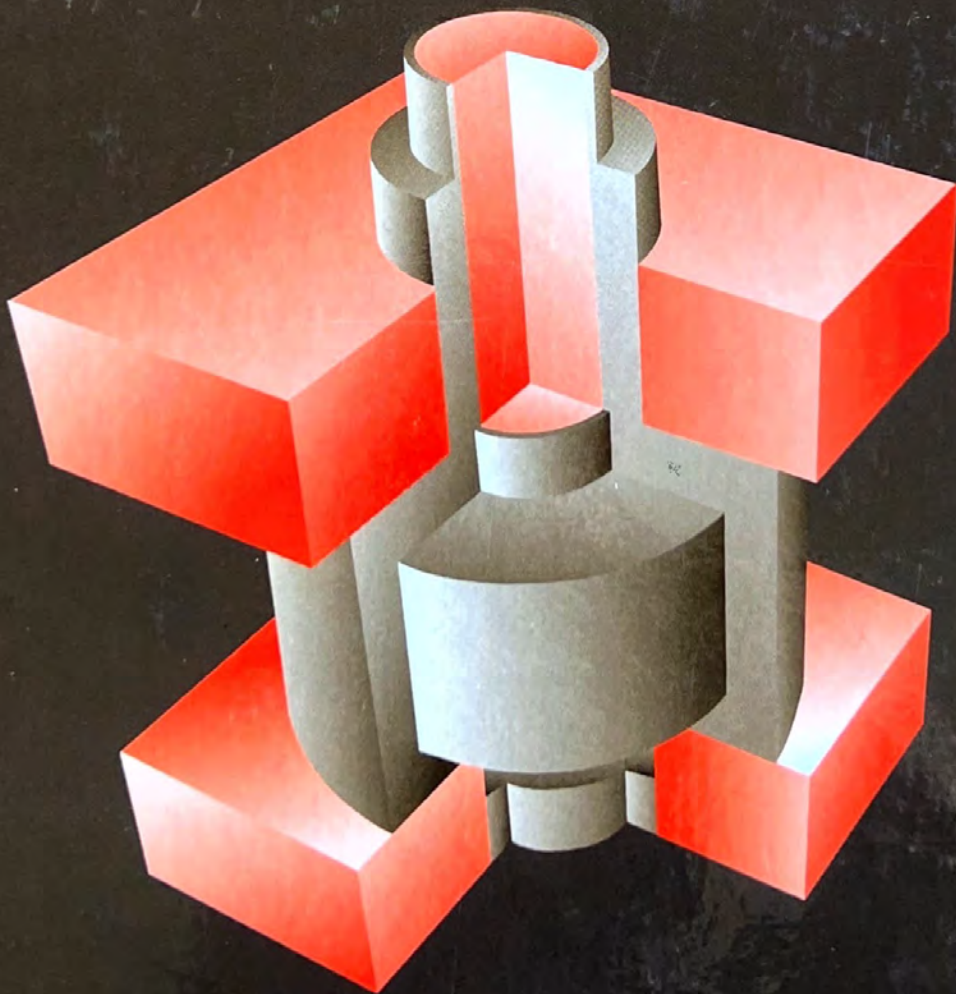


Semiconductor Devices

Physics and Technology

S. M. SZE | M. K. LEE



3RD

TSMC EX. 1024 Page 1 of 51
EDITION

3RD EDITION

Semiconductor Devices

Physics and Technology

S. M. SZE

*EtronTech Distinguished Chair Professor
College of Electrical and Computer Engineering
National Chiao Tung University
Hsinchu, Taiwan*

M. K. LEE

*Professor
Department of Electrical Engineering
National Sun Yat-sen University
Kaohsiung, Taiwan*



JOHN WILEY & SONS, INC.

Acquisitions Editor *Dan Sayre*
Marketing Manager *Christopher Ruel*
Senior Editorial Assistant *Katie Singleton*
Editorial Program Assistant *Samantha Mendel*
Production Manager *Micheline Frederick*
Cover Designer *Wendy Lai*
Pre-press Service *Robots & Cupcakes*

This book was typeset in *Times Roman* by the authors and printed and bound by *Quad Graphics/Versailles*. The cover was printed by *Quad Graphics/Versailles*.

cover photo: © 2010 IEEE. Reprinted, with permission, from IEDM Technical Digest, S. Whang et. al, "Novel 3-dimensional Dual Control-gate with Surrounding Floating-gate (DC-SF) NAND flash cell for 1Tb file storage application."

Copyright © 1985, 2002, 2012 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).

Evaluation copies are provided to qualified academics and professionals for review purposes only, for use in their courses during the next academic year. These copies are licensed and may not be sold or transferred to a third party. Upon completion of the review period, please return the evaluation copy to Wiley. Return instructions and a free of charge return shipping label are available at www.wiley.com/go/returnlabel. Outside of the United States, please contact your local representative.

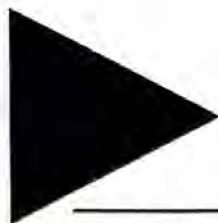
ISBN 978-0470-53794-7

Printed and bound in Singapore by Markono Print Media Pte Ltd

10 9 8 7 6 5 4 3 2 1

The inside back cover will contain printing identification and country of origin if omitted from this page. In addition, if the ISBN on the back cover differs from the ISBN on this page, the one on the back cover is correct.

TSMC EX. 1024 Page 3 of 51



Contents

Preface

Acknowledgments

► CHAPTER 0

Introduction

- 0.1 Semiconductor Devices
- 0.2 Semiconductor Technology
- Summary

PART I

SEMICONDUCTOR PHYSICS

► CHAPTER 1

Energy Bands and Carrier Concentration in Thermal Equilibrium

- 1.1 Semiconductor Materials
- 1.2 Basic Crystal Structures
- 1.3 Valence Bonds
- 1.4 Energy Bands
- 1.5 Intrinsic Carrier Concentration
- 1.6 Donors and Acceptors
- Summary

► CHAPTER 2

Carrier Transport Phenomena

- 2.1 Carrier Drift
- 2.2 Carrier Diffusion
- 2.3 Generation and Recombination Processes
- 2.4 Continuity Equation
- 2.5 Thermionic Emission Process
- 2.6 Tunneling Process
- 2.7 Space-Charge Effect
- 2.8 High-Field Effects
- Summary

PART II

SEMICONDUCTOR DEVICES

► CHAPTER 3

p-n Junction

- 3.1 Thermal Equilibrium Condition
- 3.2 Depletion Region
- 3.3 Depletion Capacitance
- 3.4 Current-Voltage Characteristics
- 3.5 Charge Storage and Transient Behavior
- 3.6 Junction Breakdown
- 3.7 Heterojunction
- Summary

vii ► CHAPTER 4

ix Bipolar Transistors and Related Devices

- 1 4.1 Transistor Action
- 1 4.2 Static Characteristics of Bipolar Transistors
- 1 4.3 Frequency Response and Switching of Bipolar Transistors
- 6 4.4 Nonideal Effects
- 12 4.5 Heterojunction Bipolar Transistors
- 4.6 Thyristors and Related Power Devices
- Summary

► CHAPTER 5

MOS Capacitor and MOSFET

- 15 5.1 Ideal MOS Capacitor
- 15 5.2 SiO₂-Si MOS Capacitor
- 17 5.3 Carrier Transport in MOS Capacitors
- 22 5.4 Charge-Coupled Devices
- 23 5.5 MOSFET Fundamentals
- 29 Summary

► CHAPTER 6

Advanced MOSFET and Related Devices

- 40 6.1 MOSFET Scaling
- 43 6.2 CMOS and BiCMOS
- 43 6.3 MOSFET on Insulator
- 53 6.4 MOS Memory Structures
- 56 6.5 Power MOSFET
- 62 Summary

► CHAPTER 7

MESFET and Related Devices

- 71 7.1 Metal-Semiconductor Contacts
- 73 7.2 MESFET
- 77 7.3 MODFET
- Summary

► CHAPTER 8

Microwave Diodes; Quantum-Effect and Hot-Electron Devices

- 82 8.1 Microwave Frequency Bands
- 83 8.2 Tunnel Diode
- 87 8.3 IMPATT Diode
- 95 8.4 Transferred-Electron Devices
- 99 8.5 Quantum-Effect Devices
- 108 8.6 Hot-Electron Devices
- 111 Summary

► CHAPTER 9 Light Emitting Diodes and Lasers		
9.1 Radiative Transitions and Optical Absorption	280	
9.2 Light-Emitting Diodes	280	
9.3 Various Light-Emitting Diodes	286	
9.4 Semiconductor Lasers	291	
Summary	302	
► CHAPTER 10 Photodetectors and Solar Cells	319	
10.1 Photodetectors	323	
10.2 Solar Cells	323	
10.3 Silicon and Compound-Semiconductor Solar Cells	336	
10.4 Third-Generation Solar Cells	343	
10.5 Optical Concentration	348	
Summary	352	
PART III SEMICONDUCTOR TECHNOLOGY	352	
► CHAPTER 11 Crystal Growth and Epitaxy		
11.1 Silicon Crystal Growth from the Melt	357	
11.2 Silicon Float-Zone Process	357	
11.3 GaAs Crystal-Growth Techniques	363	
11.4 Material Characterization	367	
11.5 Epitaxial-Growth Techniques	370	
11.6 Structures and Defects in Epitaxial Layers	377	
Summary	384	
► CHAPTER 12 Film Formation		
12.1 Thermal Oxidation	388	
12.2 Chemical Vapor Deposition of Dielectrics	392	
12.3 Chemical Vapor Deposition of Polysilicon	392	
12.4 Atom Layer Deposition	400	
12.5 Metallization	409	
Summary	412	
► CHAPTER 13 Lithography and Etching		
13.1 Optical Lithography	414	
13.2 Next-Generation Lithographic Methods	425	
13.3 Wet Chemical Etching	428	
13.4 Dry Etching	428	
Summary	441	
► CHAPTER 14 Impurity Doping		
14.1 Basic Diffusion Process	447	
14.2 Extrinsic Diffusion	450	
14.3 Diffusion-Related Processes	462	
14.4 Range of Implanted Ions	462	483
14.5 Implant Damage and Annealing		490
14.6 Implantation-Related Processes		495
Summary		501
► CHAPTER 15 Integrated Devices		
15.1 Passive Components		505
15.2 Bipolar Technology		507
15.3 MOSFET Technology		511
15.4 MESFET Technology		516
15.5 Challenges for Nanoelectronics		529
Summary		532
► APPENDIX A List of Symbols		537
► APPENDIX B International Systems of Units (SI Units)		541
► APPENDIX C Unit Prefixes		543
► APPENDIX D Greek Alphabet		544
► APPENDIX E Physical Constants		545
► APPENDIX F Properties of Important Element and Binary Compound Semiconductors at 300 K		546
► APPENDIX G Properties of Si and GaAs at 300 K		547
► APPENDIX H Derivation of the Density of States in a Semiconductor		548
► APPENDIX I Derivation of Recombination Rate for Indirect Recombination		549
► APPENDIX J Calculation of the Transmission Coefficient for a Symmetric Resonant-Tunneling Diode		553
► APPENDIX K Basic Kinetic Theory of Gases		555
► APPENDIX L Answers to Selected Problems		557
Photo credits	466	559
Index	467	563
	476	565
	480	



1

Energy Bands and Carrier Concentration in Thermal Equilibrium

- ▶ 1.1 SEMICONDUCTOR MATERIALS
- ▶ 1.2 BASIC CRYSTAL STRUCTURES
- ▶ 1.3 VALENCE BONDS
- ▶ 1.4 ENERGY BANDS
- ▶ 1.5 INTRINSIC CARRIER CONCENTRATION
- ▶ 1.6 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 solid. 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.

▶ 1.1 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 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

*A list of symbols is given in Appendix A.

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

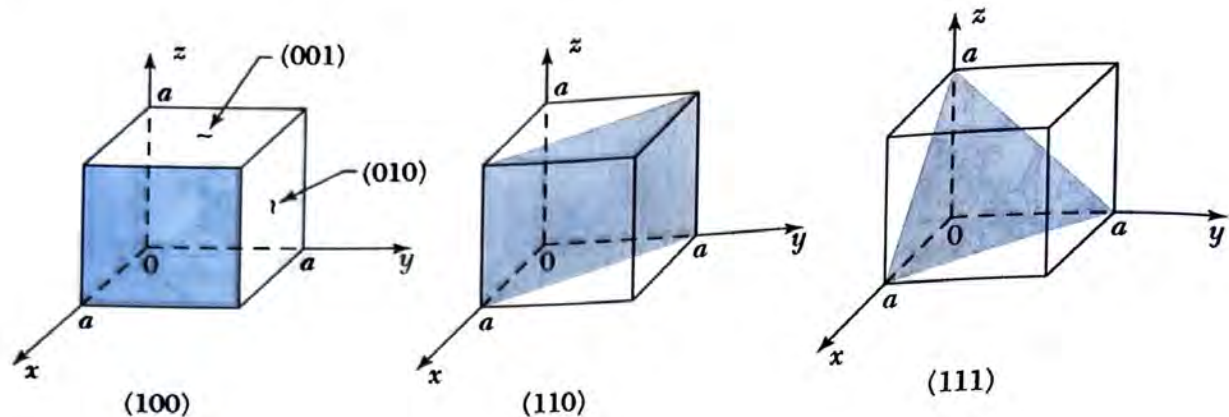


Fig. 6 Miller indices of some important planes in a cubic crystal.

Figure 6 shows the Miller indices of important planes in a cubic crystal.[§] 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 $[00\bar{1}]$.

► 1.3 VALENCE BONDS

As discussed in Section 1.2, each atom in a diamond lattice is surrounded by four nearest neighbors. Figure 7a shows the tetrahedron bonds of a diamond lattice. A simplified two-dimensional bonding diagram for the tetrahedron is shown in Fig. 7b. 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 (ionize one electron from the bond). When a bond is broken, a free electron results and can participate in current conduction.

[§] In Chapter 5, we show that the $\langle 100 \rangle$ orientation is preferred for silicon metal-oxide-semiconductor field-effect transistors (MOSFETs).

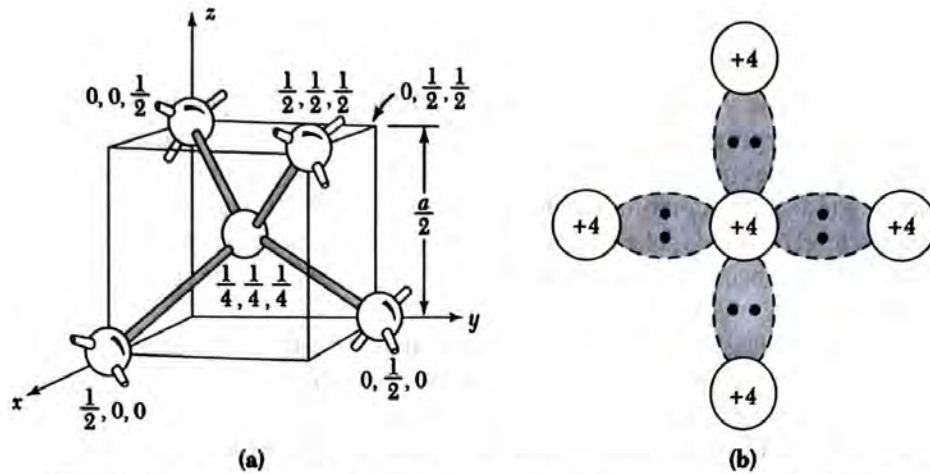


Fig. 7 (a) A tetrahedron bond. (b) Schematic two-dimensional representation of a tetrahedron bond.

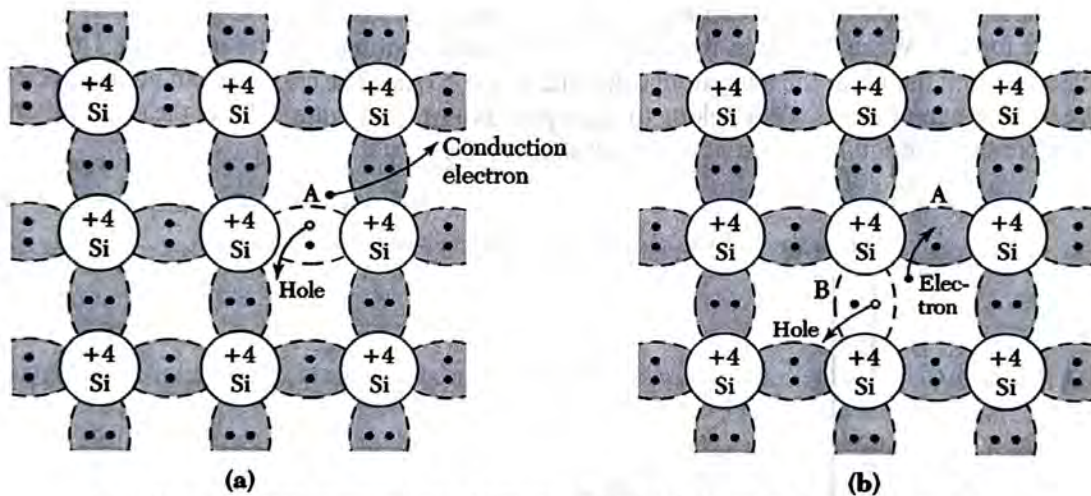


Fig. 8 The basic bond representation of intrinsic silicon. (a) A broken bond at position A, resulting in a conduction electron and a hole. (b) A broken bond at position B.

Figure 8a 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. 8b. 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.

► 1.4 ENERGY BANDS

1.4.1 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 model:⁴

► 1.6 DONORS AND ACCEPTORS

When a semiconductor is doped with impurities, the semiconductor becomes *extrinsic* and impurity energy levels are introduced. Figure 19a 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. 19b 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*.

The impurity atoms are imperfections and interrupt the perfect periodicity of the lattice; energy levels within the band gap that were forbidden are no longer disallowed. That is to say, the impurity atoms will introduce an energy level or multiple energy levels in the band gap.

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. 2:

$$E_D = \left(\frac{\epsilon_0}{\epsilon_s} \right)^2 \left(\frac{m_n}{m_0} \right) E_H. \quad (20)$$

The ionization energy for donors, measured from the conduction band edge and calculated from Eq. 20 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.

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 20 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.

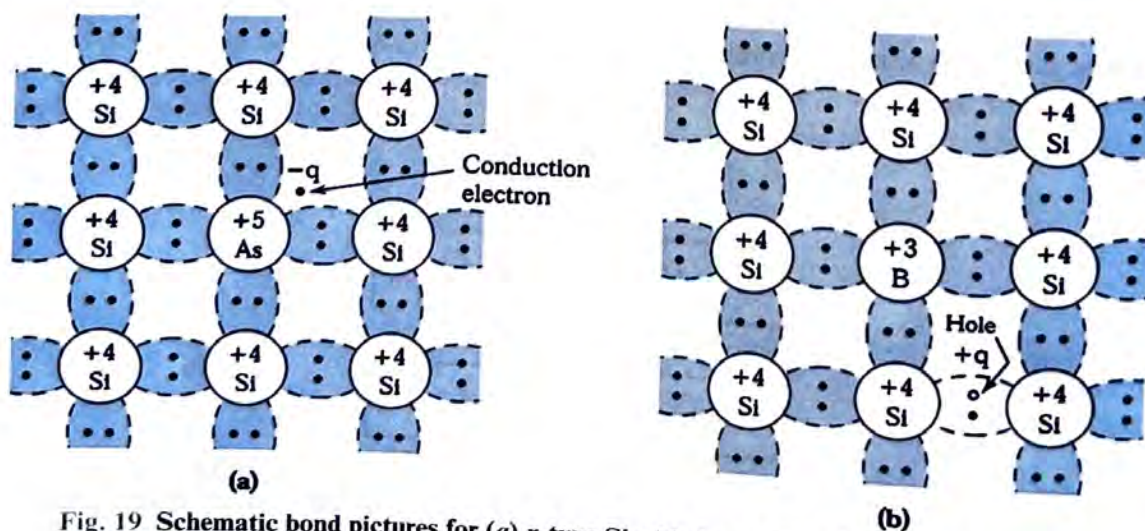


Fig. 19 Schematic bond pictures for (a) *n*-type Si with donor (arsenic) and (b) *p*-type Si with acceptor (boron).

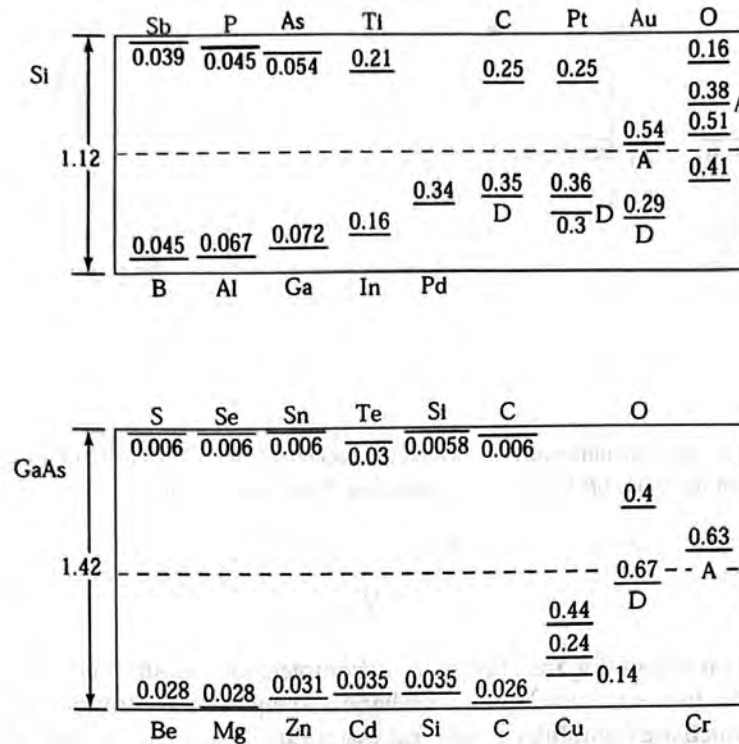


Fig. 20 Measured ionization energies (in eV) for various impurities in Si and GaAs. The levels below the gap center are measured from the top of the valence band and are acceptor levels unless labeled by *D* for donor level. The levels above the gap center are measured from the bottom of the conduction band and are donor levels unless indicated by *A* for acceptor level.⁸

1.6.1 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 . In such cases, the semiconductor is referred to as a *nondegenerate semiconductor*.

For shallow donors in *n*-type 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

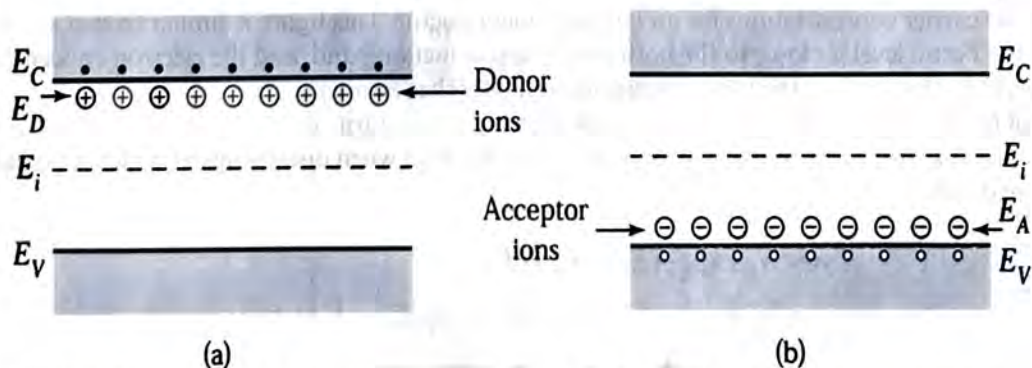


Fig. 21 Schematic energy band representation of extrinsic semiconductors with (a) donor ions and (b) acceptor ions.

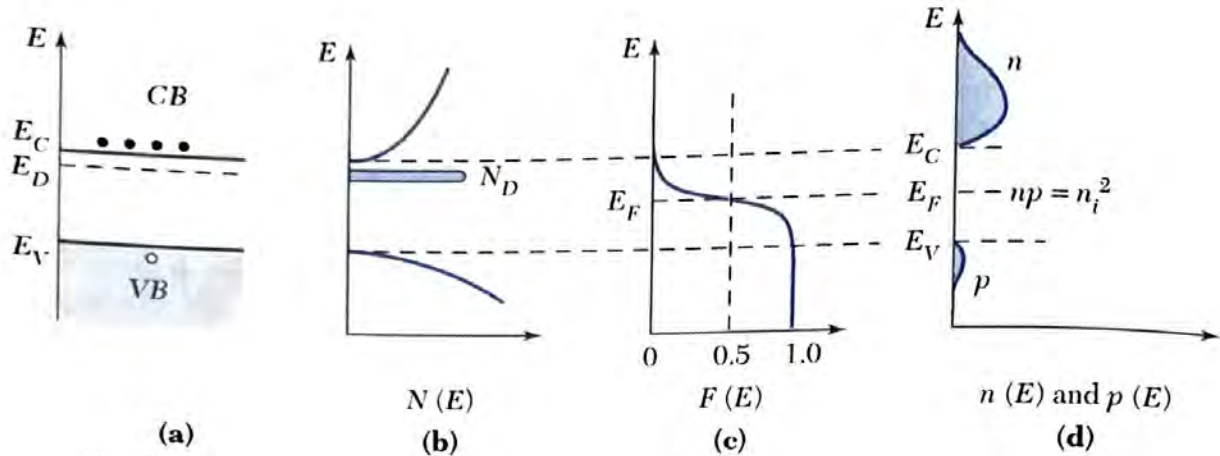


Fig. 22 *n*-Type semiconductor. (a) Schematic band diagram. (b) Density of states. (c) Fermi distribution function. (d) Carrier concentration. Note that $np = n_i^2$.

$$n = N_D \quad (21)$$

where N_D is the donor concentration. Figure 21a 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. 13 and 21, we obtain the Fermi level in terms of the effective density of states N_C and the donor concentration N_D :

$$E_C - E_F = kT \ln (N_C / N_D). \quad (22)$$

Similarly, for shallow acceptors for *p*-type semiconductors as shown in Fig. 21b, if there is complete ionization, the concentration of holes is

$$p = N_A, \quad (23)$$

where N_A is the acceptor concentration. We can obtain the corresponding Fermi level from Eqs. 14 and 23:

$$E_F - E_V = kT \ln (N_V / N_A). \quad (24)$$

From Eq. 22 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 22 illustrates the procedure for obtaining the carrier concentrations for an *n*-type semiconductor. This figure is similar to that shown in Fig. 17. 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. 13 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

$$n = n_i \exp[(E_F - E_i) / kT], \quad (25)$$

and similarly,

$$p = n_i \exp[(E_i - E_F) / kT]. \quad (26)$$

Note that the product of n and p from Eqs. 25 and 26 is n_i^2 . This result is identical to that for the intrinsic case, Eq. 17. Equation 17 is called the *mass action law*, and 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. 17,

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

The Fermi level measured from the bottom of the conduction band is given by Eq. 22:

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

The Fermi level measured from the intrinsic Fermi level is given by Eq. 25:

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

These results are shown graphically in Fig. 23.

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

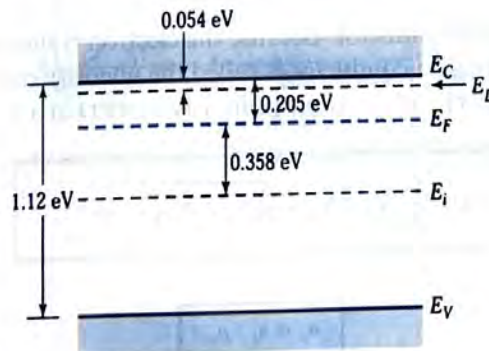


Fig. 23 Band diagram showing Fermi level E_F and intrinsic Fermi level E_i .

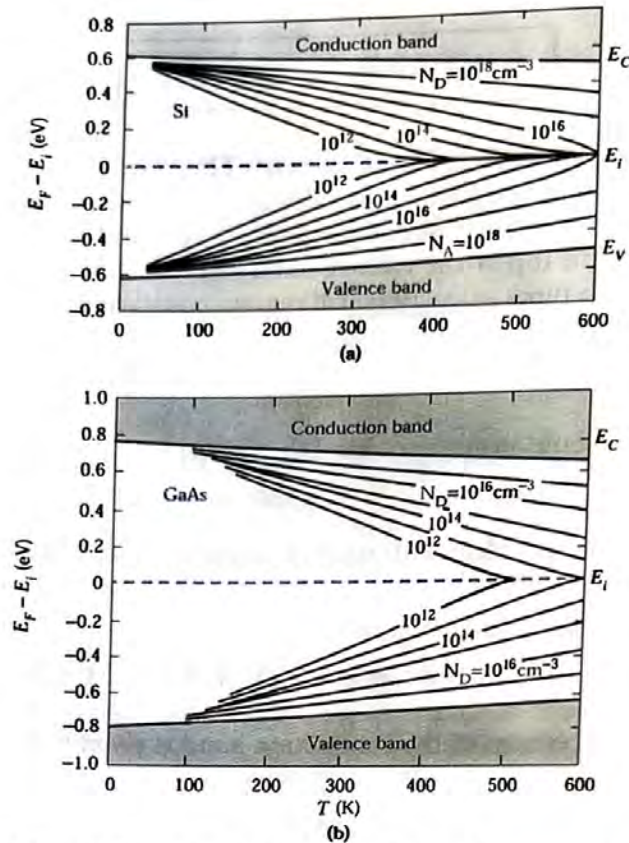


Fig. 24 Fermi level for (a) Si and (b) GaAs as a function of temperature and impurity concentration showing dependence of the bandgap on temperature.⁹

$$n + N_A = p + N_D.$$

(27)

Solving Eqs. 17 and 27 yields the equilibrium electron and hole concentrations in an n -type semiconductor:

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

(28)

$$p_n = n_i^2 / n_n.$$

(29)

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 as:

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

(30)

$$n_p = n_i^2 / p_p.$$

(31)

The subscript p refers to the p -type semiconductor.

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

$$n_n \approx N_D - N_A \quad \text{if} \quad N_D > N_A, \quad (32)$$

$$p_p \approx N_A - N_D \quad \text{if} \quad N_A > N_D. \quad (33)$$

From Eqs. 28 to 31 together with Eqs. 13 and 14, we can calculate the position of the Fermi level as a function of temperature for a given acceptor or donor concentration. Figure 24 plots these calculations for silicon⁹ and gallium arsenide. We have incorporated in the figure the variation of the bandgap with temperature (see Problem 7). Note that as the temperature increases, the Fermi level approaches the intrinsic level, that is, the semiconductor becomes intrinsic.

Figure 25 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. 18 by setting the impurity concentration equal to n_i .

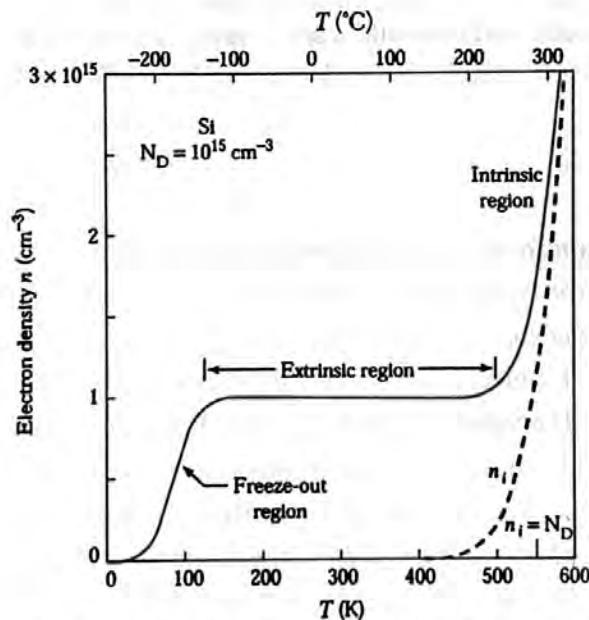


Fig. 25 Electron density as a function of temperature for a Si sample with a donor concentration of 10^{15} cm^{-3} .

1.6.2 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. 8, and the electron density (Eq. 6) 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

$$\Delta E_g = 22 \left(\frac{N}{10^{18}} \right)^{1/2} \text{ meV}, \quad (34)$$

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. A discussion of how to grow semiconductor crystals can be found in Chapter 11.

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 and K. K. Ng, *Physics of Semiconductor Devices*, 3rd ed., Wiley Interscience, Hoboken, 2007.
9. A. S. Grove, *Physics and Technology of Semiconductor Devices*, Wiley, New York, 1967.



14

Impurity Doping

- ▶ 14.1 BASIC DIFFUSION PROCESS
 - ▶ 14.2 EXTRINSIC DIFFUSION
 - ▶ 14.3 DIFFUSION-RELATED PROCESSES
 - ▶ 14.4 RANGE OF IMPLANTED IONS
 - ▶ 14.5 IMPLANT DAMAGE AND ANNEALING
 - ▶ 14.6 IMPLANTATION-RELATED PROCESSES
 - ▶ SUMMARY
-

Impurity doping is the introduction of controlled amounts of impurity dopants into semiconductor materials. The practical use of impurity doping is primarily to change the electrical properties of the semiconductors. *Diffusion* and *ion implantation* are the two key methods of impurity doping. Until the early 1970s, impurity doping was done mainly by diffusion at elevated temperatures, as shown in Fig. 1a. In this method the dopant atoms are placed on or near the surface of the wafer by deposition from the gas phase of the dopant or by using doped-oxide sources. The doping concentration decreases monotonically from the surface, and the profile of the dopant distribution is determined mainly by the temperature and diffusion time.

Since the early 1970s, many doping operations have been performed by ion implantation, as shown in Fig. 1b. In this process the dopant ions are implanted into the semiconductor by means of an ion beam. The doping concentration has a peak distribution inside the semiconductor and the profile of the dopant distribution is determined mainly by the ion mass and the implanted-ion energy. Both diffusion and ion implantation are used in fabricating discrete devices and integrated circuits because these processes generally complement each other.^{1,2} For example, diffusion is used to form a deep junction (e.g., a twin well in CMOS), whereas ion implantation is used to form a shallow junction (e.g., a source/drain junction of a MOSFET).

Specifically, we cover the following topics:

- The movement of impurity atoms in the crystal lattice under high temperature and high concentration-gradient conditions.
- Impurity profiles for constant diffusivity and concentration-dependent diffusivity.
- The impact of lateral diffusion and impurity redistribution on device characteristics.
- The process and advantages of ion implantation.
- Ion distributions in the crystal lattice and how to remove lattice damage caused by ion implantation.
- Implantation-related processes such as masking, high-energy implantation, and high-current implantation.

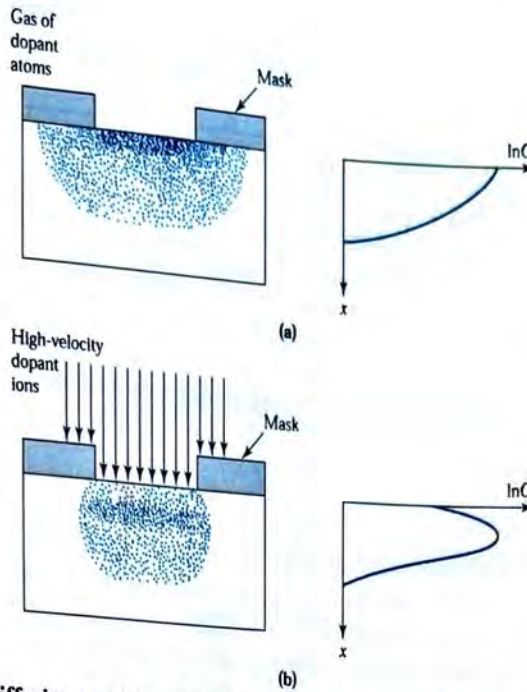


Fig. 1 Comparison of (a) diffusion and (b) ion-implantation techniques for selective introduction of dopants into the semiconductor substrate.

► 14.1 BASIC DIFFUSION PROCESS

Diffusion of impurities is typically done by placing semiconductor wafers in a carefully controlled high-temperature quartz-tube furnace and passing a gas mixture that contains the desired dopant through it. The temperature usually ranges between 800° and 1200°C for silicon and 600° and 1000°C for gallium arsenide. The number of dopant atoms that diffuse into the semiconductor is related to the partial pressure of the dopant impurity in the gas mixture.

For diffusion in silicon, boron is the most popular dopant for introducing a p -type impurity, whereas arsenic and phosphorus are used extensively as n -type dopants. These three elements are highly soluble in silicon, as they have solubilities above $5 \times 10^{20} \text{ cm}^{-3}$ in the diffusion temperature range. These dopants can be introduced in several ways, including solid sources (e.g., BN for boron, As_2O_3 for arsenic, and P_2O_5 for phosphorus), liquid sources (BBr_3 , AsCl_3 , and POCl_3), and gaseous sources (B_2H_6 , AsH_3 , and PH_3). However, liquid sources are most commonly used. A schematic diagram of the furnace and gas flow arrangement for a liquid source is shown in Fig. 2. This arrangement is similar to that used for thermal oxidation. An example of the chemical reaction for phosphorus diffusion using a liquid source is



The P_2O_5 forms a glass on a silicon wafer and is then reduced to phosphorus by silicon,



the phosphorus is released and diffuses into the silicon and Cl_2 is vented.

For diffusion in gallium arsenide, the high vapor pressure of arsenic requires special methods to prevent the loss of arsenic by decomposition or evaporation.² These methods include diffusion in sealed ampules with an overpressure of arsenic and diffusion in an open-tube furnace with a doped-oxide capping layer (e.g., silicon nitride). Most of the studies on p -type diffusion have been confined to the use of zinc in the forms of Zn-Ga-As alloys and ZnAs_2 for the sealed-ampule approach or ZnO-SiO_2 for the open-tube approach. The n -type dopants in gallium arsenide include selenium and tellurium.

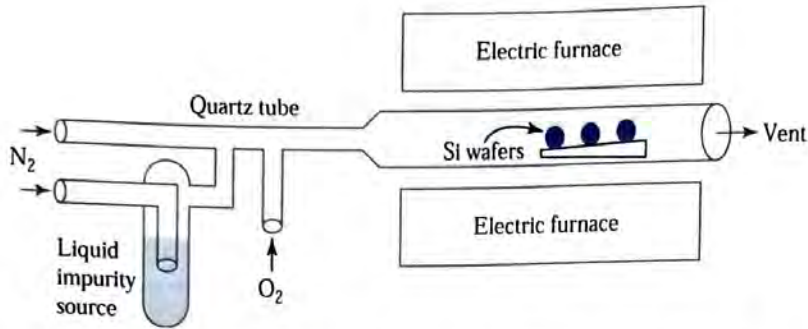


Fig. 2 Schematic diagram of a typical open-tube diffusion system.

14.1.1 Diffusion Equation

Diffusion in a semiconductor can be visualized as atomic movement of the diffusant (dopant atoms) in the crystal lattice by vacancies or interstitials. Fig.3 shows the two basic atomic-diffusion models in a solid.^{1,3} The open circles represent the host atoms occupying the equilibrium lattice positions and the solid dots represent impurity atoms. At elevated temperatures the lattice atoms vibrate around the equilibrium lattice sites. There is a finite probability that a host atom acquires sufficient energy to leave the lattice site and become an interstitial atom, thereby creating a vacancy. When a neighboring impurity atom migrates to the vacancy site as illustrated in Fig. 3a, the mechanism is called *vacancy diffusion*. If an interstitial atom moves from one place to another without occupying a lattice site (Fig. 3b), the mechanism is *interstitial diffusion*. An atom smaller than the host atom often moves interstitially.

In addition, there is extended interstitial diffusion, sometimes called *interstitialcy diffusion*. The interstitial host atom (self-interstitial) pushes the substitutional impurity atom into an interstitial site. Subsequently, the impurity atom displaces another host atom and creates a new self-interstitial. Then the process is repeated. Interstitialcy diffusion is faster than substitutional diffusion. Vacancy and interstitialcy diffusion are considered the dominant mechanisms for diffusion of P, B, As, and Sb in silicon. Phosphorus and boron diffuse via a dual (vacancy and interstitialcy) mechanism, with the interstitialcy component dominating. Arsenic and antimony diffuse predominately via a vacancy mechanism.¹

The basic diffusion process of impurity atoms is similar to that of charge carriers (electrons and holes), discussed in Chapter 2. Accordingly, we define a flux F as the number of dopant atoms passing through a unit area in a unit time and C as the dopant concentration per unit volume. From Eq. 27 in Chapter 2, we have

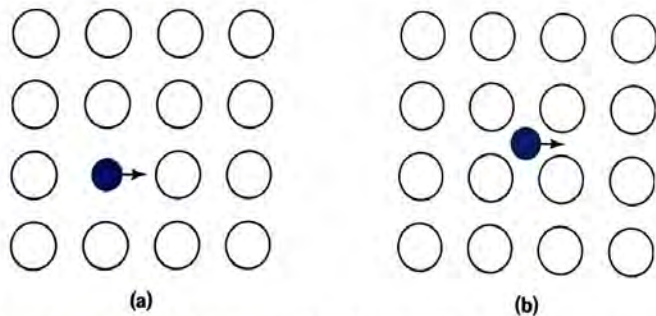


Fig. 3 Atomic diffusion mechanisms for a two-dimensional lattice.^{1,3} (a) Vacancy mechanism; (b) interstitial mechanism.

$$F = -D \frac{\partial C}{\partial x}, \quad (3)$$

where we have substituted C for the carrier concentration and the proportionality constant D is the diffusion coefficient or diffusivity. Note that the basic driving force of the diffusion process is the concentration gradient dC/dx . The flux is proportional to the concentration gradient, and the dopant atoms will move (diffuse) away from a high-concentration region toward a lower-concentration region.

If we substitute Eq. 3 into the one-dimensional continuity equation, Eq. 56 in Chapter 2, under the condition that no materials are formed or consumed in the host semiconductor (i.e., $G_n = R_n = 0$), we obtain

$$\frac{\partial C}{\partial t} = -\frac{\partial F}{\partial x} = \frac{\partial}{\partial x} \left(D \frac{\partial C}{\partial x} \right). \quad (4)$$

When the concentration of the dopant atoms is low, the diffusion coefficient can be considered independent of doping concentration, and Eq. 4 becomes

$$\boxed{\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial x^2}} \quad (5)$$

Equation 5 is often referred to as *Fick's diffusion equation*.

Figure 4 shows the measured diffusion coefficients for low concentrations of various dopant impurities in silicon and gallium arsenide.^{4,5} The logarithm of the diffusion coefficient plotted against the reciprocal of the absolute temperature gives a straight line in most cases. This implies that over the temperature range, the diffusion coefficients can be expressed as

$$\boxed{D = D_0 \exp\left(\frac{-E_a}{kT}\right)}, \quad (6)$$

where D_0 is the diffusion coefficient in cm^2/s extrapolated to infinite temperature and E_a is the activation energy in eV.

For the interstitial diffusion model, E_a is related to the energies required to move dopant atoms from one interstitial site to another. The values of E_a are found to be between 0.5 and 2 eV in both silicon and gallium arsenide. For the vacancy diffusion model, E_a is related to both the energies of motion and the energies of formation of vacancies. Thus, E_a for vacancy diffusion is larger than that for interstitial diffusion; usually between 3 and 5 eV.

For fast diffusants, such as Cu in Si and GaAs, shown in the upper portion of Fig. 4a and 4b, the measured activation energies are less than 2 eV, and interstitial atomic movement is the dominant diffusion mechanism. For slow diffusants, such as As in Si and GaAs, shown in the lower portion of Fig. 4a and 4b, E_a is larger than 3 eV, and vacancy diffusion is the dominant mechanism. For interstitialcy-dominated diffusion, such as P in silicon, E_a is also larger than 3 eV, but the diffusion coefficient is four times greater than As over the temperature range shown in Fig. 4a.

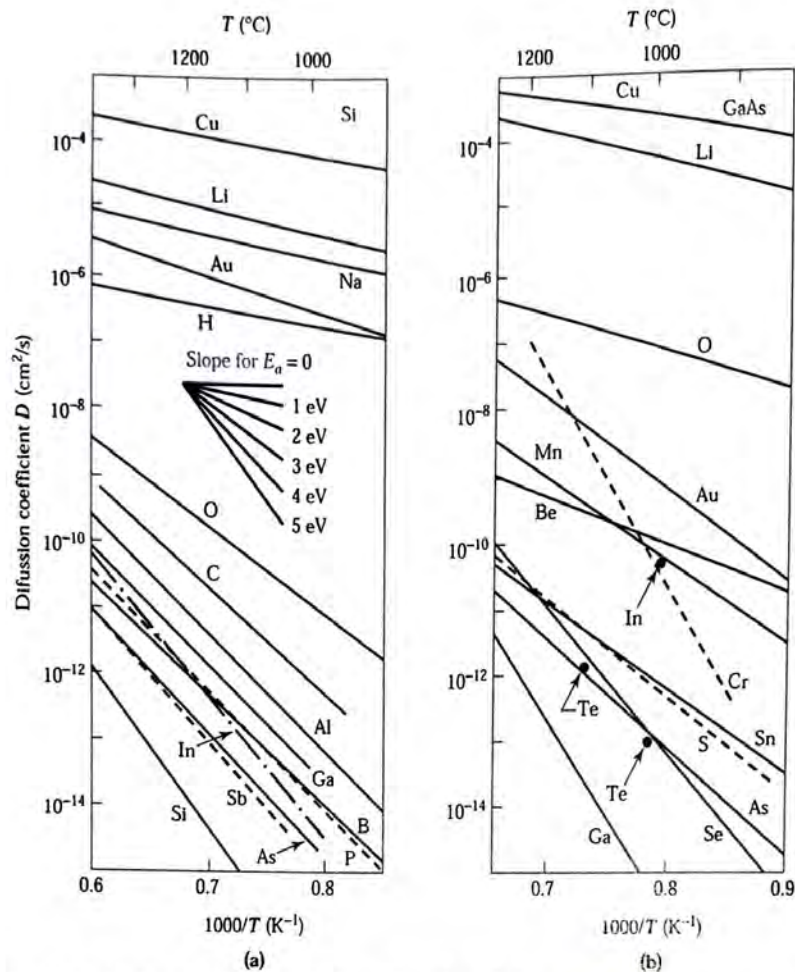


Fig. 4 Diffusion coefficient (also called diffusivity) as a function of the reciprocal of temperature for (a) silicon and (b) gallium arsenide.^{4,5}

14.1.2 Diffusion Profiles

The diffusion profile of the dopant atoms is dependent on the initial and boundary conditions. In this subsection we consider two important cases, namely, constant-surface-concentration diffusion and constant-total-dopant diffusion. In the first case, impurity atoms are transported from a vapor source onto the semiconductor surface and diffuse into the semiconductor wafers. The vapor source maintains a constant level of surface concentration during the entire diffusion period. In the second case, a fixed amount of dopant is deposited onto the semiconductor surface and is subsequently diffused into the wafers.

Constant-Surface-Concentration Diffusion

The initial condition at $t = 0$ is

$$C(x, 0) = 0, \tag{7}$$

which states that the dopant concentration in the host semiconductor is initially zero. The boundary conditions are

$$C(0, t) = C_s \tag{8a}$$

and

$$C(\infty, t) = 0 \quad (8b)$$

where C_s is the surface concentration (at $x = 0$), which is independent of time. The second boundary condition states that at large distances from the surface there are no impurity atoms.

The solution of the diffusion equation (Eq. 5) that satisfies the initial and boundary conditions is given by⁶

$$C(x, t) = C_s \operatorname{erfc} \left(\frac{x}{2\sqrt{Dt}} \right), \quad (9)$$

where erfc is the complementary error function and $\sqrt{Dt}$ is the diffusion length. The definition of erfc and some properties of the function are summarized in Table 1. The diffusion profile for the constant-surface-concentration condition is shown in Fig. 5a, where we plot, on both linear (upper) and logarithmic (lower) scales, the normalized concentration as a function of depth for three values of the diffusion length $\sqrt{Dt}$ corresponding to three consecutive diffusion times and a fixed D for a given diffusion temperature. Note that as the time progresses, the dopant penetrates deeper into the semiconductor.

The total number of dopant atoms per unit area of the semiconductor is given by

$$Q(t) = \int_0^\infty C(x, t) dx. \quad (10)$$

TABLE 1 ERROR FUNCTION ALGEBRA

$$\operatorname{erf}(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2} dy$$

$$\operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$$

$$\operatorname{erf}(0) = 0$$

$$\operatorname{erf}(\infty) = 1$$

$$\operatorname{erf}(x) \cong \frac{2}{\sqrt{\pi}} x \quad \text{for } x \ll 1$$

$$\operatorname{erfc}(x) \cong \frac{1}{\sqrt{\pi}} \frac{e^{-x^2}}{x} \quad \text{for } x \gg 1$$

$$\frac{d}{dx} \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} e^{-x^2}$$

$$\frac{d^2}{dx^2} \operatorname{erf}(x) = -\frac{4}{\sqrt{\pi}} x e^{-x^2}$$

$$\int_0^x \operatorname{erfc}(y') dy' = x \operatorname{erfc}(x) + \frac{1}{\sqrt{\pi}} (1 - e^{-x^2})$$

$$\int_0^\infty \operatorname{erfc}(x) dx = \frac{1}{\sqrt{\pi}}$$

Substituting Eq. 9 into Eq. 10 yields

$$Q(t) = \frac{2}{\sqrt{\pi}} C_s \sqrt{Dt} \cong 1.13 C_s \sqrt{Dt}. \quad (11)$$

This expression can be interpreted as follows. The quantity $Q(t)$ represents the area under one of the diffusion profiles of the linear plot in Fig. 5a. These profiles can be approximated by triangles with height C_s and base $2\sqrt{Dt}$. This leads to $Q(t) \cong C_s \sqrt{Dt}$, which is close to the exact result obtained from Eq. 11.

A related quantity is the gradient of the diffusion profile dC/dx . The gradient can be obtained by differentiating Eq. 9:

$$\left. \frac{dC}{dx} \right|_{x,t} = -\frac{C_s}{\sqrt{\pi Dt}} e^{-x^2/4Dt}. \quad (12)$$

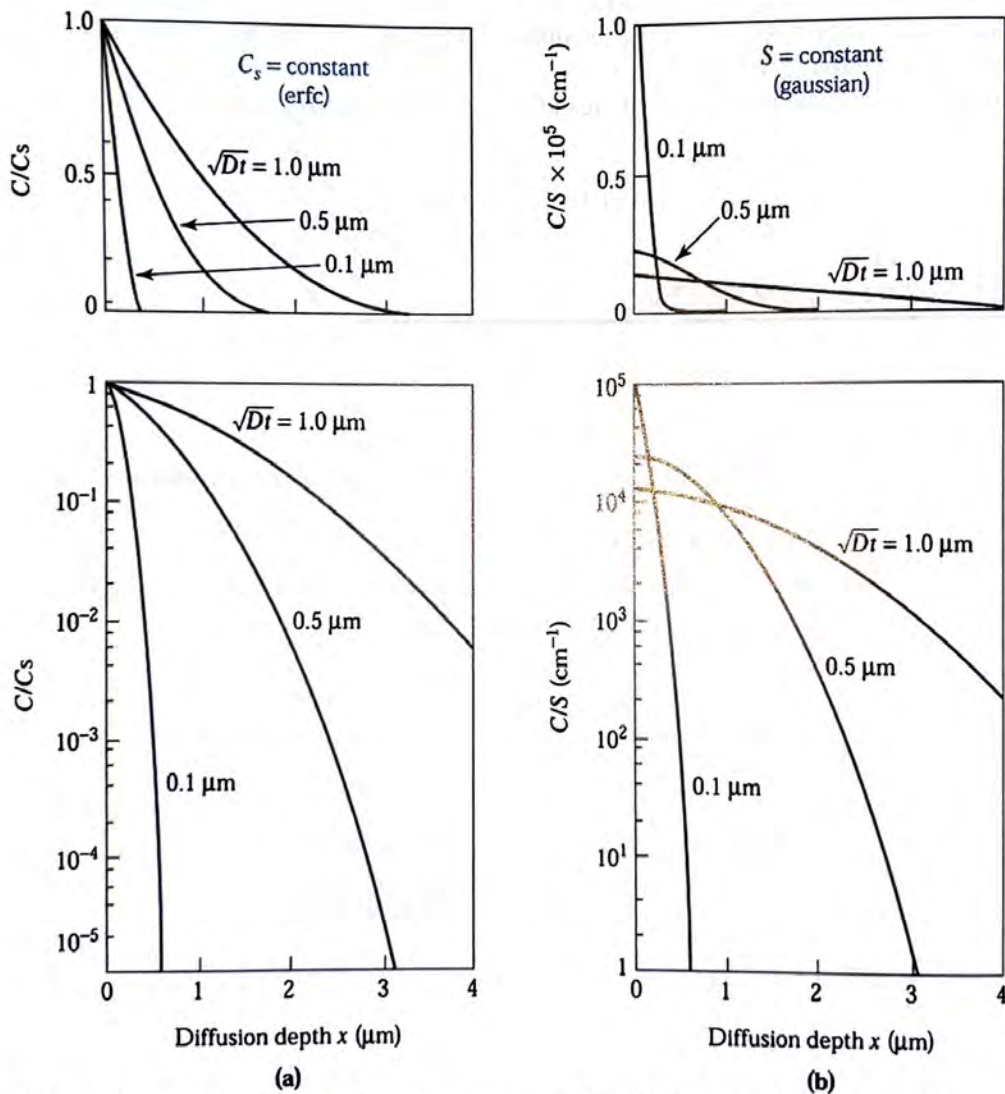


Fig. 5 Diffusion profiles. (a) Normalized complementary error function versus distance for successive diffusion times. (b) Normalized Gaussian function versus distance.

EXAMPLE 1

For a boron diffusion in silicon at 1000°C, the surface concentration is maintained at 10^{19} cm^{-3} and the diffusion time is 1 hour. Find $Q(t)$ and the gradient at $x = 0$ and at a location where the dopant concentration reaches 10^{15} cm^{-3} .

SOLUTION The diffusion coefficient of boron at 1000°C, as obtained from Fig. 4, is about $2 \times 10^{-14} \text{ cm}^2/\text{s}$, so that the diffusion length is

$$\sqrt{Dt} = \sqrt{2 \times 10^{-14} \times 3600} = 8.48 \times 10^{-6} \text{ cm},$$

$$Q(t) = 1.13 C_s \sqrt{Dt} = 1.13 \times 10^{19} \times 8.48 \times 10^{-6} = 9.5 \times 10^{13} \text{ atoms/cm}^2$$

$$\left. \frac{dC}{dx} \right|_{x=0} = -\frac{C_s}{\sqrt{\pi Dt}} = \frac{-10^{19}}{\sqrt{\pi} \times 8.48 \times 10^{-6}} = -6.7 \times 10^{23} \text{ cm}^{-4}.$$

When $C = 10^{15} \text{ cm}^{-3}$, the corresponding distance x_j is given by Eq. 9, or

$$x_j = 2\sqrt{Dt} \operatorname{erfc}^{-1} \left(\frac{10^{15}}{10^{19}} \right) = 2\sqrt{Dt} (2.75) = 4.66 \times 10^{-5} \text{ cm} = 0.466 \text{ } \mu\text{m},$$

$$\left. \frac{dC}{dx} \right|_{x=0.466 \text{ } \mu\text{m}} = -\frac{C_s}{\sqrt{\pi Dt}} e^{-x_j^2/4Dt} = -3.5 \times 10^{20} \text{ cm}^{-4}.$$

Constant-Total-Dopant Diffusion

For this case, a fixed (or constant) amount of dopant is deposited onto the semiconductor surface in a thin layer and the dopant subsequently diffuses into the semiconductor. The initial condition is the same as in Eq. 7. The boundary conditions are

$$\int_0^{\infty} C(x, t) dx = S \quad (13a)$$

and

$$C(\infty, t) = 0 \quad (13b)$$

where S is the total amount of dopant per unit area.

The solution of the diffusion equation, Eq. 5, that satisfies the above conditions is

$$C(x, t) = \frac{S}{\sqrt{\pi Dt}} \exp\left(-\frac{x^2}{4Dt}\right). \quad (14)$$

This expression is the Gaussian distribution. Since the dopant will move into the semiconductor as time increases, to keep the total dopant S constant, the surface concentration must decrease. This is indeed the case, since the surface concentration is given by Eq. 14 with $x = 0$:

$$C_s(t) = \frac{S}{\sqrt{\pi Dt}}. \quad (15)$$

Figure 5b shows the dopant profile for a Gaussian distribution where we plot the normalized concentration (C/S) as a function of the distance for three increasing diffusion lengths. Note the reduction of the surface

concentration as the diffusion time increases. The gradient of the diffusion profile is obtained by differentiating Eq. 14 and is

$$\left. \frac{dC}{dx} \right|_{x,t} = -\frac{XS}{2\sqrt{\pi}(Dt)^{3/2}} e^{-x^2/4Dt} = -\frac{X}{2Dt} C(x,t). \quad (16)$$

The gradient (or slope) is zero at $x = 0$ and at $x = \infty$ and the maximum gradient occurs at $x = \sqrt{2Dt}$.

In integrated-circuit processing, a two-step diffusion process is commonly used in which a *predeposition* diffused layer is first formed under a constant-surface-concentration condition. This step is followed by a *drive-in* diffusion (also called *redistribution* diffusion) under a constant-total-dopant condition. For most practical cases, the diffusion length $\sqrt{Dt}$ for the predeposition diffusion is much smaller than the diffusion length for the drive-in diffusion. Therefore, we can consider the predeposition profile as a delta function at the surface, and we can regard the extent of the penetration of the predeposition profile as negligibly small compared with that of the final profile that results from the drive-in step.

► EXAMPLE 2

Arsenic was predeposited by arsine gas and the resulting total amount of dopant per unit area is 1×10^{14} atoms/cm². How long would it take to drive the arsenic into a junction depth of 1 μm ? Assume a background doping of 1×10^{15} atoms/cm³ and a drive-in temperature of 1200°C. For As diffusion, $D_0 = 24$ cm²/s and $E_a = 4.08$ eV.

SOLUTION

$$D = D_0 \exp\left(\frac{-E_a}{kT}\right) = 24 \exp\left(\frac{-4.08}{8.614 \times 10^{-5} \times 1473}\right) = 2.602 \times 10^{-13} \text{ cm}^2/\text{s},$$

$$x_j^2 = 10^{-8} = 4Dt \ln\left(\frac{S}{C_B \sqrt{\pi Dt}}\right) = 1.04 \times 10^{-12} t \ln\left(\frac{1.106 \times 10^5}{\sqrt{t}}\right),$$

$$t \cdot \log t - 10.09t + 8350 = 0.$$

The solution to the above equation can be determined by the cross point of equation $y = t \cdot \log t$ and $y = 10.09t - 8350$. Therefore, $t = 1190$ seconds ≈ 20 minutes. ◀

14.1.3 Evaluation of Diffused Layers

The results of a diffusion process can be evaluated by three measurements—the junction depth, the sheet resistance, and the dopant profile of the diffused layer. The junction depth can be delineated by cutting a groove into the semiconductor and etching the surface with a solution (e.g., 100 cm³ HF and a few drops of HNO₃ for silicon) that stains the *p*-type region darker than the *n*-type region, as illustrated in Fig. 6a. If R_0 is the radius of the tool used to form the groove, then the junction depth x_j is given by

$$x_j = \sqrt{R_0^2 - b^2} - \sqrt{R_0^2 - a^2}, \quad (17)$$

where a and b are indicated in the figure. In addition, if R_0 is much larger than a and b , then

$$x_j \cong \frac{a^2 - b^2}{2R_0} \tag{18}$$

The junction depth x_j as illustrated in Fig. 6b is the position where the dopant concentration equals the substrate concentration C_B , or

$$C(x_j) = C_B \tag{19}$$

Thus, if the junction depth and C_B are known, the surface concentration C_s and the impurity distribution can be calculated, provided the diffusion profile follows one or the other simple equation derived in Section 14.1.2.

The resistance of a diffused layer can be measured by the four-point probe technique described in Chapter 2. The sheet resistance R is related to the junction depth x_j , the carrier mobility μ (which is a function of the total impurity concentration), and the impurity distribution $C(x)$ by the following expression:⁷

$$R = \frac{1}{q \int_0^{x_j} \mu C(x) dx}$$

(20)

For a given diffusion profile, the average resistivity $\bar{\rho} = Rx_j$ is uniquely related to the surface concentration C_s and the substrate-doping concentration for an assumed diffusion profile. Design curves relating C and $\bar{\rho}$ have been calculated for simple diffusion profiles, such as the erfc or Gaussian distribution.⁸ To use these curves correctly we must be sure that the diffusion profiles agree with the assumed profiles. For low concentrations and deep diffusions, the diffusion profiles generally can be represented by the aforementioned simple functions. However, as we discuss in the next section, for high concentrations and shallow diffusions, the diffusion profiles cannot be represented by these simple functions.

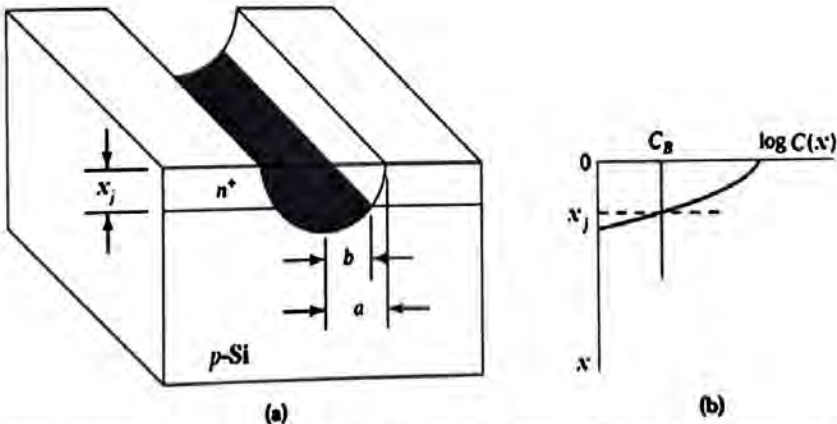


Fig. 6. Junction-depth measurement. (a) Grooving and staining; (b) position in which dopant and substrate concentration are equal.

The diffusion profile can be measured using the capacitance-voltage technique described in Chapter 7. The majority carrier profile, which is equal to the impurity profile if impurities are fully ionized, can be determined by measuring the reverse-bias capacitance of a p - n junction or a Schottky barrier diode as a function of the applied voltage. A more elaborate method is the secondary-ion-mass spectroscopy (SIMS) technique, which measures the total impurity profile. In the SIMS technique, an ion beam sputters material off the surface of a semiconductor, and the ion component is detected and mass analyzed. This technique has high sensitivity to many elements, such as boron and arsenic, and is an ideal tool for providing the precision needed for profile measurements in high-concentration or shallow-junction diffusions.⁹

► 14.2 EXTRINSIC DIFFUSION

The diffusion profiles described in Section 14.1 are for constant diffusivities. These profiles occur when the doping concentration is lower than the intrinsic-carrier concentration $n_i(T)$ at the diffusion temperature. For example, at $T = 1000^\circ\text{C}$, n_i equals $5 \times 10^{18} \text{ cm}^{-3}$ for silicon and $5 \times 10^{17} \text{ cm}^{-3}$ for gallium arsenide. The diffusivity at low concentrations is often referred to as the intrinsic diffusivity $D_i(T)$. Doping profiles that have concentrations less than $n_i(T)$ are in the *intrinsic* diffusion region as indicated in the left side of Fig. 7. In this region, the resulting dopant profiles of sequential or simultaneous diffusions of n - and p -type impurities can be determined by superposition; that is, the diffusions can be treated independently. However, when the impurity concentration, including both the substrate and the dopant, is greater than $n_i(T)$, the semiconductor becomes extrinsic and the diffusivity is considered to be extrinsic. In the extrinsic diffusion region the diffusivity becomes concentration dependent.¹⁰ In the extrinsic diffusion region the diffusion profiles are more complicated, and there are interactions and cooperative effects among the sequential or simultaneous diffusions.

14.2.1 Concentration-Dependent Diffusivity

As mentioned previously, when a host atom acquires sufficient energy from the lattice vibration to leave its lattice site, a vacancy is created. The presence of a vacancy in a crystal results in four unsatisfied and distorted bonds. The electrons of these bonds may spill into the vacancy. A neutral vacancy will act as an acceptor by acquiring a negative charge, $V^0 + e^- = V^-$. Therefore, depending on the charges associated with a vacancy, we can have a neutral vacancy V^0 , an acceptor vacancy V^- , a double-charged acceptor vacancy V^{2-} , a donor vacancy V^+ , and so forth. We expect that the vacancy density of a given charge state (i.e., the number of vacancies per unit

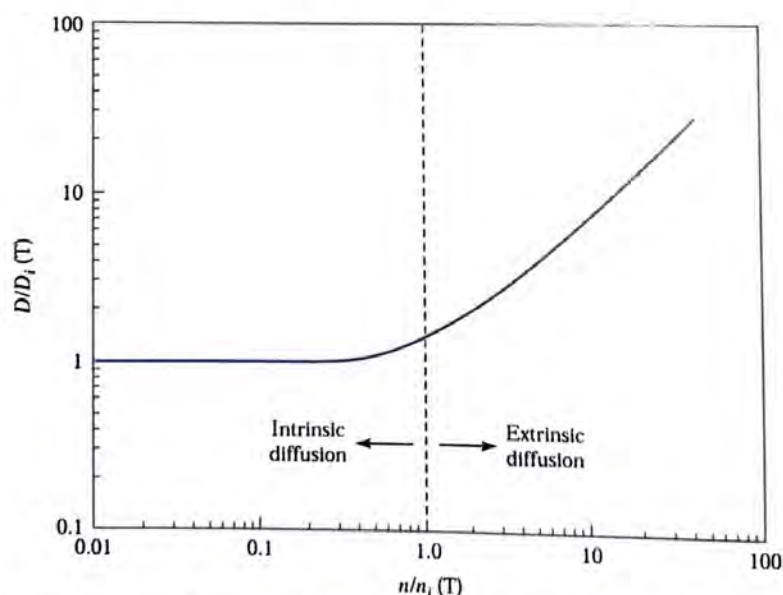


Fig. 7 Donor-impurity diffusivity versus electron concentration showing regions of intrinsic and extrinsic diffusion.¹⁰

volume, C_v) has a temperature dependence similar to that of the carrier density (see Eq. 28 in Chapter 1), that is,

$$C_v = C_i \exp\left(\frac{E_F - E_i}{kT}\right), \quad (21)$$

where C_i is the intrinsic vacancy density, E_F is the Fermi level, and E_i is the intrinsic Fermi level.

If the dopant diffusion is dominated by the vacancy mechanism, the diffusion coefficient is expected to be proportional to the vacancy density. At low doping concentrations ($n < n_i$), the Fermi level coincides with the intrinsic Fermi level ($E_F = E_i$). The vacancy density is equal to C_i and is independent of doping concentration. The diffusion coefficient, which is proportional to C_v , also is independent of doping concentration. At high concentrations ($n > n_i$), the Fermi level will move toward the conduction band edge (for donor-type vacancies), and the term $\exp[(E_F - E_i)/kT]$ becomes larger than unity. This causes C_v to increase, which in turn causes the diffusion coefficient to increase, as shown in the right side of Fig. 7.

When the diffusion coefficient varies with dopant concentration, Eq. 4 should be used as the diffusion equation instead of Eq. 5, in which D is independent of C . We consider the case where the diffusion coefficient can be written as

$$D = D_s \left(\frac{C}{C_s}\right)^\gamma, \quad (22)$$

where C_s is the surface concentration, D_s is the diffusion coefficient at the surface, C and D are the concentration and the diffusion coefficient in the bulk, and γ is a parameter to describe the concentration dependence. For such a case, we can write the diffusion equation, Eq. 4, as an ordinary differential equation and solve it numerically.

Figure 8 shows the solutions¹¹ for a constant-surface-concentration diffusion with different values of γ . For $\gamma = 0$, we have the case of constant diffusivity and the profile is the same as that shown in Fig. 5a. For $\gamma > 0$ the diffusivity decreases as the dopant concentration decreases, and increasingly steep and box-like concentration profiles result for increasing γ . Therefore, highly abrupt junctions are formed when diffusions are made into a background of an opposite impurity type due to higher impurity concentration. The abruptness of the doping profile results in a junction depth virtually independent of the background concentration. Note that the junction depth (see Fig. 8) is given by

$$x_j = 1.6\sqrt{D_s t} \quad \text{for } D \sim C \quad (\gamma = 1),$$

$$x_j = 1.1\sqrt{D_s t} \quad \text{for } D \sim C^2 \quad (\gamma = 2), \quad (23)$$

$$x_j = 0.87\sqrt{D_s t} \quad \text{for } D \sim C^3 \quad (\gamma = 3).$$

In the case of $\gamma = -2$, the diffusivity increases with decreasing concentration, which leads to a concave profile, as opposed to the convex profiles for other cases.

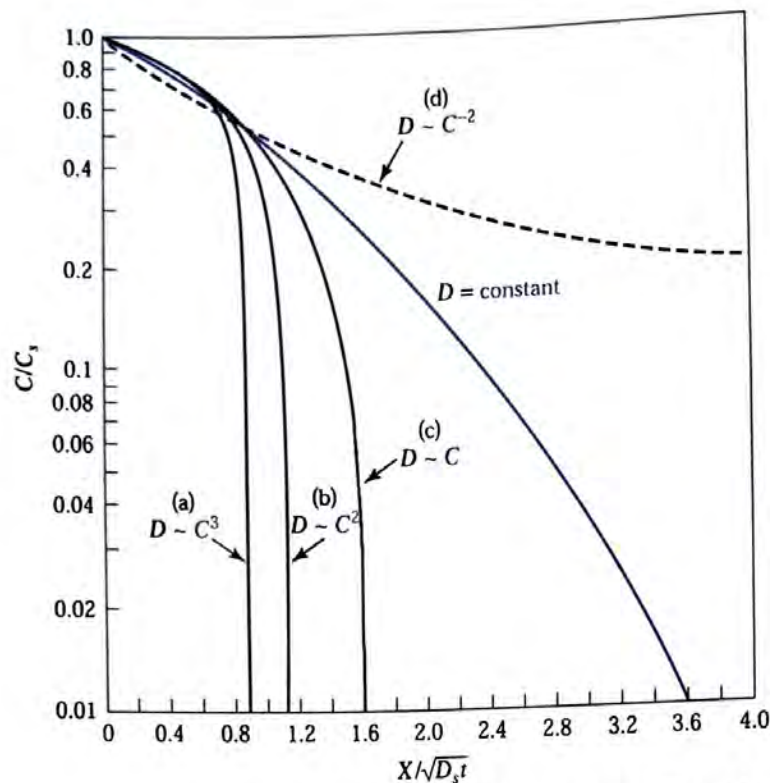


Fig. 8 Normalized diffusion profiles for extrinsic diffusion where the diffusion coefficient becomes concentration dependent.^{10,11}

14.2.2 Diffusion Profiles

Diffusion in Silicon

The measured diffusion coefficients of boron and arsenic in silicon have a concentration dependence with $\gamma \approx 1$. Their concentration profiles are abrupt, as depicted in curve *c* of Fig. 8. For gold and platinum diffusion in silicon, γ is close to -2 and their concentration profiles have the concave shape shown in curve *d* of Fig. 8.

The diffusion of phosphorus in silicon is associated with the doubly charged acceptor vacancy V^{2-} , and the diffusion coefficient at high concentration varies as C^2 . We would expect that the diffusion profile of phosphorus resembles that shown in curve *b* of Fig. 8. However, because of a *dissociation effect*, the diffusion profile exhibits anomalous behavior.

Figure 9 shows phosphorus diffusion profiles for various surface concentrations after diffusion into silicon for 1 hour at 1000°C .¹² When the surface concentration is low, corresponding to the intrinsic diffusion region, the diffusion profile is given by an erfc (curve *a*). As the concentration increases, the profile begins to deviate from the simple expression (curves *b* and *c*). At very high concentration (curve *d*), the profile near the surface is indeed similar to that in curve *b* of Fig. 8. However, at concentration n_c , a kink occurs and is followed by a rapid diffusion in the tail region. The concentration n_c corresponds to a Fermi level 0.11 eV below the conduction band. At this energy level, the coupled impurity-vacancy pair (P^+V^{2-}) dissociates to P^+ , V^- , and an electron. Thus, the dissociation generates a large number of singly charged acceptor vacancies V^- , which in turn enhances the diffusion in the tail region of the profile. The diffusivity in the tail region is over $10^{-12}\text{ cm}^2/\text{s}$, which is about two orders of magnitude larger than the intrinsic diffusivity at 1000°C . Because of its high diffusivity, phosphorus is commonly used to form deep junctions, such as the *n*-tubs in CMOS.

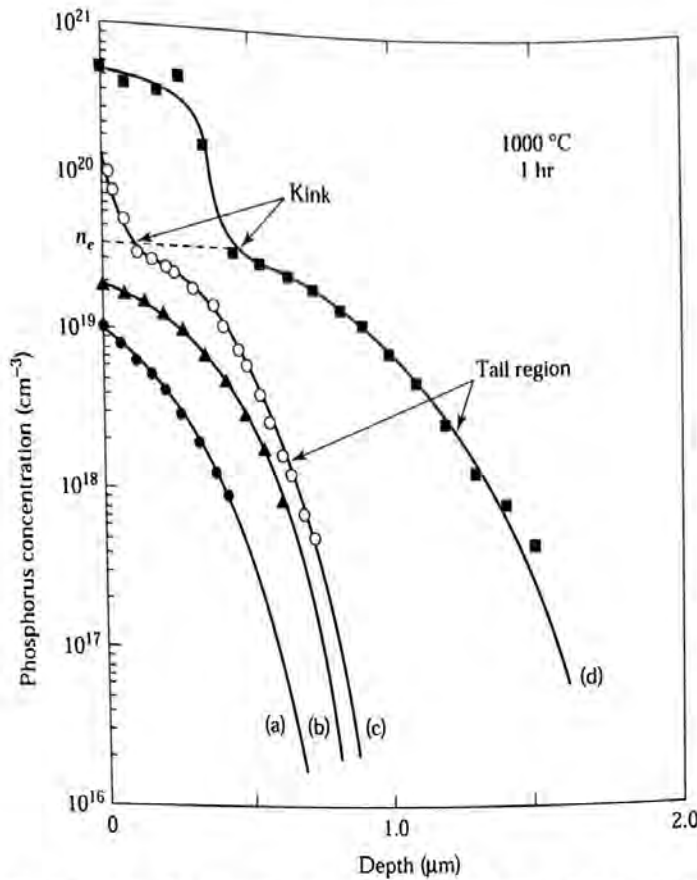


Fig. 9 Phosphorus diffusion profiles¹² for various surface concentrations after diffusion into silicon for 1 hour at 1000°C.

Zinc Diffusion in Gallium Arsenide

We expect diffusion in gallium arsenide to be more complicated than in silicon because the diffusion of impurities may involve atomic movements on both the gallium and arsenic sublattices. Vacancies play a dominant role in diffusion processes in gallium arsenide because both *p*- and *n*-type impurities must ultimately reside in lattice sites. However, the charge states of the vacancies have not been established.

Zinc is the most extensively studied diffusant in gallium arsenide. Its diffusion coefficient is found to vary as C^2 . Therefore, the diffusion profiles are steep, as shown¹³ in Fig. 10, and resemble curve *b* of Fig. 8. Note that even for the case of the lowest surface concentration, the diffusion is in the extrinsic-diffusion region, because n_i for GaAs at 1000°C is less than 10^{18} cm^{-3} . As can be seen in Fig. 10, the surface concentration has a profound effect on the junction depth. The diffusivity varies linearly with the partial pressure of the zinc vapor, and the surface concentration is proportional to the square root of the partial pressure. Therefore, from Eq. 23, the junction depth is linearly proportional to the surface concentration.

Diffusion in Strained Silicon

Strained silicon is a promising candidate for the channel of MOSFET due to the high mobility of the carriers.^{14,15} In the meantime, strain can alter the activation energy of many major steps involved in dopant diffusion, including formation of native defects and displacement of a dopant atom to form a mobile dopant complex. Strain-related bandgap narrowing can also change the charged point-defect concentration. Therefore, the diffusion associated with the defect concentration is a strong function of the strain. It impacts junction depth and effective channel length. The strain-related effects are strong for MOSFETs with scale less than 65 nm.

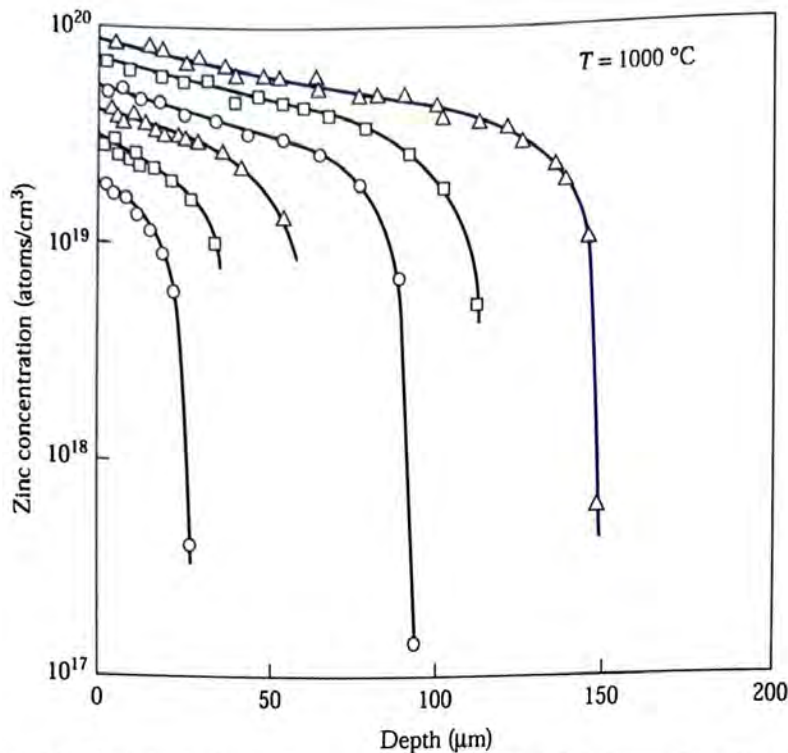


Fig. 10 Diffusion profiles¹³ of zinc in GaAs after annealing at 1000°C for 2.7 hours. The different surface concentrations are obtained by maintaining the Zn source at temperatures in the range of 600°-800°C.

Under compressive stress, the lattice constant of Si near the surface becomes smaller. To relieve it, Si atoms near the surface jump to the surface, and vacancies are created near the surface. These vacancies diffuse into the bulk and recombine with self-interstitials to decrease self-interstitials. As discussed in Section 14.1.1, since both phosphorus and boron diffuse mainly through an interstitialcy mechanism, the compressive stress results in retarded diffusion.¹⁶ Tensile stress has opposite effects to compressive stress: the lattice constant of Si becomes larger and the interstitial-mediated diffusion is enhanced. The vacancy concentration is lower due to the lattice relaxation and the vacancy-mediated diffusion is retarded.

► 14.3 DIFFUSION-RELATED PROCESSES

In this section we consider two processes in which diffusion plays an important role and the impact of these processes on device performance.

14.3.1 Lateral Diffusion

The one-dimensional diffusion equation discussed previously can describe the diffusion process satisfactorily except at the edge of the mask window. Here the impurities will diffuse downward and sideways (i.e., laterally). In this case, we must consider a two-dimensional diffusion equation and use a numerical technique to obtain the diffusion profiles under different initial and boundary conditions.

Figure 11 shows the contours of constant doping concentration for a constant-surface-concentration diffusion condition assuming that the diffusivity is independent of concentration.¹⁷ At the far right of the figure, the variation of the dopant concentration from 0.5 to $10^{-4} C/C_s$ (where C_s is the surface concentration) corresponds to the erfc distribution given by Eq. 9. The contours are in effect a map of the locations of the junctions created by diffusing into various background concentrations. For example, at $C/C_s = 10^{-4}$ (i.e., the background doping is 10^4 times lower than the surface concentration), we see from this constant-concentration curve that the vertical penetration is about 2.8 μm , whereas the lateral penetration (i.e., the penetration along the diffusion

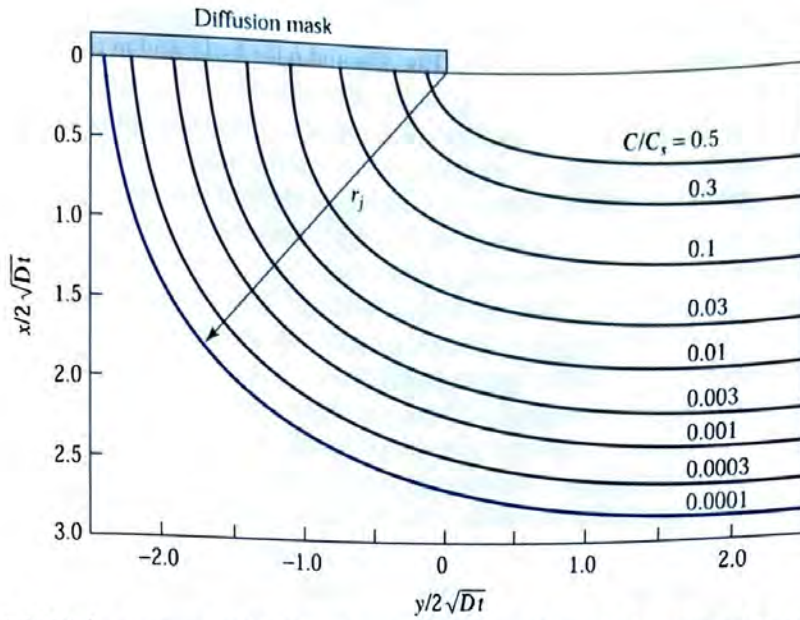


Fig. 11 Diffusion contours at the edge of an oxide window, where r_j is the radius of curvature.¹⁷

mask-semiconductor interface) is about $2.3 \mu\text{m}$. Therefore, the lateral penetration is about 80% of the penetration in the vertical direction for concentrations three or more orders of magnitude below the surface concentration. Similar results are obtained for a constant-total-dopant diffusion condition. The ratio of lateral to vertical penetration is about 75%. For concentration-dependent diffusivities, the ratio is found to be reduced slightly, to about 65%-70%.

Because of the lateral-diffusion effect, the junction consists of a central plane (or flat) region with approximately cylindrical edges with a radius of curvature r_j , as shown in Fig. 11. In addition, if the diffusion mask contains sharp corners, the shape of the junction near the corner will be roughly spherical because of lateral diffusion. Since the electric-field intensities are higher for cylindrical and spherical junction regions, the avalanche breakdown voltages of such regions can be substantially lower than that of a plane junction having the same background doping. This junction "curvature effect" was discussed in Chapter 3.

14.3.2 Impurity Redistribution During Oxidation

Dopant impurities near the silicon surface will be redistributed during thermal oxidation. The redistribution depends on several factors. When two solid phases are brought together, an impurity in one solid will redistribute between the two solids until it reaches equilibrium. This is similar to our previous discussion on impurity redistribution in crystal growth from the melt. The ratio of the equilibrium concentration of the impurity in the silicon to that in the silicon dioxide is called the *segregation coefficient* and is defined as

$$k = \frac{\text{equilibrium concentration of impurity in silicon}}{\text{equilibrium concentration of impurity in SiO}_2}. \quad (24)$$

A second factor that influences impurity distribution is that the impurity may diffuse rapidly through the silicon dioxide and escape to the gaseous ambient. If the diffusivity of the impurity in silicon dioxide is large, this factor will be important. A third factor in the redistribution process is that the oxide is growing, and thus the boundary between the silicon and the oxide advances into the silicon as a function of time. The relative rate of this advance compared with the diffusion rate of the impurity through the oxide is important in determining the extent of the redistribution. Note that even if the segregation coefficient of an impurity k equals unity, some redistribution of the impurity in the silicon will still take place. As indicated in Fig. 3 of Chapter 12, the oxide layer will be about twice as thick as the silicon layer it replaced. Therefore, the same amount of impurity will now be distributed in a larger volume, resulting in a depletion of the impurity from the silicon.

Four possible redistribution processes are illustrated⁶ in Fig. 12. These processes can be classified into two groups. In one group the oxide takes up the impurity (Fig. 12a and b for $k < 1$), and in the other the oxide rejects the impurity (Fig. 12c and d for $k > 1$). In each case, what happens depends on how rapidly the impurity can diffuse through the oxide. In group 1, the silicon surface is depleted of impurities; an example is boron with k approximately equal to 0.3. Rapid diffusion of the impurity through the silicon dioxide increases the amount of depletion; an example is boron-doped silicon heated in a hydrogen ambient because hydrogen in silicon dioxide enhances the diffusivity of boron. In group 2, k is greater than unity, so that the oxide rejects the impurity. If diffusion of the impurity through the silicon dioxide is relatively slow, the impurity piles up near the silicon surface; an example is phosphorus, with k approximately equal to 10. When diffusion through the silicon dioxide is rapid, so much impurity may escape from the solid to the gaseous ambient that the overall effect will be a depletion of the impurity; an example is gallium, with k approximately equal to 20.

The redistributed dopant impurities in silicon dioxide are seldom electrically active. However, redistribution in silicon has an important effect on processing and device performance. For example, nonuniform dopant distribution will modify the interpretation of the measurements of interface-trap properties (see Chapter 6), and a change in the surface concentration will modify the threshold voltage and device contact resistance (see Chapter 7).

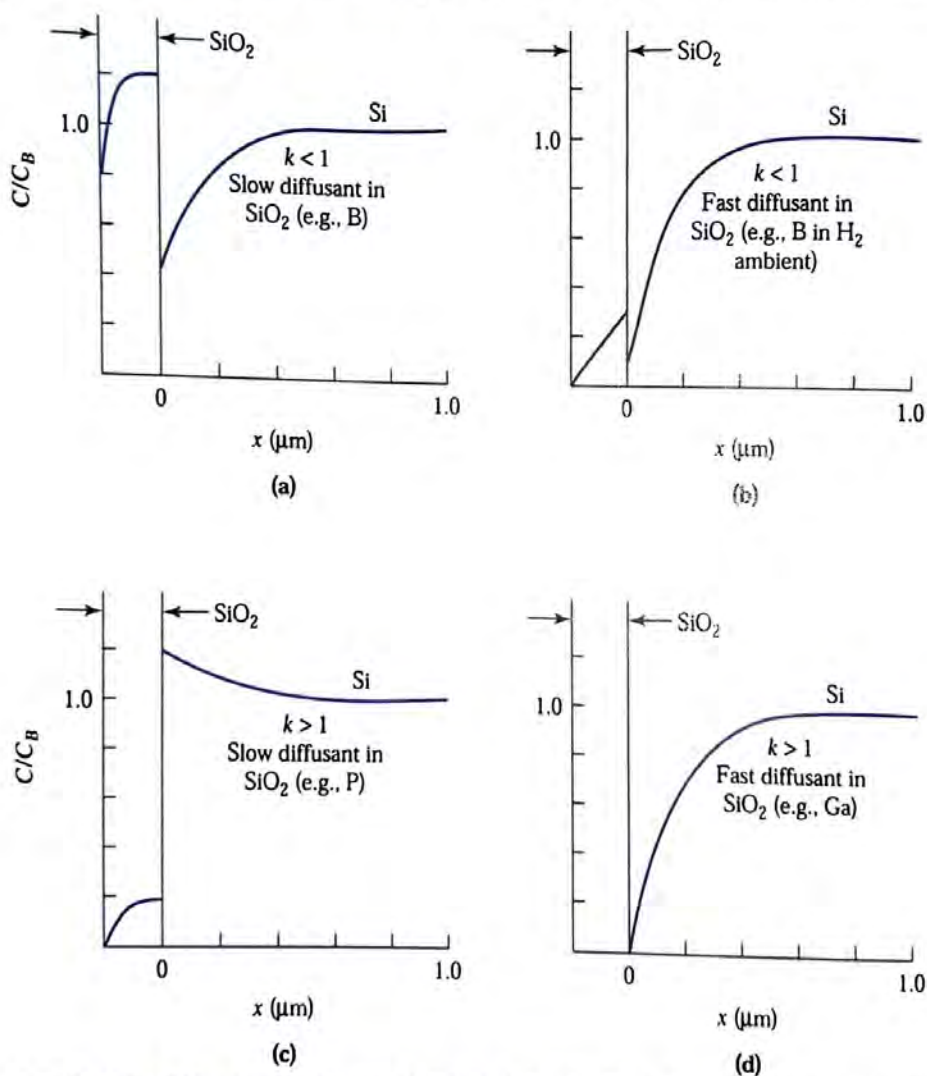


Fig. 12 Four different cases of impurity redistribution in silicon due to thermal oxidation.⁶

► 14.4 RANGE OF IMPLANTED IONS

Ion implantation is the introduction of energetic charged particles into a substrate such as silicon. Implantation energies are between 300 eV and 5 MeV, resulting in ion distributions with average depths ranging from 10 nm to 10 μm . Ion doses vary from 10^{12} ions/ cm^2 for threshold voltage adjustment to 10^{18} ions/ cm^2 for the formation of a buried insulating layer. Note that the dose is expressed as the number of ions implanted into 1 cm^2 of the semiconductor surface area. The main advantages of ion implantation are its more precise control and reproducibility of impurity dopings and its lower processing temperature than those of the diffusion process.

Basic CMOS processes usually use fifteen to seventeen ion implants per wafer. Current leading-edge CMOS processes use 20 to 23 implants, and specialized CMOS circuits (e.g., flash memory) use up to 30 implantation steps. Virtually all doping in modern CMOS devices is accomplished by ion implantation; no other technique offers comparable process control and repeatability for both the amount and position of the doping.

Figure 13 shows schematically a medium-energy ion implanter.¹⁸ The ion source has a heated filament to break up a source gas such as BF_3 or AsH_3 into charged ions (B^+ or As^+). An extraction voltage, around 40 kV, causes the charged ions to move out of the ion source chamber into a mass analyzer. The magnetic field of the analyzer is chosen such that only ions with the desired mass-to-charge ratio can travel through it without being filtered. The selected ions then enter the acceleration tube, where they are accelerated to the implantation energy as they move from high voltage to ground. Apertures ensure that the ion beam is well collimated. The pressure in the implanter is kept below 10^{-4} Pa to minimize ion scattering by gas molecules. The ion beam is then scanned over the wafer surface using electrostatic deflection plates and is implanted into the semiconductor substrate.

The energetic ions lose their energies through collision with electrons and nuclei in the substrate and finally come to rest at some depth within the lattice. The average depth can be controlled by adjusting the acceleration energy. The dopant dose can be controlled by monitoring the ion current during implantation. The principal side effect is the disruption or damage of the semiconductor lattice due to ion collisions. Therefore, a subsequent annealing treatment is needed to remove these damages.

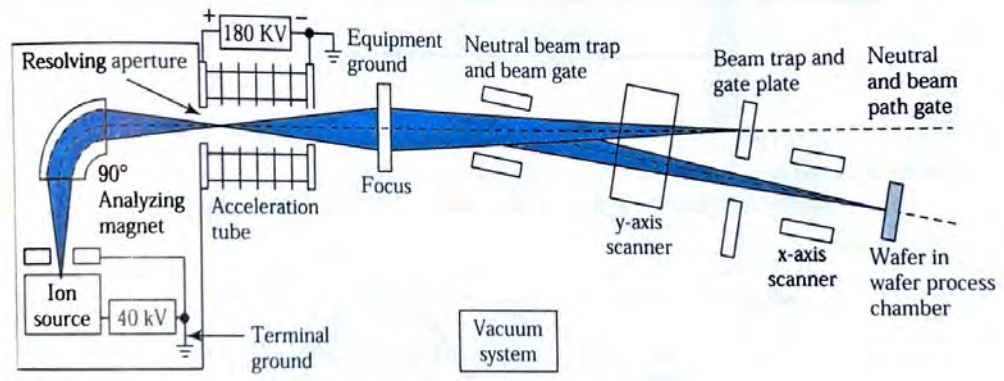


Fig. 13 Schematic of a medium-current ion implanter.

14.4.1 Ion Distribution

The total distance that an ion travels in coming to rest is called its *range* R and is illustrated¹⁹ in Fig. 14*a*. The projection of this distance along the axis of incidence is called the *projected range* R_p . Because the number of collisions per unit distance and the energy lost per collision are random variables, there will be a spatial distribution of ions having the same mass and the same initial energy. The statistical fluctuations in the projected range are called the *projected straggle* σ_p . There is also a statistical fluctuation along an axis perpendicular to the axis of incidence, called the *lateral straggle* σ_l .

Figure 14*b* shows the ion distribution. Along the axis of incidence, the implanted impurity profile can be approximated by a Gaussian distribution function:

$$n(x) = \frac{S}{\sqrt{2\pi}\sigma_p} \exp\left[-\frac{(x - R_p)^2}{2\sigma_p^2}\right],$$

(25)

where S is the ion dose per unit area. This equation is similar to Eq. 14 for constant-total-dopant diffusion, except that the quantity $4Dt$ is replaced by $2\sigma_p^2$ and the distribution is shifted along the x -axis by R_p .

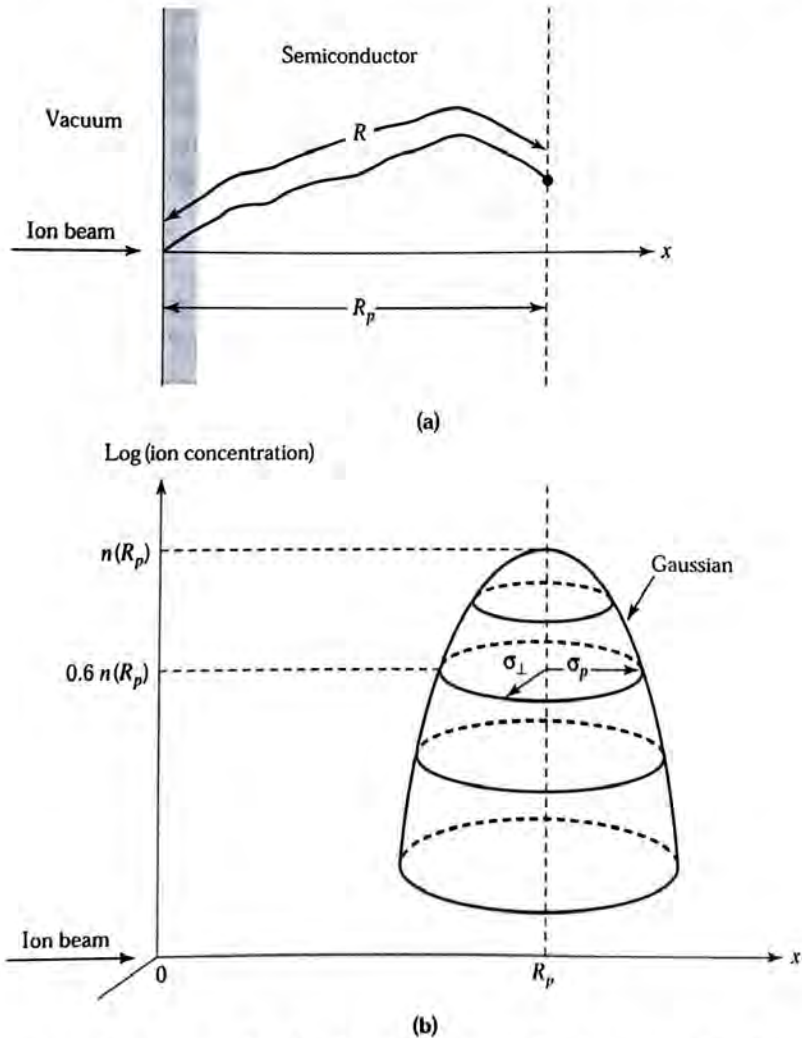


Fig. 14 (a) Schematic of the ion range R and projected range σ_p . (b) Two-dimensional distribution of the implanted ions.¹⁹

Thus, for diffusion, the maximum concentration is at $x = 0$, whereas for ion implantation the maximum concentration is at the projected range R_p . The ion concentration is reduced by 40 % from its peak value at $(x - R_p) = \pm \sigma_p$, by one decade at $\pm 2\sigma_p$, by two decades at $\pm 3\sigma_p$, and by five decades at $\pm 4.8 \sigma_p$.

Along the axis perpendicular to the axis of incidence, the distribution is also a Gaussian function of the form $\exp(-y^2/2\sigma_1^2)$. Because of this distribution, there will be some lateral implantation.²⁰ However, the lateral penetration from the mask edge (on the order of σ_1) is considerably smaller than that from the thermal diffusion process discussed in Section 14.3.

14.4.2 Ion Stopping

There are two stopping mechanisms by which an energetic ion, on entering a semiconductor substrate (also called the target), can be brought to rest. The first is by transferring its energy to the target nuclei. This causes deflection of the incident ion and also dislodges many target nuclei from their original lattice sites. If E is the energy of the ion at any point x along its path, we can define a nuclear stopping power $S_n(E) \equiv (dE/dx)_n$ to characterize this process. The second stopping mechanism is the interaction of the incident ion with the cloud of electrons surrounding the target's atoms. The ion loses energy in collisions with electrons through Coulombic interaction. The electrons can be excited to higher energy levels (excitation), or they can be ejected from the atom (ionization). We can define an electronic stopping power $S_e(E) \equiv (dE/dx)_e$ to characterize this process.

The average rate of energy loss with distance is given by a superposition of the above two stopping mechanisms:

$$\frac{dE}{dx} = S_n(E) + S_e(E). \quad (26)$$

If the total distance traveled by the ion before coming to rest is R , then

$$R = \int_0^R dx = \int_0^{E_0} \frac{dE}{S_n(E) + S_e(E)}, \quad (27)$$

where E_0 is the initial ion energy. The quantity R has been defined previously as the range.

We can visualize the nuclear stopping process by considering the elastic collision between an incoming hard sphere (energy E_0 and mass M_1) and a target hard sphere (initial energy zero and mass M_2),

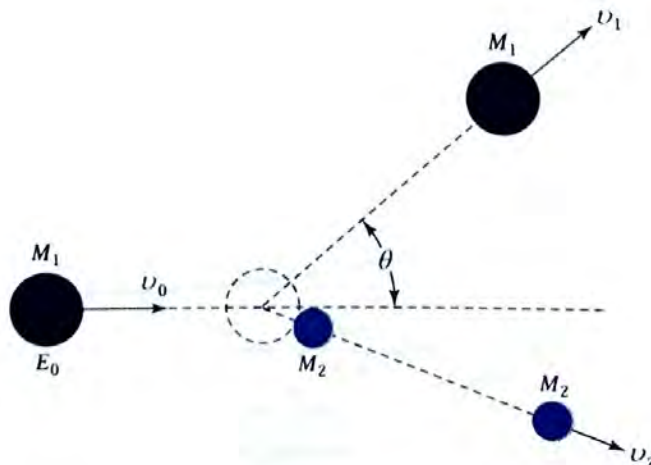


Fig. 15 Collision of hard spheres.

as illustrated in Fig. 15. When the spheres collide, momentum is transferred along the centers of the spheres. The deflection angle θ and the velocities v_1 and v_2 can be obtained from the requirements for conservation of momentum and energy. The maximum energy loss is in a head-on collision. For this case, the energy loss by the incident particle M_1 or the energy transferred to M_2 is

$$\frac{1}{2} M_2 v_2^2 = \left[\frac{4 M_1 M_2}{(M_1 + M_2)^2} \right] E_0. \quad (28)$$

Since M_2 is usually of the same order of magnitude as M_1 , a large amount of energy can be transferred in the nuclear stopping process.

Detailed calculations show that the nuclear stopping power increases linearly with energy at low energies (similar to Eq. 28), and $S_n(E)$ reaches a maximum at some intermediate energy. At high energies, $S_n(E)$ becomes smaller because fast particles may not have sufficient interaction time with the target atoms to achieve effective energy transfer. The calculated values of $S_n(E)$ for arsenic, phosphorus, and boron in silicon at various energies are shown in Fig. 16 (solid line, where the superscript indicates the atomic weight).²¹ Note that heavier atoms, such as arsenic, have larger nuclear stopping power, that is, larger energy loss per unit distance.

The electronic stopping power is found to be proportional to the velocity of the incident ion, or

$$S_e(E) = k_e \sqrt{E} \quad (29)$$

where the coefficient k_e is a relatively weak function of atomic mass and atomic number. The value of k_e is approximately 10^7 (eV)^{1/2}/cm for silicon and 3×10^7 (eV)^{1/2}/cm for gallium arsenide. The electronic stopping power in silicon is plotted in Fig. 16 (dotted line). Also shown in the figure are the crossover energies at which $S_n(E)$

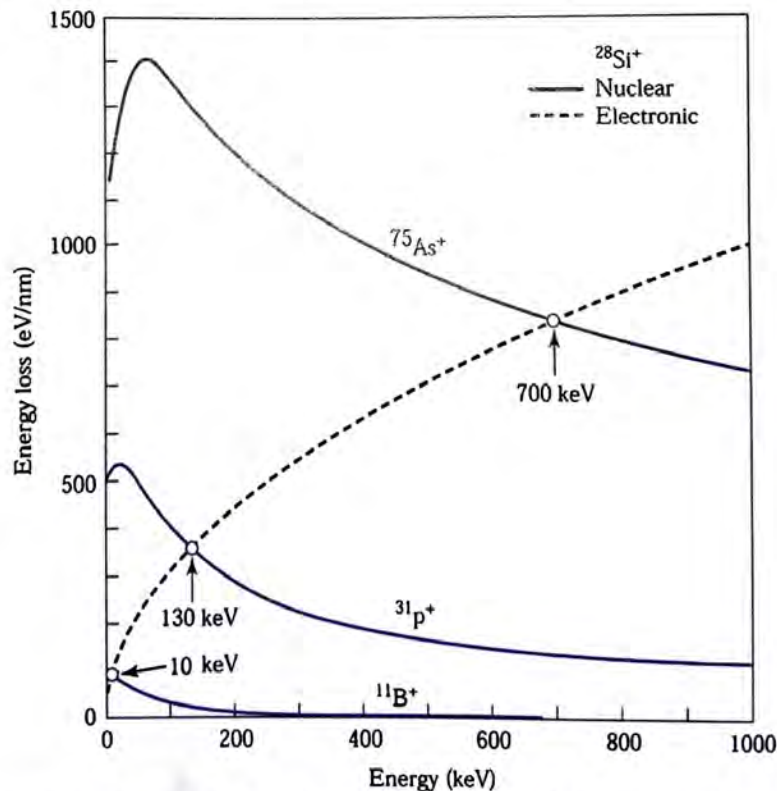


Fig. 16 Nuclear stopping power $S_n(E)$, and electronic stopping power $S_e(E)$ for As, P, and B in Si. The points of intersection of the curves correspond to the energy at which nuclear and electronic stopping are equal.²¹

equal $S_n(E)$. For boron, which has a relatively low ion mass compared with the target silicon atom, the crossover energy is only 10 keV. This means that over most of the implantation energy range of 1 keV to 1 MeV, the main energy loss mechanism is electronic stopping. On the other hand, for arsenic with relatively high ion mass, the crossover energy is 700 keV. Thus, nuclear stopping dominates over most of the energy range. For phosphorus, the crossover energy is 130 keV. For an E_0 less than 130 keV, nuclear stopping will dominate; for higher energies, electronic stopping will take over.

Once $S_n(E)$ and $S_e(E)$ are known, we can calculate the range from Eq. 27. This in turn can give us the projected range and projected straggle with the help of the following approximate equations¹⁸:

$$R_p \cong \frac{R}{1 + (M_2 / 3M_1)}, \quad (30)$$

$$\sigma_p \cong \frac{2}{3} \left[\frac{\sqrt{M_1 M_2}}{M_1 + M_2} \right] R_p. \quad (31)$$

Figure 17a shows the projected range (R_p), the projected straggle (σ_p), and the lateral straggle (σ_L) for arsenic, boron, and phosphorus in silicon.²² As expected, the larger the energy loss, the smaller the range. Also, the projected range and straggles increase with ion energy. For a given element at a specific incident energy, σ_p and σ_L are comparable and usually within $\pm 20\%$. Figure 17b shows the corresponding values for hydrogen, zinc, and tellurium in GaAs.

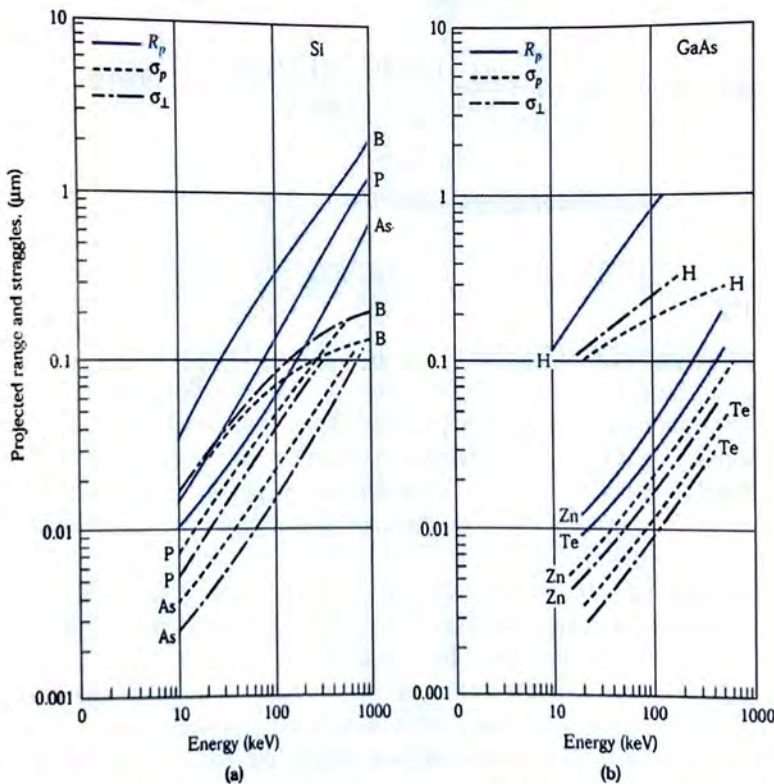


Fig. 17 Projected range, projected straggle, and lateral straggle for (a) B, P, and As in Si, and (b) H, Zn, and Te in GaAs.^{20,22}

tellurium in gallium arsenic.²⁰ Comparing Fig. 17a with Fig. 17b, we see that most of the popular dopants (except hydrogen) have larger projected ranges in silicon than they have in gallium arsenic.

► EXAMPLE 3

Assume 100 keV boron implants on a 200 mm silicon wafer at a dose of 5×10^{14} ions/cm². Calculate the peak concentration and the required ion beam current for 1 minute of implantation.

SOLUTION From Fig. 17a, we obtain 0.31 and 0.07 μm for the projected range and project straggle, respectively.

$$\text{From Eq. 25, } n(x) = \frac{S}{\sqrt{2\pi}\sigma_p} \exp\left[-\frac{(x - R_p)^2}{2\sigma_p^2}\right],$$

$$\frac{dn}{dx} = -\frac{S}{\sqrt{2\pi}\sigma_p} \frac{2(x - R_p)}{2\sigma_p^2} \exp\left[-\frac{(x - R_p)^2}{2\sigma_p^2}\right] = 0.$$

The peak concentration is at $x = R_p$, and $n(x) = 2.85 \times 10^{19}$ ions/cm³.

The total number of implanted ions = $Q = 5 \times 10^{14} \times \pi \times \left(\frac{20}{2}\right)^2 = 1.57 \times 10^{17}$ ions.

The required ion current = $I = \frac{qQ}{t} = \frac{1.6 \times 10^{-19} \times 1.57 \times 10^{17}}{60} = 4.19 \times 10^{-4}$ A.

$$= 0.42 \text{ mA.}$$

14.4.3 Ion Channeling

The projected range and straggle of the Gaussian distribution discussed previously give a good description of the implanted ions in amorphous or fine-grained polycrystalline substrates. Both silicon and gallium arsenide behave as if they were amorphous semiconductors, provided the ion beam is misoriented from the low-index crystallographic direction (e.g., $\langle 111 \rangle$). In this situation, the doping profile described by Eq. 25 is followed closely near the peak and extended to one or two decades below the peak value. This is illustrated¹⁹ in Fig. 18. However, even for a misorientation of 7° from the $\langle 111 \rangle$ -axis, there still is a tail that varies exponentially with distance as $\exp(-x/\lambda)$, where λ is typically on the order 0.1 μm .

The exponential tail is related to the ion-channeling effect. Channeling occurs when incident ions align with a major crystallographic direction and are guided between rows of atoms in a crystal. Figure 19 illustrates a diamond lattice viewed along a $\langle 110 \rangle$ -direction.²³ Ions implanted in the $\langle 110 \rangle$ -direction will follow trajectories that will not bring them close enough to a target atom to lose significant amounts of energy in nuclear collisions. Thus, for channeled ions, the only energy loss mechanism is electronic stopping and the range of channeled ions can be significantly larger than it would be in an amorphous target. Ion channeling is particularly critical for low-energy implant and heavy ions.

Channeling can be minimized by several techniques: a blocking amorphous surface layer, wafer misorientation, and creating a damage layer in the wafer surface. The usual blocking amorphous layer is simply a thin layer of grown silicon dioxide (Fig. 20a). The layer randomizes the direction of the ion beam so that the

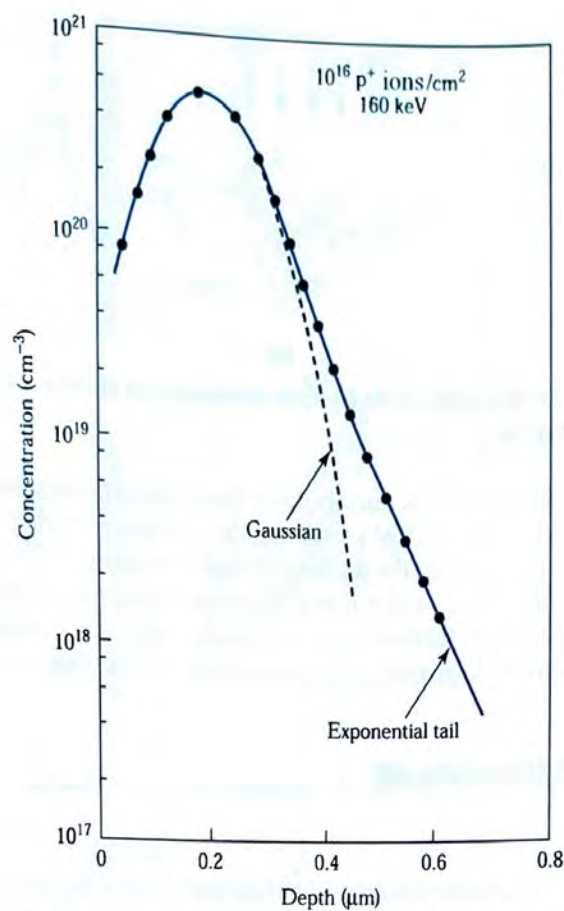


Fig. 18 Impurity profile obtained in a purposely misoriented target. Ion beam is incident 7° from the <111>-axis.¹⁹

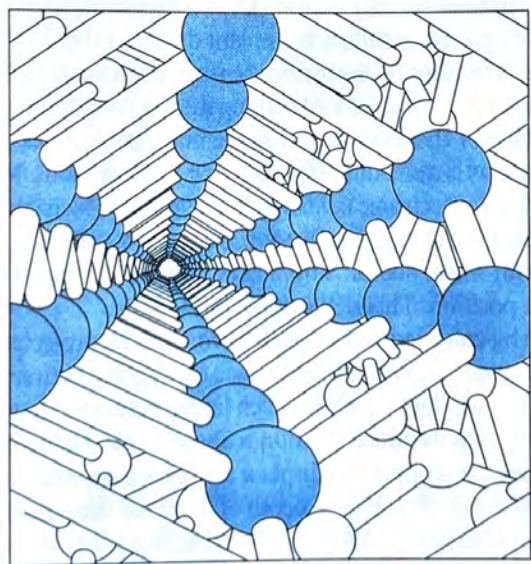


Fig. 19 Model for a diamond structure, viewed along a <110> axis.²³

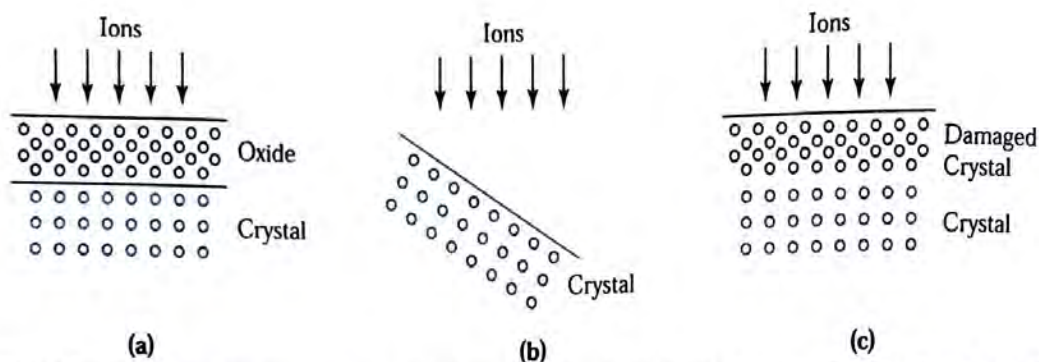


Fig. 20 (a) Implant through an amorphous oxide layer, (b) misorientation of the beam direction to all crystal axes, and (c) predamage on the crystal surface.

ions enter the wafer at different angles and not directly down the crystal channels. Misorientation of the wafers 5° – 10° from the major plane also has the effect of preventing the ions from entering the channels (Fig. 20b). With this method, most implantation machines tilt the wafer by 7° and then apply a 22° twist from the flat to prevent channeling. Predamaging the wafer surface with a heavy silicon or germanium implant creates a randomizing layer in the wafer surface (Fig. 20c). This method, however, increases the use of the expensive ion implanter and produces point defects that become leakage paths during subsequent processing.

► 14.5 IMPLANT DAMAGE AND ANNEALING

14.5.1 Implant Damage

When energetic ions enter a semiconductor substrate, they lose their energy in a series of nuclear and electronic collisions and finally come to rest. The electronic-energy loss can be accounted for in terms of electronic excitations to higher energy levels or of the generation of electron-hole pairs. However, electronic collisions do not displace semiconductor atoms from their lattice positions. Only nuclear collisions can transfer sufficient energy to the lattice that host atoms are displaced, resulting in implant damage (also called lattice disorder).²⁴ These displaced atoms may possess large fractions of the incident energy, and can in turn cause cascades of secondary displacement of nearby atoms to form a *tree of disorder* along the ion path. When the displaced atoms per unit volume approach the atomic density of the semiconductor, the material becomes amorphous.

The tree of disorder for light ions is quite different from that for heavy ions. Much of the energy loss for light ions, $^{11}\text{B}^+$ in silicon (e.g.) is due to electronic collisions (see Fig. 16), which do not cause lattice damage. The ions lose their energies as they penetrate deeper into the substrate. Eventually, the ion energy is reduced below the crossover energy (10 keV for boron) where nuclear stopping becomes dominant. Therefore, most of the lattice disorder occurs near the final ion position. This is illustrated in Fig. 21a.

We can estimate the damage by considering a 100 keV boron ion. Its projected range is $0.31\text{ }\mu\text{m}$, (Fig. 17a), and its initial nuclear energy loss is only $3\text{ eV}/\text{\AA}$ (Fig. 16). Since the spacing between lattice planes in silicon is about $2.5\text{ }\text{\AA}$, this means that the boron ion will lose 7.5 eV for each lattice plane because of nuclear stopping. The energy required to displace a silicon atom from its lattice position is about 15 eV . Therefore, the incident boron ion does not release enough energy from nuclear stopping to displace a silicon atom when it first enters the silicon substrate. When the ion energy is reduced to about 50 keV (at a depth of $1500\text{ }\text{\AA}$), the energy loss due to nuclear stopping increases to 15 eV for each lattice plane (i.e., $6\text{ eV}/\text{\AA}$), sufficient to create a lattice disorder. Assuming that one atom is displaced per lattice plane for moves roughly $25\text{ }\text{\AA}$ from its original position, the damage volume is

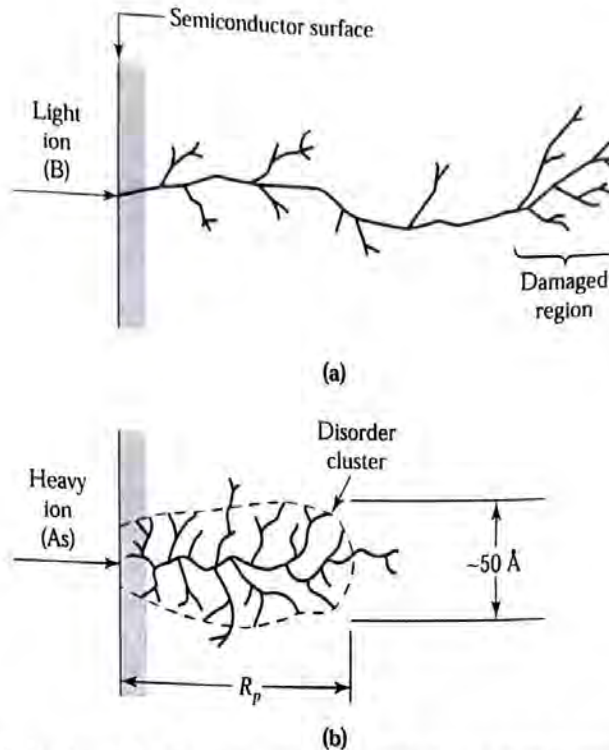


Fig. 21 Implantation disorder caused by (a) light ions and (b) heavy ions.^{2,18}

given by $V_D \equiv \pi (25 \text{ Å})^2 (1500 \text{ Å}) = 3 \times 10^{-18} \text{ cm}^3$. The damage density is $600/V_D = 2 \times 10^{20} \text{ cm}^{-3}$, which is only 0.4% of the atoms. Thus, very high doses of light ions are needed to create an amorphous layer.

For heavy ions, the energy loss is primarily due to nuclear collisions; therefore, we expect substantial damage. Consider a 100 keV arsenic ion with a projected range of 0.06 μm or 60 nm. The average nuclear energy loss over the entire energy range is about 1320 eV/nm (Fig. 16). This means that the arsenic ion loses about 330 eV for each lattice plane on the average. Most of the energy is given to one primary silicon atom. Each primary atom will subsequently displace 22 target atoms (i.e., 330 eV/15 eV). The total number of displaced atoms is 5280. Assuming a range of 2.5 nm for the displaced atoms, the damage volume is $V_D \equiv \pi (25 \text{ nm})^2 (60 \text{ nm}) = 10^{-18} \text{ cm}^3$. The damage density is then $5280/V_D \equiv 5 \times 10^{21} \text{ cm}^{-3}$, or about 10% of the total number of atoms in V_D . As a result of the heavy-ion implantation, the material has become essentially amorphous. Figure 21b illustrates the situation where the damage forms a disordered cluster over the entire projected range.

To estimate the dose required to convert a crystalline material to an amorphous form, we can use the criterion that the energy density is of the same order of magnitude as that needed for melting the material (i.e., 10^{21} keV/cm^3). For 100 keV arsenic ions, the dose required to make amorphous silicon is then

$$S = \frac{(10^{21} \text{ keV/cm}^3) R_p}{E_0} = 6 \times 10^{13} \text{ ions/cm}^2, \quad (32)$$

For 100 keV boron ions, the dose required is $3 \times 10^{14} \text{ ions/cm}^2$ because R_p for boron is five times larger than for arsenic. However, in practice, higher doses ($>10^{16} \text{ ions/cm}^2$) are required for boron implantation into a target at room temperature because of the nonuniform distribution of the damage along the ion path.

14.5.2 Annealing

Because of the damaged region and the disorder cluster that result from ion implantation, semiconductor parameters such as mobility and lifetime are severely degraded. In addition, most of the ions as implanted are not located in substitutional sites. To activate the implanted ions and restore mobility and other material parameters,

we must anneal the semiconductor at an appropriate combination of time and temperature. Annealing is a heat treatment that alters the microstructure of a material, causing changes in properties.

In conventional annealing, we use an open-tube batch-furnace system similar to that for thermal oxidation. The wafers are in an isothermal environment: the furnace walls are at the same temperature as the wafers. This process requires long time and high temperature to remove the implant damages. However, conventional annealing may cause substantial dopant diffusion and cannot meet the requirement for shallow junctions and narrow doping profiles. Rapid thermal annealing (RTA) is an annealing process that employs a variety of energy sources with a wide range of times, from 100 seconds down to nanoseconds—all short compared with conventional annealing. RTA can activate the dopant fully with minimal redistribution.

Conventional Annealing of B and P

Annealing characteristics depend on the dopant type and the dose involved. Figure 22 shows the annealing behaviors of boron and phosphorus implantation into silicon substrates.²² The substrate is held at room temperature (T_s) during implantation. For a given ion dose, the annealing temperature is defined as the temperature at which 90% of the implanted ions are activated by a 30-minute annealing in a conventional annealing furnace. For boron implantation, higher annealing temperatures are needed for higher doses. For phosphorus at lower doses, the annealing behavior is similar to that for boron. However, when the dose is greater than 10^{15} cm^{-2} , the annealing temperature drops to about 600°C . This phenomenon is related to the solid-phase epitaxy process. At phosphorus doses greater than $6 \times 10^{14} \text{ cm}^{-2}$ the silicon surface layer becomes amorphous. The single-crystal semiconductor underneath the amorphous layer serves as a seeding area for recrystallization of the amorphous layer. The epitaxial-growth rate along the $\langle 100 \rangle$ direction is 10 nm/min at 550°C and 50 nm/min at 600°C , with an activation energy at 2.4 eV . Therefore, a $100\text{--}500 \text{ nm}$ amorphous layer can be recrystallized in a few minutes. During the solid-phase epitaxial process, the impurity dopant atoms are incorporated into the lattice sites along with the host atoms; thus, full activation can be obtained at relatively low temperatures.

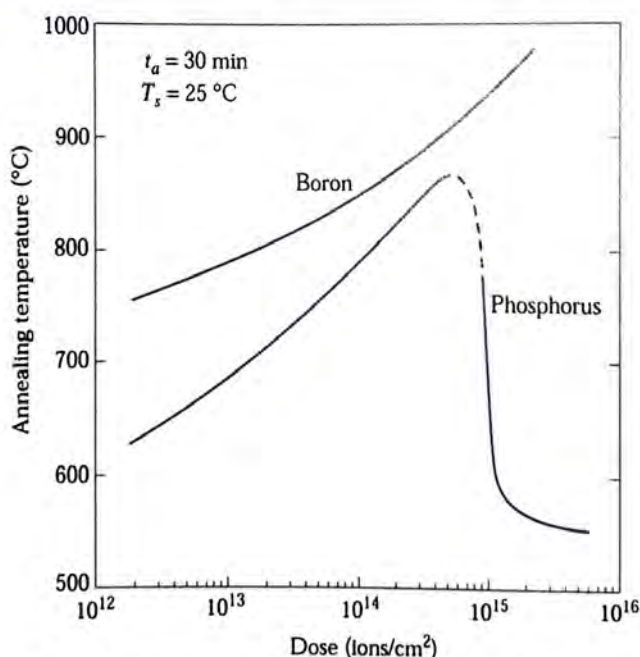


Fig. 22 Annealing temperature versus dose for 90% activation of boron and phosphorus.

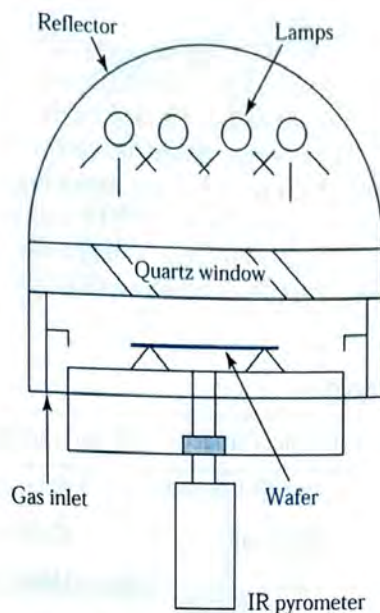


Fig. 23 Rapid thermal annealing system that is optically heated.

Rapid Thermal Annealing

A rapid thermal annealing system with transient lamp heating is shown in Fig. 23. Typical lamps in a RTA system are tungsten filaments or arc lamps. The processing chamber is made of quartz, silicon carbide, stainless steel, or aluminum, and has quartz windows through which the optical radiation passes to illuminate the wafer. The wafer holder is often made of quartz and contacts the wafer in a minimum number of places. A measurement system is placed in a control loop to set wafer temperature. The RTA system interfaces with a gas-handling system and a computer that controls system operation. Typically, wafer temperature in a RTA system is measured with a noncontact optical pyrometer that determines wafer's temperature.

In RTA, wafer is heated quickly under atmospheric conditions or at low pressure. However, the wafer is not in thermal equilibrium with its environment. The tungsten-halogen lamps ($1500\sim 2500\text{ }^{\circ}\text{C}$) are much hotter than the wafers ($600\sim 1100\text{ }^{\circ}\text{C}$)^{25,26} and the chamber walls ($20\sim 500\text{ }^{\circ}\text{C}$) are usually much cooler than the wafer. It is these temperature differences that permit rapid heating and cooling of the wafer. Because of the high temperatures of both the wafer and the lamps, the physics of RTA is dominated by radiation heat transfer, and the optical properties of the wafer and the chamber play an important part in the behavior.

One key advantage of RTA for ion implanted layers is the ability to reduce transient-enhanced diffusion. Transient-enhanced diffusion is the very large increase in dopant diffusivity in ion-implanted silicon, which is from the large excess of point defects that result from the ion implantation process. The phenomenon is especially severe for boron doping, since boron is already a fast diffuser, and its diffusivity is increased by silicon interstitials. Transient-enhanced diffusion effects are more severe at lower temperatures, because the degree of the excess of silicon interstitials over the equilibrium value (super-saturation) is greater at lower temperatures than higher ones. As a result, annealing at higher temperatures reduces the effects of transient-enhanced diffusion, so long as the heating cycle can be kept sufficiently short.

A comparison between conventional furnace and RTA technologies is shown in Table 2. To achieve short processing times using RTA, trade-offs must be made in temperature and process uniformity, temperature measurement and control, and wafer stress and throughput. In addition, there are concerns over the introduction of electrically active wafer defects during the very fast ($100^{\circ}\sim 300^{\circ}\text{C/s}$) thermal transients. Rapid heating with temperature gradients in the wafers can cause wafer damage in the form of slip dislocations induced by thermal stress. On the other hand, conventional furnace processing brings with it significant problems, such as particle generation from the hot walls, limited ambient control in an open system, and a large thermal mass that restricts controlled heating times to tens of minutes. In fact, requirements on contamination, process control, and cost of manufacturing floor space have resulted in the paradigm shift to the RTA process.

Other Applications of Rapid Thermal Processing

Rapid thermal processing (RTP), which includes RTA, is a key technology in the fabrication of advanced integrated circuits, with a wide range of applications. In addition, for ion implantation damage annealing and dopant activation, RTP is also used for metal silicide and nitride formation, dielectric formation and annealing, and reflow of deposited oxides.²⁶ Typical RTP systems use radiant energy sources, often tungsten-halogen lamps, to heat a wafer to a high temperature for a period of less than a minute. Shrinking device dimensions and increasing wafer diameters are expected to make the use of RTP even more widespread as a result of its low thermal budget, fast cycle time, and compatibility with single-wafer processing. New applications, including gate dielectric formation and rapid thermal chemical vapor deposition (RTCVD), are also emerging.

TABLE 2 TECHNOLOGY COMPARISON

Determinant	Conventional furnace	Rapid thermal annealing
Process	Batch	Single-wafer
Furnace	Hot-wall	Cold-wall
Heating rate	Low	High
Cycle time	High	Low
Temperature monitor	Furnace	Wafer
Thermal budget	High	Low
Particle problem	Yes	Minimal
Uniformity and repeatability	High	Low
Throughput	High	Low

Millisecond Annealing

As devices have scaled down to below 40 nm technology node, even 1 nm of diffusion is significant for ultra-shallow junctions. The requirement of much less diffusion leads to the need for millisecond-duration heating cycles during annealing. Adequate implant damage annealing and dopant activation in millisecond heating push the annealing peak temperature somewhat below the melting point of silicon.

The peak-temperature time for conventional RTP systems is usually limited by the maximum cooling rate of the wafer and the time to switch off the heating energy source. These factors typically limit the peak-temperature time to 1 second. With the use of arc lamp energy sources, which can be switched off very fast, the spike width can be reduced to ~ 0.3 second, but this is not short enough to meet future requirements.

For millisecond annealing, the wafer is rapidly heated with a fast ramp to an intermediate temperature and then a very short, high-energy pulse from an array of powerful water-wall flash-lamps produces a temperature jump on the whole front surface of the wafer. This condition allows extremely fast cooling through conduction of heat away from the surface into the bulk of the wafer. The nature of the heating cycle is illustrated in Fig. 24. The ability to adjust the intermediate temperature and the magnitude of the temperature jump provide flexibility in tuning the amount of diffusion relative to the amount of activation.

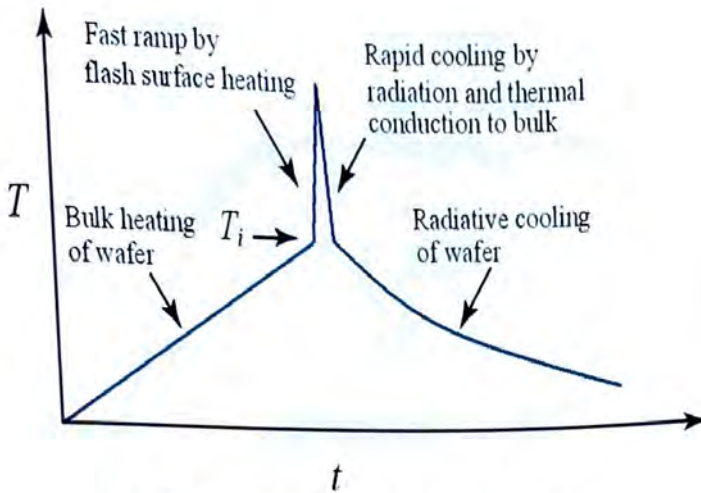


Fig. 24 The heating cycle of millisecond annealing.

► 14.6 IMPLANTATION-RELATED PROCESSES

In this section we consider a few implantation-related processes, such as multiple implantation, masking, high-angle implantation, high-energy implantation, and high-current implantation.

14.6.1 Multiple Implantation and Masking

In many applications, doping profiles other than the simple Gaussian distribution are required. One such case is the preimplantation of silicon with an inert ion to make the silicon surface region amorphous. This technique allows close control of the doping profile and permits nearly 100% dopant activation at low temperatures, as discussed previously. In such a case, a deep amorphous region may be required. To obtain this type of region, we must make a series of implants at varying ion energies and doses.

Multiple implantation can also be used to form a flat doping profile, as shown in Fig. 25. Here, four boron implants into silicon are used to provide a composite doping profile. The measured carrier concentration and that predicted using range theory are shown in the figure. Other doping profiles, unavailable from diffusion techniques, can be obtained by using various combinations of impurity dose and implantation energy. Multiple implants have been used to preserve stoichiometry during the implantation and annealing of GaAs. This approach, whereby equal amount of gallium and n -type dopant (or arsenic and p -type dopant) are implanted prior to annealing, has resulted in higher carrier activation.

To form p - n junctions in selected areas of the semiconductor substrate, an appropriate mask should be used for the implantation. Because implantation is a low-temperature process, a large variety of masking materials can be used. The minimum thickness of masking material required to stop a given percentage of incident ions can be estimated from the range parameters for ions. The inset in Fig. 26 shows the profile of an implant in a masking material. The dose implanted in the region beyond a depth d (shown shaded) is given by integration of Eq. 25 as

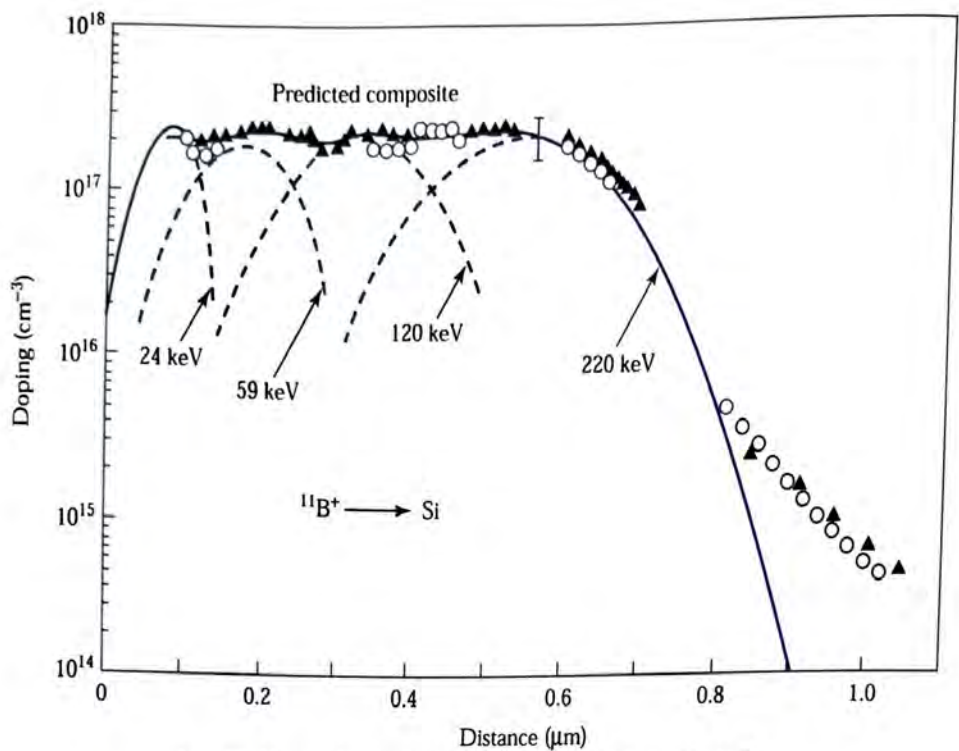


Fig. 25 Composite doping profile using multiple implants.²⁷

$$S_d = \frac{S}{\sqrt{2\pi}\sigma_p} \int_d^\infty \exp\left[-\left(\frac{x-R_p}{\sqrt{2}\sigma_p}\right)^2\right] dx. \tag{33}$$

From Table 1 we can derive the expression

$$\int_x^\infty e^{-y^2} dy = \frac{\sqrt{\pi}}{2} \operatorname{erfc}(y). \tag{34}$$

Therefore, the fraction of the dose that has “transmitted” beyond a depth d is given by the transmission coefficient T :

$$T \equiv \frac{S_d}{S} = \frac{1}{2} \operatorname{erfc}\left(\frac{d-R_p}{\sqrt{2}\sigma_p}\right).$$

(35)

Once T is given, we can obtain the mask thickness d from Eq. 35 for any given R_p and σ_p . The values of d to stop 99.99% of incident ions ($T = 10^{-4}$) are shown in Fig. 26 for SiO_2 , Si_3N_4 , and a photoresist as masking materials.^{19,25} Mask thicknesses given in this figure are for boron, phosphorus, and arsenic implanted into silicon. These mask thicknesses also used as guidelines for impurity masking in gallium arsenide. The dopants are shown in parentheses. Since both R_p and σ_p vary approximately linearly with energy, the

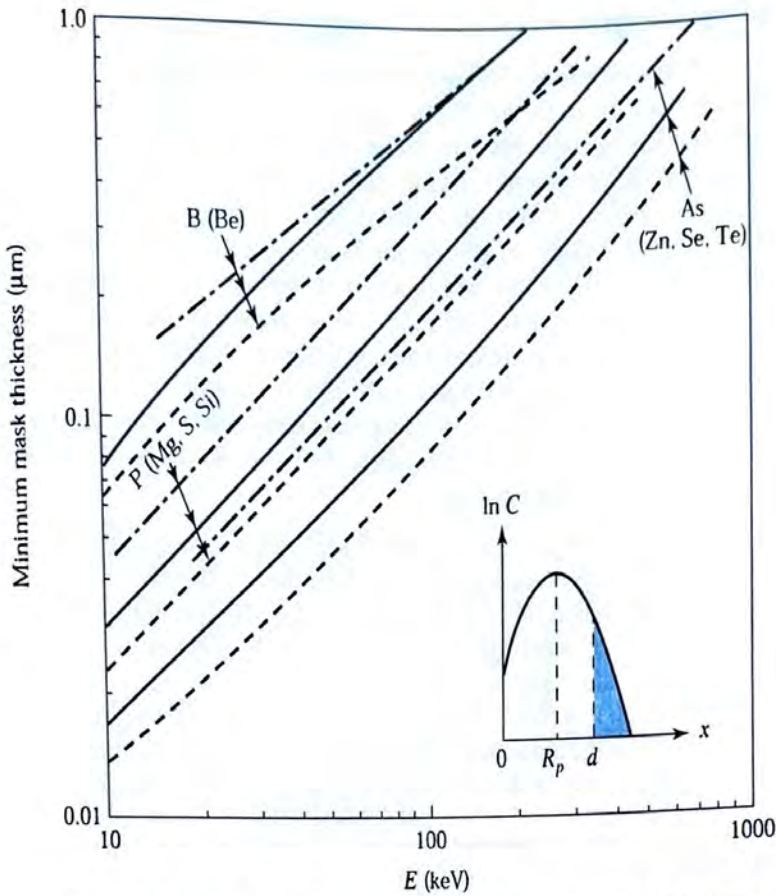


Fig. 26 Minimum thickness²⁸ of SiO_2 (—), Si_3N_4 (---), and photoresist (— · — · —) to produce a masking effectiveness of 99.99%.

minimum thickness of the masking material also increases linearly with energy. In certain applications, instead of totally stopping the beam, the masks can be used as attenuators, that can provide an amorphous surface layer to the incident ion beam to minimize the channeling effect.

► EXAMPLE 4

When boron ions are implanted at 200 keV, what thickness of SiO_2 will be required to mask 99.996% of the implanted ions ($R_p = 0.53 \mu\text{m}$, $\sigma_p = 0.093 \mu\text{m}$)?

SOLUTION The complementary error function in Eq. 35 can be approximated if the argument is large (see Table 1):

$$T \cong \frac{1}{2\sqrt{\pi}} \frac{e^{-u^2}}{u},$$

where the parameter u is given by $(d - R_p)/\sqrt{2}\sigma_p$. For $T = 10^{-4}$, we can solve the above equation to give $u = 2.8$. Thus,

$$d = R_p + 3.96\sigma_p = 0.53 + 3.96 \times 0.093 = 0.898 \mu\text{m}.$$

14.6.2 Tilt-Angle Ion Implantation

In scaling devices to submicron dimensions, it is important also to scale dopant profiles vertically. We need to produce junction depths less than 15 nm for the 28 nm technology node, including diffusion during dopant activation and subsequent processing steps. Modern device structures, such as the lightly doped drain (LDD), require precise control of dopant distributions vertically and laterally.

It is the ion velocity perpendicular to the surface that determines the projected range of an implanted ion distribution. If the wafer is tilted at a large angle to the ion beam, then the effective ion energy is greatly reduced. Figure 27 illustrates this for 60 keV arsenic ions as a function of the tilt angle, showing that it is possible to achieve extremely shallow distributions using a high tilt angle (86°). In tilt-angle ion implantation, we should consider the shadow effect (inset in Fig. 27) for the patterned wafer. A lower tilt angle leads to a small shadow area. For example, if the height of the patterned mask is 0.5 μm , with vertical sidewalls, a 7° incident ion beam will induce a 61 nm shadow. This shadow effect may introduce an unexpected series resistance in the device.

14.6.3 High-Energy and High-current Implantation

High-energy implantors, capable of energies as high as 1.5-5 MeV, are available and have been used for some novel applications. The majority of these depend on the ability to dope the semiconductor to many micrometers in depth, without the need for long diffusion times at high temperatures. High-energy implantors can also be used to produce low-resistivity buried layers. For example, a buried layer 1.5-3 μm below the surface for a CMOS device can be achieved by high-energy implantation.

High-current implantors (10-20 mA) operating in the 25-30 keV range are routinely used for the predeposition step in diffusion technology because the amount of total dopant can be controlled precisely. After the predeposition, the dopant impurities can be driven in by a high-temperature diffusion step at the same time that

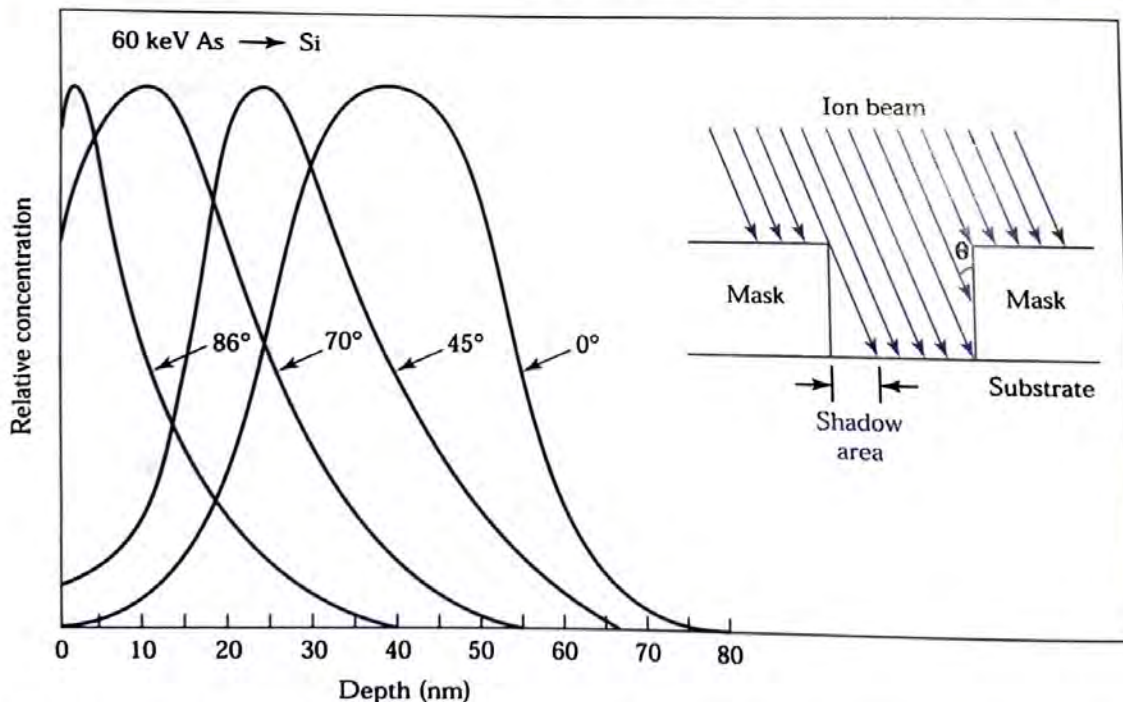


Fig. 27 60 keV arsenic implanted into silicon, as a function of beam tilt angle. Inset shows the shadow area for tilt-angle ion implantation.

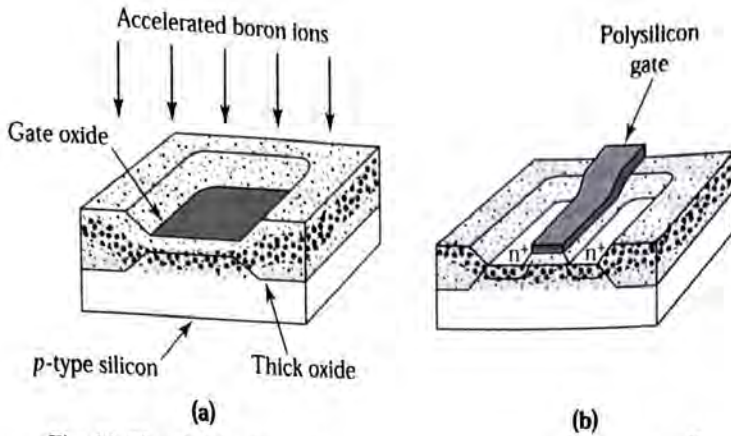


Fig. 28 Threshold voltage adjustment using boron ion implantation.²⁸

implant damage at the surface region is annealed out. Another application is the threshold voltage adjustment in MOS devices. A precisely controlled amount of dopant (e.g., boron) is implanted through the gate oxide to the channel region²⁹ (Fig. 28a). Because the projected range of boron in silicon and silicon oxide are comparable, if we choose a suitable incident energy, the ions will penetrate just the thin gate oxide, not the thicker field oxide. The threshold voltage will vary approximately linearly with the implanted dose. After boron implantation, polysilicon can be deposited and patterned to form the gate electrode of the MOSFET. The thin oxide surrounding the gate electrode is removed, and the source and drain regions are formed as shown in Fig. 28b by another high-dose arsenic implantation.

High-current implantors with energies in the 150-200 keV range are now available. A major use for these machines is to form a buried silicon dioxide layer by implanting oxygen ions into a silicon substrate followed by a thermal annealing process. This separation by implantation of oxygen (SIMOX) is a key SOI (silicon-on-insulator) technology. The SIMOX process as shown in Fig. 29 uses a high-energy O^+ beam, typically in the 150 to 200 keV range, so that the oxygen ions have projected ranges of 100-200 nm. A heavy dose, $1-2 \times 10^{18}$ ions/cm², is used to produce an insulating layer of SiO_2 that is 100-500 nm thick. The use of SIMOX material leads to a significant reduction of source/drain capacitances in MOS devices. Moreover, it reduces coupling between devices and thus allows tighter packing without the problem of latchup. It is widely proposed as the material of choice for advanced, high-speed CMOS circuits.

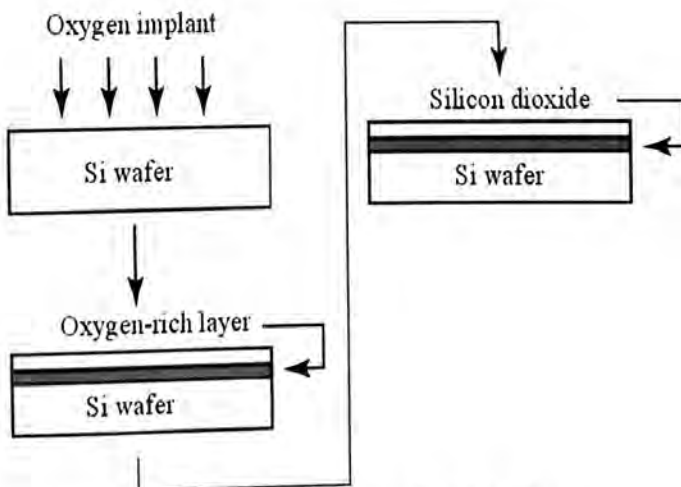


Fig. 29 The SIMOX process for SOI wafer.

Currently, 10%-20% of SOI wafers are made by the SIMOX process and 80%-90% are made by Smart Cut technology, which utilizes wafer bonding and layer transfer from a donor or seed wafer to a support handle wafer. The Smart Cut technology is shown in Fig. 30. A seed wafer is oxidized to a desired thickness. The next step is a hydrogen implantation through the oxide and into Si with a dose that is typically greater than $5 \times 10^{16} \text{ cm}^{-2}$. After the implantation, the seed wafer and the handle wafer are carefully cleaned in order to eliminate any particles and surface contaminants, and to make both surfaces hydrophilic. Wafer pairs are aligned and contacted. The wafer bonding relies on chemical assistance through hydrogen bonds and water molecules. There are a few monolayers of water trapped between two native oxide films immediately after the wafers are fused together. When the bonded wafer pair is heated to high temperature, the water will diffuse through the thin oxide to the Si interfaces to form more oxide. Finally, a complete closure of the interface occurs by coupling of the interface hydroxyl species and liberates hydrogen into the Si bulk.

The bonded wafer pairs is loaded into a furnace and heated to a temperature of 400°C - 600°C , at which point the wafers split along the hydrogen implanted plane. The mechanism for splitting is that sufficient high-dosage hydrogen implantation will produce a sufficiently high density of platelets or microcavities. Microcavities grow to form microcracks and eventually lead to splitting of a thin layer away from the main substrate. The as-split wafer surface has a mean roughness of a few nanometers. A light touch-polish or other surface treatment brings the same surface roughness as in the standard bulk Si. The SOI technology can reduce not only the parasitic device capacitance but also the short channel effects, and thereby improve performance of scaled devices.

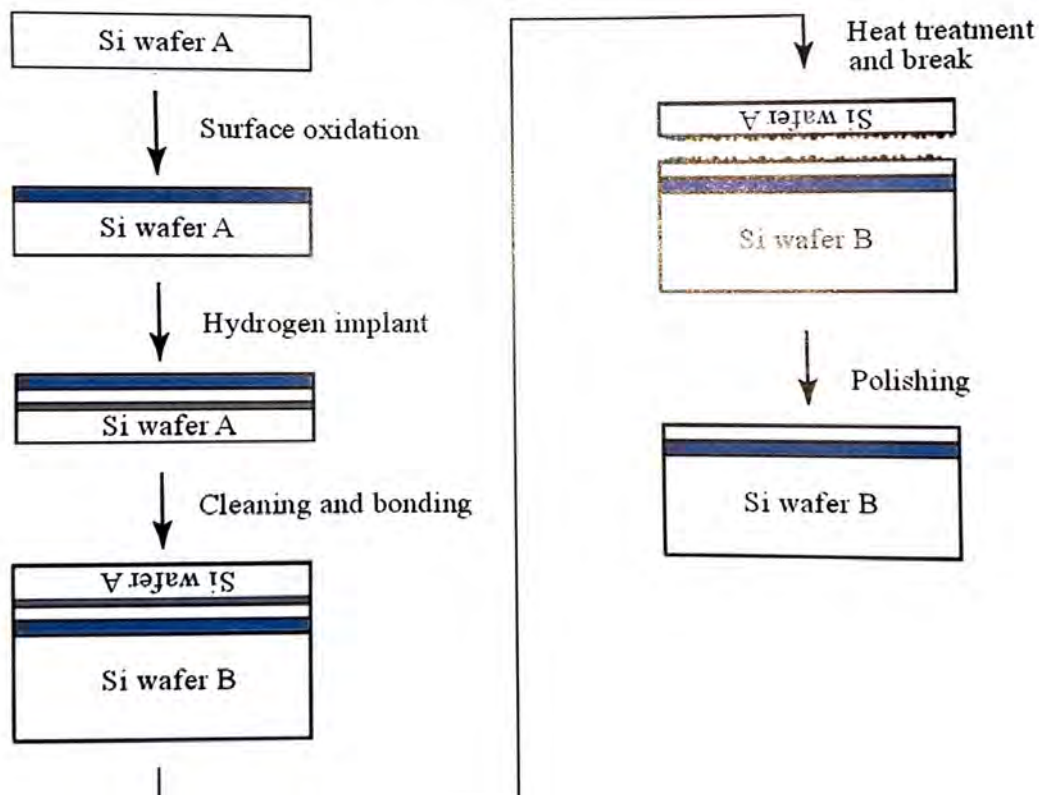


Fig. 30 The Smart Cut process for SOI wafer.

SUMMARY

Diffusion and ion implantation are the two key methods for impurity doping. We first considered the basic diffusion equation for constant diffusivity. We obtain the complementary error function (erfc) and the Gaussian function for the constant- surface-concentration case and constant-total-dopant case, respectively. The results of a diffusion process can be evaluated by measurements of the junction depth, the sheet resistance, and the dopant profile.

When the doping concentration is higher than the intrinsic carrier concentration n_i at the diffusion temperature, the diffusivity becomes concentration dependent. This dependence has profound effects on the resulting doping profile. For example, arsenic and boron diffusivities in silicon vary linearly with the impurity concentration. Their doping profiles are much more abrupt than the erfc profile. Phosphorus diffusivity in silicon varies as the square of concentration. This dependence and a dissociation effect give rise to a phosphorus diffusivity that is 100 times larger than its intrinsic diffusivity.

Lateral diffusion at the edge of a mask and impurity redistribution during oxidation are two processes in which diffusion can have an important impact on device performance. The former can substantially reduce the breakdown voltage, and the latter will influence the threshold voltage as well as the contact resistance.

The key parameters for ion implantation are the projected range R_p and its standard deviation σ_p , also called at R_p from the surface of the semiconductor substrate. The implantation profile can be approximated by a Gaussian distribution with peak located at R_p . The advantages of ion implantation process are more precise control of the amount of dopant, a more reproducible doping profile, and lower processing temperature compared with the diffusion process.

We considered R_p and σ_p for various elements in silicon and gallium arsenide and discussed the channeling effect and ways to minimize this effect. However, implantation may cause severe damage to the crystal lattice. To remove the implant damage and to restore mobility and other device parameters, we must anneal the semiconductor at an appropriate combination of time and temperature. Currently, rapid thermal annealing (RTA) is preferred to conventional furnace annealing because RTA can remove implant damage without thermal broadening of the doping profile.

Ion implantation has wide applications for advanced semiconductor devices. These include (a) multiple implantation to form novel distributions, (b) selection of masking materials and thickness to stop a given percentage of incident ions from reaching the substrate. (c) tilt-angle implantation to form ultrashallow junctions, (d) high-energy implantation to form buried layers, and (e) high-current implantation for predeposition and threshold voltage adjustment and to form an insulating layer for SOI applications.

REFERENCES

1. S. M. Sze, Ed., *VLSI Technology*, 2nd Ed., McGraw-Hill, New York, 1988, Ch. 7, 8.
2. S. K. Ghandhi, *VLSI Fabrication Principles*, 2nd Ed., Wiley, New York, 1994, Ch. 4, 6.
3. W. R. Runyan and K. E. Bean, *Semiconductor Integrated Circuit Processing Technology*, Addison-Wesley, Reading, MA, 1990, Ch. 8.
4. H. C. Casey and G. L. Pearson, "Diffusion in Semiconductors," in J. H. Crawford and L. M. Slifkin, Eds., *Point Defects in Solids*, Vol. 2, Plenum, New York, 1975.
5. J. P. Joly, "Metallic Contamination of Silicon Wafers," *Microelectron. Eng.*, **40**, 285 (1998).
6. A. S. Grove, *Physics and Technology of Semiconductor Devices*, Wiley, New York, 1967.
7. ASTM Method F374-88, "Test Method for Sheet Resistance of Silicon Epitaxial, Diffused, and Ion-implanted Layers Using a Collinear Four-Probe Array," **V10**, 249 (1993).
8. J. C. Irvin, "Evaluation of Diffused Layers in Silicon," *Bell Syst. Tech. J.*, **41**, 2 (1962).