0) - p_{no}] \frac{D_p}{L_p} \frac{1}{\sinh(W/L_p)} \quad (64)$$

3.4.2 Minority Carriers at the Surface

When surface recombination is introduced at one end of a semiconductor sample under illumination (Fig. 17), the hole current density flowing into the surface from the bulk of

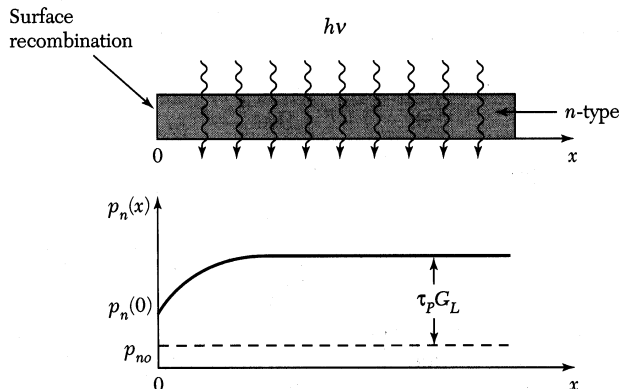


Fig. 17 Surface recombination at $x = 0$. The minority carrier distribution near the surface is affected by the surface recombination velocity.⁶

the semiconductor is given by qU_s . In this example, it is assumed that the sample is uniformly illuminated with uniform generation of carriers. The surface recombination leads to a lower carrier concentration at the surface. This gradient of hole concentration yields a diffusion current density that is equal to the surface recombination current. Therefore, the boundary condition at $x = 0$ is

$$qD_p \left. \frac{dp_n}{dx} \right|_{x=0} = qU_s = qS_{lr}[p_n(0) - p_{no}]. \quad (65)$$

The boundary condition at $x = \infty$ is given by Eq. 45a. At steady state the differential equation is

$$\frac{\partial p_n}{\partial t} = 0 = D_p \frac{\partial^2 p_n}{\partial x^2} + G_L - \frac{p_n - p_{no}}{\tau_p}. \quad (66)$$

The solution of the equation, subject to the boundary conditions above, is⁶

$$p_n(x) = p_{no} + \tau_p G_L \left(1 - \frac{\tau_p S_{lr} e^{-x/L_p}}{L_p + \tau_p S_{lr}} \right). \quad (67)$$

A plot of this equation for a finite S_{lr} is shown in Fig. 17. When $S_{lr} \rightarrow 0$, then $p_n(x) \rightarrow p_{no} + \tau_p G_L$, as obtained previously (Eq. 45a). When $S_{lr} \rightarrow \infty$, then

$$p_n(x) = p_{no} + \tau_p G_L (1 - e^{-x/L_p}). \quad (68)$$

From Eq. 68 we can see that at the surface the minority carrier density approaches its thermal equilibrium value p_{no} .

3.4.3 The Haynes–Shockley Experiment

One of the classic experiments in semiconductor physics is the demonstration of drift and diffusion of minority carriers, first made by Haynes and Shockley.⁷ The experiment allows independent measurement of the minority carrier mobility μ and diffusion coefficient D . The basic setup of the Haynes–Shockley experiment is shown in Fig. 18a. The voltage source, V_1 , establishes an electric field in the $+x$ direction in the n -type semiconductor bar. Excess carriers are produced and effectively injected into the semiconductor bar at contact (1) by a pulse. Contact (2) will collect a fraction of the excess carriers as they drift through the semiconductor. After a pulse, the transport equation is given by Eq. 59 by setting $G_p = 0$ and $\partial \mathcal{E} / \partial x = 0$ (i.e., the applied field is constant across the semiconductor bar):

$$\frac{\partial p_n}{\partial t} = -\mu_p \mathcal{E} \frac{\partial p_n}{\partial x} + D_p \frac{\partial^2 p_n}{\partial x^2} - \frac{p_n - p_{no}}{\tau_p}. \quad (69)$$

If no field is applied along the sample, the solution is given by

$$p_n(x, t) = \frac{N}{\sqrt{4\pi D_p t}} \exp\left(-\frac{x^2}{4D_p t} - \frac{t}{\tau_p}\right) + p_{no}, \quad (70)$$

where N is the number of electrons or holes generated per unit area. Figure 18b shows this solution as the carriers diffuse away from the point of injection and recombine.

If an electric field is applied along the sample, the solution is in the form of Eq. 70, except that x is replaced by $x - \mu_p \mathcal{E} t$ (Fig. 18c). Thus, all the excess carriers move with the drift velocity $\mu_p \mathcal{E}$. At the same time, the carriers diffuse outward and recombine as in the field-free case.

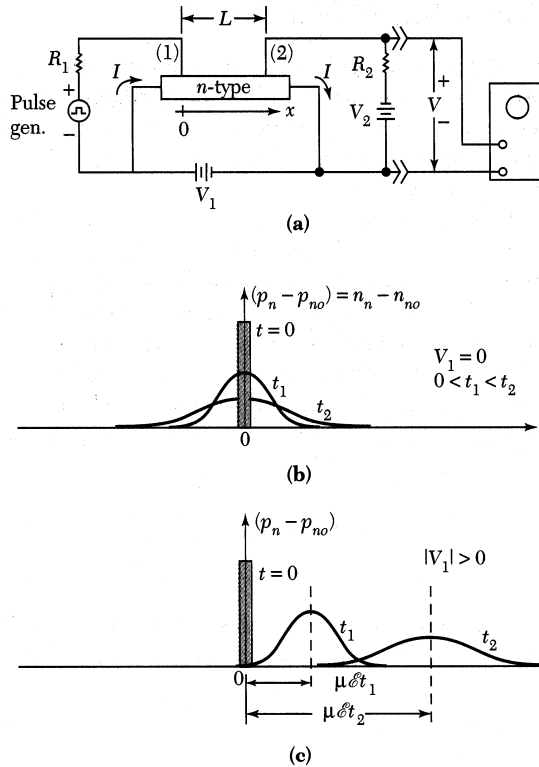


Fig. 18 The Haynes–Shockley experiment. (a) Experimental setup. (b) Carrier distributions without an applied field. (c) Carrier distributions with an applied field.⁷

EXAMPLE 7

In a Haynes–Shockley experiment, the maximum amplitudes of the minority carriers at $t_1 = 100 \mu\text{s}$ and $t_2 = 200 \mu\text{s}$ differ by a factor of 5. Calculate the minority carrier lifetime.

SOLUTION When an electric field is applied, the minority carrier distribution is given by

$$\Delta p \equiv p_n - p_{no} = \frac{N}{\sqrt{4\pi D_p t}} \exp\left(-\frac{(x - \mu_p \mathcal{E} t)^2}{4D_p t} - \frac{t}{\tau_p}\right).$$

At the maximum amplitude

$$\Delta p = \frac{N}{\sqrt{4\pi D_p t}} \exp\left(-\frac{t}{\tau_p}\right).$$

Therefore

$$\frac{\Delta p(t_1)}{\Delta p(t_2)} = \frac{\sqrt{t_2}}{\sqrt{t_1}} \frac{\exp(-t_1/\tau_p)}{\exp(-t_2/\tau_p)} = \sqrt{\frac{200}{100}} \exp\left(\frac{200 - 100}{\tau_p (\mu\text{s})}\right) = 5$$

$$\therefore \tau_p = \frac{200 - 100}{\ln(5/\sqrt{2})} = 79 \mu\text{s}.$$

► 3.5 THERMIONIC EMISSION PROCESS

In previous sections, we considered carrier transport phenomena inside the bulk semiconductor. At the semiconductor surface, carriers may recombine with the recombination centers due to the dangling bonds of the surface region. In addition, if the carriers have sufficient energy, they may be “thermionically” emitted into the vacuum. This is called the *thermionic emission process*.

Figure 19a shows the band diagram of an isolated n -type semiconductor. The electron affinity, $q\chi$, is the energy difference between the conduction band edge and the vacuum level in the semiconductor; and the work function, $q\phi_s$, is the energy between the Fermi level and the vacuum level in the semiconductor. From Fig. 19b, it is clear that an electron can be thermionically emitted into the vacuum if its energy is above $q\chi$.

The electron density with energies above $q\chi$ can be obtained from an expression similar to that for the electron density in the conduction band (Eqs. 9 and 16 of Chapter 2) except that the lower limit of the integration is $q\chi$ instead of E_C :

$$n_{th} = \int_{q\chi}^{\infty} n(E) dE = N_C \exp\left[-\frac{q(\chi + V_n)}{kT}\right], \quad (71)$$

where N_C is the effective density of states in the conduction band, and V_n is the difference between the bottom of the conduction band and the Fermi level.

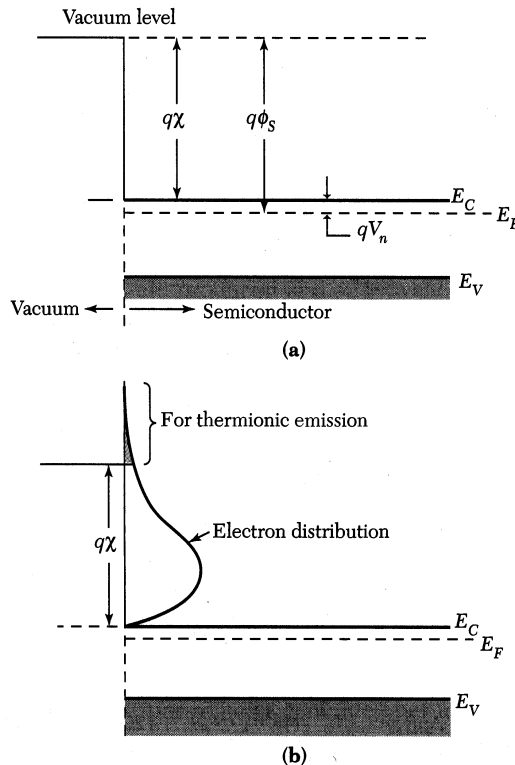


Fig. 19 (a) The band diagram of an isolated n -type semiconductor. (b) The thermionic emission process.

EXAMPLE 8

Calculate the thermionically emitted electron density, n_{th} , at room temperature for an n -type silicon sample with an electron affinity of $q\chi = 4.05$ eV and $qV_n = 0.2$ eV. If we reduce the effective $q\chi$ to 0.6 eV, what is n_{th} ?

$$\begin{aligned}
 \text{SOLUTION } n_{th}(4.05 \text{ eV}) &= 2.86 \times 10^{19} \exp\left(-\frac{4.05 + 0.2}{0.0259}\right) = 2.86 \times 10^{19} \exp(-164) \\
 &\cong 10^{-52} \approx 0 \\
 n_{th}(0.6 \text{ eV}) &= 2.86 \times 10^{19} \exp\left(-\frac{0.8}{0.0259}\right) = 2.86 \times 10^{19} \exp(-30.9) \\
 &= 1 \times 10^6 \text{ cm}^{-3}.
 \end{aligned}$$

From the above example, we see that at 300 K there is no emission of electrons into vacuum for $q\chi = 4.05$. However, if we can lower the effective electron affinity to 0.6 eV, a substantial number of electrons can be thermionically emitted. The thermionic emission process is of particular important for metal-semiconductor contacts to be considered in Chapter 7.

3.6 TUNNELING PROCESS

Figure 20a shows the energy band diagram when two isolated semiconductor samples are brought close together. The distance between them is d , and the potential barrier height qV_0 is equal to the electron affinity $q\chi$. If the distance is sufficient small, the electrons in the left-side semiconductor may transport across the barrier and move to the right-side semiconductor, even if the electron energy is much less than the barrier height. This process is associated with the *quantum tunneling phenomenon*.

Based on Fig. 20a, we have redrawn the one-dimensional potential barrier diagram in Fig. 20b. We first consider the transmission (or tunneling) coefficient of a particle (e.g., electron) through this barrier. In the corresponding classic case, the particle is always reflected if its energy E is less than the potential barrier height qV_0 . We see that in the quantum case, the particle has finite probability to transmit or “tunnel” through the potential barrier.

The behavior of a particle (e.g., a conduction electron) in the region where $qV(x) = 0$ can be described by the Schrödinger equation:

$$-\frac{\hbar^2}{2m_n} \frac{d^2\psi}{dx^2} = E\psi \quad (72)$$

or

$$\frac{d^2\psi}{dx^2} = -\frac{2m_n E}{\hbar^2} \psi, \quad (73)$$

where m_n is the effective mass, $\hbar$ is the reduced Planck constant, E is the kinetic energy, and ψ is the wave function of the particle. The solutions are

$$\psi(x) = Ae^{jkx} + Be^{-jkx} \quad x \leq 0, \quad (74)$$

$$\psi(x) = Ce^{jkx} \quad x \geq d, \quad (75)$$

where $k \equiv \sqrt{2m_n E/\hbar^2}$. For $x \leq 0$, we have an incident-particle wave function (with amplitude A) and a reflected wave function (with amplitude B); for $x \geq d$, we have a transmitted wave function (with amplitude C).

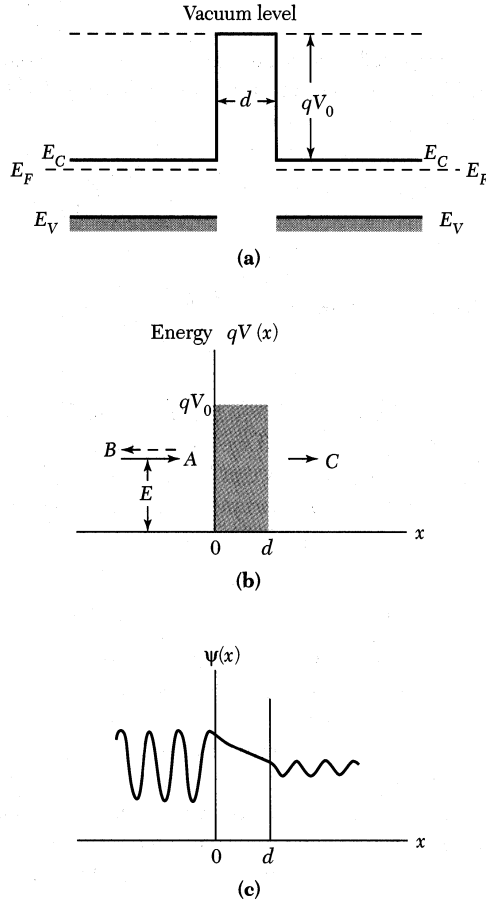


Fig. 20 (a) The band diagram of two isolated semiconductors with a distance d . (b) One-dimensional potential barrier. (c) Schematic representation of the wave function across the potential barrier.

Inside the potential barrier, the wave equation is given by

$$-\frac{\hbar^2}{2m_n} \frac{d^2\psi}{dx^2} + qV_0\psi = E\psi \quad (76)$$

or

$$\frac{d^2\psi}{dx^2} = \frac{2m_n(qV_0 - E)}{\hbar^2} \psi. \quad (77)$$

The solution for $E < qV_0$ is

$$\psi(x) = Fe^{\beta x} + Ge^{-\beta x}, \quad (78)$$

where $\beta \equiv \sqrt{2m_n(qV_0 - E)/\hbar^2}$. A schematic representation of the wave functions across the barrier is shown in Fig. 20c. The continuity of ψ and $d\psi/dx$ at $x = 0$ and $x = d$, which is required by the boundary conditions, provides four relations between the five coefficients (A, B, C, F, and G). We can solve for $(C/A)^2$, which is the *transmission coefficient*:

$$\left(\frac{C}{A}\right)^2 = \left[1 + \frac{(qV_0 \sinh \beta d)^2}{4E(qV_0 - E)}\right]^{-1} \quad (79)$$

The transmission coefficient decreases monotonically as E decreases. When $\beta d \gg 1$, the transmission coefficient becomes quite small and varies as

$$\left(\frac{C}{A}\right)^2 \sim \exp(-2\beta d) = \exp\left[-2d\sqrt{2m_n(qV_0 - E)/\hbar^2}\right] \quad (80)$$

To have a finite transmission coefficient, we require a small tunneling distance d , a low-potential barrier qV_0 , and a small effective mass. These results will be used for tunnel diodes in Chapter 8.

3.7 HIGH-FIELD EFFECTS

At low electric-fields, the drift velocity is linearly proportional to the applied field. We assume that the time interval between collision, τ_c , is independent of the applied field. This is a reasonable assumption as long as the drift velocity is small compared with the thermal velocity of carriers, which is about 10^7 cm/s for silicon at room temperature.

As the drift velocity approaches the thermal velocity, its field dependence on the electric field will begin to depart from the linear relationship given in Section 3.1. Figure 21 shows the measured drift velocities of electrons and holes in silicon as a function of the electric field. It is apparent that initially the field dependence of the drift velocity is linear, corresponding to a constant mobility. As the electric field is further increased, the drift velocity increases less rapidly. At sufficiently large fields, the drift velocity approaches a saturation velocity. The experimental results can be approximated by the empirical expression⁸

$$v_n, v_p = \frac{v_s}{[1 + (\mathcal{E}_0 / \mathcal{E})^\gamma]^{1/\gamma}} \quad (81)$$

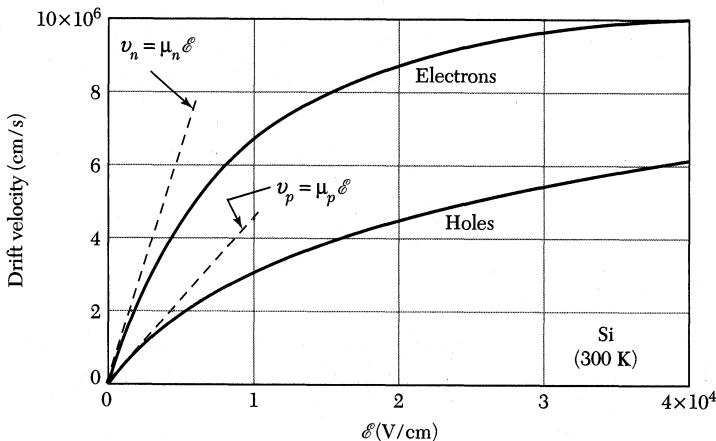


Fig. 21 Drift velocity versus electric field in Si.⁸

where v_s is the saturation velocity (10^7 cm/s for Si at 300 K), $\mathcal{E}_0$ is a constant, equal to 7×10^3 V/cm for electrons and 2×10^4 V/cm for holes in high-purity silicon materials, and γ is 2 for electrons and 1 for holes. Velocity saturation at high fields is particularly likely for field-effect transistors (FETs) with very short channels. Even moderate voltage can result in a high field along the channel. This effect is discussed in Chapter 6.

The high-field transport in n -type gallium arsenide is different from that of silicon.⁹ Figure 22 shows the measured drift velocity versus field for n -type and p -type gallium arsenide. The results for silicon are also shown in this log-log plot for comparison. Note that for n -type GaAs, the drift velocity reaches a maximum, then decreases as the field further increases. This phenomenon is due to the energy band structure of gallium arsenide that allows the transfer of conduction electrons from a high-mobility energy minimum (called a valley) to low-mobility, higher-energy satellite valleys, that is, electron transfer from the central valley to the satellite valleys along the [111] direction shown in Chapter 2.

To understand this phenomenon, consider the simple two-valley model of n -type gallium arsenide shown in Fig. 23. The energy separation between the two valleys is $\Delta E = 0.31$ eV. The lower valley's electron effective mass is denoted by m_1 , the electron mobility by μ_1 , and the electron density by n_1 . The upper-valley quantities are denoted by m_2 , μ_2 , and n_2 , respectively; and the total electron concentration is given by $n = n_1 + n_2$. The steady-state conductivity of the n -type GaAs can be written as

$$\sigma = q(\mu_1 n_1 + \mu_2 n_2) = qn\bar{\mu}, \quad (82)$$

where the average mobility is

$$\bar{\mu} \equiv (\mu_1 n_1 + \mu_2 n_2) / (n_1 + n_2). \quad (83)$$

The drift velocity is then

$$v_n = \bar{\mu} \mathcal{E}. \quad (84)$$

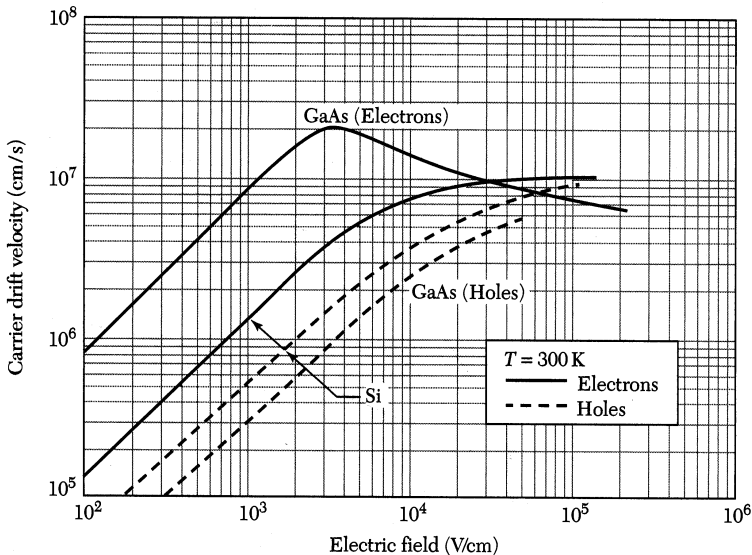


Fig. 22 Drift velocity versus electric field in Si and GaAs. Note that for n -type GaAs, there is a region of negative differential mobility.^{8,9}

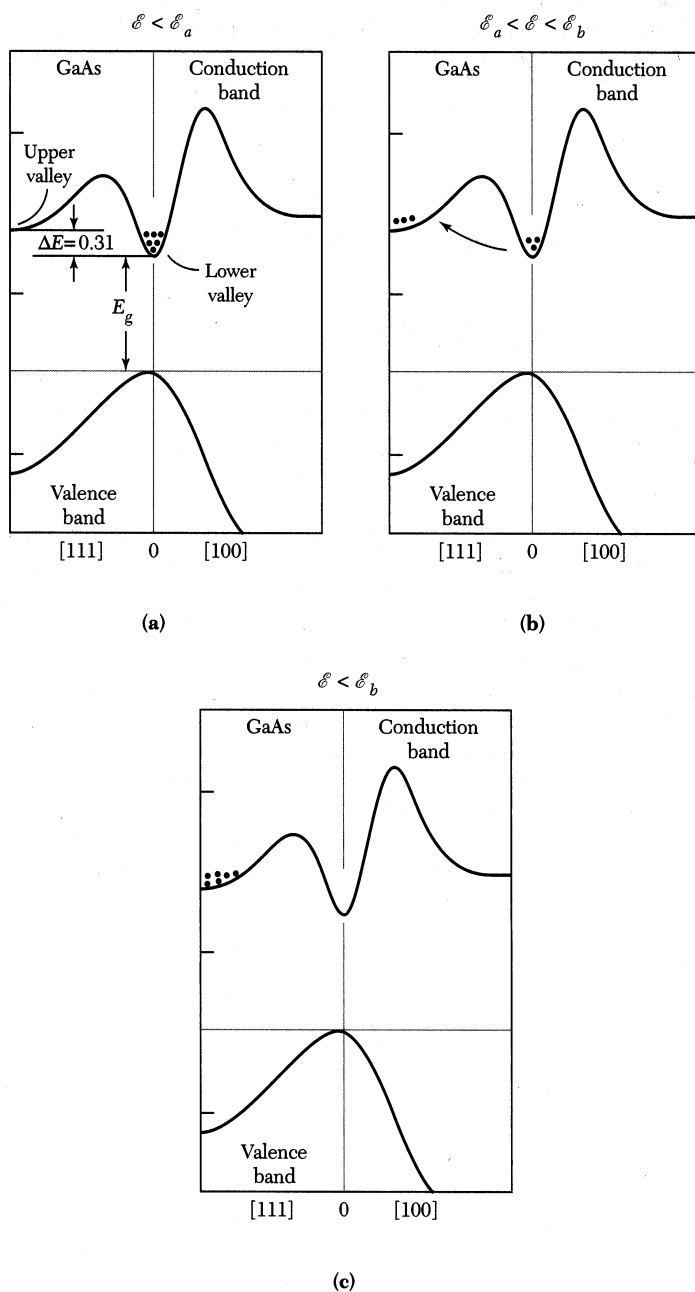


Fig. 23 Electron distributions under various conditions of electric fields for a two-valley semiconductor.

For simplicity we make the following assignments for the electron concentrations in the various ranges of electric-field values illustrated in Fig. 23. In Fig. 23a, the field is low and all electrons remain in the lower valley. In Fig. 23b, the field is higher and some electrons gain sufficient energies from the field to move to the higher valley. In Fig. 23c, the field is high enough to transfer all electrons to the higher valley. Thus, we have

$$\begin{aligned} n_1 &\equiv n & \text{and} & & n_2 &\equiv 0 & \text{for } 0 < \mathcal{E} < \mathcal{E}_a, \\ n_1 + n_2 &\equiv n & & & & & \text{for } \mathcal{E}_a < \mathcal{E} < \mathcal{E}_b, \\ n_1 &\equiv 0 & \text{and} & & n_2 &\equiv n & \text{for } \mathcal{E} > \mathcal{E}_b. \end{aligned} \quad (85)$$

Using these relations, the effective drift velocity takes on the asymptotic values

$$\begin{aligned} v_n &\equiv \mu_1 \mathcal{E} & \text{for } 0 < \mathcal{E} < \mathcal{E}_a, \\ v_n &\equiv \mu_2 \mathcal{E} & \text{for } \mathcal{E} > \mathcal{E}_b. \end{aligned} \quad (86)$$

If $\mu_1 \mathcal{E}_a$ is larger than $\mu_2 \mathcal{E}_b$, there is a region in which the drift velocity decreases with an increasing field between $\mathcal{E}_a$ and $\mathcal{E}_b$, as shown in Fig. 24. Because of the characteristics of the drift velocity in *n*-type gallium arsenide, this material is used in microwave transferred-electron devices discussed in Chapter 8.

When the electric field in a semiconductor is increased above a certain value, the carriers gain enough kinetic energy to generate electron-hole pairs by an *avalanche process* that is shown schematically in Fig. 25. Consider an electron in the conduction band (designated by 1). If the electric field is high enough, this electron can gain kinetic energy before it collides with the lattice. On impact with the lattice, the electron imparts most of its kinetic energy to break a bond, that is, to ionize a valence electron from the valence band to the conduction band and thereby generate an electron-hole pair (designated by 2 and 2'). Similarly, the generated pair now begins to accelerate in the field and collides with the lattice as indicated in the figure. In turn, they will generate other electron-hole pairs (e.g., 3 and 3', 4 and 4'), and so on. This process is called the *avalanche process*; it is also referred to as the *impact ionization* process. This process will result in breakdown in *p-n* junction, which is discussed in Chapter 4.

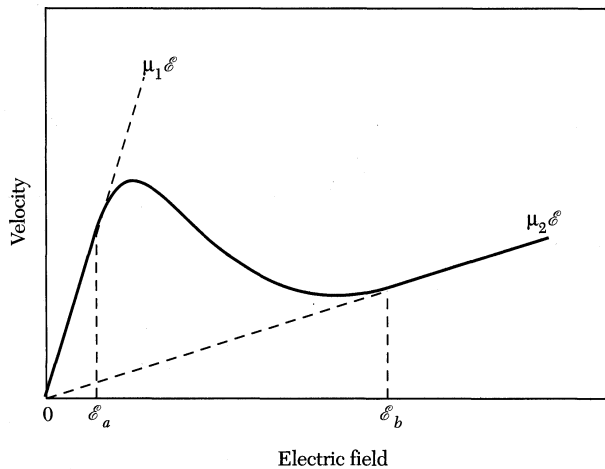


Fig. 24 One possible velocity-field characteristic of a two-valley semiconductor.

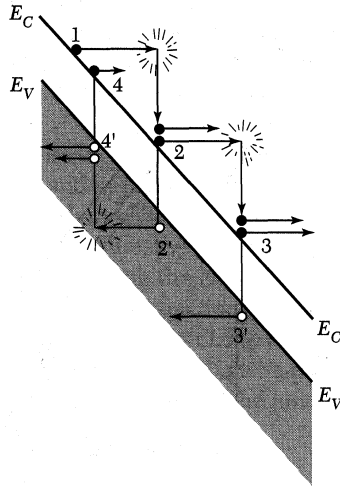


Fig. 25 Energy band diagram for the avalanche process.

To gain some ideas about the ionization energy involved, let us consider the process leading to 2–2' shown in Fig. 25. Just prior to the collision, the fast-moving electron (no. 1) has a kinetic energy $\frac{1}{2}m_1v_s^2$ and a momentum m_1v_s , where m_1 is the effective mass and v_s is the saturation velocity. After collision, there are three carriers: the original electron plus an electron–hole pair (no. 2 and no. 2'). If we assume that the three carriers have the same effective mass, the same kinetic energy, and the same momentum, the total kinetic energy is $\frac{3}{2}m_1v_f^2$, and the total momentum is $3m_1v_f$, where v_f is the velocity after collision. To conserve both energy and momentum before and after the collision, we require that

$$\frac{1}{2}m_1v_s^2 = E_g + \frac{3}{2}m_1v_f^2 \quad (87)$$

and

$$m_1v_s = 3m_1v_f, \quad (88)$$

where in Eq. 87 the energy E_g is the bandgap corresponding to the minimum energy required to generate an electron–hole pair. Substituting Eq. 88 into Eq. 87 yields the required kinetic energy for the ionization process:

$$E_0 = \frac{1}{2}m_1v_s^2 = 1.5E_g. \quad (89)$$

It is obvious that E_0 must be larger than the bandgap for the ionization process to occur. The actual energy required depends on the band structure. For silicon, the value for E_0 is 3.6 eV ($3.2E_g$) for electrons and 5.0 eV ($4.4E_g$) for holes.

The number of electron–hole pairs generated by an electron per unit distance traveled is called the *ionization rate* for the electron, α_n . Similarly, α_p is the ionization rate for the holes. The measured ionization rates for silicon and gallium arsenide are shown⁹ in Fig. 26. We note that both α_n and α_p are strongly dependent on the electric field. For a substantially large ionization rate (say 10^4 cm^{-1}), the corresponding electric field is $\geq 3 \times 10^5 \text{ V/cm}$ for silicon and $\geq 4 \times 10^5 \text{ V/cm}$ for gallium arsenide. The electron–hole pair generation rate G_A from the avalanche process is given by

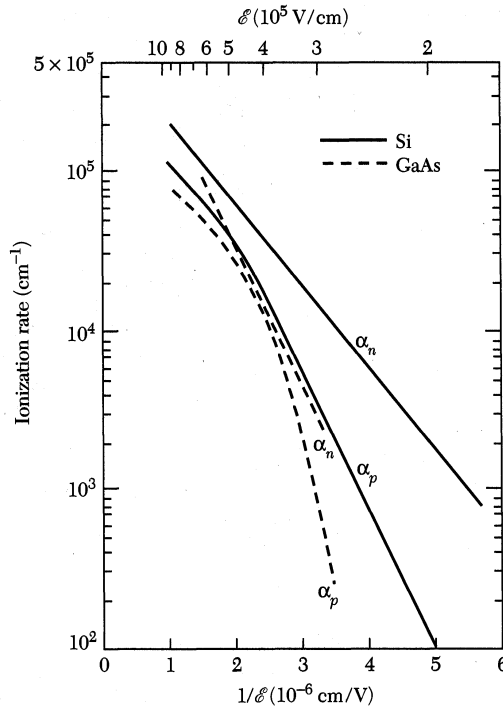


Fig. 26 Measured ionization rates versus reciprocal field for Si and GaAs.⁹

$$G_A = \frac{1}{q} (\alpha_n |J_n| + \alpha_p |J_p|), \quad (90)$$

where J_n and J_p are the electron and hole current densities, respectively. This expression can be used in the continuity equation for devices operated under an avalanche condition.

► SUMMARY

Various transport processes are at work in semiconductor devices. These include drift, diffusion, generation, recombination, thermionic emission, tunneling, and impact ionization.

One of the key transport processes is the carrier drift under the influence of an electric field. At low fields, the drift velocity is proportional to the electric field. This proportionality constant is called mobility. Another key transport process is the carrier diffusion under the influence of the carrier concentration gradient. The total current is the sum of the drift and diffusion components.

Excess carriers in a semiconductor cause a nonequilibrium condition. Most semiconductor devices operate under nonequilibrium conditions. Carriers can be generated by various means such as forward biasing a p - n junction, incident light, and impact ionization. The mechanism that restores equilibrium is the recombination of the excess minority carriers with the majority carriers by direct band-to-band recombination or via localized energy states in the forbidden energy gap. The governing equation for the rate of change of charge carriers is the continuity equation.

Among other transport processes, thermionic emission occurs when carriers in the surface region gain enough energy to be emitted into the vacuum level. Another, the tunneling process, is based on the quantum tunneling phenomena that results in the transport of electrons across a potential barrier even if the electron energy is less than the barrier height.

As the electric field becomes higher, the drift velocity departs from its linear relationship with the applied field and approaches a saturation velocity. This effect is particularly important in the study of short-channel field-effect transistors discussed in Chapter 6. When the field exceeds a certain value, the carriers gain enough kinetic energy to generate electron-hole pair by colliding with the lattice and breaking a bond. This effect is particularly important in the study of p - n junctions. The high field accelerates these new electron-hole pairs, which collide with the lattice to create more electron-hole pairs. As this process, called impact ionization or the avalanche process, continues, the p - n junction breaks down and conducts a large current. The junction breakdown is discussed in Chapter 4.

REFERENCES

1. R. A. Smith, *Semiconductors*, 2nd ed., Cambridge Univ. Press, London, 1978.
2. J. L. Moll, *Physics of Semiconductors*, McGraw-Hill, New York, 1964.
3. W. F. Beadle, J. C. C. Tsai, and R. D. Plummer, Eds., *Quick Reference Manual for Semiconductor Engineers*, Wiley, New York, 1985.
4. (a) R. N. Hall, "Electron-Hole Recombination in Germanium," *Phys. Rev.*, **87**, 387 (1952); (b) W. Shockley and W. T. Read, "Statistics of Recombination of Holes and Electrons," *Phys. Rev.*, **87**, 835 (1952).
5. M. Prutton, *Surface Physics*, 2nd ed., Clarendon, Oxford, 1983.
6. A. S. Grove, *Physics and Technology of Semiconductor Devices*, Wiley, New York, 1967.
7. J. R. Haynes and W. Shockley, "The Mobility and Life of Injected Holes and Electrons in Germanium," *Phys. Rev.*, **81**, 835 (1951).
8. D. M. Caughey and R. E. Thomas, "Carrier Mobilities in Silicon Empirically Related to Doping and Field," *Proc. IEEE*, **55**, 2192 (1967).
9. S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed., Wiley, New York, 1981.

PROBLEMS (* INDICATES DIFFICULT PROBLEMS)

FOR SECTION 3.1 CARRIER DRIFT

1. Find the resistivities of intrinsic Si and intrinsic GaAs at 300 K.
2. Assume that the mobility of electrons in silicon at $T = 300$ K is $\mu_n = 1300$ cm²/V-s. Also assume that the mobility is mainly limited by lattice scattering. Determine the electron mobility at (a) $T = 200$ K and (b) $T = 400$ K.
3. Two scattering mechanisms exist in a semiconductor. If only the first mechanism is present, the mobility will be 250 cm²/V-s. If only the second mechanism is present, the mobility will be 500 cm²/V-s. Determine the mobility when both scattering mechanisms exist at the same time.
4. Find the electron and hole concentrations, mobilities, and resistivities of silicon samples at 300 K, for each of the following impurity concentrations: (a) 5×10^{15} boron atoms/cm³; (b) 2×10^{16} boron atoms/cm³ and 1.5×10^{16} arsenic atoms/cm³; and (c) 5×10^{15} boron atoms/cm³, 10^{17} arsenic atoms/cm³, and 10^{17} gallium atoms/cm³.
- *5. Consider a compensated n -type silicon at $T = 300$ K, with a conductivity of $\sigma = 16$ (Ω -cm)⁻¹ and an acceptor doping concentration of 10^{17} cm⁻³. Determine the donor

concentration and the electron mobility. (A compensated semiconductor is one that contains both donor and acceptor impurity atoms in the same region.)

6. For a semiconductor with a constant mobility ratio $b \equiv \mu_n / \mu_p > 1$ independent of impurity concentration, find the maximum resistivity ρ_m in terms of the intrinsic resistivity ρ_i and the mobility ratio.
7. A four-point probe (with probe spacing of 0.5 mm) is used to measure the resistivity of a p -type silicon sample. Find the resistivity of the sample if its diameter is 200 μm and its thickness is 50 μm . The contact current is 1 mA, and the measured voltage between the inner two probes is 10 mV.
8. Given a silicon sample of unknown doping, Hall measurement provides the following information: $W = 0.05$ cm, $A = 1.6 \times 10^{-3}$ cm² (refer to Fig. 8), $I = 2.5$ mA, and the magnetic field is 30 nT ($1 \text{ T} = 10^{-4} \text{ Wb/cm}^2$). If a Hall voltage of +10 mV is measured, find the Hall coefficient, conductivity type, majority carrier concentration, resistivity, and mobility of the semiconductor sample.
9. A semiconductor is doped with N_D ($N_D \gg n_i$) and has a resistance R_1 . The same semiconductor is then doped with an unknown amount of acceptors N_A ($N_A \gg N_D$), yielding a resistance of $0.5 R_1$. Find N_A in terms of N_D if $D_n/D_p = 50$.
- *10. Consider a semiconductor that is nonuniformly doped with donor impurity atoms $N_D(x)$. Show that the induced electric field in the semiconductor in thermal equilibrium is given

$$\text{by } \mathcal{E}(x) = -\left(\frac{kT}{q}\right) \frac{1}{N_D(x)} \frac{dN_D(x)}{dx}.$$

FOR SECTION 3.2 CARRIER DIFFUSION

11. An intrinsic Si sample is doped with donors from one side such that $N_D = N_0 \exp(-ax)$. (a) Find an expression for the built-in field $\mathcal{E}(x)$ at equilibrium over the range for which $N_D \gg n_i$. (b) Evaluate $\mathcal{E}(x)$ when $a = 1 \mu\text{m}^{-1}$.
12. An n -type Si slice of a thickness L is inhomogeneously doped with phosphorus donor whose concentration profile is given by $N_D(x) = N_0 + (N_L - N_0)(x/L)$. What is the formula for the electric potential difference between the front and the back surfaces when the sample is at thermal and electric equilibria regardless of how the mobility and diffusivity varies with position? What is the formula for the equilibrium electric field at a plane x from the front surface for a constant diffusivity and mobility?

FOR SECTION 3.3 GENERATION AND RECOMBINATION PROCESS

13. Calculate the electron and hole concentration under steady-state illumination in an n -type silicon with $G_L = 10^{16} \text{ cm}^{-3}\text{s}^{-1}$, $N_D = 10^{15} \text{ cm}^{-3}$, and $\tau_n = \tau_p = 10 \mu\text{s}$.
14. An n -type silicon sample has 2×10^{16} arsenic atoms/cm³, 2×10^{15} bulk recombination centers/cm³, and 10^{10} surface recombination centers/cm². (a) Find the bulk minority carrier lifetime, the diffusion length, and the surface recombination velocity under low-injection conditions. The values of σ_p and σ_s are 5×10^{-15} and $2 \times 10^{-16} \text{ cm}^2$, respectively. (b) If the sample is illuminated with uniformly absorbed light that creates 10^{17} electron-hole pairs/cm²-s, what is the hole concentration at the surface?
15. Assume that an n -type semiconductor is uniformly illuminated, producing a uniform excess generation rate G . Show that in steady state the change in the semiconductor conductivity is given by $\Delta\sigma = q(\mu_n + \mu_p)\tau_p G$.

FOR SECTION 3.4 CONTINUITY EQUATION

16. The total current in a semiconductor is constant and is composed of electron drift current and hole diffusion current. The electron concentration is constant and equal to 10^{16} cm^{-3} .

The hole concentration is given by

$$p(x) = 10^{15} \exp\left(\frac{-x}{L}\right) \text{ cm}^{-3} \quad (x \geq 0),$$

where $L = 12 \text{ } \mu\text{m}$. The hole diffusion coefficient is $D_p = 12 \text{ cm}^2/\text{s}$ and the electron mobility is $\mu_n = 1000 \text{ cm}^2/\text{V}\cdot\text{s}$. The total current density is $J = 4.8 \text{ A/cm}^2$. Calculate (a) the hole diffusion current density versus x , (b) the electron current density versus x , and (c) the electric field versus x .

- *17. Excess carriers are injected on one surface of a thin slice of n -type silicon with thickness W and extracted at the opposite surface where $p_n(W) = p_{n0}$. There is no electric field in the region $0 < x < W$. Derive the expression for current densities at the two surfaces.
- 18. In Prob. 17, if carrier lifetime is $50 \text{ } \mu\text{s}$ and $W = 0.1 \text{ mm}$, calculate the portion of injected current that reaches the opposite surface by diffusion ($D = 50 \text{ cm}^2/\text{s}$).
- *19. An n -type semiconductor has excess carrier holes 10^{14} cm^{-3} , and a bulk minority carrier lifetime 10^{-6} s in the bulk material, and a minority carrier lifetime 10^{-7} s at the surface. Assume zero applied electric field and let $D_p = 10 \text{ cm}^2/\text{s}$. Determine the steady-state excess carrier concentration as a function of distance from the surface ($x = 0$) of the semiconductor.

FOR SECTION 3.5 THERMIONIC EMISSION PROCESS

- 20. A metal, with a work function $\phi_m = 4.2 \text{ V}$, is deposited on an n -type silicon semiconductor with affinity $\chi = 4.0 \text{ V}$ and $E_g = 1.12 \text{ eV}$. What is the potential barrier height seen by electrons in the metal moving into the semiconductor?
- 21. Consider a tungsten filament with metal work function ϕ_m inside a high vacuum chamber. Show that if a current is passed through the filament to heat it up sufficiently, the electrons with enough thermal energy will escape into the vacuum and the resulted thermionic current density is

$$J = A^* T^2 \exp\left(\frac{-q\phi_m}{kT}\right)$$

where A^* is $4\pi qmk^2/h^3$ and m is free electron mass. The definite integral

$$\int_{-\infty}^{\infty} e^{-ax^2} dx = \left(\frac{\pi}{a}\right)^{1/2}.$$

FOR SECTION 3.6 TUNNELING PROCESS

- 22. Consider a electron with an energy of 2 eV impinging on a potential barrier with 20 eV and a width of $3 \text{ } \text{\AA}$. What is the tunneling probability?
- 23. Evaluate the transmission coefficient for an electron of energy 2.2 eV impinging on a potential barrier of height 6.0 eV and thickness 10^{-10} meters . Repeat the calculation for a barrier thickness of 10^{-9} meters .

FOR SECTION 3.7 HIGH FIELD EFFECTS

- 24. Use the velocity-field relations for Si and GaAs shown in Fig. 22 to determine the transit time of electrons through a $1 \text{ } \mu\text{m}$ distance in these materials for an electric field of (a) 1 kV/cm and (b) 50 kV/cm .
- 25. Assume that a conduction electron in Si ($\mu_n = 1350 \text{ cm}^2/\text{V}\cdot\text{s}$) has a thermal energy kT , related to its mean thermal velocity by $E_{th} = m_0 v_{th}^2/2$. This electron is placed in an electric field of 100 V/cm . Show that the drift velocity of the electron in this case is small compared to its thermal velocity. Repeat for a field of 10^4 V/cm , using the same value of μ_n . Comment on the actual mobility effects at this higher value of field.