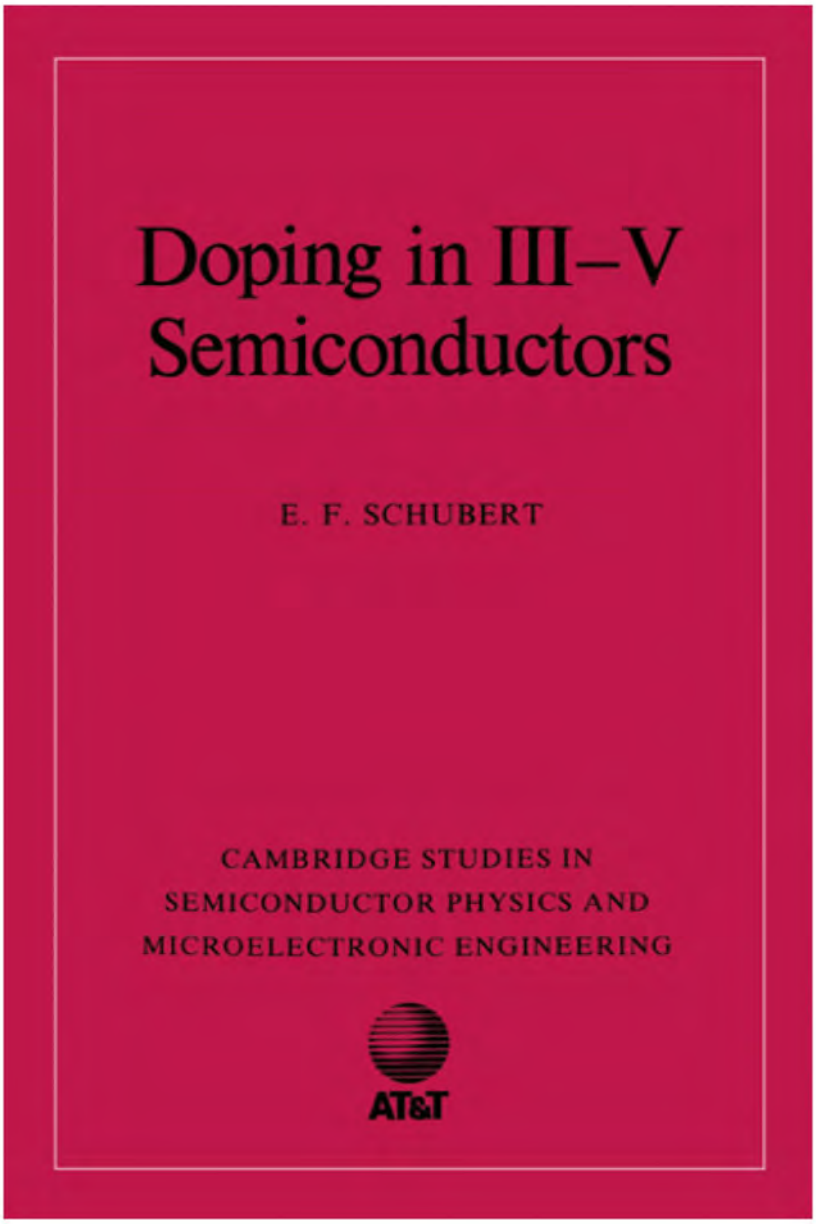


INNOSCIENCE EXHIBIT 1032
INNOSCIENCE v. EPC
IPR2023-01381

This is the first book to describe thoroughly the many facets of doping in compound semiconductors. Equal emphasis is given to the fundamental materials physics and to the technological aspects of doping.

The author describes various doping techniques, including doping during epitaxial growth, doping by implantation, and doping by diffusion. The key characteristics of all dopants that have been employed in III–V semiconductors are discussed. In addition, general characteristics of dopants are analyzed, including the electrical activity, saturation, amphotericity, autocompensation, and maximum attainable dopant concentration. Redistribution effects are important in semiconductor microstructures. Linear and non-linear diffusion, different microscopic diffusion mechanisms, surface segregation, surface drift, surface migration, impurity-induced disordering, and the respective physical driving mechanisms are illustrated.

Topics related to basic impurity theory include the hydrogenic model for shallow impurities, linear screening, density of states, classical and quantum statistics, the law of mass action, as well as many analytic approximations for the Fermi–Dirac integral for three-, two- and one-dimensional systems. The timely topic of highly doped semiconductors, including band tails, impurity bands, bandgap renormalization, the Mott transition, and the Burnstein–Moss shift, is discussed as well.

Doping is essential in many semiconductor heterostructures including high-mobility selectively doped heterostructures, quantum well and quantum barrier structures, doping superlattice structures and δ-doping structures. Technologically important deep levels are summarized, including Fe, Cr, and the DX-center, the EL2 defect, and rare-earth impurities. The properties of deep levels are presented phenomenologically, including emission, capture, Shockley–Read recombination, the Poole–Frenkel effect, lattice relaxation, and other effects. The final chapter is dedicated to the experimental characterization of impurities.

This book will be of interest to graduate students, researchers and development engineers in the fields of electrical engineering, materials science, physics, and chemistry working on semiconductors. The book may also be used as a text for graduate courses in electrical engineering and materials science.

Published online by Cambridge University Press

Published online by Cambridge University Press

Cambridge Studies in Semiconductor Physics and
Microelectronic Engineering: 1

EDITED BY
Haroon Ahmed
Cavendish Laboratory, University of Cambridge
Michael Pepper
Cavendish Laboratory, University of Cambridge
Alec Broers
Department of Engineering, University of Cambridge

Doping in III–V Semiconductors

Published online by Cambridge University Press

Published online by Cambridge University Press

Doping in III–V Semiconductors

E. Fred Schubert
AT&T Bell Laboratories





Published online by Cambridge University Press

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 2RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521419192

© AT&T 1993

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 1993
This digitally printed first paperback version 2005

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data
Schubert, E. Fred
Doping in III–V semiconductors / E. Fred Schubert.
p. cm. – (Cambridge studies in semiconductor physics and
microelectronic engineering: 1)
Includes bibliographical references and index.
ISBN 0 521 41919 0
1. Compound semiconductors. 2. Semiconductor doping. I. Title
II. Title: Doping in 3–5 semiconductors. III. Series.
QC611.8.C64S34 1993
621.3815'2–dc20 92-35998 CIP

ISBN-13 978-0-521-41919-2 hardback
ISBN-10 0-521-41919-0 hardback

ISBN-13 978-0-521-01784-8 paperback
ISBN-10 0-521-01784-X paperback

Published online by Cambridge University Press

To Jutta Maria
and to
Anne, Martin and Ursula

Published online by Cambridge University Press

Published online by Cambridge University Press

Contents

Foreword		*page* xi
Preface		xv
List of symbols		xviii
Introduction		1
1.	Shallow impurities	6
	1.1 Hydrogenic impurities	6
	1.2 Screening of impurity potentials	27
	1.3 High doping effects	34
2.	Phenomenology of deep levels	54
	2.1 General characteristics	55
	2.2 Capture, emission, and recombination	60
	2.3 Effects relating to experimental properties	67
3.	Semiconductor statistics	78
	3.1 Density of continuum states	80
	3.2 Classical and quantum statistics	94
	3.3 Carrier concentrations	118
	3.4 Law of mass action	130
4.	Growth technologies	137
	4.1 Molecular-beam epitaxy	137
	4.2 Gas-source molecular-beam epitaxy	146
	4.3 Chemical-beam epitaxy	149
	4.4 Organometallic vapor-phase epitaxy	153
	4.5 Vapor-phase epitaxy	161
5.	Doping with elemental sources	165
	5.1 Doping techniques and calibration	165
	5.2 Doping homogeneity and uniformity	177
	5.3 Elemental dopants	181

Published online by Cambridge University Press

337-TA-1366 RX-0255

6. Gaseous doping sources 208
6.1 General considerations 209
6.2 Doping homogeneity and uniformity 213
6.3 Gaseous doping precursors 219

7. Impurity characteristics 248
7.1 Doping incorporation during growth 249
7.2 Doping incorporation by implantation 262
7.3 Doping incorporation by diffusion 272
7.4 Amphotericity and autocompensation 277
7.5 Maximum doping concentration 290

8. Redistribution of impurities 306
8.1 Diffusion and brownian motion 307
8.2 Fickian diffusion 311
8.3 Diffusion mechanisms 316
8.4 Some specific impurities 325
8.5 Surface segregations and surface migration 337
8.6 Impurity-induced layer disordering 349

9. Deep centers 354
9.1 DX-type centers 355
9.2 The EL2 defect 365
9.3 Hydrogen 374
9.4 Chromium, iron, and oxygen 384
9.5 Rare-earth impurities 390

10. Doping in heterostructures, quantum wells, and superlattices 392
10.1 Selectively doped heterostructures 392
10.2 Doping aspects of selectively doped heterostructures 404
10.3 Doping superlattices 412
10.4 Impurities in quantum wells and quantum barriers 422

11. Delta doping 433
11.1 Electronic structure 434
11.2 Growth, localization, and redistribution 446
11.3 Transport properties 458
11.4 Electronic devices 463
11.5 Optical devices 471

12. Characterization techniques 482
12.1 Electronic characterization techniques 482
12.2 Optical characterization techniques 508
12.3 Chemical and structural characterization techniques 531

Appendix A: Properties of III–V semiconductors 541

Appendix B: Constants and conversions 545

References 546

Index 600

Published online by Cambridge University Press

RX-0255.11

Foreword

Technology revolutions mark their origin from a single breakthrough such as the demonstration of the transistor. Subsequent technological advances make the revolution a reality. The historical evolution from Si transistor to integrated circuits to high speed computers and telecommunications is one excellent example.

Active workers in the field know each advance represents a major expenditure of time and effort (really 'blood and sweat'). Progress comes about in a competitive atmosphere involving creative ideas, personality forces and technology prognostications.

We are presently in the midst of the microelectronics revolution. Insatiable demands exist for greater data rates, for consumer electronics and for superior telecommunications. We envision a world of wireless communications, video transmissions and displays, and numerous applications of high speed data transmission. Satisfying this demand is the goal of current research.

For our field of microelectronics this means a systems analysis from the final product to the atomic configurations of the materials that make the product. This analysis itself is remarkable; we can precisely relate the macroscopic system properties – how fast will the system operate – to the microscopic atomic structure – where do the atoms sit in the solid.

Analysis of today's devices reveal the limitations of today's semiconductors – mostly silicon. Silicon has been the workhorse for the last thirty years. The 'silicon community' continues to squeeze all that is obtainable from this most robust and manufacturable semiconductor. Nevertheless an all silicon technology has foreseeable and predictable limitations.

An electrical engineering student can open almost any textbook and see that the path is clear – III–V semiconductors can do all that Si can do and more – they

https://doi.org/10.1017/CBO9780511599828.001 Published online by Cambridge University Press

emit light! Not only do III–V materials have fundamental characteristics which make faster electronics, but they are efficient emitters and detectors of photons. Combined with a transmission medium – the optical fiber – these III–V semiconductor devices gave rise to the era of optical communications. The intrinsically high speed in an optical system is the basis of our vision of ultra high speed data transmission, yielding electronically controlled visual communications, high definition TV and personal video.

The 'cost effective' realization of this technology, on a worldwide basis, depends on our ability to fabricate the materials that make the system. Modern silicon technology demonstrates that basic materials manipulation has governed the advance of the technology. Near perfect crystalline growth, a well behaved passivating oxide, controlled doping, and advanced lithography made the Si revolution a reality. The extension of the semiconductor revolution to III–V devices depends on these materials manipulations, which eventually determine their manufacturability.

Tomorrow's technology will be based on today's research; the body of information that allows the 'systems evaluation' and meets the 'manufacturability requirements'. As a starting point a systems evaluation may assume an ideal structure; however, no material is perfect. Better systems evaluation includes far more realistic material configurations and the picture of this realistic material is one goal of the semiconductor materials scientist. Manufacturing requirements will also be determined by this research in terms of yields, process margins, etc.

Both systems and manufacturability evaluations require a deep understanding of the numerous steps that go into the fabrication of a semiconductor. Each of these 'steps' becomes a sub-field unto itself – with hundreds of investigators, numerous conferences and publications in an ever expanding network of scientific research.

This book focuses on such an element of III–V devices – the doping of these semiconductors. Basically it addresses the question of the behavior of dopants in III–V semiconductors – how are they incorporated, where do the atoms sit? As explained in the Preface 'doping' is an essential element to make an active device from a semiconductor. Clever and controlled manipulation of 'doping' can enhance the device properties. Doping processes will be one critical factor in determining manufacturability. Thus the knowledge represented in this book is the critical input needed to provide insight into our technology directions.

One further point, lest it get lost in the current world focus on applied research. Semiconductor science is a field that spans the space of exciting fundamental science to immediate and current applications. Unique solid state structures with controlled doping have revealed new phenomena of the most

https://doi.org/10.1017/CBO9780511599828.001 Published online by Cambridge University Press

fundamental character. Such reports have led to recent Nobel prizes and exciting new concepts. These 'unique solid state' structures combine the latest in materials fabrication and dopant control. Thus, it is worthwhile to note that this slice of science is not only crucial to technology but enables forefront basic research. It has opened the fascinating world of quantum structures and low dimensional configurations.

Those of us who have lived through this semiconductor science revolution look back with pride at the accomplishments of the field and look forward with excitement at what is to come. The science and art of fabricating active III–V semiconductors is a critical element in determining the future.

L. C. Feldman
AT&T Bell Laboratories
January, 1993

https://doi.org/10.1017/CBO9780511599828.001 Published online by Cambridge University Press

https://doi.org/10.1017/CBO9780511599828.001 Published online by Cambridge University Press

Preface

It was realized in the infancy of semiconductor science that impurities are the essential ingredient in these materials. The first usage of the term 'impurities' dates back to 1931 (Wilson, 1931):

> Electrons on a foreign atom [in a semiconductor] do not take part directly in [electrical] conduction. They must be transferred by the effect of lattice vibrations to an atom of the pure substance. In this case, the main function of the 'impurities' is to provide electrons for the upper unoccupied energy bands of the crystal, while acting as scatterers is only a secondary function.

The knowledge of the significant consequences of impurities motivated researchers to, first, purify the semiconductor materials, for example by zone refining, and, second, deliberately add 'foreign atoms' to purified semiconductors, for example by thermal diffusion of such 'foreign atoms' into the semiconductor. The deliberate doping of semiconductors with impurities required that the usage of the term 'impurities' was broadened. Such an extended definition was offered by Shockley (1950):

> The conductivity arises from the presence of [ ] atoms which are termed 'impurities', even though added to the otherwise pure [ ] [semiconductor].

The term 'dopant' was non-existent in the English language until its introduction in 1963 by Seidman and Marshall (1963), who wrote about crystal growth from the melt:

> The melt consists of intrinsic polycrystalline Ge and Si mixed with

https://doi.org/10.1017/CBO9780511599828.002 Published online by Cambridge University Press

337-TA-1366 RX-0255

suitable 'dopants' such as P, which produces an n-type melt, or Ga, which produces a p-type melt.

Thus, the term 'impurities' is used for unintentional as well as intentional foreign atoms in the semiconductor, while the term 'dopants' is used exclusively for deliberately added impurities.

This book is devoted to the many aspects of doping in III–V semiconductors. The book covers, with equal emphasis, fundamental concepts and basic theory of impurities, as well as the practical and technological aspects of the doping process. The basic conceptual theory of shallow impurities and deep levels and their effect on the free carrier concentration is given in Chaps. 1, 2, and 3. Chapters 4, 5, and 6 cover the characteristics of individual impurity elements as well as the doping process during epitaxy with solid elemental sources as well as gaseous sources. Chapter 7 covers characteristics of impurities incorporated during growth, ion implantation, and diffusion, as well as impurity characteristics such as amphotericity, autocompensation, and the maximum attainable doping concentration. Redistribution effects during growth and thermal processing are covered in Chap. 8 and include diffusion, segregation, migration, and other effects. Chapter 9 describes deep levels that are relevant to III–V semiconductor technology, such as the DX-center, EL2, O, Cr, Fe, and rare-earth impurities. Doping in semiconductor microstructures, which require highly controlled doping profiles, is described in Chaps. 10 and 11. Finally, Chap. 12 summarizes electronic, optical, and structural impurity characterization techniques.

The book is written for graduate students, researchers, and development engineers of the disciplines electrical engineering, physics, materials science, and chemistry working at universities, industrial research laboratories, government laboratories and production facilities. It is suited as a textbook for graduate courses in electrical engineering and material science. The book is self-contained and covers all major concepts in detail. Units of the *Système International* (S.I. units) are used throughout the book. Boxes are drawn around those equations which are frequently used in semiconductor science.

The author has benefited significantly from suggestions, discussions and collaborations with his colleagues at AT&T Bell Laboratories, especially with G. A. Baraff, T. Y. Chang, A. Y. Cho, L. C. Feldman, G. H. Gilmer, A. M. Glass, H. J. Gossmann, R. D. Grober, T. D. Harris, W. D. Johnston, Jr., R. F. Kazarinov, R. F. Kopf, M. Lax, A. F. J. Levi, H. S. Luftman, R. M. Lum, R. J. Malik, D. P. Monroe, M. B. Panish, L. Pfeiffer, D. Ritter, S. Schmitt-Rink, H. L. Stormer, V. Swaminathan, W. T. Tsang, J. S. Weiner, and J. L. Zilko. I also thank Professor D. G. Deppe (Univ. Texas, Austin), Professor Y. H. Yao (Univ. Nebraska,

https://doi.org/10.1017/CBO9780511599828.002 Published online by Cambridge University Press

RX-0255.17

Lincoln), and Professor K. Ploog (Paul Drude Institute, Berlin) for many suggestions and valuable discussions. The encouragement of F. Capasso and D. V. Lang during the course of this work was highly appreciated. I am indebted to L. C. Feldman for writing the foreword.

E. F. Schubert
AT&T Bell Laboratories
February, 1993

https://doi.org/10.1017/CBO9780511599828.002 Published online by Cambridge University Press

Symbols

α	(cm^{-1})	absorption coefficient
α	(-)	Sommerfeld's fine structure constant
α		variational parameter
a_B	(Å)	Bohr radius
a_B^*	(Å)	effective Bohr radius
a_0	(Å)	lattice constant
A	(cm^2)	area
A_1, A_2, A_3		normalization constants
$\mathcal{B}$	(T)	magnetic induction
c	(m/s)	velocity of light in vacuum
c		constant
c_n (c_p)	(s^{-1})	electron (hole) capture probability
c_n^* (c_p^*)	($cm^{-3}s^{-1}$)	electron (hole) capture rate
C	(F/cm^2)	capacitance per unit area
δ		delta function
Δ	(cm^{-2})	delta operator $(\partial^2/\partial x^2+\partial^2/\partial y^2+\partial^2/\partial z^2)$
ΔE	(eV)	activation energy
ΔE_C (ΔE_V)	(eV)	discontinuity in the conduction (valence) band
ΔE_{PF}	(eV)	lowering of emission barrier (Poole–Frenkel energy)
ΔR_L	(cm)	longitudinal straggle
ΔR_P	(cm)	projected straggle
D	(cm^2/s)	diffusion constant (also called diffusion coefficient)
D_i	(cm^2/s)	diffusion constant in intrinsic semiconductors
D_0	(cm^2/s)	pre-exponential factor of diffusion constant
ε	(A s/V m)	permittivity ($\varepsilon = \varepsilon_r\varepsilon_0$)
ε_0	(A s/V m)	permittivity in vacuum
ε_r	(-)	dielectric constant
η_a	(-)	doping activation efficiency
η_D	(-)	doping efficiency

Published online by Cambridge University Press

η_F	(-)	normalized Fermi energy
e	(C)	elementary charge
e_n (e_p)	(s^{-1})	electron (hole) emission probability
e_n^* (e_p^*)	($cm^{-3}s^{-1}$)	electron (hole) emission rate
E	(J)	energy
E_a	(eV)	activation energy
E_a	(eV)	acceptor ionization energy
E_d	(eV)	donor ionization energy
E_{dd}	(eV)	deep donor ionization energy
E_g	(eV)	bandgap energy
E_{kin}	(J)	kinetic energy
E_{pot}	(J)	potential energy
E_{sd}	(eV)	shallow donor ionization energy
E_t	(eV)	trap ionization energy
E_A	(eV)	acceptor energy
E_{BE}	(eV)	energy of bound exciton
E_C	(eV)	energy of conduction band edge
E_D	(eV)	donor energy
E_{DD}	(eV)	deep donor energy
E_F	(eV)	Fermi energy
E_{FE}	(eV)	energy of free exciton
E_{SD}	(eV)	shallow donor energy
E_{Ryd}	(eV)	Rydberg energy
E_{Ryd}^*	(eV)	effective Rydberg energy
E_T	(eV)	trap energy
E_V	(eV)	energy of valence band edge
E_0	(eV)	energy of quantized ground-state subband
$\mathcal{E}$	(V/cm)	electric field
ϕ	(V)	potential
ϕ_B	(V)	barrier height
ϕ	(°)	cylindrical or spherical coordinate (azimuthal angle)
f	(Hz)	frequency
f_B	(-)	Boltzmann distribution function
f_F	(-)	Fermi distribution function
f_M	(-)	Maxwell distribution function
f_{MB}	(-)	Maxwell–Boltzmann distribution function
F	(N)	force
$F_{1/2}$	(-)	Fermi–Dirac integral of order ½
g	(-)	state degeneracy
g_0	(-)	ground-state degeneracy
G	(J)	Gibb's free energy
h	(Js)	Planck's constant
$\hbar$	(Js)	Planck's constant divided by 2π
H	(J)	enthalpy
H	(J)	hamiltonian operator $(-(\hbar^2/2m^*)\Delta + eV)$

Published online by Cambridge University Press

I		intensity
j	(A/cm^2)	current density
j	(cm^{-2}s^{-1})	particle flow density
k	(J/K)	Boltzmann's constant
k	(cm^{-1})	wave vector ($2\pi/\lambda$)
K	(-)	equilibrium constant
K_c	(-)	compositional distribution coefficient
K_t	(-)	thermodynamic distribution coefficient
λ	(cm)	wavelength
λ	(cm)	transition region thickness
l	(cm)	length
l	(-)	orbital angular momentum quantum number
l_h	(cm)	hop distance
L_z	(cm)	thickness of quantum well
L_D	(cm)	diffusion length
μ	(cm^2/V s)	mobility
μ_0	(V s/A m)	permeability in vacuum
m	(kg)	mass
m_0	(kg)	electron mass
m^*	(kg)	effective mass
m_e^*	(kg)	electron effective mass
m_h^*	(kg)	hole effective mass
m_{hh}^*	(kg)	heavy hole effective mass
m_l	(-)	azimuthal quantum number
m_l^*	(kg)	effective mass for motion along a longitudinal direction
m_{lh}^*	(kg)	light hole effective mass
m_t^*	(kg)	effective mass for motion along a transverse direction
m_x^*	(kg)	effective mass for motion along the x-direction
m_y^*	(kg)	effective mass for motion along the y-direction
m_z^*	(kg)	effective mass for motion along the z-direction
M	(kg)	mass of lattice atoms
M_c	(-)	number of equivalent minima in the conduction band
∇	(cm^{-1})	nabla operator, $(\partial/\partial x,\ \partial/\partial y,\ \partial/\partial z)$
n	(cm^{-3})	electron concentration
n	(-)	principal quantum number
n_i	(cm^{-3})	intrinsic electron concentration
n_{2DEG}	(cm^{-2})	density of two-dimensional electron gas
n^{2D}	(cm^{-2})	two-dimensional electron concentration
n^{1D}	(cm^{-1})	one-dimensional electron concentration
N_c	(cm^{-3})	effective density of states at conduction band edge
N_c^{2D}	(cm^{-2})	two-dimensional effective density of states at conduction band ed
N_c^{1D}	(cm^{-1})	one-dimensional effective density of states at conduction band ed
N_{crit}	(cm^{-3})	critical Mott concentration
N_v	(cm^{-3})	effective density of states at valence band edge
N_A	(cm^{-3})	acceptor concentration

Published online by Cambridge University Press

N_{Avo}	(mol^{-1})	Avogadro's number
N_{CV}	(cm^{-3})	CV concentration
N_D	(cm^{-3})	donor concentration
N_{DD}	(cm^{-3})	deep donor concentration
N_{SD}	(cm^{-3})	shallow donor concentration
N^0	(cm^{-3})	concentration of neutral centers
N^+	(cm^{-3})	concentration of positively charged centers
N^-	(cm^{-3})	concentration of negatively charged centers
N^{2D}	(cm^{-2})	two-dimensional impurity density
N^{1D}	(cm^{-1})	one-dimensional impurity density
ω	(s^{-1})	angular frequency ($2\pi/f$)
Ω	(steradian)	solid angle
ψ	($cm^{-3/2}$)	time-independent three-dimensional wave function
Ψ	($cm^{-3/2}$)	time-dependent three-dimensional wave function
p	(cm^{-3})	hole concentration
p	(kg m/s)	momentum
p		probability
p^{2D}	(cm^{-2})	two-dimensional hole concentration
p^{1D}	(cm^{-1})	one-dimensional hole concentration
P	(N/m^2)	pressure
Q	(C/cm^2)	charge per unit area
ρ_{DOS}	($cm^{-3}eV^{-1}$)	density of states
ρ_{DOS}^{2D}	($cm^{-2}eV^{-1}$)	two-dimensional density of states
ρ_{DOS}^{1D}	($cm^{-1}eV^{-1}$)	one-dimensional density of states
r	(cm)	radius
r	(cm)	polar coordinate
r_s	(cm)	screening radius
r_D	(cm)	Debye screening radius
r_H	(-)	Hall factor
r_{TF}	(cm)	Thomas–Fermi screening radius
R	($J\ mol^{-1}K^{-1}$)	universal gas constant
R	(Ω)	resistance
R_H	(cm^3/C)	Hall coefficient
R_n (R_p)	($cm^{-3}s^{-1}$)	net capture rate of electrons (holes)
R_p	(cm)	projected range
R_{SR}	($cm^{-3}s^{-1}$)	Shockley–Read recombination rate
σ	($\Omega^{-1}cm^{-1}$)	conductivity
σ	(cm^2)	capture cross section
σ		standard deviation of a gaussian distribution
σ		step function
s	(-)	spin quantum number
S	(eV/K)	entropy
θ	(°)	spherical coordinate (polar angle)
τ	(s)	time constant
τ_c	(s)	capture time constant

Published online by Cambridge University Press

τ_e	(s)	emission time constant
τ_{inel}	(s)	mean time between inelastic scattering events
τ_n (τ_p)	(s)	minority carrier lifetime
t	(s)	time
t	(cm)	thickness
t_{dep}	(cm)	depletion thickness
t_{spacer}	(cm)	spacer layer thickness
T	(K)	absolute temperature
v	(cm/s)	velocity
v_d	(cm/s)	drift velocity
v_g	(cm/s)	growth rate
$\overline{v}$	(cm/s)	mean velocity
v_{rms}	(cm/s)	root-mean-square velocity
v_x	(cm/s)	velocity along the x-direction
V	(V)	potential
V	(V)	voltage (potential difference)
V	(m^3)	volume
V_{exc}	(cm^3)	excitonic volume
W_D	(cm)	depletion layer width
x	(cm)	cartesian coordinate
x	(-)	composition of alloy semiconductors
y	(cm)	cartesian coordinate
y	(-)	composition of alloy semiconductors
z	(cm)	cartesian coordinate
z_D	(cm)	depth of doped layer
z_p	(cm)	period of superlattice
z_0	(cm)	width of ground-state subband
Z	(-)	charge state of center

Some quantities have no dimension and these are marked by '(-)'. If a quantity can have different dimensions, then the dimensions are omitted. Infrequently used symbols are defined in the text.

Published online by Cambridge University Press

Introduction

Impurities are the lifeblood of virtually all semiconductor devices. Impurities determine the semiconductive properties of materials and allow us to vary the conductivity from the semi-insulating via the semi-conductive to the semi-metallic range of the conductivity spectrum. The species and concentration of impurities determine the conductivity type and the free carrier concentration of semiconductors. For p-type and n-type impurities the electronic conduction occurs predominantly in the valence and conduction band, respectively. Employment of both conductivity types in one semiconductor makes possible pn-junctions, i.e. rectifying devices. Employment of layered n-p-n-type conductivity allows amplification and switching of electronic signals in transistors. Finally, free carriers must be injected or extracted in all optoelectronic devices. Thus, impurities providing free carriers form the basis of light-emitting diodes, current-injection lasers, photodetectors, and solar cells.

Impurities incorporated into a semiconductor lattice predominantly occupy substitutional lattice sites. In compound semiconductors, the lattice site can be either a cation or an anion site. Which of the two sites is preferred by an impurity depends on a number of factors, including the valence electron correlation between impurity and host, the bond strength between the impurity atom and the surrounding host lattice, and the size of the impurity (strain effects). Examples of impurities incorporated into a III–V semiconductor lattice are schematically shown in Fig. 1. Such compound semiconductors have a zincblende structure, in which cations and anions are bonded with tetrahedral symmetry. Intentional impurities are selected from columns adjacent to column III or column V of the Periodic Table, i.e. from column II, IV, or VI. Group-II impurities predominantly occupy group-III sites due to their valence electron configuration correlation. Substitution of the group-III element by a group-II impurity provides a 7 valence

https://doi.org/10.1017/CBO9780511599828.003 Published online by Cambridge University Press

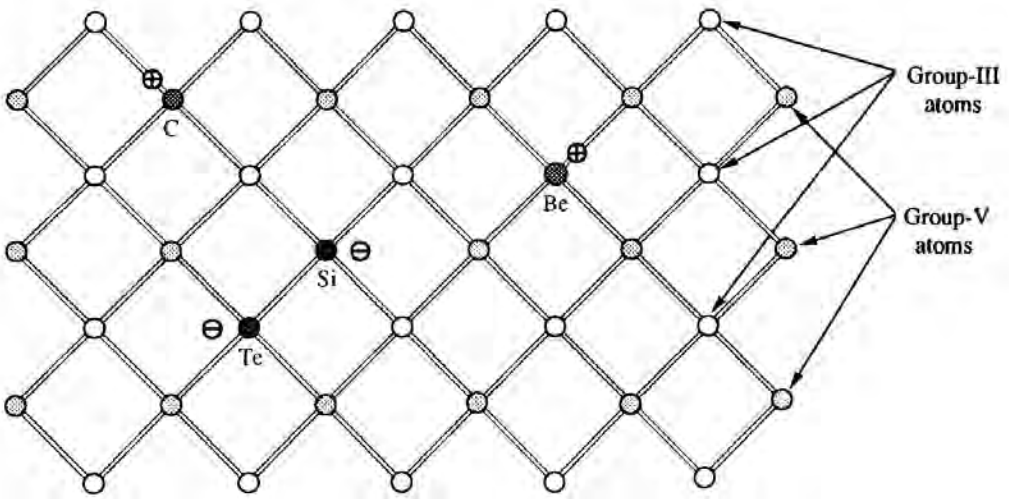


Fig. 1. Schematic illustration of the zincblende lattice consisting of group-III and group-V elements. The lattice is doped with donors (Si, Te) and acceptors (Be, C) which have an excess electron and a deficit electron, respectively.

electron configuration which is close to the 8 electron noble gas electron shell. Substitution of a group-V element by a group-II impurity would result in a 5 valence electron configuration, i.e. far from the 8 electron noble-gas configuration. For analogous reasons, group-VI impurities occupy group-V sites. Group-IV impurities can occupy both cation as well as anion sites. Thus, substitutional group-IV impurities have either one electron less or one electron more than the lattice atom they replaced, i.e. they can have an *excess electron* or a *deficit electron.*

It is a fundamental property of semiconductors that they have, in the intrinsic state, a completely filled valence band and an empty conduction band. Valence band states are formed from atomic states of p-orbital symmetry, while conduction band states derive themselves from atomic states of s-orbital symmetry. We first consider an impurity with an excess electron. If such an impurity is ionized in an otherwise pure semiconductor, the empty continuum state of lowest energy is a conduction band state. In the process of an electronic excitation, such impurities can *donate* their electron to the conduction band and are therefore called *donors*. We next consider an impurity with a deficit electron (hole). In an otherwise pure semiconductor, the hole state of such an impurity can be filled only by an electron from a filled band, e.g. the valence band. That is, such impurities can *accept* electrons from the valence band and are therefore called *acceptors*. Both types of impurities are schematically shown in Fig. 1.

https://doi.org/10.1017/CBO9780511599828.003 Published online by Cambridge University Press

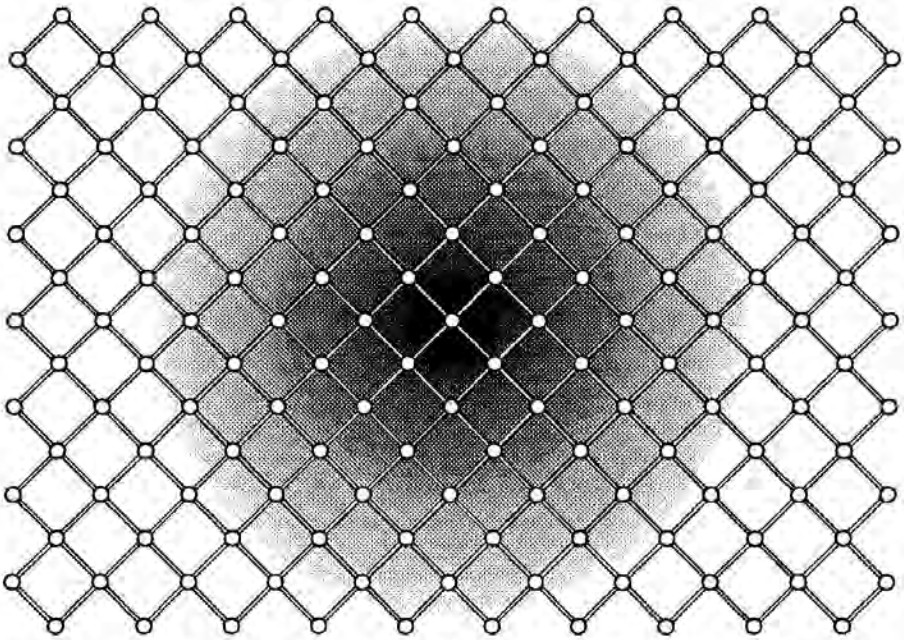


Fig. 2. Schematic illustration of a hydrogenic wave function of a donor. The effective Bohr radius is much larger than the period of the lattice.

The energy required to ionize a donor or acceptor is called the activation energy of the impurity. As we will see in Chap. 1, the activation energy is mainly determined by the coulombic interaction between the impurity atom and the charge carrier. This interaction is relatively weak due to dielectric screening and due to the light effective mass of the free carriers. The weak binding of charged carriers to an impurity has two notable consequences. First, the wave function of the bound state of the impurity is much larger than the unit cell of the lattice. This situation is shown in Fig. 2, which shows a schematic electronic wave function of s-orbital symmetry which extends over many lattice periods. In III–V semiconductors, the effective Bohr radius of donors can exceed 100 Å, i.e. the bound state wave function extends over more than 10 000 unit cells. Second, the activation energy of such shallow impurities is smaller than or comparable to the thermal energy kT at room temperature. This very much simplifies the treatment of the activation of shallow impurities. At room temperature, all shallow impurities can be considered as activated, i.e. the free carrier concentration coincides, in the absence of compensation effects, with the impurity concentration.

https://doi.org/10.1017/CBO9780511599828.003 Published online by Cambridge University Press

In practice, the characteristics of impurities are more complex than the simple coulombic substitutional center described above. An impurity can occupy sites other than substitutional sites, e.g. interstitial sites. Moreover, an impurity may form complexes with lattice atoms, other impurities, and native defects. Electrically, an impurity can be inactive, or compensating, which affects the efficiency of the dopant. The saturation of the electrical activity at high doping concentrations, and the amphoteric nature of some impurity species determine the properties of doped semiconductors as well. Furthermore, the characteristics of doped semiconductors change quite drastically at high doping concentrations. Effects associated with high doping concentrations are the Mott transition, the Burstein–Moss shift, band tail formation, impurity band formation, and bandgap shrinkage. These effects can change the characteristics of doped semiconductors drastically and therefore need to be taken into account.

Doping in compound semiconductor microstructures and heterostructures requires an unprecedented control of the spatial distribution of dopants. To illustrate the necessity of highly controlled doping profiles, we consider a hetero-bipolar transistor as an example. The band diagram of a hetero-bipolar transistor is shown in Fig. 3(a). The structure consists of a wide-gap n-type emitter, a narrow-gap p-type base and an n-type collector. One of the key advantages of a npn hetero-bipolar transistor is the reduction of the hole current from the base to the emitter. The hole current is reduced as compared to homojunction transistors due to the band discontinuity in the valence band. Under forward bias conditions of the emitter–base junction, the band edges are flat as shown in Fig. 3(a). Then, holes must overcome the energy discontinuity denoted as ΔE in Fig. 3(a). We next consider a hetero-bipolar transistor whose emitter–base pn-junction is displaced with respect to the heterojunction, as shown in Fig. 3(b). In this case, the effective energy barrier for the injection of holes into the emitter is reduced, which increases the magnitude of the hole current from the base into the emitter. The calculation of the barrier height for hole injection shows that even a minor displacement of the pn-junction by 20–50 Å has a deleterious effect. Accurate control of the spatial distribution of dopants is not only required in hetero-bipolar transistors but in many other structures, for example in selectively doped heterostructures.

The unprecedented control of the spatial distribution of impurities caused much interest in well-defined distributions including δ-function-like doping distributions. Impurities can redistribute during epitaxial growth as well as during post-growth or post-implantation processing. Such redistribution effects include diffusion, surface segregation, and surface migration. The assessment of the properties of impurities requires the characterization of semiconductor

https://doi.org/10.1017/CBO9780511599828.003 Published online by Cambridge University Press

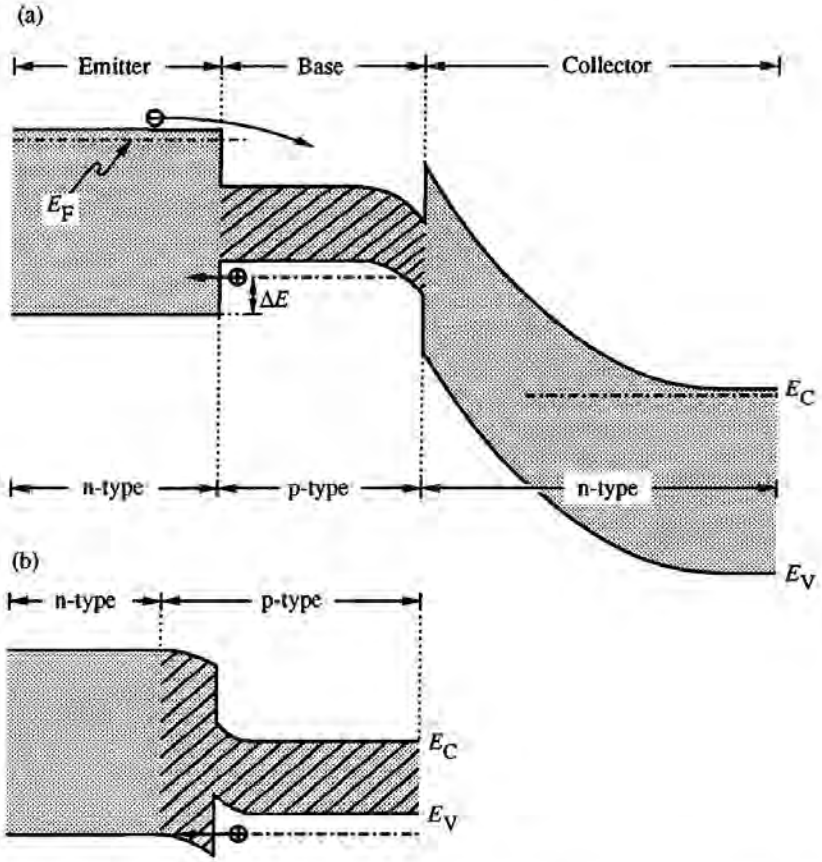


Fig. 3. Schematic band diagram of a hetero-bipolar transistor consisting of wide-gap emitter, base, and wide-gap collector. The locations of the emitter–base pn-junction and of the heterojunction coincide and are displaced in (a) and (b), respectively.

materials by means of electronic, optical, and structural measurement techniques. The techniques frequently reveal that III–V semiconductors contain shallow levels but also a wide variety of deep levels which can be due to impurity atoms, native defects, and complexes.

During the last decade, doping of III–V semiconductors has made tremendous progress. During this time, doping has emerged as an entire discipline of III–V semiconductor research and development. This discipline has many facets which are of great importance for further development and for production of photonic and electronic components and integrated microsystems. This book is dedicated to the many aspects of doping and may provide further insight into this discipline and also serve as a reference for those working with compound semiconductors.

https://doi.org/10.1017/CBO9780511599828.003 Published online by Cambridge University Press

1

Shallow impurities

Shallow impurities are of great technological importance in semiconductors since they determine the conductivity and the carrier type of the semiconductor. Shallow impurities are defined as impurities which are ionized at room temperature. This condition limits the ionization energy of such impurities to values << 100 meV. Shallow impurities can be either acceptors or donors, i.e. 'accept' electrons from the valence band or 'donate' electrons to the conduction band.

The hydrogen atom model can serve as the basis for the calculation of many properties of shallow impurities such as ionization energy and state wave functions. In this chapter, the hydrogen atom is analyzed in terms of Bohr's semi-classical model and in terms of a quantum mechanical approach. The hydrogen atom model is then applied to shallow impurities. Properties such as ionization energies, wave functions, central cell correction terms, and screening of impurity potentials by free carriers are summarized. Effects associated with high impurity or high free carrier concentrations are also discussed including the Mott transition, the Burstein–Moss shift, band tails, impurity bands, and bandgap shrinkage.

1.1 Hydrogenic impurities

Impurities in semiconductors can be incorporated on substitutional sites, interstitial sites, or as impurity complexes. Here, we restrict ourselves to *substitutional, shallow* impurities. Examples for such impurities are Be, Zn, Si, and Sn. These impurities are shallow, i.e. their ionization energy is comparable to the thermal energy kT at room temperature. As a consequence, shallow impurities are fully ionized at room temperature. The hydrogen atom model has proven to

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

predict accurately many properties of shallow impurities. This section summarizes the properties of the semi-classical hydrogen atom, the quantum mechanical hydrogen atom, hydrogenic donors, and hydrogenic acceptors. Finally, central-cell corrections to the hydrogenic model are discussed.

1.1.1 Coulomb potentials and Bohr's hydrogen atom model

The electrostatic potential of a point charge is called the Coulomb potential or the $1/r$ potential. The Coulomb potential of a positive point charge $(+e)$ in vacuum located at $r = 0$ is obtained from Poisson's equation and is given in spherical coordinates by

$$V(r) = \frac{e}{4\pi\varepsilon_0 r} \tag{1.1a}$$

where e is the elementary charge and ε_0 is the permittivity of vacuum. Analogously, the Coulomb potential of a positively charged impurity located at $r = 0$ in a semiconductor with the dielectric constant $\varepsilon_r = \varepsilon/\varepsilon_0$ is given by

$$V(r) = \frac{e}{4\pi\varepsilon r} \tag{1.1b}$$

where ε is the permittivity of the semiconductor.

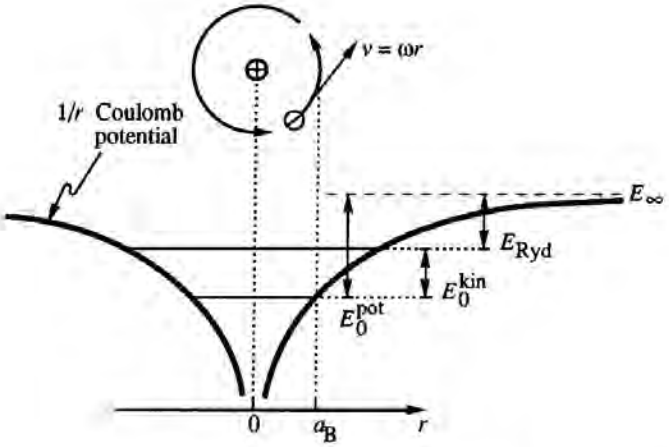


Fig. 1.1. Illustration of Bohr's hydrogen atom model. The electron orbits the positive proton with a $1/r$ Coulomb potential at a radius r_B. The ionization energy of the atom is the Rydberg energy $E_{Ryd} = E_\infty - |E_0^{pot}| + |E_0^{kin}|$.

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

The hydrogen atom model developed by Bohr is based on (i) classical mechanics of an electron in the Coulomb potential of a positive point charge and on (ii) the quantization of the electron angular momentum. The Bohr model predicts many of the physical properties of the hydrogen atom most notably the emission spectra of the atom. The model is a fascinating example of the simplicity and the power of quantum mechanics. For the classical motion of an electron in the Coulomb potential, the potential energy is given by

$$E_{\text{pot}} = \frac{-e^2}{4\pi\varepsilon_0 r}. \tag{1.2}$$

For the hydrogen atom, the permittivity is that of vacuum since a vacuum is assumed between the proton and the electron. The schematic Coulomb potential and an electron orbiting the proton at a distance r are shown in Fig. 1.1. The *attractive* electrostatic Coulomb force F_c on the proton and the electron towards each other is given by

$$F_c = e\mathcal{E}_c = \frac{1}{4\pi\varepsilon_0}\frac{e^2}{r^2} \tag{1.3}$$

where $\mathcal{E}_c = dE_{\text{pot}}/dr$ is the Coulomb field. The *repulsive* centrifugal force F_z on the orbiting electron with mass m_0 and velocity v is given by

$$F_z = \frac{m_0 v^2}{r} = m_0 r\omega^2 \tag{1.4}$$

where ω is the angular frequency of the electron. Equation (1.4) is valid, only if the electron mass, m_0, is much smaller than the proton mass, m_p. This condition is fulfilled, since $m_0/m_p \cong 1/1840$. The first classical condition in Bohr's hydrogen atom model is $F_c = F_z$, i.e. the balance of attractive and repulsive forces.

The second condition for the Bohr atom is the quantization of the angular momentum of the electron which is given by

$$\boxed{m_0 vr = m_0 \omega r^2 = n\hbar \quad (n = 1,2,3\cdots)} \tag{1.5}$$

The validity of the angular momentum quantization can be visualized by recalling the wave character of the electron. An electron wave around the positive proton is shown in Fig. 1.2. The electron wave is stable, only if the wave is interfering constructively with itself, i.e. when the length of the electron orbit equals integer multiples of the electron wavelength, λ,

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

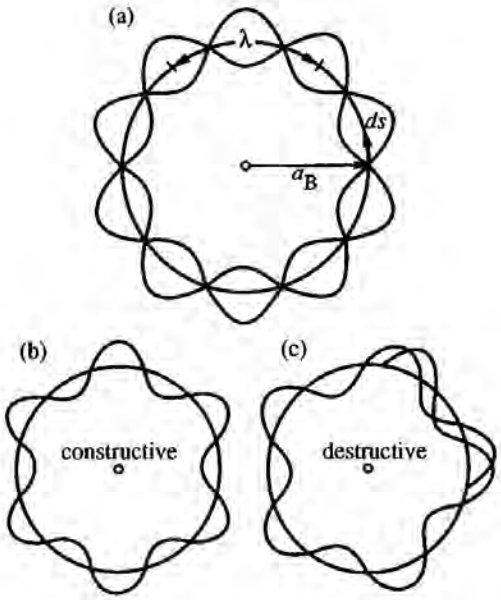


Fig. 1.2. Quantization of the electron orbit of an atom in terms of integer multiples of the electron de Broglie wavelength. Only if $2\pi r_B = n\lambda$, does constructive interference of the electron wave occur.

$$2\pi r = n\lambda \quad (n = 1,2,3\cdots) . \tag{1.6}$$

The reader can easily verify that Eqs. (1.5) and (1.6) are identical by recalling that the kinetic energy of a particle is given by $E = \hbar^2 k^2/2m_0$ where $k = 2\pi/\lambda$ and m_0 are the electron wave vector and mass, respectively.

Elimination of force and velocity from Eqs. (1.3)–(1.5) yields the radii of the allowed electron orbits in the hydrogen atom

$$\boxed{a_{B,n} = \frac{4\pi\varepsilon_0 n^2\hbar^2}{m_0 e^2} \quad (n = 1,2,3\cdots)} \tag{1.7}$$

The radius of the ground state orbit ($n=1$) is given by

$$a_B = 0.53 \text{ Å} \tag{1.8}$$

which is called the ***Bohr radius*** of the hydrogen atom.

Insertion of the Bohr radii into Eq. (1.2) yields the potential energy of the electron

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

$$E_{\text{pot},n} = \frac{-1}{(4\pi\varepsilon_0)^2}\,\frac{e^4 m_0}{n^2\hbar^2} \quad (n = 1,2,3\cdots)\,. \tag{1.9}$$

Furthermore, the kinetic energy is obtained via the electron velocity of Eq. (1.5) according to

$$E_{\text{kin},n} = \frac{1}{2}\,\frac{1}{(4\pi\varepsilon_0)^2}\,\frac{e^4 m_0}{n^2\hbar^2} \quad (n = 1,2,3,\cdots)\,. \tag{1.10}$$

Comparison of Eq. (1.9) with Eq. (1.10) reveals that the kinetic energy is just half of the potential energy, i.e.

$$\boxed{E_{\text{kin},n} = \tfrac{1}{2}\,|E_{\text{pot},n}|} \tag{1.11}$$

The energy required to move the electron from the nth state energy, E_n, to the vacuum level at infinite distance from the proton, i.e. $E_\infty = E(r\to\infty)$, is given by

$$\boxed{E_{\text{Ryd},n} = \frac{1}{2}\,\frac{1}{(4\pi\varepsilon_0)^2}\,\frac{e^4 m_0}{n^2\hbar^2} \quad (n = 1,2,3\cdots)} \tag{1.12a}$$

which, for $n = 1$, is called the ***Rydberg energy***. This energy is required to ionize a hydrogen atom. For $n = 1$ the Rydberg energy is given by

$$E_{\text{Ryd}} = 13.6\ \text{eV}\,. \tag{1.12b}$$

In the classical orbital motion of the electron around the proton, the ratio of the electron velocity and the velocity of light c can be calculated from the Bohr model. The ratio is obtained as

$$\alpha = \frac{v}{c} = \frac{e^2}{4\pi\varepsilon_0\hbar c} \cong \frac{1}{137} \tag{1.13}$$

which is called the ***Sommerfeld fine structure constant***. Evaluation of Eq. (1.13) yields that the electron orbits the proton with a velocity of approximately 2200 km/s.

The magnetic field generated by the circular current of the orbiting electron can be calculated from Bohr's model using the Maxwell equations. It is given by

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

$$\mu_B = \frac{1}{2}\,\mu_0\,\hbar\,\frac{e}{m_0} \tag{1.14}$$

and is called the ***Bohr magneton***.

The above calculation demonstrates that the relatively simple Bohr model, i.e. classical mechanics and angular momentum quantization, provides many physical quantities of the hydrogen atom. The calculated state energies of the hydrogen atom were found to agree with hydrogen emission spectra. The Bohr model and its prediction of the electron energies was one of the first successes of the quantum theory. Further refinement of the model is obtained by considering not only circular orbits but also elliptical orbits. On such an elliptical orbit the velocity of the electron is a function of the position, i.e. the velocity is not constant as in the circular orbit. (The position dependence of the velocity is analogous to the planetary motion around the sun).

Using the momentum $p = \hbar k$ and $k = 2\pi/\lambda$, the angular momentum quantization condition, which for circular motion is given by Eq. (1.6), can be written for any orbit as

$$\boxed{\frac{1}{2\pi}\int_0^{2\pi} p_\phi \,\mathrm{d}\phi = n_\phi \hbar \quad (n_\phi = 1,2,3,\cdots)} \tag{1.15}$$

where $p_\phi = m\omega_\phi r^2$ is the position-dependent (that is angle-dependent) angular momentum. For a circular orbit, the angular momentum is a constant and Eq. (1.15) reduces to Eq. (1.5).

The condition of classical mechanics for motion on an elliptical orbit and the angular momentum quantization condition lead to the total energy of the electron. The total energy is given by

$$E_{n_\phi} = \frac{1}{2}\,\frac{1}{(4\pi\varepsilon_0)^2}\,\frac{e^4 m_0}{n_\phi^2 \hbar^2} \quad (n_\phi = 1,2,3) \tag{1.16}$$

which is identical to Eq. (1.12). Thus, elliptical orbits for the electron exist and have the same energy as electrons on circular orbits. The total energy of a particle on an elliptical orbit in a $1/r$ potential can be calculated by classical mechanics and depends only on the main axis a of the ellipsis. The main axis is then given by

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

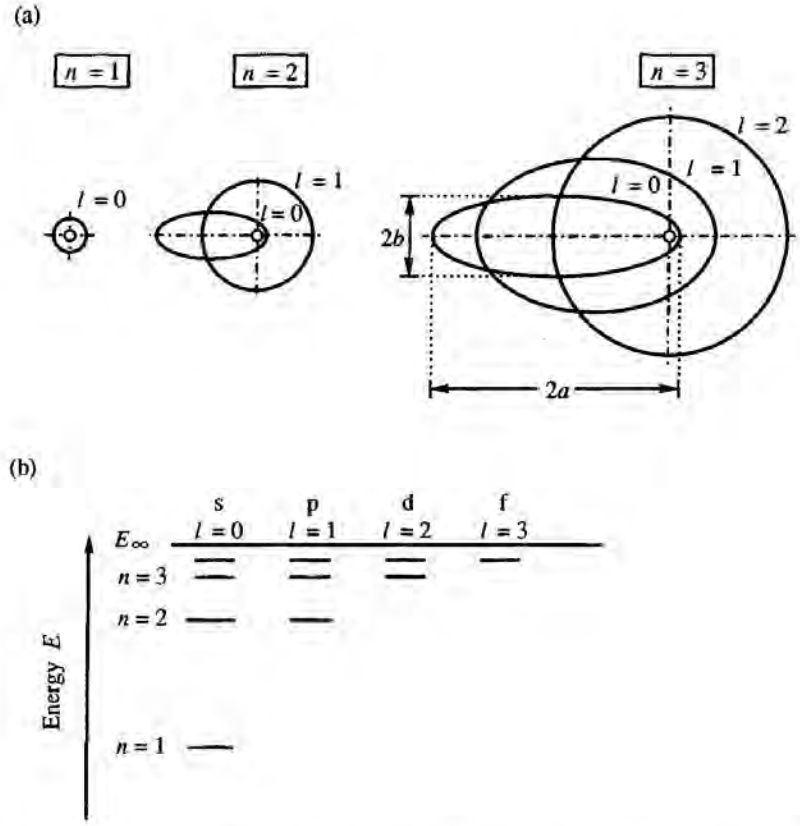


Fig. 1.3. (a) Allowed orbits of an electron in a hydrogen atom. The energies of different electron states are identical for electron orbits with the same main axis of the ellipses. (b) Energies of the electron in the hydrogen atom for different principal and angular momentum quantum numbers.

$$a = \frac{4\pi\hbar^2\,\varepsilon_0}{m_0 e^2}\, n_\phi^2 \,. \tag{1.17}$$

The angular momentum of the particle on the elliptical orbit is given by

$$p_\phi^2 = \frac{a(1-\varepsilon^2)\, e^2 m_0}{4\pi\varepsilon_0} \tag{1.18a}$$

where ε is the eccentricity of the ellipsis. Since the angular momentum is quantized according to $p_\phi = n_\phi \hbar$, one obtains with Eq. (1.18a)

$$n_\phi^2 \hbar^2 = \frac{a(1-\varepsilon^2) e^2 m_0}{4\pi\varepsilon_0} \,. \tag{1.18b}$$

Inserting the main axis, a (see Eq. (1.17)) into Eq. (1.18b) yields

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

$$\frac{b}{a} = \frac{n_\phi}{n} \tag{1.19}$$

where a and b are the axes of the ellipsis. The angular momentum quantum number n_ϕ can assume values of 1, 2, 3, ⋯, n, which represents a family of ellipses. If the angular quantum number coincides with the principal quantum number, i.e. $n_\phi = n$, the previously calculated circular orbit is obtained. The ellipses for the $n = 1,2$ and 3 states are shown in Fig. 1.3(a).

If the angular momentum quantum number is formally introduced as $l = n_\phi - 1$, then l can assume values of

$$l = 0\ ,1\ ,2\ ,\ \cdots,\ n-1\ . \tag{1.20}$$

The value of $l = 0$ corresponds to the ellipsis with the largest eccentricity. The value of $l = n-1$ represents the circular orbit.

Each orbit of an electron around the proton of a hydrogen atom is fully determined by the principle quantum number, n, and the angular quantum number, l. The orbits of the electron for different quantum numbers and the corresponding energies are shown in Fig. 1.3. Frequently, the $l = 0$, 1, 2 and 3 orbitals are historically denoted as the s, p, d and f orbitals (White, 1934). If several quantum states have the same energy the states are called ***degenerate states***. For example, the two states determined by $n = 2, l = 0$ and $n = 2, l = 1$ are degenerate. In addition to the principal quantum number and the angular momentum number, the quantum number m describes the quantization of the azimuthal angular momentum in units of $\hbar$. For a discussion of the azimuthal quantum number, we refer to the literature (see, for example, Bohm, 1951).

Upon any perturbation of the hydrogen atom, different electron orbits respond in different ways to the perturbation. Therefore, degenerate electron states will split and become non-degenerate upon a suitable perturbation. The perturbation of the hydrogen atom can be achieved, for example, by an electric field (Stark effect) or a magnetic field (Zeeman effect).

Even though the Bohr model explains many characteristics of the hydrogen atom it is limited in its applicability. For example, if the principles of the Bohr model are applied to the helium atom, incorrect results are obtained for the energy levels in that atom. In addition, the deterministic Bohr hydrogen model violates the quantum mechanical uncertainty principle. That is, momentum and position of the electron are exactly determined at all times in Bohr's model, which contradicts $\Delta x \Delta p \cong \hbar$. Nevertheless, due to its simplicity and clarity, the Bohr model has not lost its attractiveness.

The Bohr model was refined in 1925 by inclusion of the electron spin. The spin of an electron is also called its intrinsic angular momentum and can be

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

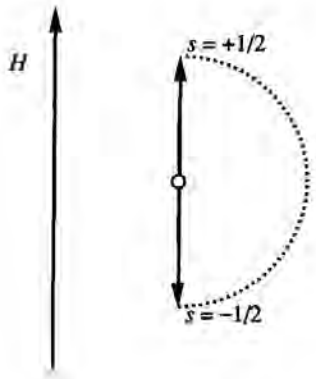


Fig. 1.4. Alignement of the spin of an electron in an external magnetic field H. The spin is either parallel (spin 'up') or antiparallel (spin 'down') to the magnetic field, as represented by the spin quantum number $s = \pm 1/2$. The difference in angular momentum between the two orientations is $\hbar$.

visualized as the rotation of an electron around its own symmetry axis. Goudsmit and Uhlenbeck postulated the spin when conducting Zeeman effect experiments on hydrogen. Due to angular momentum quantization the difference between intrinsic angular momenta is $\hbar$. The intrinsic angular momentum is then given by

$$p_s = s\hbar \quad \text{with} \quad s = \pm ½ \tag{1.21}$$

where s is the spin quantum number which can assume values of $s = \pm 1/2$. If an electron with spin is subjected to an external magnetic field, the spin-axis will orientate (align) itself in a parallel or antiparallel manner, as shown in Fig. 1.4. All other orientations are of transient nature since the repulsive and attractive magnetic forces of the spin and external field tend to reorientate the spin to the parallel or antiparallel orientation.

1.1.2 Wave functions of the hydrogen atom

The Bohr model predicts the energy levels of the hydrogen atom with amazing accuracy. However, the wave functions ψ_{nlm} and the probability distributions $\psi_{nlm}\,\psi^*_{nlm}$ cannot be obtained from the Bohr model. The Schrödinger equation must be solved in order to obtain the exact solutions. Although the exact solution of the hydrogen atom model goes beyond the scope of this book, the results of the exact hydrogen atom are summarized in the following. For further study, the reader is referred to textbooks on quantum mechanics (Bohm, 1951; Sherwin, 1959; Davydov, 1965; Borowitz, 1967; Saxon, 1968; Merzbacher, 1970).

The wave functions of the hydrogen atom can be obtained by solving the Schrödinger equation. For a particle with mass m and charge e in a potential V, the Schrödinger equation is given by

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

$$-\frac{\hbar^2}{2m}\,\Delta\Psi + eV\Psi = -\frac{\hbar}{\mathrm{i}}\,\frac{\partial}{\partial t}\,\Psi \tag{1.22}$$

where $\Psi = \Psi(x, y, z, t) = \psi(x, y, z)\mathrm{e}^{\mathrm{i}\omega t}$ is the time-dependent wave function and $\Delta = (\partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2)$ is the delta operator. Not being interested in the time dependence of the solution, we use the kinetic energy operator $E = -(\hbar/\mathrm{i})(\partial/\partial t)$ to obtain the time-independent Schrödinger equation

$$\boxed{-\frac{\hbar^2}{2m}\,\Delta\psi + (eV-E)\psi = 0} \tag{1.23}$$

where $\psi = \psi(x,y,z)$ is the time-independent wave function. The Coulomb potential is a spherically symmetric potential and can be expressed solely as a function of the radius r, i.e. $V = V(r)$ (see Eq. (1.1)). It is useful to convert the Δ-operator into spherical coordinates (r, θ, ϕ), i.e.

$$\begin{aligned}\Delta &= \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \\ &= \frac{1}{r^2}\,\frac{\partial}{\partial r}\left[r^2\,\frac{\partial}{\partial r}\right] + \frac{1}{r^2\sin\theta}\,\frac{\partial}{\partial\theta}\left[\sin\theta\,\frac{\partial}{\partial\theta}\right] + \frac{1}{r^2\sin^2\theta}\,\frac{\partial^2}{\partial\phi^2}\end{aligned} \tag{1.24}$$

where r, θ and ϕ are the radius, the polar angle, and azimuthal angle, respectively. The Schrödinger equation is a separable linear differential equation and can be solved by employing the product method. The wave functions can then be written as

$$\psi(r,\theta,\phi) = R(r)\Theta(\theta)\Phi(\phi)\,. \tag{1.25}$$

Since $\psi\psi^*$ is the quantum mechanical probability density of the particle, the wave function ψ must satisfy the condition

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi\psi^*\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = \int_0^{\infty}\int_0^{2\pi}\int_0^{\pi}\psi\psi^*\,r^2\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi\,\mathrm{d}r = 1\,. \tag{1.26}$$

The solution of the Schrödinger equation in spherical coordinates for the Coulomb potential is a set of orthogonal functions which are usually classified by the four quantum numbers n, l, m_l, and s. The quantum numbers are

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

n = the principal quantum number
l = the orbital angular momentum number
m_l = the azimuthal quantum number
s = the spin quantum number

The quantum number can assume values of

$$n = 1,2,3,\ldots \tag{1.27a}$$

$$l = 0,1,2,\ldots,n-2,\ n-1 \tag{1.27b}$$

$$m_l = -l,-l+1,\ldots,l-1,+l \tag{1.27c}$$

$$s = -\tfrac{1}{2},\ +\tfrac{1}{2}\ . \tag{1.27d}$$

Two electrons with different spin can occupy an orbit defined by the three quantum numbers n, l, and m_l (Pauli principle).

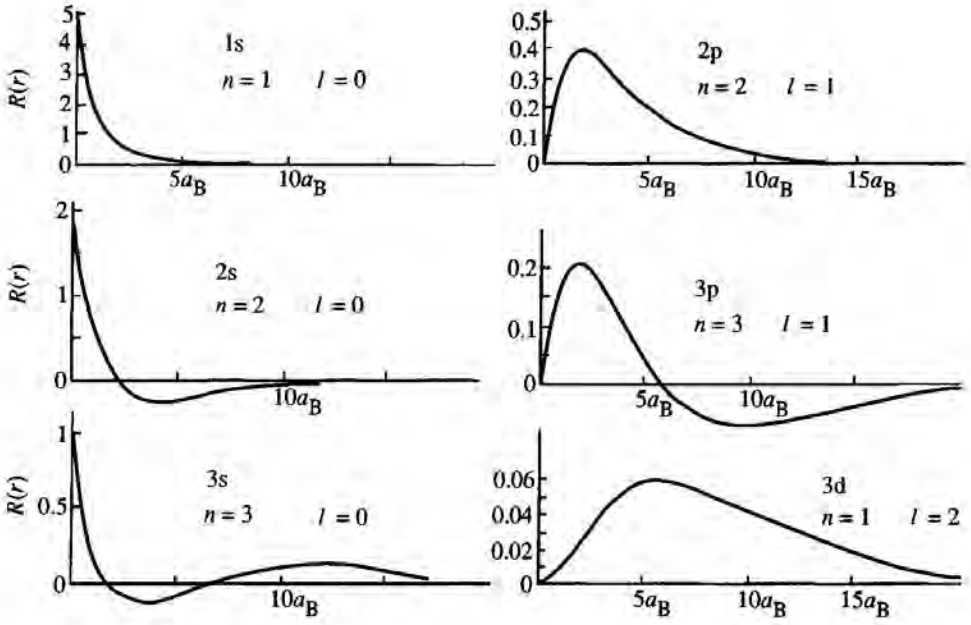


Fig. 1.5. Calculated radial parts of the wave functions for some elementary states of the hydrogen atom.

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

The wave functions corresponding to the three quantum numbers n, l, and m_l are designated as ψ_{nlm_l}. Correspondingly, the radial part of the wave functions are denoted as R_{nlm_l} (see (Eq. 1.25)). The wave functions for some of the lowest states of the hydrogen atom are (Sherwin, 1959)

$$\psi_{100} = \frac{1}{\sqrt{\pi}} \left[\frac{1}{a_B}\right]^{3/2} e^{-r/a_B} \tag{1.28a}$$

$$\psi_{200} = \frac{1}{4\sqrt{2\pi}} \left[\frac{1}{a_B}\right]^{3/2} \left[2 - \frac{r}{a_B}\right] e^{-r/2a_B} \tag{1.28b}$$

$$\psi_{210} = \frac{1}{4\sqrt{2\pi}} \left[\frac{1}{a_B}\right]^{3/2} (\cos\theta) \frac{r}{a_B} e^{-r/2a_B} \tag{1.28c}$$

$$\psi_{211} = \frac{1}{4\sqrt{2\pi}} \left[\frac{1}{a_B}\right]^{3/2} \frac{e^{i\phi}}{\sqrt{2}} (\sin\theta) \frac{r}{a_B} e^{-r/2a_B} \tag{1.28d}$$

$$\psi_{21-1} = \frac{1}{4\sqrt{2\pi}} \left[\frac{1}{a_B}\right]^{3/2} \frac{e^{-i\phi}}{\sqrt{2}} (\sin\theta) \frac{r}{a_B} e^{-r/2a_B} . \tag{1.28e}$$

Further solutions of the hydrogen atom state can be found in the literature (Bohm, 1951). The radial parts R_{nl0} ($m_l = 0$) for some of the lowest hydrogen atom states are shown in Fig. 1.5. States of 's-type' (i.e. $l = 0$) symmetry have a maximum of the wave function at $r = 0$. States of 'p-type' (i.e. $l = 1$) symmetry have a node at $r = 0$.

The probability of finding an electron at radius r can be obtained by integration over all angles θ and ϕ

$$p(r)\,dr = \int_0^{2\pi} \int_0^{\pi} \psi\psi^* r^2 \sin\theta \, d\theta \, d\phi \, dr . \tag{1.29}$$

For wave functions of s-type symmetry, ψ does not depend on θ and ϕ. Eq. (1.29) then yields

$$p(r) = \psi\psi^* 4\pi r^2 . \tag{1.30}$$

Using $\psi = \psi_{100}$ according to Eq. (1.28) one obtains

$$p(r) = 4 \frac{r^2}{a_B^3} e^{-2r/a_B} . \tag{1.31}$$

The probability $p(r)$ has a maximum at $r = a_B$. Thus, the classical Bohr radius

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

is the radius of maximum probability in the quantum mechanical picture.

1.1.3 Hydrogenic donors

The hydrogen atom model can be applied to shallow donors in III–V semiconductors. Properties predicted by the hydrogen atom model agree amazingly well with experimentally determined properties of shallow donors. Such donors are called ***effective-mass-like donors, hydrogen-like donors***, or briefly ***hydrogenic donors***. Ionization energy, wave functions, and effective Bohr radius are well predicted for such donors. The similarity of the hydrogen atom and donor impurities originates in the $1/r$ coulombic potential of both entities.

Two modifications are required in order to apply Bohr's hydrogen atom model to shallow donors. These corrections are related to the effective mass of carriers and of the dielectric constant of the semiconductor (Bethe, 1942). First, the effective mass of electrons in semiconductors, m_e^*, differs from the electron mass quite significantly. The dispersion relation of a semiconductor with a spherically symmetric band structure is given by

$$E = \frac{\hbar^2 k^2}{2m_e^*} . \tag{1.32}$$

Thus, the dispersion relation $E(k)$ allows one to determine the effective mass according to

$$m_e^* = \hbar^2 \left(\frac{\mathrm{d}^2 E}{\mathrm{d}k^2} \right)^{-1} . \tag{1.33}$$

The dispersion relation of Eq. (1.32) differs from the dispersion relation of a free electron just by the magnitude of the electron mass. In order to apply the hydrogen atom model, the electron mass must be replaced by the effective electron mass. The second correction arises from the dielectric properties of semiconductors. The Coulomb potential of a positive point-charge in a semiconductor located at $r = 0$ is given by

$$V(r) = \frac{e}{4\pi\varepsilon r} \tag{1.34}$$

which differs from Eq. (1.1a) by the static dielectric constant $\varepsilon_r = \varepsilon/\varepsilon_0$. How does the potential change in the presence of an electron orbiting the impurity charge? The polarization of the valence electrons is then more complicated and cannot be taken into account by the substitution of ε for ε_0. To answer the question we first make the simplifying assumption that the electronic charge can be described by a diffuse electron cloud with a spatial extent much larger than the lattice constant. In the limit of an infinitely large electron cloud, the potential of

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

the positive impurity is correctly described by Eq. (1.34). The situation changes, however, if the electron were to orbit the impurity atom with a radius comparable to the lattice constant. In this case the polarization of the lattice depends on the donor as well as on the electron charge. The true potential is not given by the dielectrically screened potential of Eq. (1.34). For a finite size of the electron orbit, the polarization of lattice atoms is overestimated by Eq. (1.34). A smaller dielectric constant $\varepsilon_r^* < \varepsilon_r$ can be used to account for the reduction in polarization. It should therefore be noted that Eq. (1.34) assumes a Coulomb potential screened by the dielectric properties of the semiconductor (i.e. by polarization of tightly bound valence electrons and nuclei of the lattice) and that the equation can only be used if the electron can be described by a diffuse electron cloud with a large spatial extent. The true potential $V(r)$ which arises from the positive donor ion, the electron bound to the donor ion, and the polarization of the surrounding semiconductor is rather complicated and cannot be expressed in terms of a simple $1/r$ potential (Kohn, 1957a, 1957b). Employing the approximate $1/r$ potential, the Schrödinger equation is given by

$$-\frac{\hbar^2}{2m_e^*}\Delta\psi + \frac{e}{4\pi\varepsilon r}\psi = E\psi \tag{1.35}$$

which is called the effective-mass equation for a hydrogenic impurity.

Using the effective-mass and the dielectric constant corrections, the following properties of hydrogenic impurities can be derived. The ***effective Bohr radius*** is obtained from Eq. (1.7) and is given by

$$a_{B,n}^* = \frac{4\pi\varepsilon n^2\hbar^2}{m_e^* e^2} \quad (n=1,2,3,\cdots)\,. \tag{1.36}$$

The radius of the donor ground state ($n = 1$) is then given by

$$a_B^* = \frac{4\pi\varepsilon\hbar^2}{m_e^* e^2} = \frac{\varepsilon_r}{m_e^*/m_0}\, a_B = \frac{\varepsilon_r}{m_e^*/m_0}\, 0.53\ \text{Å} \tag{1.37}$$

The effective Bohr radius is also called the *donor Bohr radius*. As an example, we consider a hydrogenic donor in GaAs with $\varepsilon_r = 13.1$ and $m_e^* = 0.067\ m_0$. Insertion of these values into Eq. (1.37) yields $a_B^* = 103$ Å which is the effective Bohr radius of donors in GaAs.

The ***effective Rydberg energy*** is obtained by applying the effective-mass and the dielectric constant corrections to Eq. (1.12).

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

$$E^*_{\mathrm{Ryd},n} = \frac{1}{2}\,\frac{1}{(4\pi\varepsilon)^2}\,\frac{e^4 m_e^*}{n^2\hbar^2} \quad (n=1,2,3,\ \ldots)\,. \tag{1.38}$$

The ***donor ionization energy*** is required for a transition from $n=1$ to $n\to\infty$ and is given by

$$\boxed{E_\mathrm{d} = \frac{e^4 m_e^*}{2\,(4\pi\varepsilon\hbar)^2} = \frac{m_e^*/m_0}{\varepsilon_\mathrm{r}^2}\,E_\mathrm{Ryd} = \frac{m_e^*/m_0}{\varepsilon_\mathrm{r}^2}\,13.6\ \mathrm{eV}} \tag{1.39}$$

The donor ionization energy is occasionally also referred to as donor Rydberg energy. As an example, we consider a hydrogenic donor in GaAs and obtain $E_\mathrm{d} = 5.3$ meV which is in agreement with experimental results.

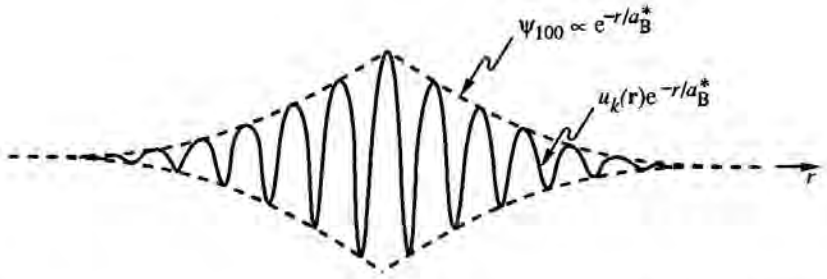


Fig. 1.6. Schematic illustration of a hydrogenic donor wave function with the quantum numbers $n = 1$ and $l = m_l = 0$. The wave function is the product of the lattice-periodic Bloch function $u_k(\mathbf{r})$ and the envelope function ψ_{100}.

Finally, the wave functions of hydrogenic donors can be obtained from Eq. (1.28) by substituting the effective Bohr radius for the Bohr radius. The ground-state envelope wave function is then obtained as

$$\psi_{100}(r) = \frac{1}{\sqrt{\pi}}\left[\frac{1}{a_\mathrm{B}^*}\right]^{3/2} e^{-r/a_\mathrm{B}^*}\,. \tag{1.40}$$

It should be noted that Eq. (1.40) describes the donor *envelope* function rather than the donor wave function. The actual donor ground-state wave function is given by (Kohn, 1957a)

$$\psi_{\mathrm{d},100}(\mathbf{r}) = \psi_{100}(r)\,u_k(\mathbf{r}) \tag{1.41}$$

where $u_k(\mathbf{r})$ is the lattice-periodic factor of the well-known Bloch function of conduction band electrons. The function $u_k(\mathbf{r})$ has translational symmetry with

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

337-TA-1366 RX-0255

respect to the semiconductor lattice constant. The ground-state wave function according to Eq. (1.41) is schematically shown in Fig. 1.6. The dashed curve represents the impurity envelope function of Eq. (1.40).

It should be noted that the use of Bohr's hydrogen atom model for shallow impurities is not self-sufficient. The *ab initio* assumption of a 'large' electron cloud and the substitution of effective electron mass and dielectric constant cannot be justified solely on the basis of the hydrogenic model. Even though the hydrogenic model yields a relatively large electron orbit, this result does not justify the initial assumptions. However, more rigorous calculations (Kohn 1957a, 1957b; Madelung, 1978; Altarelli and Bassani, 1982) indeed demonstrate that the electron distribution has a spatial extent much larger than the lattice constant. The substitution of the electron mass, m_0, by the effective mass, m_e^*, is therefore justified since the electron orbit around the donor extends over many lattice constants. The effective mass of electrons in semiconductors is a direct consequence of the periodic potential of the lattice. Thus, electrons bound to donors are subject to the periodic potential, since the effective Bohr radius is much larger than the lattice constant, $a_B^* >> a_0$. If, in contrast, electrons were tightly bound ($a_B^* \cong a_0$) the effective-mass correction could not be applied. Similar arguments apply to the substitution of the permittivity of the semiconductor, ε, for the permittivity of vacuum, ε_0. The lattice atoms are polarized by the Coulomb field which results in its reduction as compared to the field without polarization. The effect of the polarization is taken into account via the dielectric constant. Since the effective Bohr radius extends over many lattice constants, the use of the dielectrically screened Coulomb potential is justified. Despite the simplicity of the hydrogen atom model for shallow donors, the model yields quite accurate results.

The ***degeneracy*** of the donor ground state is a quantity required for the occupancy probability of the donor state (see Chap. 3). The ground state has the quantum numbers $n = 0$, $l = 0$, $m_l = 0$ and $s = \pm ½$. Thus, since the donor ground state can be occupied by an electron with spin $+½$ or $-½$, the ground state degeneracy is $g = 2$.

1.1.4 Hydrogenic acceptors

The application of the hydrogenic model to acceptors in III–V semiconductors is complicated by their degenerate valence band structure. For hydrogenic donors, the effective electron mass was substituted for the electron mass. The substitution was possible, since the conduction band was assumed to be parabolic, isotropic, and non-degenerate (as for most III–V semiconductors). Such a simple substitution is not possible for acceptors, since the valence band structure of III–V

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

RX-0255.44

337-TA-1366 RX-0255

semiconductors is much more complicated than the conduction band structure. The electronic band structure of several III–V semiconductors with zincblende structure was calculated by Chelikowsky and Cohen (1976). The band structure near the center of the Brillouin zone is schematically shown in Fig. 1.7. The highest point of the valence band is located at $k = 0$. Without spin-orbit coupling this point would be sixfold degenerate with three dispersion relations and twofold spin degeneracy (Kohn, 1957a). The simplest way to understand this degeneracy is to consider the tight binding limit, in which the wave functions corresponding to the highest point go over into atomic 3p functions. The spin-orbit coupling lifts the degeneracy partially and leads to the situation shown in Fig. 1.7. The top of the valence band remains fourfold degenerate at $k = 0$. The corresponding dispersion relations are called the heavy hole (hh) and light hole (lh) dispersion relations. The top of the twofold (spin) degenerate split-off (so) band is at $k = 0$ at an energy E_{so} below the valence band maximum where E_{so} is the spin-orbit coupling energy. The top of the valence band corresponds to atomic $j = 3/2$ states (Kohn, 1957a), where j is the total angular momentum (obit + spin), i.e. $j = l+s = l + ½ = 3/2$. The *inner quantum number* j is formally not necessary since it can be expressed by l and s. The introduction of j by Sommerfeld (1920) has historic reasons (Finkelnburg, 1958). The top of the valence band has Γ_8 symmetry. The split-off band corresponds to atomic $j = l-s = 1 - ½ = ½$ states which have Γ_7 symmetry.

For strong spin-orbit coupling, the split-off band is far removed from the top of the valence band, i.e. $\Delta E_{so} >> E_a$ where E_a is the acceptor binding energy. In this case, both the heavy hole and light hole band must be taken into account and the Hamiltonian is, therefore, a 4×4 matrix (Kohn, 1957a). In the limit of weak spin-orbit coupling, i.e. $\Delta E_{so} \equiv E_a$, all three valence bands must be taken into account and the Hamiltonian is a 6×6 matrix.

Kohn (1957a) used a 6×6 Hamiltonian matrix to calculate acceptor energies in cubic semiconductors. The author used variational envelope wave functions for acceptors of the form

$$\psi_i(r) = A_i e^{-r/r_i} \tag{1.42}$$

where r_i is a variational parameter. Subsequently, Baldareschi and Lipari (1973) developed a now widely accepted model for shallow acceptor states in cubic semiconductors with degenerate valence bands. In their approach, the Hamiltonian is written as the sum of a spherical term and a cubic correction, thus pointing out the relevance of the spherical symmetry in the acceptor problem and the strong similarity to the case of atoms with spin-orbit interaction. Neglecting the cubic term, Hamiltonians with radial symmetry were obtained. Variational

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

Innoscience1366-00138718

RX-0255.45

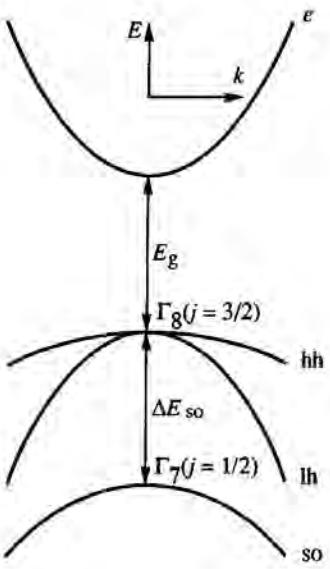


Fig. 1.7. Schematic illustration of the electron, heavy hole (hh), light hole (lh) and split-off (so) dispersion relations near the center of the Brillouin zone ($k = 0$).

wave functions were used to calculate acceptor ionization energies.

In the limit of strong spin-orbit interaction, that is for spin-orbit splittings (ΔE_{so}) much larger than the acceptor energy (E_a), Baldareschi and Lipari (1973) calculated the effective Bohr radius. For an effective hole mass m_h^*, they obtained the effective Bohr radius

$$a_B^* = \frac{4\pi\varepsilon\hbar^2\gamma_1}{e^2 m_h^*}, \tag{1.43}$$

the effective Rydberg energy

$$E_{Ryd}^* = \frac{e^4 m_h^*}{2(4\pi\varepsilon\hbar)^2\gamma_1}, \tag{1.44}$$

and the acceptor ionization energy

$$E_a = E_{Ryd}^* f(\mu), \tag{1.45}$$

with

$$\mu = \frac{6\gamma_3 + 4\gamma_2}{5\gamma_1}. \tag{1.46}$$

The parameters γ_1, γ_2, and γ_3 are the so-called Luttinger parameters which describe the hole dispersion relation near the center of the Brillouin zone (Luttinger, 1956). The function $f(\mu)$ relates the acceptor energy with the

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

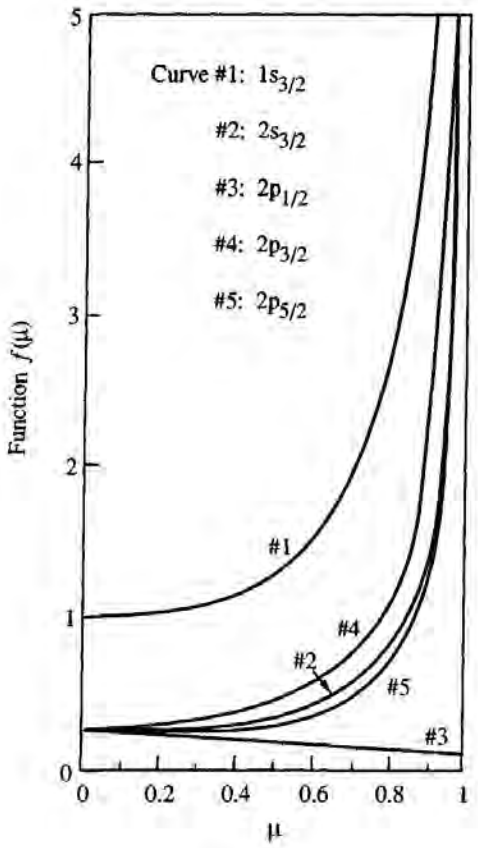


Fig. 1.8. Calculated function $f(\mu)$ versus μ in the limit of strong spin-orbit coupling for the spherical acceptor model (after Baldareschi and Lipari, 1973).

effective Rydberg energy (see Eq. (1.45)). The function $f(\mu)$ is shown in Fig. 1.8. For the Luttinger parameters of III–V semiconductors, μ assumes values of $\mu \cong 0.6-0.9$ and the function $f(\mu)$ assumes values of $f(\mu) \cong 1.5-4$ for the ground state energy of acceptors. Baldareschi and Lipari (1973) obtained the numerical values of $f(\mu)$ using a variational approach and using spherical trial functions (Kohn, 1957a) for the acceptor wave functions (see Eq. (1.42)). The Luttinger parameters, the variable μ, the effective Rydberg energy, and the ground state ($1s_{3/2}$) and excited state energies of hydrogenic acceptors in several III–V semiconductors are shown in Table 1.1 (Luttinger, 1956; Baldareschi and Lipari, 1973). The calculated and experimental acceptor energies agree reasonably well.

In subsequent work, Baldareschi and Lipari (1974) investigated the contribution of cubic symmetry terms of the Hamiltonian to the spherical model for acceptor states. The effects of the cubic symmetry were studied using perturbation theory which allowed the authors to reproduce all the details of

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

acceptor spectra. The quantitative changes in acceptor energy caused by the cubic term are small, i.e. less than 1 meV for the III–V semiconductors listed in Table 1.1.

Table 1.1. *Values for the Luttinger parameters (γ, γ_2, γ_3), the parameter μ, the effective Rydberg energy, and ground and excited state energies (in meV) for acceptors in III–V semiconductors.*

	γ_1	γ_2	γ_3	μ	E^*_{Ryd} (meV)	$E_{1s_{3/2}}$	$E_{2s_{3/2}}$	$E_{2p_{1/2}}$	$E_{2p_{3/2}}$	$E_{2p_{5/2}}$
AlSb	4.15	1.01	1.75	0.701	22.8	42.4	12.4	3.3	17.5	10.5
GaP	4.20	0.98	1.66	0.661	28.0	47.5	13.7	4.2	19.1	11.7
GaAs	7.65	2.41	3.28	0.767	11.3	25.6	7.6	1.6	11.1	6.5
GaSb	11.80	4.03	5.26	0.808	4.7	12.5	3.8	0.65	5.6	3.2
InP	6.28	2.08	2.76	0.792	14.1	35.2	10.5	2.0	15.5	8.9
InAs	19.67	8.37	9.29	0.907	3.2	16.6	5.1	0.4	7.9	4.4
InSb	35.08	15.64	16.91	0.935	1.2	8.6	2.7	0.2	4.2	2.3

The ***degeneracy*** of the acceptor ground states in III–V semiconductors is $g = 4$. Typical acceptor energies are much smaller than the spin-orbit splitting energy, i.e. $E_a << \Delta E_{so}$. The top of the valence band is fourfold degenerate due to heavy and light hole dispersion and due to twofold spin degeneracy. Since the acceptor wave functions are composed of valence band wave functions near the top of the band (Kohn, 1957a), the acceptor degeneracy is $g = 4$ as well.

1.1.5 Central cell corrections

The ionization energy of hydrogenic donors and acceptors as calculated from effective-mass theory does not depend on the chemical nature of the impurity atom. On the other hand, experimental values of the ionization energy do depend on the chemical nature especially for acceptors in III–V semiconductors. The difference in ionization energy between chemically different impurities is attributed to ***central cell potentials***. The central cell potential is assumed to be due to the chemical characteristics (e.g. electronegativity) of the impurity atom and thus the potential leads to a correction of the hydrogenic ionization energy. This correction is frequently referred to as ***chemical shift*** (Pantelides, 1975).

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

The total impurity potential is the sum of the Coulomb potential and the central cell potential V_{cc} and can be written as

$$V(r) = \frac{e}{4\pi\varepsilon r} + V_{cc}(r) \,. \tag{1.47}$$

The central cell potential is a short range potential and has a spatial extent of no more than the unit cell (central cell) of the host semiconductor. *Donor* wave functions are usually quite delocalized in III–V semiconductors. Therefore, the central cell corrections play a minor role for donors and their ionization energy is well described by the hydrogen model. In contrast, the *acceptor* Bohr radius is usually much smaller resulting in a significant central cell correction. This difference between donor and acceptor states is indeed observed experimentally, for example in GaAs. Several model potentials have been used for the central potential including a constant potential extending over the unit cell (Abarenkov and Heine, 1965), and δ-function-like potentials. Various models for central cell potentials were reviewed by Pantelides (1978), Stoneham (1975, 1986), and Altarelli and Bassani (1982). Finally, it is worthwhile to note that *isoelectronic* impurities lack the Coulomb term in Eq. (1.47). For isoelectronic impurities, the central cell potential is the only potential that can bind electrons (Thomas and Hopfield, 1966).

A simple model explaining the chemical shift of impurity ionization energies was proposed by Phillips (1970a, 1970b). The model is based on the local strain around the impurity atom. The strain is caused by the mismatch of the valence bonds of the impurity with valence bonds of the host lattice. Using this model, the chemical trend in donor ionization energies of Te, S, and Se in GaP were qualitatively explained (Phillips, 1970b). Phillips (1973) developed a second model in which the chemical shift in ionization energy is based on the difference in electronegativity, ΔX, between the impurity atom and the host lattice. The author showed that a large difference in electronegativity between impurity and the atom replaced by the impurity results in a large chemical shift. The chemical shift was assumed to be proportional to the heat of formation, i.e.

$$\Delta E \propto (\Delta X)^2 \tag{1.48}$$

where ΔE is the difference between the calculated effective mass impurity energy and the actual impurity energy (i.e. ΔE is the chemical shift). Using Eq. (1.48), differences in chemical shifts of impurities occupying cation and anion sites in GaP were explained.

Note that chemical shifts are expected to be larger for impurity states with s-type symmetry as compared to states with p-type symmetry. The central cell

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

potential is spatially restricted to the atomic vicinity of the impurity atom. In this region the amplitude of s-type wave functions is large while p-type wave functions have a node. Thus, perturbation theory predicts greater corrections for s-type symmetry states as compared to p-symmetry states.

1.1.6 Impurities associated with subsidiary minima

The conduction band structure of III–V semiconductors consists of three local minima, which occur at the L, Γ, and X-point of the Brillouin zone. The Γ minimum is located at the center of the Brillouin zone at $k=0$, while the L and X minima occur at finite wave vectors. Donors, which by their very nature are associated with the conduction band, can form donor levels with all local minima of the conduction band. Impurity states associated with subsidiary minima of the conduction band were first analyzed by Bassani et al. (1969). The theoretical study revealed that the impurity states are generally formed from Bloch functions of many Brillouin zones and their respective contributions depend on the particular band structure and on the strength and nature of the impurity potential. The existence of several conduction band minima with large effective masses can increase the number of bound states as compared to the single valley hydrogenic model. It was further shown that resonant states in the continuum of one local minima can be produced by impurity states associated with another minima. This situation is shown in Fig. 1.9 which depicts two donor levels associated with two conduction band minima and a resonant state in the continuum of the energetically lower minimum (Altarelli and Bassani, 1982).

Donor levels associated with subsidiary conduction band minima were experimentally observed in III–V semiconductors. Adler (1969) used hydrostatic pressure to study the effect of donors on the electron transfer in n-type GaAs. Onton et al. (1972) directly observed a subsidiary conduction band minimum and its associated donor levels in InP by optical absorption measurements.

1.2 Screening of impurity potentials

Variations of the electrostatic potential are reduced in magnitude by the spatial redistribution of free carriers. Variations in the band-edge potential can occur due to local doping concentration changes or local compositional changes of a semiconductor. An example of a potential fluctuation is shown in Fig. 1.10. At some location, an excess positive donor charge causes a dip in the band-edge potential. The potential dip attracts electrons and results in a locally higher concentration of electrons at the potential dip. The potential generated by the negative charge of the excess electrons reduces the original fluctuation and smooths the potential, i.e. free carriers *screen* the potential fluctuation. This is

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

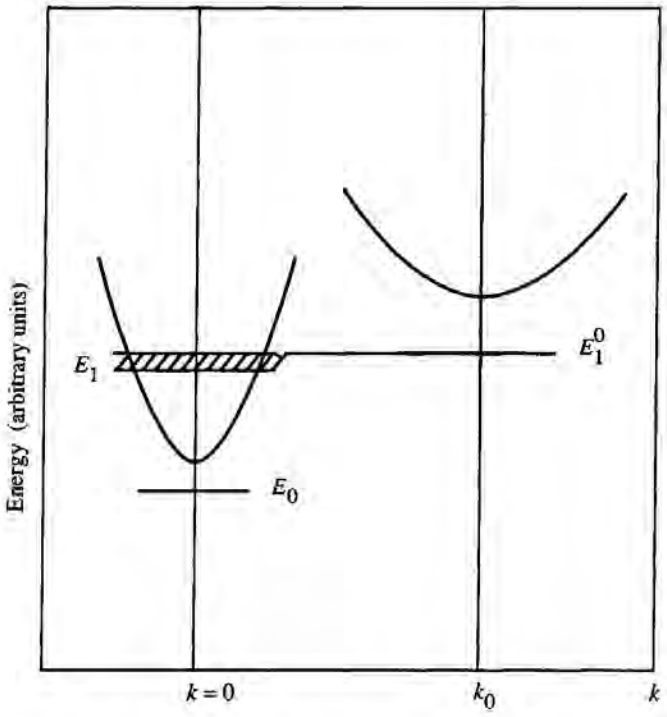


Fig. 1.9. Schematic representation of a bound state E_0 and a resonant state E_1 associated with the absolute minimum and a subsidiary minimum of the band structure, respectively. The resonant state E_1 (originating from the unperturbed state E_1^0) has a finite width due to its degeneracy with band states (after Altarelli and Bassani, 1982).

what screening of potential fluctuations by free carriers is all about. As an example of a potential fluctuation, consider the Coulomb potential of an impurity. As we recall, the impurity potential has a $1/r$ dependence. In the presence of free carriers the Coulomb potential is screened. The resulting potential is called the ***screened Coulomb potential*** which does *not* have the $1/r$ dependence of the unscreened Coulomb potential.

The potential variation is modeled in terms of the spatially non-uniform electrostatic potential $V(\mathbf{r})$, where $\mathbf{r} = (x,y,z)$ is the spatial coordinate. To find the energy levels in the perturbed potential $V(\mathbf{r})$, Schrödinger's and Poisson's equation must be solved simultaneously in order to find the free carrier distribution and the (screened) potential.

This procedure to calculate the screened potential is quite elaborate and it is usually possible to circumvent it, if the potential $V(\mathbf{r})$ is relatively smooth, i.e. it changes only little over the length of the electron wavelength. The electrons then

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

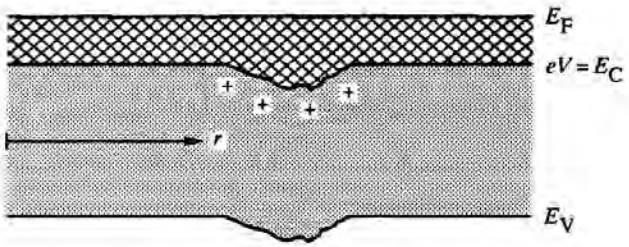


Fig. 1.10. Perturbation of the conduction band edge of a semiconductor by excessive positive charges. Electrons accumulate in the potential minimum and screen (i.e. compensate) the excess charge.

'see' only the potential at their own location. The total energy of the electron then follows the classical sum of potential and kinetic energy, i.e.

$$E = eV(\mathbf{r}) + \hbar^2 k^2 / 2m_e^* \,. \tag{1.49}$$

In other words, we assume that the spatial dimensions of the potential perturbation are so large that size quantization of the carrier system does not need to be considered.

The electron concentration in the potential perturbation shown in Fig. 1.10 is next calculated using classical quantization, i.e. using the semiclassical density of states. The conduction band edge sufficiently far away from the perturbation has a potential V_0 and an electron concentration of n_0. The electron concentration at the potential perturbation is given by (see Chap. 3)

$$n(\mathbf{r}) = \int_{eV(\mathbf{r})}^{\infty} \mathrm{f}(E)\, \rho_{\mathrm{DOS}}(E)\, \mathrm{d}E \tag{1.50}$$

where $\mathrm{f}(E)$ and $\rho_{\mathrm{DOS}}(E)$ are the Fermi–Dirac distribution and the density of states, respectively. In the following, we differentiate between a degenerate and a non-degenerate electron gas. The results obtained for the two cases are, as will be seen, quite different. For a *non-degenerate* electron gas, Boltzmann statistics is employed and the local electron concentration is obtained as (see Chap. 3)

$$n(\mathbf{r}) = \frac{1}{\sqrt{2}} \left[\frac{m_e^* kT}{\pi \hbar^2} \right]^{3/2} \exp\left[\frac{E_F - eV(\mathbf{r})}{kT} \right] \tag{1.51}$$

where E_F is the Fermi level. For a *degenerate* electron gas, Fermi–Dirac statistics must be employed and the electron concentration in the limit of *extreme degeneracy* ($E_F - eV(\mathbf{r}) >> kT$) is given by (see Chap. 3)

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

$$n(\mathbf{r}) = \frac{(2m_e^*)^{3/2}}{3\pi^2\hbar^3}(E_F - eV(\mathbf{r}))^{3/2}\,. \tag{1.52}$$

If the potential perturbation is a 'dip' in the conduction band then $E_F - eV(\mathbf{r}) > E_F - eV_0$ and electrons accumulate in the dip. If the perturbation is a 'bump' in the conduction band then $E_F - eV(\mathbf{r}) < E_F - eV_0$ and electrons deplete at the location of the perturbation. In the case of a 'dip', the potential generated by the excess electrons reduces the magnitude of the dip, i.e. smooths the potential. The resulting potential is obtained from Poisson's equation, which relates the charge density and the potential $V(\mathbf{r})$ according to

$$\nabla^2 V(\mathbf{r}) = -\frac{e}{\varepsilon}[\xi(\mathbf{r}) - n(\mathbf{r}) - n_0] \tag{1.53}$$

where $e\xi(\mathbf{r})$ is the concentration of fixed (positive) charge, $\nabla = (\partial/\partial x, \partial/\partial y, \partial/\partial z)$ is the Nabla operator, and $\nabla^2 = \Delta = (\partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2)$. The concentration $\xi(\mathbf{r})$ is homogeneous except at the location of interest. The average concentration of electrons is n_0, while $n(\mathbf{r})$ is the deficiency or excess of charge which depends on $\mathbf{r}$. Equation (1.53) is called a quasi-classical equation of the Thomas–Fermi type. Since $n(\mathbf{r})$ depends in a non-linear manner on $V(\mathbf{r})$, the differential equation is non-linear. It is the basic equation of non-linear screening theory. Unfortunately, the equation has no general solution. However, in the limit of small variations $n(\mathbf{r})$, it is possible to linearize the differential equation. Suppose the potential fluctuations of $V(\mathbf{r})$ are small as compared to $E_F - eV(\mathbf{r})$. Then Eqs. (1.51) and (1.52) can be linearized, i.e.

$$\boxed{n(\mathbf{r}) = n_0 + \left.\frac{dn}{dV(\mathbf{r})}\right|_{n=n_0} V(\mathbf{r})} \tag{1.54}$$

which greatly simplifies screening theory. The linearization of $n(\mathbf{r})$, i.e. our restriction to *small* potential fluctuations is the basis of ***linear screening theory***. The linearization of Eq. (1.54) allows one to write the Poisson equation (Eq. (1.53)) as

$$\nabla^2 V(\mathbf{r}) = \frac{V(\mathbf{r})}{r_s^2} - \frac{e\xi(\mathbf{r})}{\varepsilon} \tag{1.55a}$$

where

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

$$r_s = \left[-\left(\frac{dn}{dV} \right)_{V=0} \frac{e}{\varepsilon} \right]^{-1/2} \tag{1.55b}$$

is called the screening radius. The screening radius is usually referred to as the ***Debye screening radius*** and the ***Thomas–Fermi screening radius*** for non-degenerate and degenerate electron systems, respectively. Using Eq. (1.55b) and non-degenerate and degenerate statistics for an isotropic and parabolic conduction band (see Chap. 3), the screening radii are obtained as:

Debye screening radius

$$\boxed{r_D = (\varepsilon kT/e^2 n)^{1/2}} \tag{1.56}$$

Thomas–Fermi screening radius

$$\boxed{r_{TF} = \pi^{2/3} \sqrt{\frac{\varepsilon \hbar^2}{e^2 m^* (3n)^{1/3}}}} \tag{1.57}$$

Note the different functional dependences of the Debye and the Thomas–Fermi screening radius on temperature and free carrier concentration. While r_D depends on temperature, r_{TF} is temperature-independent. Furthermore, the Thomas–Fermi screening radius depends very weakly on the electron concentration, i.e. $n^{-1/6}$. Now the analytic solution of the screened potential can be obtained by integration of Eq. (1.55a) and the solution is given by

$$V(\mathbf{r}) = \int_r K(\mathbf{r} - \mathbf{r}')\, \xi(\mathbf{r}')\, d\mathbf{r}' \tag{1.58}$$

with $K(r) = (e/4\pi\varepsilon r)\, e^{-r/r_s}$.

We now calculate the screened potential of a single ionized impurity. The charge distribution of such a single impurity located at the origin of the coordinate system is $\xi(\mathbf{r}) = \delta(\mathbf{r})$. Insertion into Eq. (1.58) yields the ***screened Coulomb potential***

$$\boxed{V(r) = \frac{e}{4\pi\varepsilon r} e^{-r/r_s}} \tag{1.59}$$

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

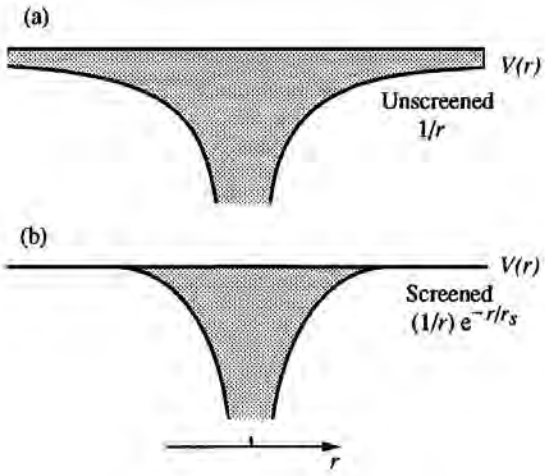


Fig. 1.11. (a) Unscreened and (b) screened Coulomb potentials. The unscreened and screened Coulomb potentials coincide for $r \to 0$ but strongly deviate for large radii.

where the screening radius, r_s, can be either the Debye or the Thomas–Fermi radius depending on the degeneracy of the screening electron gas. The screened Coulomb potential (Debye and Hückel, 1923) is also called the ***Yukawa potential*** in analogy to a potential in meson theory. The screened Coulomb potential is modified as compared to the unscreened Coulomb potential by the factor e^{-r/r_s}. Note that for $r \to 0$ the screened and the unscreened Coulomb potential become identical. For $r \to \infty$ the screened and unscreened Coulomb potential are strongly diverging due to the e^{-r/r_s} factor. The screened and the unscreened Coulomb potential are illustrated in Fig. 1.11 for an impurity located at $r = 0$. The unscreened potential is a long range potential while the screened one has a shorter spatial range.

There are several major results which are inferred from linear screening theory which are summarized in the following. *First*, the screening of potential perturbations leads to an exponential decay of the perturbing potential with distance, i.e. e^{-r/r_s}, where r_s is the screening radius. *Second*, in the case of several potential perturbations, each perturbation is screened separately. According to Eq. (1.58), a potential fluctuation can be expressed as a superposition of (unscreened) Coulomb potentials. The screened potential is then given by the superposition of the *screened Coulomb potentials*. *The applicability of the superposition principle is a consequence of the linearity* of the screening theory. *Third*, the screening radius r_s does not depend on the *magnitude* of the

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

337-TA-1366 RX-0255

potential perturbation to be screened. Instead the screening radius depends on the properties of the screening electron (hole) gas only. The independence of the screening radius on the magnitude of the potential fluctuation is again due to the linearization of screening theory.

The linear screening theory becomes invalid for very large potential fluctuations. The linearization of screening theory given in Eq. (1.54) is based on the assumption of *small* fluctuations. As an example of a large potential perturbation, consider a metal–semiconductor junction (Schottky barrier). Such junctions can induce perturbations of e.g. 1 eV (Schottky barrier height). For such large potential perturbations ***non-linear screening theory*** must be employed to obtain realistic results. Linear screening theory cannot be applied to Schottky barriers. The dependence of the depletion region thickness (i.e. screening length) on the magnitude of the Schottky barrier height of a metal–semiconductor interface is a result of non-linear screening theory.

The nature of screening changes drastically, if the screening carriers are confined to a two-dimensional plane. This situation is referred to as ***two-dimensional screening***. Ando et al. (1982) considered a Coulomb charge located at the cylindrical coordinates $r = 0$ and $z = z_0$ which is screened by electrons confined to the plane $z = 0$. The screened Coulomb potential in the electron plane ($z = 0$) is given by

$$V(r, z = 0) = \int_0^\infty q\, A(q) J_0(qr)\, \mathrm{d}q \tag{1.60}$$

where J_0 is the Bessel function of zero order. The constant $A(q)$ is

$$A(q) = \frac{e}{4\pi\varepsilon} \frac{\mathrm{e}^{qz_0}}{q + 1/r_s^{2D}} \tag{1.61}$$

where r_s^{2D} is the two-dimensional screening radius

$$\boxed{r_s^{2D} = \frac{2\varepsilon}{e^2} \frac{\pi\hbar^2}{m^*}} \tag{1.62}$$

Note that the two-dimensional screening radius depends neither on temperature, nor on the sheet charge density. This characteristic is a result of the two-dimensional density of states which is constant and does not depend on energy (see Chap. 3). For large values of r, where $r/r_s^{2D} >> 0$, the asymptotic form of the potential seen by the electrons is (Stern, 1967)

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

RX-0255.56

$$V(r,z=0) \equiv \frac{e(1+z_0/r_s^{2D})}{4\pi\epsilon r^3/(r_s^{2D})^2} \,. \tag{1.63}$$

This inverse-cube dependence of the potential on distance is much weaker than the exponential decay found in the three-dimensional case, and is one of the principal qualitative differences between two-dimensional and three-dimensional screening (Ando et al., 1982).

1.3 High doping effects

At high doping concentrations, many characteristics of semiconductors change. The changes are due to either the high concentration of impurities or due to the high free carrier concentration accompanying the high doping concentrations. Among the characteristics changed at high doping concentrations are the impurity ionization energy, the fundamental absorption edge, the density of states in the vicinity of the band edges, and the energy of the fundamental gap. Effects causing these changes are the Mott transition, the Burstein–Moss shift, band tailing effects, and bandgap renormalization.

1.3.1 The Mott transition

The Mott transition refers to a ***insulator-to-metal*** transition occurring in semiconductors at high doping concentrations. Consider an n-type semiconductor with low doping concentration. At low temperatures ($T \to 0$), shallow impurities are neutral, i.e. electrons occupy the ground state of the donor impurities. All continuum states in the conduction band are unoccupied. In this case, the semiconductor has the properties of an *insulator*. As the doping concentration increases, the Coulomb potentials of impurities overlap as schematically shown in Fig. 1.12. As a result of the overlapping impurity potentials, electrons can transfer more easily from one donor to another donor. Electrons transfer from one donor state to a state of an adjacent donor by either *tunneling* or by *thermal emission* over the barrier. The probability of both processes increases with decreasing donor separation. In other words, the activation energy for electron transport is reduced. In the extreme case, the activation energy approaches zero, i.e. the conductivity remains finite even for $T \to 0$. The semiconductor then has *metal*-like properties.

Screening of impurity potentials is a second contribution to the reduction of the impurity ionization energy. Impurity potentials are effectively screened at high free carrier concentrations. Screened potentials are less capable of binding electrons. Thus, the effective ionization energy decreases due to screening.

The insulator-to-metal transition occurs at the impurity concentration at which the distance between impurities becomes comparable to the Bohr radius. If

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

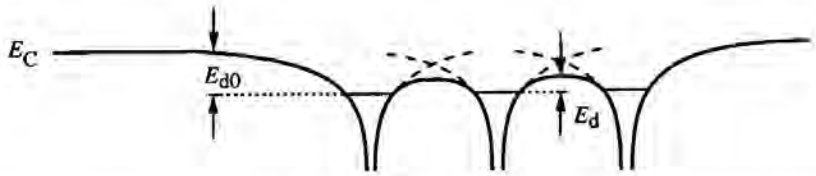


Fig 1.12. Conduction band edge with three donor potentials. For high donor concentrations the Coulomb potentials overlap and the ionization energy E_{d0} is reduced to E_d.

donors with concentration N were to occupy sites of a simple cubic lattice, their separation would be $N^{-1/3}$. The Mott transition would then occur at a concentration

$$2a_B^* = N_{crit}^{-1/3} \tag{1.64a}$$

where a_B^* is the effective Bohr radius and N_{crit} is called the critical concentration. However, Eq. (1.64a) does not give the correct result, because impurities are distributed *randomly* in semiconductors. Using a poissonian distribution of impurities, one can show that the Bohr orbital of an impurity is likely to overlap with the orbitals of one, two, or three neighboring impurities if

$$2a_B^* = \frac{3}{2\pi} N_{crit}^{-1/3} . \tag{1.64b}$$

Rearrangement of the equation yields the ***Mott criterion***

$$\boxed{a_B^* \, N_{crit}^{1/3} \cong 0.24} \tag{1.64c}$$

As an example, we calculate the critical concentration of donors in GaAs with an effective Bohr radius of $r_B^* = 103$ Å. Equation (1.64c) yields a critical density of $N_{crit} = 1.2 \times 10^{16}$ cm^{-3}.

The Mott criterion can be also obtained by a fundamentally different approach, i.e. by considering the binding of electrons to screened Coulomb potentials (see, for example, Mott, 1990). At low concentrations, electrons are bound to the (essentially) unscreened Coulomb potentials of the impurities. At higher free carrier concentrations, screening becomes relevant and Coulomb potentials must be replaced by screened Coulomb (Yukawa) potentials. The binding energy of screened Coulomb potentials is smaller than the binding energy of Coulomb potentials. Furthermore, the binding energy decreases with

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

337-TA-1366 RX-0255

increasing screening. The insulator-to-metal transition occurs, if the binding energy of electrons bound to screened Coulomb potentials becomes zero.

To solve this problem quantitatively, we use the Coulomb potential $V(r) = (e/4\pi\varepsilon r)\exp(-r/r_{TF})$, where r_{TF} is the screening radius. Furthermore, we use the variational wave function $\psi(r) = c\exp(-r/\alpha)$, where c is a normalization constant and α is the variational parameter (Flügge, 1971). The binding energy, E, and the spatial extent of the wave function, α, can be obtained by the variational method. From the condition $E = 0$ one obtains the variational parameter $\alpha = 2a_B^*$ and $r_{TF}/a_B^* = 1$. Using Eq. (1.57) for r_{TF} and Eq. (1.37) for a_B^* yields

$$\boxed{a_B^* N_{crit}^{1/3} = \frac{1}{4}\left(\frac{\pi}{3}\right)^{1/3} \cong 0.25} \tag{1.65}$$

which is similar to Eq. (1.64c).

There is yet another approach to calculate the Mott density. The calculation is based on the reduction of the bandgap energy due to many-body effects. Haug and Schmitt-Rink (1984) showed that the energy level of an exciton merges with the conduction band due to the lowering of the conduction band edge at high concentrations (bandgap renormalization, see Sect. 1.3.5). The authors further showed that the energy of the exciton is remarkably constant with carrier concentration due to the charge neutrality of the exciton. The critical concentration, estimated from the bandgap reduction due to many-body effects (see Sect. 1.3.5), agrees well with the Mott criterion of Eq. (1.65). Even though the result of Haug and Schmitt-Rink (1984) was obtained for excitons, it also applies to neutral donors which can be thought of as excitons with an infinitely heavy hole mass.

The insulator-to-metal transition does not occur abruptly at the critical concentration. Instead the transition is gradually evolving with increasing impurity concentration. Quantitatively, the gradual nature of the Mott transition can be expressed in terms of a continuously changing impurity activation energy. Experimental donor activation energies have been described by the equation (Debye and Conwell, 1954)

$$\boxed{E_d = E_{d0}[1 - (N_D/N_{crit})^{1/3}]} \tag{1.66}$$

where E_{d0} is the donor activation energy for $N_D \ll N_{crit}$. The reduction in donor ionization energy is thus proportional to the distance between the donor

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

RX-0255.59

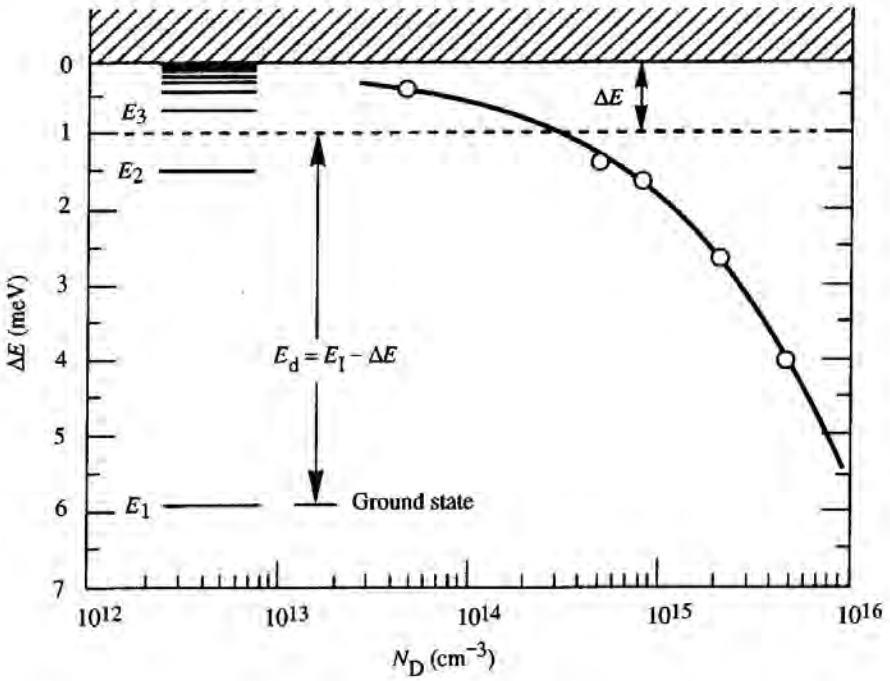


Fig. 1.13. Hydrogenic donor levels in GaAs, E_n, and donor ionization energy E_d as a function of doping concentration (Stillman et al., 1982).

atoms. The ionization energy of donors in GaAs as a function of the donor concentration is shown in Fig. 1.13 (Stillman et al., 1982). The effective ionization energy is measured from the impurity level to the quasi-continuum. The effective ionization energy approaches zero at the critical density which is $N_{crit} \equiv 10^{16}$ cm^{-3} for donors in GaAs.

1.3.2 The Burstein–Moss shift

The shift of the absorption edge to higher energies occurring at high doping concentrations is referred to as the ***Burstein–Moss effect*** or ***shift*** (Burstein, 1954; Moss, 1961). This shift is also called ***band filling*** or ***phase space filling***. The up-shift of the absorption edge is related to band filling which is schematically shown in Fig. 1.14 for n-type doping. The conduction band becomes significantly filled at high doping concentrations due to the finite density of states. Due to band filling, absorption transitions cannot occur from the top of the valence band to the bottom of the conduction band. As a result, the fundamental edge of absorption transitions shifts from $E_C - E_V = E_g$ for undoped semiconductors, to

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

$E_F - E_V > E_g$ in heavily doped n-type semiconductors. This shift was first observed in n-type InSb (Moss, 1961) and has since been used to make semiconductors transparent in the near-band-edge region (Verie, 1967; Dapkus et al., 1969, Deppe et al., 1990). It should be pointed out that the Burstein–Moss shift competes with the formation of impurity bands (Sect. 1.3.3) and band tails (Sect. 1.3.4). As will be seen, the Burstein–Moss shift dominates in semiconductors with light effective carrier mass.

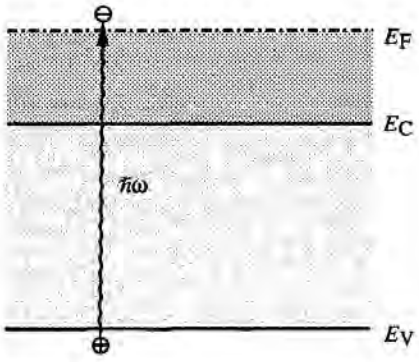


Fig. 1.14. Schematic illustration of conduction band filling due to heavy n-type doping. The absorption edge occurs at an energy E_F–E_C which can be significantly larger than the bandgap energy E_C–E_V.

Quantitatively, the Burstein–Moss shift can be calculated from the filling of the conduction or valence band in n-type and p-type semiconductors, respectively. Assuming a degenerately doped n-type semiconductor, the band filling for a single-valley, isotropic, and parabolic band is in the limit of extreme degeneracy given by (see Chap. 3)

$$E_F - E_C = \frac{\hbar^2}{2m_e^*}(3\pi^2 n)^{2/3} \tag{1.67}$$

where n is the free carrier density and m_e^* is the effective carrier mass. The absorption edge occurs then at the energy

$$E = E_g + (E_F - E_C)\,. \tag{1.68}$$

Note that the Burstein–Moss shift is inversely proportional to the effective mass. This explains the fact that the Burstein–Moss effect is more prominent in semiconductors with light carrier masses. For example, the shift clearly manifests itself in n-type GaAs ($m_e^* = 0.067\, m_0$) but not in p-type GaAs ($m_{hh}^* = 0.45\, m_0$) as illustrated in Figs. 1.15 and 1.22.

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

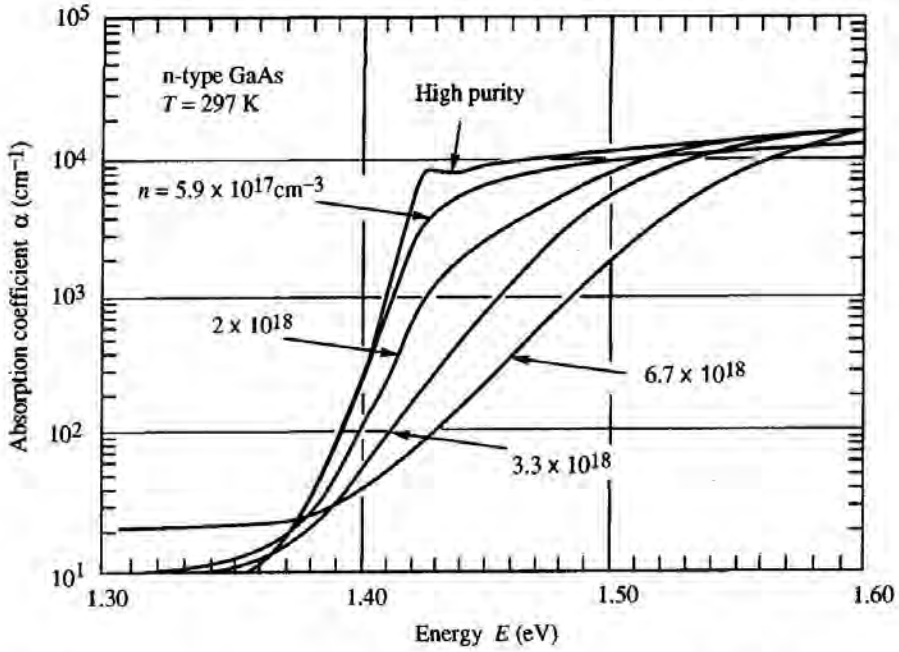


Fig. 1.15. Absorption coefficient in n-type GaAs at T = 297 K for different doping concentrations. The absorption edge shifts to higher energies with increasing doping concentration (after Casey et al., 1975).

For the sake of completeness, we consider the Burstein–Moss effect in *two-dimensional systems*. In such systems, free carriers are confined to a thin sheet. The confinement can be achieved in terms of quantum well structures or at the heterojunctions of two semiconductors. The band filling of a semiconductor with a single, parabolic subband is, in the high doping limit, given by (see Chap. 3)

$$E_F - E_0 = \frac{\pi\hbar^2}{m_e^*} n^{2D} \tag{1.69}$$

where E_0 is the bottom of the subband and n^{2D} is the two-dimensional carrier concentration per cm^2. The energy given by Eq. (1.69) represents the shift with respect to the absorption edge in an undoped two-dimensional semiconductor.

The Burstein–Moss shift is illustrated in Fig. 1.15 for n-type GaAs (Casey et al., 1975). The graph shows the absorption coefficient as a function of photon energy for n-type doping concentrations up to 6.7×10^{18} cm^{-3}. As the carrier concentration exceeds 6×10^{17} cm^{-3}, the Burstein–Moss shift due to the filling of the conduction band begins to have a significant effect. This shift of the absorption coefficient to higher energies is readily shown in Fig. 1.15 for samples

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

with $5.9 \times 10^{17} < n < 6.7 \times 10^{18}$ cm^{-3}. These curves tend to converge at 1.6 eV to within ±10% of the α of the high-purity sample. For the highly doped sample, absorption of $\alpha = 20$ cm^{-1} occurs for $E < 1.38$ eV, which is probably related to band tails (see Sect. 1.3.4). Note that the absorption spectra of n-type GaAs are in stark contrast to absorption in p-type GaAs (Casey et al., 1975). The difference is due to the much heavier mass of holes as compared to electrons, which makes the Burstein–Moss shift less important in p-type GaAs.

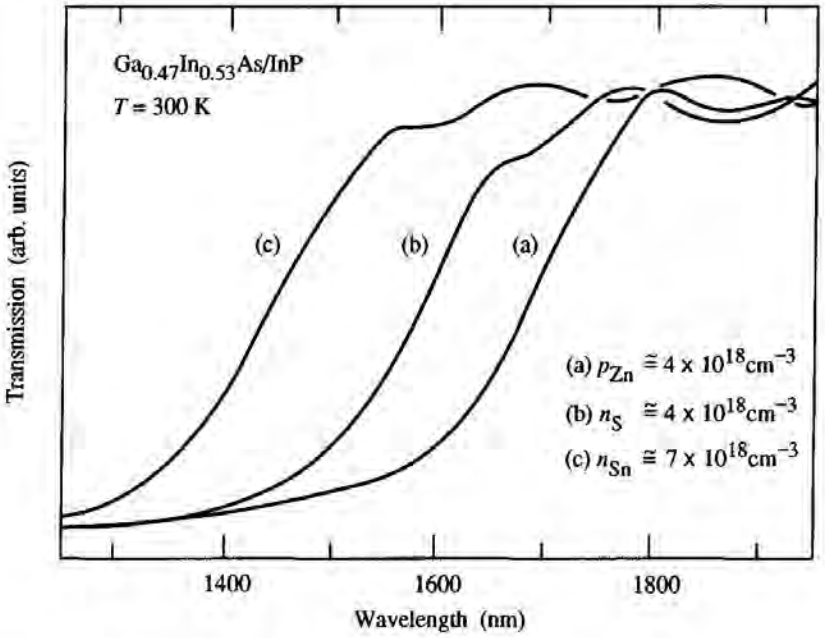


Fig. 1.16. Optical transmission versus wavelength for p- and n-type $Ga_{0.47}In_{0.53}As$ epitaxial layers grown on InP. The layers are doped with Zn, S, and Sn. Near-band-edge absorption decreases as the n-type doping level increases (after Deppe et al., 1990).

Transmission spectra of a $Ga_{0.47}In_{0.53}As$ lattice matched to InP are shown in Fig. 1.16 for various doping concentrations (Deppe et al., 1990). For a p-type Zn doping level of $N_{Zn} \cong 4 \times 10^{18}$ cm^{-3}, no appreciable shift of the absorption edge is observed as compared to undoped bulk $Ga_{0.47}In_{0.53}As$ ($\lambda^{300K}_{GaInAs} \cong 1.65$ μm). However, a significant shift is observed for n-type $Ga_{0.47}In_{0.53}As$ at a doping level of 4×10^{18} cm^{-3} and an even larger shift at a doping level of 7×10^{18} cm^{-3}. At the highest n-type doping level, the semiconductor is virtually transparent at a wavelength of 1.55 μm. The highly n-type doped $Ga_{0.47}In_{0.53}As$ was used for quarter-wave reflectors operating at

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

1.55 μm, which require optical transparency at that wavelength (Deppe et al., 1990).

1.3.3 Impurity bands

At impurity concentrations well below the critical Mott concentration, impurities can be considered as isolated, non-interacting entities. As the concentration increases but is still well below the Mott concentration, impurities begin to *interact*. Carrier transport at low temperatures occurs via thermally assisted tunneling between impurity states. This transport process is called ***hopping conduction***. At still higher impurity concentrations but below the critical Mott concentration, overlapping impurity states form an ***impurity band***. At low temperatures, carriers can propagate within the impurity band without entering the conduction band. This transport process is known as ***impurity band conduction***. This section summarizes the elementary characteristics of impurity-assisted conduction mechanisms for concentrations below the Mott concentration. Extensive reviews of the topic were given by Mott (1987, 1990) and Shklovskii and Efros (1984).

Consider a semiconductor with a donor density well below the critical Mott transition density. Upon cooling the semiconductor to low temperature, the conductivity is expected to decrease as free electrons freeze out onto localized donor states. For $kT << E_d$ the conductivity of an n-type semiconductor is expected to become vanishingly small. Experimentally, zero conductivity is not observed in semiconductors containing a net concentration of shallow impurities. Instead, the temperature dependence of the conductivity is less drastic than expected from free carrier freeze-out. The conductivity in this regime is not given by electrons excited to the conduction band but rather by electrons hopping from neutral donors to ionized donors. The conductivity is referred to as ***hopping conductivity*** and can be described by the general hopping conductivity formula

$$\sigma_{hop} = (\alpha/T^{\beta}) \exp\left[-\frac{E_{hop}}{kT}\right]^{\gamma} \tag{1.70}$$

where α and β are constants, E_{hop} is the thermal activation energy for the hopping process and γ determines the functional dependence of the exponential factor. For simplicity, the factor γ is frequently assumed to be unity. Austin and Mott (1969) showed for gaussian localization, in which transport occurs via tunneling to remote but energetically similar donors, that the value of $\gamma = ¼$. Efros and Shklovskii (1975) showed that $\gamma = ½$, if Coulomb interaction between adjacent donors determines the hopping transport between adjacent donors. In this case the thermal activation energy for the hopping process is given by

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

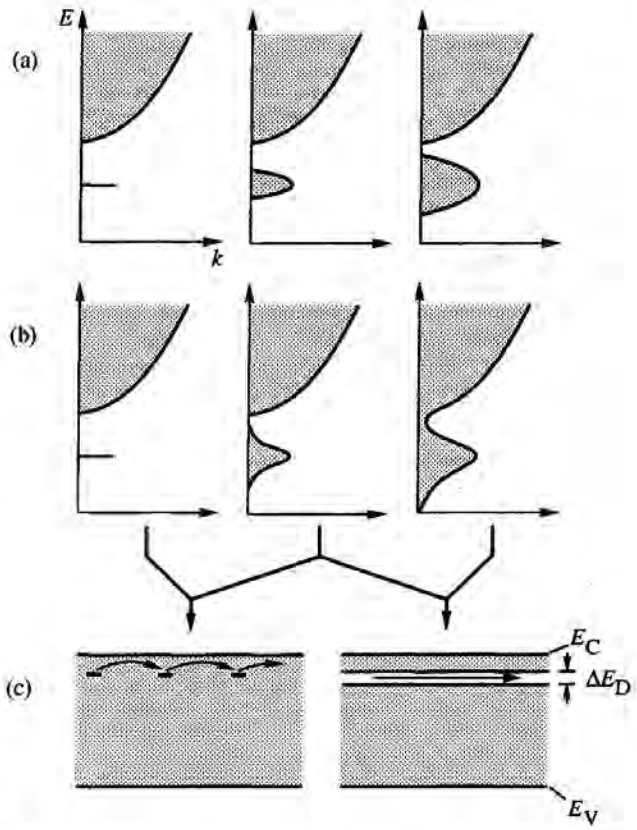


Fig. 1.17. Donor impurity level/band at low, medium and high doping concentrations for (a) an ordered impurity distribution and (b) a random impurity distribution. (c) Band diagram illustrating hopping conduction and impurity band conduction.

$$E_{hop} \equiv \alpha \frac{e^2}{4\pi\varepsilon} \left[\frac{4\pi}{3} N_D\right]^{1/3} \tag{1.71}$$

where the factor α has a numerical value of about 0.60. The qualitative physical explanation of the hopping conduction process is as follows. Consider a lightly compensated n-type semiconductor at low temperatures. Most of the donors are neutral, some donors ionized due to the slight compensation. When the donor impurities are closely spaced, their energy levels split. Electrons can tunnel from a donor state to an empty state of an adjacent donor. A so-called ***Coulomb gap*** develops between filled donor states and empty donor states. The Coulomb gap is caused by the long-range Coulomb interaction of localized electrons (Pollak and Knotek, 1974; Efros and Shklovskii, 1975), and occurs at the Fermi level. The

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

337-TA-1366 RX-0255

tunneling from filled donor states to adjacent empty donor states therefore requires a small thermal activation energy given by Eq. (1.71). The activation energy can be interpreted as the energy from the Fermi level to the energy of the maximum of the density of empty state distribution (Böer, 1990). Typically the activation energy is much smaller than the donor ionization energy.

The regime of hopping conduction is schematically illustrated in Fig. 1.17 which shows the dispersion relations and donor impurity states at different doping densities. At low doping density, states of adjacent impurities do not interact ($N << N_{crit}$) and impurity ground states are discrete and energetically well-defined levels. As the doping concentration increases, states of adjacent impurities interact, split, and finally form an impurity band ($N \leq N_{crit}$). At even higher doping concentrations, the impurity band widens and merges with the continuum band ($N \geq N_{crit}$).

It is instructive to consider the effect of *ordered* and *random* impurity distribution on the formation of an impurity band, which is shown in Fig. 1.17(a) and (b), respectively. In the case of an *ordered* impurity distribution, the potentials of the impurities are strictly periodic, similar to the periodic potential assumed in the Kronig–Penney model. As a result, the impurity band has a well-defined width and well-defined edges, as shown in Fig. 1.17(a). However, the case of *random* impurity distribution in semiconductors is a more realistic assumption (Shockley, 1961). For a random impurity distribution, the impurity band does not have well-defined band edges but the impurity states will tail into the forbidden gap, as shown in Fig. 17(b). Tail states of impurities occurring at high doping density are discussed in Sect. 1.3.4. Figure 1.17(c) schematically shows hopping conduction and impurity band conduction.

Impurity bands are formed at sufficiently high doping concentrations. The impurity band is formed by the quasi-periodic Coulomb potentials of the impurities. Since the impurity potentials are not strictly periodic, a periodic band calculation (e.g. Kronig–Penney) cannot be used. The hydrogenic potentials of donors can be used to form a Hubbard band (Hubbard, 1963). In the Hubbard model, the center of the impurity band coincides with the energy level of the unperturbed impurity state. If the donors forming the impurity band are partially compensated by acceptors, the band is only partially filled and impurity band conduction can occur within the impurity band, i.e. electrons need not occupy conduction band states for carrier transport (Adler, 1982). The impurity band width, ΔE_D, is approximately given by the overlap integral between donors separated by the average distance $N_D^{-1/3}$. The band width is approximately equal to the interaction energy

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

RX-0255.66

$$\Delta E_D \cong \frac{e^2}{4\pi\varepsilon N_D^{1/3}} . \tag{1.72}$$

As an example, we consider GaAs with a donor concentration of 1×10^{16} cm^{-3} and $\varepsilon_r = 13.1$ and obtain $\Delta E_D \cong 2.4$ meV.

The mobility associated with impurity band conduction are very low, typically < 1 cm^2/V s. The low mobility is due to the heavy dispersion mass. In order to qualitatively relate the low impurity band mobility with the narrow width of the impurity band, let us recall some of the basic results of one-dimensional band theory. The dispersion mass of a parabolic band is given by

$$m^* = \frac{\hbar^2}{d^2E/dk^2} . \tag{1.73}$$

In the simple Kronig–Penney model, the dispersion (i.e. effective) mass is given by

$$m^* = 2\hbar^2/z_p^2 \Delta E_D \tag{1.74}$$

where z_p is the period of the periodic potential and ΔE_D is the width of the band. Although the result of the one-dimensional Kronig–Penney model cannot be rigorously applied to an impurity band, the model provides a qualitative understanding of the functional dependences. It is therefore reasonable to conclude that a heavy effective mass is associated with transport of carriers in a narrow impurity band. Since the mobility and the effective mass are related according to the Drude model by

$$\mu = \frac{e\tau}{m^*} , \tag{1.75}$$

a low mobility results for carrier transport in a partially filled impurity band.

1.3.4 Band tails

The random distribution of charged impurities results in potential fluctuations of the band edges. In an *undoped* semiconductor such potential fluctuations are absent and the band edges are well defined. In a highly doped semiconductor the potential fluctuations cause the band edges to vary spatially. This situation is schematically shown in Fig. 1.18. States with energy below the unperturbed conduction band edge or above the unperturbed valence band edge are called ***tail states***. The tail states significantly change the density of states in the vicinity of the band edge.

The magnitude of band-edge energy fluctuations caused by the random distribution of charged donors and acceptors was first calculated by Kane (1963). For an ionized donor concentration of N_D and an ionized acceptor concentration

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

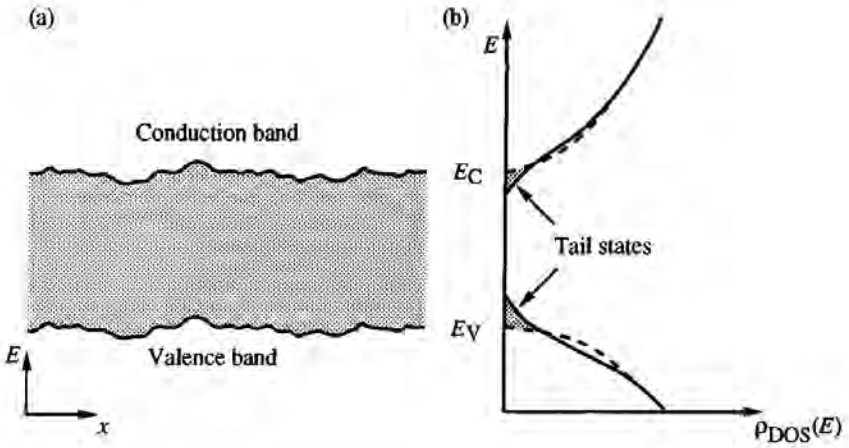


Fig. 1.18. (a) Spatially fluctuating band edges caused by random distribution of impurities. (b) Resulting densities of states in the conduction and valence band with tail states extending into the forbidden gap. The dashed lines show the parabolic densities of states in undoped semiconductors.

of N_A, the root-mean-square (energy) fluctuation of the band edges is given by (Kane, 1963; Morgan, 1965)

$$\sigma_E = \frac{e^2}{4\pi\varepsilon} [2\pi(N_D + N_A) r_s]^{1/2} \tag{1.76}$$

where r_s is the screening radius given by (see Eq. 1.55b)

$$r_s = \left[- \left(\frac{dn}{dV} \right)_{V=0} \frac{e}{\varepsilon} \right]^{-1/2} . \tag{1.77}$$

Note that the square-root dependence of the energy fluctuation, σ_E, on the doping concentration, $N_D + N_A$, is a result of the Poisson distribution assumed for impurities. The Poisson distribution can be replaced by a gaussian distribution, if the energy fluctuations are small. Small fluctuations occur if the random variation of the number of donor atoms within a spherical volume defined by the screening radius is much smaller than the average number of donors within the sphere (Kane, 1963), i.e.

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

$$\sqrt{(N_D + N_A)\, r_s^3} \ll (N_D + N_A)\, r_s^3 \,. \tag{1.78}$$

The fluctuation of the potential, e.g. the conduction band edge potential, can then be expressed in terms of a gaussian distribution. The probability for the conduction band edge energy to occur at an energy E_C, is given by

$$p(E_C) = \frac{1}{\sqrt{2\pi}\,\sigma_E}\, e^{-\frac{1}{2}\left[\frac{E_C}{\sigma_E}\right]^2} \tag{1.79}$$

where we defined the mean energy of the conduction band edge as $E_C = 0$ and σ_E is given by Eq. (1.76). The unperturbed density of states in the conduction band is given by (see Chap. 3)

$$\rho_{DOS}(E) = \frac{1}{2\pi^2}\left[\frac{2m^*}{\hbar^2}\right]^{3/2}\sqrt{E - E_C}\,. \tag{1.80}$$

The density of states in a perturbed potential is obtained by a summation over all locations in space, i.e. by an integral over the probability distribution of the band edge

$$\rho_{DOS,Kane} = \frac{1}{2\pi^2}\left[\frac{2m^*}{\hbar^2}\right]^{3/2}\int_{-\infty}^{E}\sqrt{E - E_C}\,\frac{1}{\sqrt{2\pi}\,\sigma_E}\, e^{-\frac{1}{2}\left[\frac{E_C}{\sigma_E}\right]^2} dE_C \tag{1.81}$$

where $\rho_{DOS,Kane}$ is the density of states according to Kane (1963), or briefly the Kane function. The density of states given by Eq. (1.81) is valid for the conduction band. The corresponding density of states of the valence band can be obtained by replacing E_C by $-E_V$, E by $-E$, and the electron effective mass by the heavy-hole effective mass.

The parabolic density of states, the gaussian probability distribution of the band edge, and the Kane function are shown in Fig. 1.19. The density of states shown in Fig. 1.19(c) tails into the gap. Tail states have energies lower than the average band-edge energy. It is important to note that the density of states of the Kane function does not change the *bandgap* energy, i.e. the spatially *averaged* position of the conduction band edge does not change. The tailing of the band into the forbidden gap changes the density of state drastically in the vicinity of the band edge. However, the density of states is practically unchanged for energies $E - E_C \gg \sigma_E$, as can be easily inferred from Eq. (1.81). The Kane function is of great practical use for the simple, quantitative description of band edges. For example, Casey and Stern (1976) used the Kane function to calculate absorption and spontaneous emission in doped GaAs.

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

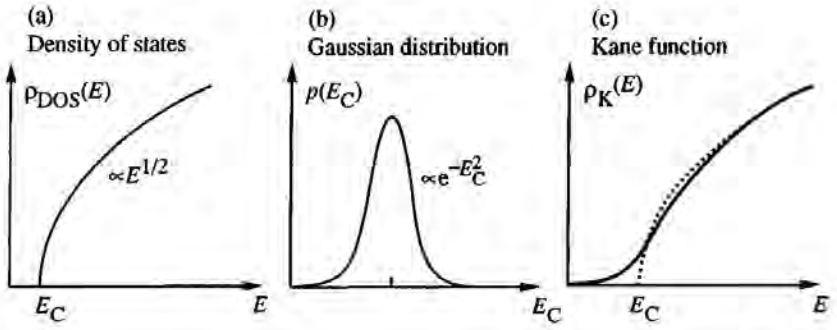


Fig. 1.19. (a) Density of states of a parabolic, spherical valley. (b) Gaussian probability distribution. (c) The Kane function describes the density of states of a parabolic band including tail states at $E < E_C$ which arise from potential fluctuations caused by randomly distributed impurities.

The density of states according to Kane (Eq. (1.81)) overestimates the extent of band tailing since tunneling of carriers through the potential barriers is not taken into account. In addition, the quantization of carriers in the potential minima is not taken into account in the Kane model. However, in the limit of carriers with a large effective mass, the Kane function applies, since quantum effects can be neglected for such heavy carriers.

Halperin and Lax (1966, 1967) and Lax and Halperin (1966) calculated the density of states in band tails taking into account (i) the quantization of carriers in the potential minima as well as (ii) tunneling of carriers through potential barriers. The importance of carrier quantization can be estimated by comparing the magnitude of the potential fluctuations with the quantization energy of carriers. The former is given by Eq. (1.76) for poissonian (random) impurity distributions. The latter can be estimated from

$$E_Q \equiv \frac{\hbar^2}{2m^* r_s^2} \tag{1.82}$$

where r_s is the screening radius. The density of states calculated by Halperin and Lax (1966) is shown in Fig. 1.20 along with the density of states of the unperturbed bands and the density of states of the Kane model for GaAs with doping concentrations of $N_A = 1.1 \times 10^{19}$ cm^{-3} and $N_D = 9 \times 10^{18}$ cm^{-3} (Casey and Stern, 1976). The density of states calculated according to the Kane model shown in Fig. 1.20 was used to interpolate between the unperturbed band and the Halperin–Lax result. The parameter σ_E (see Eq. (1.81)) was adjusted in order to obtain a good fit between the two functions.

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

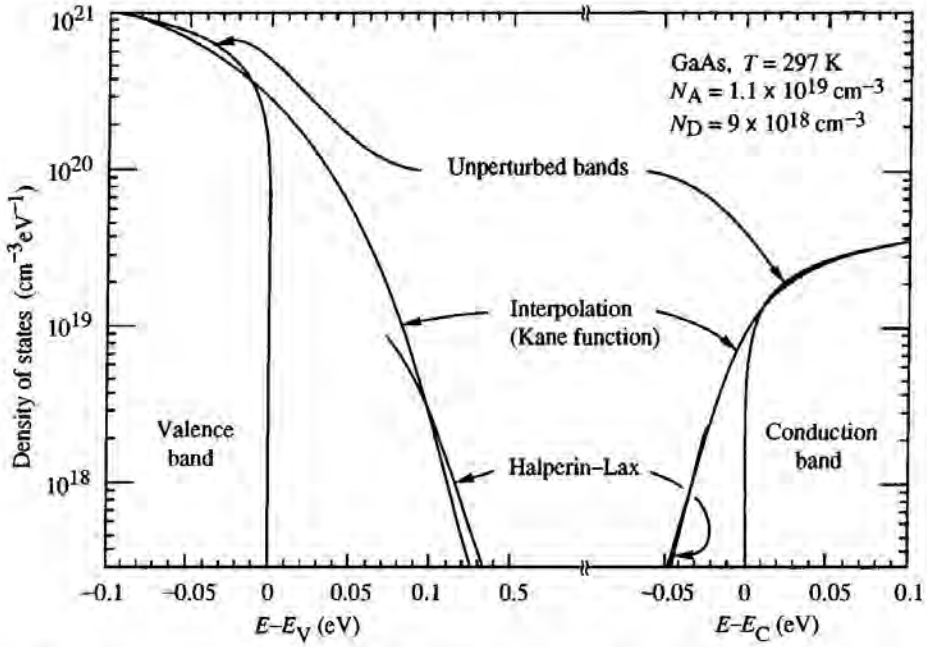


Fig. 1.20. Densities of states in the conduction and valence band versus energy for GaAs with a net acceptor concentration of 2×10^{18} cm^{-3}. The curves show the densities of states in the unperturbed bands and the densities of states in the band tails calculated by Halperin and Lax. The curves which join them are interpolated Kane functions (after Casey and Stern, 1976).

1.3.5 Bandgap narrowing

At high doping concentrations the bandgap energy of semiconductors decreases. The magnitude of the reduction increases with doping concentration and is usually referred to as ***bandgap narrowing***, ***bandgap shrinkage***, or ***bandgap renormalization***. There are many reasons for the reduction of the band gap which have been reviewed by Abram et al. (1978). The most important reasons for bandgap narrowing are many-body effects of free carriers which lower the electron energies as compared to a non-interacting carrier system. Many-body effects describe the interaction of free carriers. The interaction becomes important at small carrier-to-carrier distances, i.e. at high free carrier concentrations. Electrons can interact with each other either by their long-range Coulomb potential or via their spin. We first consider coulombic electron–electron interactions. Consider an electron which is added to a highly doped, neutral semiconductor. When the electron is added to the semiconductor, other electrons in the vicinity of the added electron spatially redistribute in order

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

to reduce the long-range coulombic interaction energy. The energy of the electron added to the semiconductor is *reduced* by the redistribution of neighboring electrons. Many-body effects are thus related to screening, i.e. the spatial redistribution of carriers in the presence of potential perturbations. Carriers can also interact via their spin (Mahan, 1990). Due to the Fermion nature of electrons, each volume element in phase space can be occupied by at most two electrons with opposite spin (Pauli principle). Electrons with like spin have a repulsive interaction while electrons with opposite spin have an attractive interaction. If electrons were distributed uniformly throughout the crystal, the attractive and repulsive energies would exactly cancel each other. However, due to the interaction, electrons with like spin tend to stay away from each other and electrons with opposite spin tend to stay closer. As a result, the interaction energy reduces the total energy of the electron system.

Many-body interactions can occur between free carriers and between free carriers and ionized impurities. Such interactions are called carrier–carrier and carrier–impurity interactions. Thus, bandgap narrowing occurs in highly doped semiconductors as well as in undoped, but highly excited semiconductors with high free carrier concentrations. We summarize the many-body interactions in n- and p-type semiconductors as follows:

i. *Electron–electron interactions.* The repulsive and attractive interactions of electrons with like and opposite spin and the long-range coulombic interactions lead to a net attractive term. As a consequence, the conduction band edge is lowered (i.e. lowering of gap energy).

ii. *Electron–donor interactions.* The interaction of electrons and ionized donors is attractive and leads to another lowering of the conduction band edge.

iii. *Hole–hole interactions.* The spin interaction energy and the coulombic interaction result in a net attractive energy and an increase in the valence band edge (i.e. lowering of gap energy).

iv. *Hole–acceptor interactions.* The interaction between holes and donors is attractive and leads to an increase of the valence band edge (i.e. decrease in gap energy).

v. *Electron-hole interactions.* Highly excited semiconductors have a large concentration of electrons and holes. Interaction effects lead to a reduction of the energy gap in highly excited semiconductors.

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

Next we calculate the magnitude of bandgap renormalization due to electron–electron interaction effects and follow the calculation of Haug and Schmitt-Rink (1985). At room temperature, the magnitude of bandgap renormalization can be approximated by the classical self energy of an electron interacting with an electron gas, i.e. interacting with its own polarization field. Using the classical analogy, we define the self energy of an electron interacting with an electron gas as

$$\Delta E_g \equiv e \lim_{r \to 0} [V_s(r) - V(r)] \tag{1.83}$$

where V(r) is the Coulomb potential and V_s(r) is the screened Coulomb potential of an electron in an electron gas. Equation (1.83) thus represents the change in electrostatic energy of an electron before and after the electron gas has spatially redistributed itself to reduce the Coulomb interaction energy. In other words, ΔE_g describes the energy that the electrons gain by avoiding each other. Assuming that the screened potential is of the Yukawa form, we obtain

$$\Delta E_g \equiv -\frac{e^2}{4\pi\varepsilon r_s} \tag{1.84}$$

The screening radius, r_s, is given by the Debye and the Thomas–Fermi radii in non-degenerate and degenerate semiconductors, respectively. Insertion of the screening radii into Eq. (1.84) yields the bandgap renormalization energies

$$\Delta E_g = -\frac{e^3\sqrt{n}}{4\pi\varepsilon^{3/2}\sqrt{kT}} \quad \text{(Debye)} \tag{1.85a}$$

$$\Delta E_g = -\frac{e^3\sqrt{m^*(3n)^{1/3}}}{4\,\pi^{5/3}\,\varepsilon^{3/2}\,\hbar} \quad \text{(Thomas–Fermi)} \tag{1.86b}$$

The momentum-dependence of the renormalization energy is relatively small (Wolf, 1962; Haug and Schmitt-Rink, 1985). If the momentum-dependence is neglected, bandgap renormalization produces a *rigid downward movement* of the conduction band dispersion (or, in p-type semiconductors, a rigid upward movement of the valence band). Finally, bandgap renormalization also occurs in undoped but highly excited semiconductors. Brinkman and Rice (1973) examined the effect of electron–hole interactions in highly excited semiconductors and showed that bandgap renormalization occurs due to electron–hole interactions. Phenomenological expressions for the carrier–carrier interaction and the carrier–impurity interaction were given by Mahan (1980) and Landsberg et al. (1985). Assuming that the interaction energies are proportional to the carrier–carrier distance, the change in energy gap follows a 1/3 power of

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

the doping concentration, that is

$$\Delta E_g \propto (N_D)^{1/3} \tag{1.86}$$

where full activation of the donors was assumed, i.e. $n = N_D$. Similar considerations are valid for p-type semiconductors. In this case carrier-carrier and carrier–acceptor interactions must be considered. Note that the phenomenological dependence of ΔE_g on n given in Eq. (1.86), i.e. $\Delta E_g \propto n^{1/3}$, is weaker than the Debye result, $\Delta E_g \propto n^{1/2}$ (see Eq. (1.85a)), but stronger than the Thomas–Fermi result, $\Delta E_g \propto n^{1/6}$ (see Eq. (1.85b)).

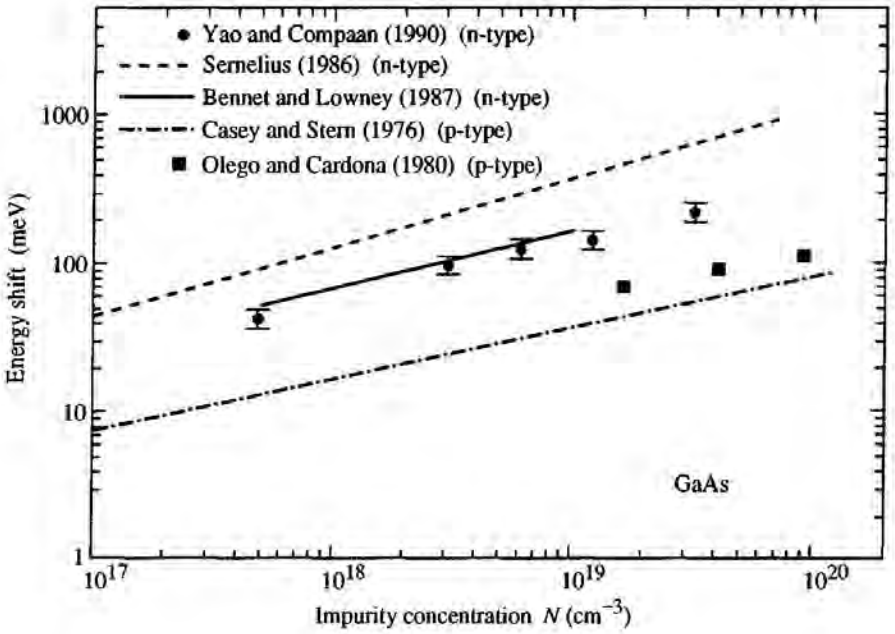


Fig. 1.21. Bandgap narrowing for n-type and p-type GaAs as a function of impurity concentration (after Yao and Compaan, 1990).

Experimental and theoretical results of bandgap narrowing as a function of doping concentration are shown in Fig. 1.21 for n-type and p-type doping of GaAs (Yao and Compaan, 1990). The graph includes experimental data for n-type (Yao and Compaan, 1990), and p-type GaAs (Olego and Cardona, 1980), as well as theoretical data for n-type (Sernelius, 1986; Bennet and Lowney, 1987) and p-type GaAs (Casey and Stern, 1976). The data of Fig. 1.21 shows that bandgap narrowing can assume quite large values, such as 200 meV, at high n-type doping concentrations. Furthermore, the change in bandgap energy follows

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

the $N^{1/3}$ dependence predicted by Eq. (1.86). The magnitudes of bandgap narrowing can be expressed by the following simple formulas:

$$\text{n-type GaAs:} \quad \Delta E_g(\text{meV}) \cong -6.6 \times 10^{-5} \sqrt[3]{N_D(\text{cm}^{-3})} \tag{1.87}$$

$$\text{p-type GaAs:} \quad \Delta E_g(\text{meV}) \cong -2.4 \times 10^{-5} \sqrt[3]{N_A(\text{cm}^{-3})} \,. \tag{1.88}$$

These equations represent a fit to the experimental data shown in Fig. 1.21. As expected, the bandgap narrowing for n-type GaAs is larger as compared to p-type GaAs (Yao and Compaan, 1990). Bandgap narrowing was studied to a smaller extent in other III–V semiconductors. An approximate formula for bandgap narrowing in n-type InP was given by Böer (1990)

$$\text{n-type InP:} \quad \Delta E_g(\text{meV}) \cong -2.25 \times 10^{-5} \sqrt[3]{N_D(\text{cm}^{-3})} \tag{1.89}$$

where it is assumed that all donors are active ($n = N_D$). Further phenomenological expressions and parameters for other III–V semiconductors were given by Jain et al. (1990).

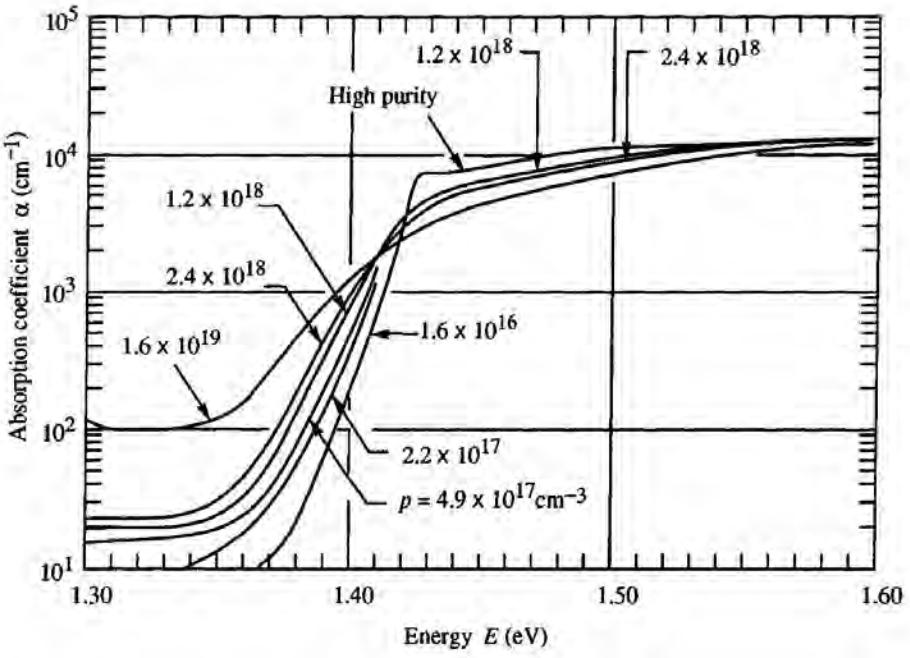


Fig. 1.22. Absorption coefficient for p-type GaAs at 297 K for different doping concentrations (after Casey and Stern, 1976).

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

The Burstein–Moss shift and bandgap narrowing are two phenomena which cause the Fermi level of highly doped semiconductors to change in opposite directions. While bandgap narrowing causes the Fermi level of an n-type semiconductor to decrease, the Burstein–Moss shift is due to an increase of the Fermi level. Absorption measurements allow one to measure $E_F - E_V$ and $E_C - E_F$ for n-type and p-type semiconductors, respectively. In n-type GaAs, the Burstein–Moss shift prevails, resulting in a blue-shift of the absorption edge (see Sect. 1.3.2). In p-type GaAs, bandgap narrowing prevails and results in a red-shift of the absorption edge. The absorption coefficient of p-type GaAs is shown in Fig. 1.22 for different p-type doping levels (Casey and Stern, 1976). High-purity GaAs exhibits a rapidly decreasing absorption coefficient at the fundamental gap. However, as the doping concentration increases, the absorption below the gap energy increases as well. Furthermore, the absorption coefficient decreases less rapidly as compared to the high-purity GaAs. Both characteristics indicate that bandgap narrowing and the formation of tail states dominate the near-band-edge optical absorption rather than the Burstein–Moss shift. The strongly different absorption characteristics of p-type (see Fig. 1.22) and n-type GaAs (see Fig. 1.15) are due to the heavier hole mass as compared to the effective electron mass. The Burstein–Moss shift is inversely proportional to the carrier effective mass, which results in less band filling in p-type GaAs.

https://doi.org/10.1017/CBO9780511599828.004 Published online by Cambridge University Press

2

Phenomenology of deep levels

Deep levels are states within the forbidden gap of a semiconductor that are far removed from either conduction or valence band. Many of the deep levels are closer to the center of the gap than to either of the band edges. The name ***midgap center*** is frequently used for such deep levels. Due to their ability to capture free electrons and holes, deep levels are also called ***traps*** or ***deep traps***. Deep levels caused by substitutional, non-hydrogenic impurities are referred to as ***deep impurities***. Chromium in GaAs and Fe in InP are examples of such deep impurities. Finally, deep levels can be due to point defect centers. Examples of such ***deep centers*** are anti-site defects and interstitials.

Deep levels can be caused not only by point defects but also by spatially extended defects. Examples for extended defects are dislocations such as threading dislocations, misfit dislocations, or screw dislocations. Another example of extended 'defects' are semiconductor surfaces. Electronic states at semiconductor surfaces are called ***surface states*** or ***Bardeen states*** after Bardeen (1947) who demonstrated the influence of such states on the surface potential. In many semiconductors, Bardeen states are located in the vicinity of the center of the gap (e.g. GaAs and InP) and have properties similar to deep levels occurring in the bulk.

Deep centers can be classified according to their charge state. Centers with a neutral and positively charged state are called *donor-like states*. Centers with a neutral and negatively charged state are called *acceptor-like states*. However, deep levels can have many levels within the forbidden gap as well as within the valence or conduction band. Examples for such multileveled deep centers are Ga and As vacancies in GaAs (Baraff and Schlüter, 1985).

https://doi.org/10.1017/CBO9780511599828.005 Published online by Cambridge University Press

It is not the purpose of this chapter to discuss exhaustively experimental and theoretical characteristics of deep levels since deep levels are only indirectly related to the topic of doping. A near-complete discussion of the properties of deep levels can be found in the literature (Queisser, 1971; Jaros, 1982; Pantelides, 1986). Nevertheless, deep impurities can cause the properties of doped semiconductors to deteriorate strongly and are therefore of importance for the understanding of such properties. First, deep levels can compensate shallow dopants and, as a consequence, reduce the free carrier concentration. A concomitant effect is a reduction in carrier mobility, since the concentration of charged levels increases while the free carrier concentration decreases. Second, deep centers are recombination centers and decrease the radiative efficiency as well as the minority carrier lifetime. Recombination via deep traps (Shockley–Read recombination) can drastically degrade the performance of minority carrier devices such as light-emitting diodes, lasers, and bipolar transistors. Due to this characteristic, deep levels are dubbed ***lifetime killers*** and ***luminescence killers.*** To summarize, the negative consequences of traps are (i) reduced luminescence efficiency, (ii) shorter minority carrier lifetime, (iii) a shorter diffusion length of minority carriers, (iv) compensation of intentional shallow impurities, and (v) reduced drift mobility.

This chapter contains a phenomenological description of deep levels and the physical consequences of deep levels. The reader is referred to Chap. 9 for experimental characteristics of specific deep centers in III–V semiconductors.

2.1 General characteristics

The energy level or energy levels of deep traps are far removed from the band edges. Consider a donor-like deep level with the ground-state energy E_T. Such a donor is considered to be *deep*, if the condition

$$E_t >> kT \quad (T = 300\ \text{K}) \tag{2.1}$$

is fulfilled, where $E_t = E_C - E_T$ is the activation energy of the deep level (see Fig. 2.1a). An analogous definition can be made for acceptor-like deep levels. As a consequence of Eq. (2.1), donor- and acceptor-like deep centers are not ionized at room temperature. That is, donor-like and acceptor-like deep centers are neutral in n-type and p-type semiconductors, respectively. Note, however, that donor-like deep centers are ionized in p-type semiconductors and the free hole concentration is reduced due to the presence of donor-like traps.

The confining potential of a deep level is a short-range potential rather than the long-range Coulomb potential of shallow impurities. The short-range potential of deep impurities typically extends over one or a few atomic radii

https://doi.org/10.1017/CBO9780511599828.005 Published online by Cambridge University Press

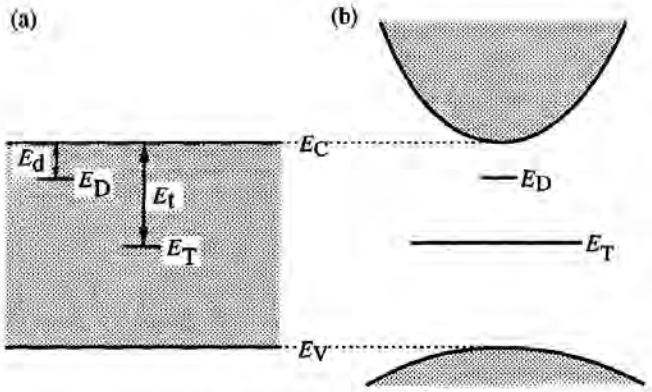


Fig. 2.1. Illustration of a deep level and a shallow level in (a) real space and in (b) momentum space.

(Madelung, 1978). The spatial extent of the impurity wave function decreases as the ionization energy of the center increases. According to the WKB approximation, the impurity wave function decay length, L, and the trap ionization energy are related by $L \propto \exp -\sqrt{E_t^{-1}}$. Electrons bound to deep levels are typically confined to the volume of one or a few unit cells. This characteristic of deep centers is in marked contrast to shallow impurities. Hydrogenic shallow impurities (see Chap. 1) confine carriers to a sphere defined by the Bohr radius which can be on the order of 100 Å, i.e. a sphere comprising thousands of unit cells. Due to this striking difference, the properties of deep levels are mainly determined by the chemical nature of the center, while the chemical nature is of minor importance for shallow impurities with a large effective Bohr radius.

The different degrees of spatial confinement of deep and shallow impurities can be represented as different degrees of uncertainty in momentum space. Figure 2.1(b) shows the dispersion relations as well as impurity levels which are represented as horizontal lines of various lengths. For shallow impurities the horizontal line has a width of approximately $1/a_B^*$ where a_B^* is the effective Bohr radius

$$a_B^* = \frac{\varepsilon}{\varepsilon_0} \frac{m_0}{m_e^*} a_B \,, \tag{2.2}$$

and a_B is the Bohr radius of the hydrogen atom. The width of the line is a result of the position–momentum uncertainty relation $\Delta x \Delta p \cong h$ which, using $p = \hbar k$,

https://doi.org/10.1017/CBO9780511599828.005 Published online by Cambridge University Press

can be written as $\Delta x \Delta k \cong 2\pi$. Here Δx represents the position uncertainty of an electron bound to the center. For deep impurities, the position uncertainty Δx is much smaller as compared to hydrogenic impurities, which results in a larger momentum uncertainty. Thus, the wave function of a deep impurity state requires contributions from a large range of k values, i.e. the deep impurity wave function contains contributions from a large range of the Brillouin zone (Seeger, 1982). Transitions between the conduction or valence band and the deep level are therefore possible for a large range of carrier momenta rather than only for $k \cong 0$ as in the case of shallow levels.

Another consequence of the strong localization of electrons on deep levels is the breakdown of effective mass theory for deep levels. The effective mass of carriers is a result of the periodic lattice potential. In effective mass theory, the periodic crystal potential is taken into account by the employment of the effective mass. Due to the strong localization to only one or a few lattice constants, the carrier bound to the deep center is *not* subject to a periodic potential. Thus, the basis for the effective mass concept is not given, which complicates the theoretical treatment of deep levels.

Due to the breakdown of effective mass theory, the crystal potential must be taken into account explicitly in addition to the potential of the deep impurity. Madelung (1978) pointed out that in the limit of a strong deep impurity potential, the crystal potential can be considered as a perturbation of the impurity potential. In this case, the theoretical treatment of deep centers is simplified. The energy levels of the deep impurity would then depend mostly on the nature of the center rather than on the host semiconductor.

The chemical origin of deep levels is frequently unknown. This is because deep levels can have many origins, e.g. impurities, point defects, or extended defects. Due to the uncertainty of their origin, deep levels are frequently labeled by their activation energy, e.g. the '0.8 eV level' in GaAs. Martin et al. (1977) introduced a systematic nomenclature for deep levels in n-type GaAs. Subsequently, Mitonneau et al. (1977) provided a systematic nomenclature for deep levels in p-type GaAs. An Arrhenius plot of the emission rates and the activation energy of the deep levels are shown in Figs. 2.2 and 2.3 for n-type and p-type GaAs, respectively. The first two letters in the trap label refer to the type of trap (E = electron trap, H = hole trap) and to the institution of publication (e.g. B = Bell Laboratories). The following number is a 'serial' number of the trap. Some of the different labels are most likely a trap with the same chemical origin (e.g. EB5 = EL4). Further discussion on specific deep levels in III–V semiconductors is referred to Chap. 9.

https://doi.org/10.1017/CBO9780511599828.005 Published online by Cambridge University Press

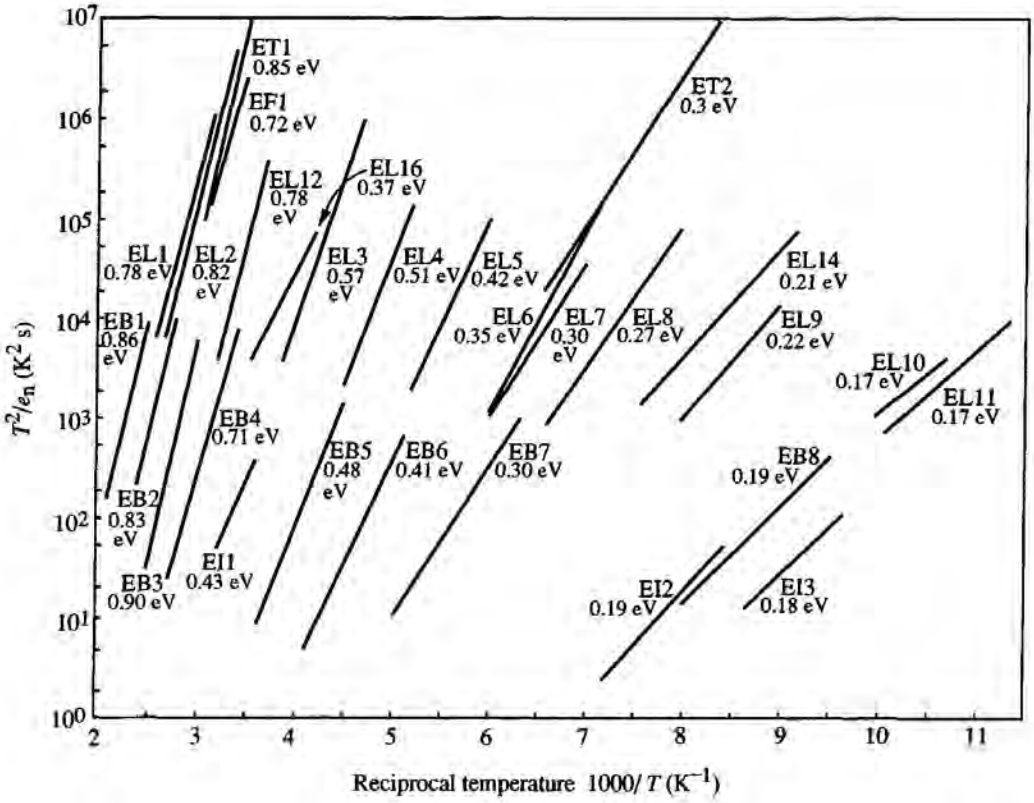


Fig. 2.2. Arrhenius plots and activation energies of deep electron traps in GaAs. The levels are labeled as ET, EF, EL, EB, and EI (after Martin et al., 1977).

Two important classes of deep substitutional impurity atoms are impurities with the tetrahedral sp^3 bond configuration and impurities with d-orbital participation in the bonding to host atoms. The sp^3 bond configuration is a common configuration and arises from the hybridization of spherically symmetric s orbitals and three p orbitals. The hybridized electron configuration is called sp^3-type and consists of four orbitals directed to the corners of a tetrahedron. Thus, sp^3-bonded impurities match the tetrahedral symmetry of atoms in the zincblende lattice of III–V semiconductors. Theoretical models for sp^3 bonding were reviewed by Lannoo and Bourgoin (1981) and Bourgoin and Lannoo (1983). A particularly simple model for sp^3-bonded deep impurities was developed by Hjalmarson et al. (1980). In this model, the impurity has an electrically inactive filled 'hyperdeep' level which forms the bond between the impurity and the host. The antibonding state of this level is predicted to be in the forbidden gap and forms the electrically active deep level. Based on this model, Vogl (1981)

https://doi.org/10.1017/CBO9780511599828.005 Published online by Cambridge University Press

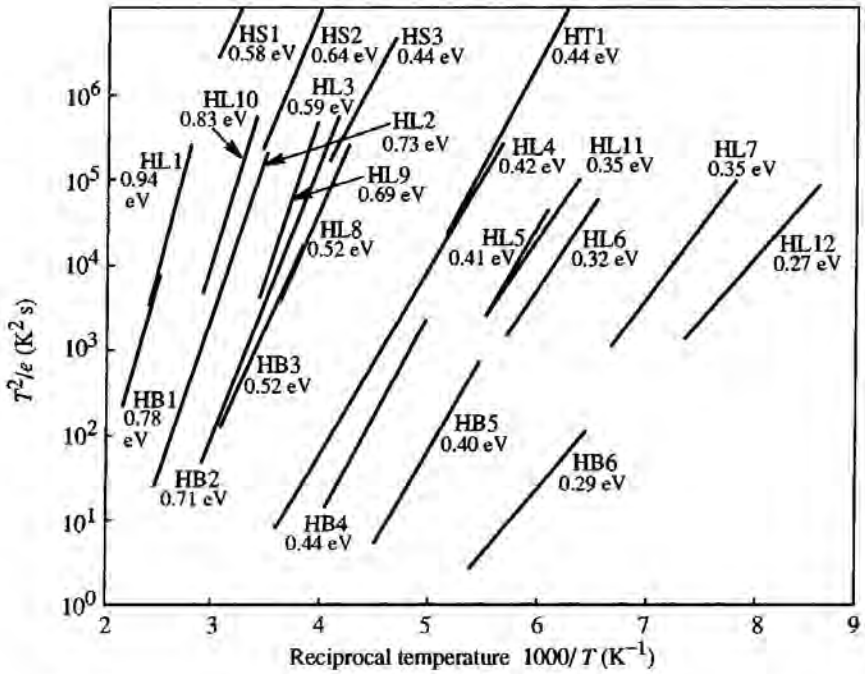


Fig. 2.3. Arrhenius plot and activation energies of deep hole traps in GaAs. The levels are labeled as HS, HT, HL,and HB (after Mitonneau et al., 1977).

predicted chemical trends for sp^3-bonded substitutional impurities in covalent semiconductors. In this approach, the trends were predicted based solely on the energy bands of the undisturbed host semiconductor and on the impurities' atomic structure. The response of the host electrons and the lattice to the perturbation caused by the impurity were not taken into account. Major chemical trends in deep trap energies were well explained by the model.

Transition metals are a second important class of deep impurities in III–V semiconductors. Transition metals with the 3d orbitals partially filled include Sc, Ti, V, Cr, Mn, Fe, Co, and Ni. For these impurities, d-orbital electrons participate in the bonding to the host semiconductor. Chromium and Fe are of great technological importance, since they are used to render III–V semiconductors semi-insulating (see Chap. 9). In addition, transition metals are frequently found as background impurities in III–V semiconductors. The oxidation states and charge states of transition metal impurities were categorized by Zunger (1986). Transition metal impurities are unique since internal optical transitions can occur between d orbitals. These transitions are typically found in the infra-red range. For example, internal d-to-d transitions occur in Fe^{2+} impurities in InP:Fe

https://doi.org/10.1017/CBO9780511599828.005 Published online by Cambridge University Press

(Bishop, 1986). The main luminescence peak occurs at an energy of 0.351 eV corresponding to $\lambda \cong 3.5$ μm. Vogl (1985) studied systematical chemical trends in transition metal impurities in semiconductors.

2.2 Capture, emission, and recombination

The dynamic process of capture and emission of electrons and holes by deep traps allows considerable insight in the phenomenology of deep levels. The consequences of capture and emission of carriers by deep levels are widespread and affect the transport as well as the optical properties of semiconductors. Capture of carriers by deep traps reduces the free carrier concentration as well as the mobility. Transient effects can be expected after rapid potential changes in semiconductor devices due to the presence of traps. Finally, traps shorten the minority carrier lifetime and thus the minority carrier diffusion length L_D (since $L_D \cong \sqrt{D\tau}$). The optical properties of semiconductors can be strongly affected by deep traps. Non-radiative recombination via traps competes with radiative band-to-band recombination. As a result, the luminescence efficiency of semiconductors is reduced by traps. The analysis of capture, emission, and recombination via deep traps allows one to quantitatively evaluate the deleterious effects alluded to.

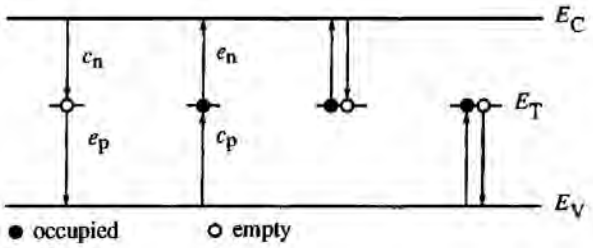


Fig. 2.4. Schematic illustration of emission and capture of electrons and holes. Traps that interact predominantly with the conduction band are called electron traps. Hole traps interact predominantly with the valence band.

The capture and emission process of a deep trap is schematically shown in Fig. 2.4. An empty level can either capture an electron from the conduction band or emit a hole into the valence band (i.e. bind an electron from the valence band). A filled trap can either emit an electron to the conduction band or capture a hole from the valence band (i.e. release an electron into the valence band). Traps that interact predominantly with the conduction band are called electron traps. The following inequalities allow one to differentiate between electron and hole traps

https://doi.org/10.1017/CBO9780511599828.005 Published online by Cambridge University Press

(Sah, 1967)

$$c_n >> e_p \quad \text{and} \quad e_n >> c_p \quad \text{for electron traps}, \tag{2.3}$$

$$e_p >> c_n \quad \text{and} \quad c_p >> e_n \quad \text{for hole traps}. \tag{2.4}$$

where c_n and c_p are the electron and hole capture rates and e_n and e_p are the electron and hole emission rates, respectively. Since the capture rates depend on the free carrier concentration, the same centers can act as electron traps and hole traps in n-type and p-type semiconductors, respectively.

The capture of carriers by a trap can be described in terms of a ***capture cross section***, σ, which has the dimension of an area. If a free carrier traverses the area of the size of the capture cross section, the carrier will be captured by the trap. Assuming an n-type semiconductor with electron concentration of n and an average carrier velocity v, the ***capture probability*** of one deep trap, i.e. the probability of *one* trap capturing a free carrier per unit time is given by

$$\boxed{c_n = nv\sigma} \tag{2.5}$$

The velocity of carriers can either be the *thermal* velocity (for a non-degenerate distribution) or the *Fermi* velocity (for a degenerate carrier density). The mean time that an initially unoccupied deep trap remains unoccupied is

$$\tau_c = 1/c_n . \tag{2.6}$$

The total ***capture rate*** is determined by the concentration of traps. If the concentration of unoccupied traps is N_t^0, then the capture rate is given by

$$\boxed{c_n^* = N_t^0 c_n = N_t^0\, nv\sigma} \tag{2.7}$$

Thus the capture rate gives the *number of carriers captured per unit volume and per unit time*. The reader is cautioned that the capture ***probability*** and the capture ***rate*** are frequently not distinguished in the literature, even though they are different quantities.

Under thermal equilibrium conditions, the emission probability and the capture probability are identical. The equality of emission and capture probabilities is a result of the ***principle of detailed balance*** (Blakemore, 1987). This principle states *for systems in equilibrium that the rate (and probability) of a process and its inverse are equal and balance in detail irrespective of other processes*. The equilibrium occupancy of traps and of the conduction band is

https://doi.org/10.1017/CBO9780511599828.005 Published online by Cambridge University Press

given by the Boltzman distribution. If we consider electrons of concentration n in the conduction band and occupied traps of concentration N_t^-, then the Boltzmann distribution yields

$$n/N_t^- = (N_c/N_t^0)\, g_0 \exp\left[-\frac{E_C - E_T}{kT}\right] \tag{2.8}$$

where g_0 is the ground-state degeneracy factor of the trap and N_t^0 is the concentration of empty traps. Assuming that all traps are initially filled with electrons, i.e. $N_t^- = N_t$, with the principle of detailed balance, and with Eq. (2.5), one obtains the ***emission probability*** per trap and per unit time according to

$$\boxed{e_n = \nu\sigma N_c\, g_0 \exp\left[-\frac{E_C - E_T}{kT}\right]} \tag{2.9}$$

The mean time that a trap is occupied by an electron is given by the inversion of the emission probability, i.e.

$$\tau_e = 1/e_n \tag{2.10}$$

Accordingly, the ***emission rate***, that is the number of electrons emitted per unit volume in the unit time interval is given by

$$\boxed{e_n^* = N_t^-\, e_n = N_t^-\, \nu\sigma N_c\, g_0 \exp\left[-\frac{E_C - E_T}{kT}\right]} \tag{2.11}$$

which should not be confused with the emission probability.

The knowledge of the emission and capture rates allows us to formulate the ***rate equations*** for transient emission and transient capture of electrons. Consider first the simple case of an n-type semiconductor with a concentration of N_t^- filled traps. Neglecting any capture processes, the rate equation for electron emission is given by

$$\frac{dN_t^-}{dt} = -N_t^-\, e_n\,. \tag{2.12a}$$

Assuming that all traps are initially filled with electrons [i.e. $N_t^-\,(t=0) = N_t$] the differential equation has the solution

https://doi.org/10.1017/CBO9780511599828.005 Published online by Cambridge University Press

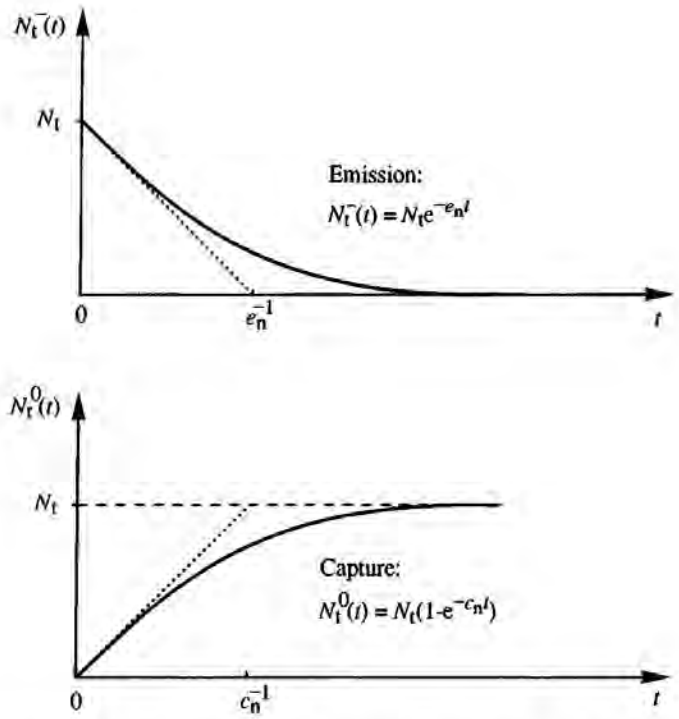


Fig. 2.5. Transient emission of electrons from traps and transient capture of electrons by traps. At $t = 0$ all traps are assumed to be filled and empty for emission and capture, respectively.

$$N_t^-(t) = N_t e^{-e_n t} . \tag{2.12b}$$

Consider next the opposite case in which electron capture dominates and any emission processes can be neglected. The rate equation is then given by

$$\frac{dN_t^0}{dt} = N_t^0 c_n . \tag{2.13a}$$

Assuming that all traps are initially unoccupied [i.e. $N_t^0(t=0) = N_t$] the differential equation has the solution

$$N_t^0(t) = N_t(1 - e^{-c_n t}) . \tag{2.13b}$$

The emission and capture transients of Eqs. (2.12b) and (2.13b) are shown in Fig. 2.5. If both processes, i.e. emission and capture, are considered, the rate equation is

https://doi.org/10.1017/CBO9780511599828.005 Published online by Cambridge University Press

$$\frac{dN_t^-}{dt} = -N_t^- e_n + N_t^0 c_n \tag{2.14a}$$

where the identity

$$\frac{dN_t^-}{dt} = -\frac{dN_t^0}{dt} \tag{2.14b}$$

was used. Under steady state conditions, $dN_t^-/dt = 0$, and one obtains

$$\boxed{\frac{N_t^-}{N_t^0} = \frac{c_n}{e_n}} \tag{2.15}$$

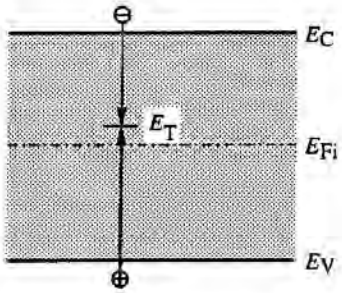


Fig. 2.6. Recombination of electrons and holes via a Shockley–Read recombination center.

In semiconductor structures with an appreciable number of minority carriers, electron–hole recombination can occur via deep traps. The electron–hole recombination via a deep level is schematically shown in Fig. 2.6. As a result, the minority carrier lifetime as well as the luminescence efficiency is reduced. Properties of minority carrier devices such as bipolar transistors, light-emitting diodes, and lasers are strongly affected by traps. Recombination via deep levels was first analyzed by Shockley and Read (1952) and is therefore frequently referred to as ***Shockley–Read recombination***.

For the calculation of the Schockley–Read recombination (Shockley and Read, 1952; Hall 1951, 1952) we assume an equilibrium electron and hole concentration of n_0 and p_0, respectively. We first consider the capture of electrons and subsequently the capture of holes. We assume that the trap has a concentration of N_t and the charge states of 0 and −1 when unoccupied and occupied by an electron, respectively. We further assume that the trap has a single level at an energy E_t below the conduction band edge. The capture rate of electrons is given by Eq. (2.7) according to

https://doi.org/10.1017/CBO9780511599828.005 Published online by Cambridge University Press

$$c_n^* = N_t^0 \, n v_n \sigma_n$$

$$= N_t (1 - f_t) n v_n \sigma_n \tag{2.16}$$

where N_t^0 is the concentration of unoccupied traps. The Fermi–Dirac factor f_t is the probability that the trap is filled, i.e. $(1 - f_t)$ is the probability that the trap is empty. The Fermi–Dirac distribution (see Chap 3) is given by

$$f_t = \left[1 + e^{(E_T - E_F)/kT}\right]^{-1} . \tag{2.17}$$

The emission rate of electrons into the conduction band is given by Eq. (2.11) according to

$$e_n^* = N_t^- \, v_n \sigma_n \, N_c g_0 e^{-E_t/kT}$$

$$= N_t f_t \, v_n \, \sigma_n N_c \, e^{-E_t/kT} \tag{2.18}$$

where the ground-state degeneracy g_0 is assumed to be unity, and $E_t = E_C - E_T$. Using the Boltzmann distribution (see Chap. 3) for the occupancy of the conduction band

$$n = N_c \, e^{(E_F - E_C)/kT} , \tag{2.19}$$

Eq. (2.18) can be rewritten as

$$e_n^* = N_t f_t \, v_n \sigma_n \, n_1 \tag{2.20}$$

where n_1 is the free electron concentration if the Fermi level coincides with the trap level. Under equilibrium conditions, the emission rate and the capture rate coincide, that is

$$c_n^* = e_n^* ,$$

$$N_t (1 - f_t) n v_n \sigma_n = N_t \, f_t n_1 \, v_n \sigma_n . \tag{2.21}$$

However, if conditions depart from equilibrium, then $c_n^* \neq e_n^*$ and a ***net capture rate***, R_n, can be defined as

$$R_n = c_n^* - e_n^* = N_t v_n \sigma_n [(1 - f_t) n - f_t n_1] . \tag{2.22}$$

The net capture rate of holes, R_p, is obtained in an analogous way and is given by

$$R_p = N_t \, v_p \sigma_p [f_t p - (1 - f_t) p_1] \tag{2.23}$$

where $p_1 = n_i^2 / n_1$, σ_p is the capture cross section of the trap for holes and v_p is either the thermal or the Fermi velocity of holes. The trap alternatively captures

https://doi.org/10.1017/CBO9780511599828.005 Published online by Cambridge University Press

337-TA-1366 RX-0255

electrons and holes. In a steady state process the net hole capture rate equals the net electron capture rate, that is $R = R_n = R_p$. This condition gives us the following equation for f_t

$$f_t = \frac{n v_n \sigma_n + p_1 v_p \sigma_p}{v_n \sigma_n (n + n_1) + v_p \sigma_p (p + p_1)} . \tag{2.24}$$

With $R = R_n = R_p$ and the insertion of Eq. (2.24) into Eq. (2.22) or (2.23) one obtains, using $n_1 p_1 = n_i^2$, the net recombination rate

$$R_{SR} = R = \frac{N_t v_n v_p \sigma_n \sigma_p (np - n_i^2)}{v_n \sigma_n (n + n_1) + v_p \sigma_p (p + p_1)} \tag{2.25}$$

which gives us the net non-radiative ***Schockley–Read recombination rate*** in semiconductors. Note that for equilibrium conditions $np = n_i^2$ and the Shockley–Read rate equals zero. The Shockley–Read rate has the dimension cm^{-3}/s. The non-equilibrium carrier concentrations is conveniently written in terms of $n = n_0 + \Delta n$ and $p = p_0 + \Delta p$ with $\Delta n = \Delta p$ due to neutrality. Using this notation, the Shockley–Read recombination rate is

$$\boxed{R_{SR} = \frac{n_0 \Delta p + p_0 \Delta p + \Delta p^2}{(N_t v_p \sigma_p)^{-1} (n_0 + n_1 + \Delta p) + (N_t v_n \sigma_n)^{-1} (p_0 + p_1 + \Delta p)}} \tag{2.26}$$

The lifetime of excess carriers is given by the equation $R_{SR} = \Delta p/\tau$. Consequently the lifetime is given by

$$\frac{1}{\tau} = \frac{n_0 + p_0 + \Delta p}{(N_t v_p \sigma_p)^{-1} (n_0 + n_1 + \Delta p) + (N_t v_n \sigma_n)^{-1} (p_0 + p_1 + \Delta p)} . \tag{2.27}$$

We now differentiate between majority carriers and minority carriers and assume that electrons are in the majority, i.e. $n_0 >> p_0$ and $n_0 >> n_1$. If we further assume a small deviation from equilibrium i.e. $\Delta p << n$, then

$$\boxed{\frac{1}{\tau} = \frac{1}{\tau_{p_0}} = N_t v_p \sigma_p} \tag{2.28}$$

If holes are the majority carriers, the lifetime is obtained in an analogous way.

$$\frac{1}{\tau} = \frac{1}{\tau_{n_0}} = N_t v_n \sigma_n . \tag{2.29}$$

https://doi.org/10.1017/CBO9780511599828.005 Published online by Cambridge University Press

RX-0255.89

The results show that the Shockley–Read recombination rate is limited by the rate of capture of minority carriers. This result suggests itself since the capture of majority carriers is a much more likely event than the capture of minority carriers. Equation (2.27) can then be written as

$$\frac{1}{\tau} = \frac{n_0 + p_0 + \Delta p}{\tau_{p_0}(n_0 + n_1 + \Delta n) + \tau_{n_0}(p_0 + p_1 + \Delta p)} . \tag{2.30}$$

For small deviations from equilibrium, i.e. $\Delta p << n_0$, the equation simplifies to

$$\tau = \tau_{p_0} \frac{n_0 + n_1}{n_0 + p_0} + \tau_{n_0} \frac{p_0 + p_1 + \Delta p}{n_0 + p_0} \equiv \tau_{p_0} \frac{n_0 + n_1}{n_0 + p_0} . \tag{2.31}$$

Inspection of the equation reveals that the lifetime does not change for small deviations from equilibrium in an extrinsic semiconductor.

For further insight we assume that the trap captures electrons and holes at the same rate, i.e. $v_n \sigma_n = v_p \sigma_p$ and $\tau_{p_0} = \tau_{n_0}$. One obtains from Eq. (2.31)

$$\tau = \tau_{p_0} \left[1 + \frac{n_1 + p_1}{n_0 + p_0} \right] . \tag{2.32}$$

For the special case of intrinsic material, i.e. $n_0 = p_0 = n_i$, the equation simplifies to

$$\tau_i = \tau_{p_0} \left[1 + \frac{n_1 + p_1}{2n_i} \right] = \tau_{p_0} \left[1 + \cosh \left(\frac{E_T - E_{F_i}}{kT} \right) \right] \tag{2.33}$$

where E_{F_i} is the intrinsic Fermi level which is typically close to the middle of the gap. Inspection of the equation reveals that the lifetime is minimized, if $E_T - E_{F_i}$ is small, that is if the trap level is close to midgap. For such midgap levels the lifetime is given by $\tau = 2\tau_{p_0}$. This result demonstrates that deep levels are effective recombination centers if they are in the middle of the gap.

2.3 Effects relating to experimental properties

Several properties of deep levels directly relate to characteristics that can be observed experimentally. For example, if the capture of an electron from the conduction band onto the trap level is an optical transition, then the transition can be detected by luminescence measurements. Another example is the optical excitation of traps. What is the optical ionization threshold for traps? How does the threshold depend on the atomic structure and configuration of the trap? Other properties relate to transport properties of traps. The ionization energy of traps is

https://doi.org/10.1017/CBO9780511599828.005 Published online by Cambridge University Press

known to depend on the strength of an electric field applied to the semiconductor. What is the functional dependence of the trap ionization energy on the electric field? These questions can be answered using phenomenological models of traps. The following section summarizes several effects and characteristics of deep levels that are relevant to experiments. The characteristics include (i) recombination characteristics, (ii) configurational relaxation of the trap and the surrounding lattice, and (iii) thermal ionization of traps in the presence of an electric field (Poole–Frenkel effect).

2.3.1 Recombination mechanisms

For shallow impurities, thermal ionization and recombination is primarily assisted by a single phonon. That is, the ionization and neutralization of an impurity is associated by the absorption and the emission of a phonon. For deep levels, a recombination process involving one phonon is not possible. Energies of acoustical and optical phonons in III–V semiconductors are typically limited to energies < 50 meV. Thus, the conservation of energy cannot be satisfied by the emission of a single phonon for deep level energies that are removed by > 50 meV from one of the band edges. Free carriers cannot recombine with deep levels unless there is a way of disposing of the carrier's energy. The recombination process of deep levels is therefore fundamentally different from the recombination process of a shallow level and has been the subject of considerable interest. Since multiphonon recombination has a very small probability (Lax, 1959), other mechanisms are required to explain the relatively large capture cross sections of deep traps which were observed experimentally.

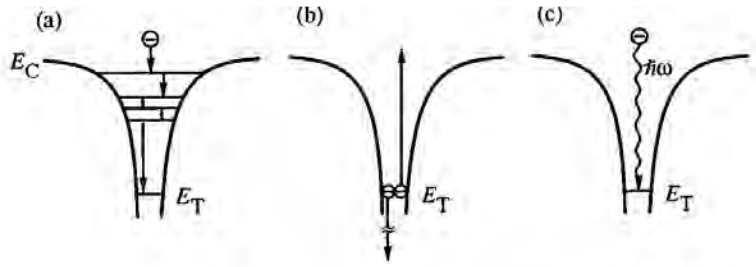


Fig. 2.7. Capture of an electron in the conduction band by a deep trap (a) via a cascade process, (b) assisted by an Auger process, and (c) by a radiative transition.

https://doi.org/10.1017/CBO9780511599828.005 Published online by Cambridge University Press

There are four processes by which recombination of free carriers with deep levels can occur. These processes are ***optical*** recombination, ***cascade*** recombination, ***Auger*** recombination, and ***multiphonon*** recombination. The former three of the mechanisms are schematically shown in Fig. 2.7. The mechanisms will be discussed in the following.

In the ***cascade mechanism*** for non-radiative recombination with a deep level, electrons cascade down excited states of the deep level via single phonon emission (Lax, 1959, 1960). To illustrate the cascade mechanism, consider a donor-like deep trap that is unoccupied and positively charged. Electrons in the conduction band are attracted by the long-range Coulomb potential of the deep trap. The deep level generally has a ground state and excited states associated with the short-range potential and a hydrogenic series of levels associated with the long-range Coulomb potential. In the initial stages of the capture process, the electron will be captured by the long-range potential, i.e. occupy states of the hydrogenic series. Within the hydrogenic series of states, the electron can jump to lower states by the emission of a single phonon. Since the probability of one-phonon transitions is substantial, such transitions enable the electron to escape from the conduction band (Jaros, 1982). The electron descends down the ladder of excited states, losing its energy by the successive emission of single phonons until the separation between the excited states becomes too large. Although it is obviously impossible to bring the electron down to the ground state without a multiphonon transition (or without the emission of a photon), it is sufficient that the electrons are trapped in the excited states and become unavailable for conduction. Lax argued that if a center is to behave as an efficient trap at a given temperature, its ionization energy of the hydrogenic series must be substantially larger than kT so that there are enough excited states from which the electron cannot be thermally ejected back into the conduction band. The motion up and down among the excited states is so rapid compared with the rate at which the electron disappears from the conduction band that the corresponding delay may be neglected and be accounted for by expressing the fraction of electrons that 'stick'. The sticking probability $p(U)$ simply depends upon the probability than an electron with energy U will emit a phonon of required energy. To obtain the capture cross section, one must calculate the total probability, i.e. (Jaros, 1982)

$$\sigma(E) = \int \sigma(E,U)p(U)\mathrm{d}U \tag{2.34}$$

where E is the kinetic energy of the incoming (free) electron and $\sigma(E,U)$ is the cross section for the corresponding collision event. Lax argued that Eq. (2.34) leads to realistic values for the calculated capture cross section. He pointed out that multiphonon transitions to a deep level gives capture cross sections that are

https://doi.org/10.1017/CBO9780511599828.005 Published online by Cambridge University Press

10 orders of magnitude too small. In III–V semiconductors the energy of the hydrogenic donor series is $\lesssim kT$. This enables carriers to jump to the hydrogenic ground state by emission of a single phonon. Even though the cascade capture is less effective if the hydrogenic ground state is shallow, the cascade process is more likely than multiphonon capture (Lax, 1992). Capture cross sections of centers capturing carriers by the cascade process are typically on the order of the Bohr radius. For an effective Bohr radius of 100 Å, the capture cross section is on the order of $\sigma = a_B^{*2}\,\pi \cong 10^{-12}$ cm^2. The cascade process does not work equally well for *neutral* centers since neutral centers do not have the attractive long-range Coulomb potential. Lax (1959) argued that the large cross sections found for neutral centers can be explained on the same basis, the attractive potential in this case being provided by the large polarizability of the neutral center. This problem was further discussed by Jaros and Brand (1976). They concluded that neutral centers may indeed have several energy levels within the forbidden gap. Cascade recombination processes can occur via such energy levels. However, the authors also concluded that some neutral centers may have one single energy level only. In this case, a cascade recombination process cannot occur.

The second non-radiative recombination mechanism is the ***Auger recombination mechanism*** which is schematically shown in Fig. 2.7(b). Sheinkman (1965) proposed this mechanism for multiply charged deep centers. As an example, consider a double negatively charged center (see Fig. 2.7b). The total energy can be preserved if one of the electrons transfers to the valence band and the other is lifted to the conduction band. Thus, neither phonons nor photons are required for the recombination process. Similarly, the level can be doubly charged by capturing an electron from the conduction band and another carrier simultaneously from the valence band (this situation can be obtained from Fig. 2.7(b) by reversing the direction of the arrows). Auger band-to-band recombination was analyzed by Landsberg (1970). The probability of such Auger band-to-band processes increases with the square of the free carrier concentration, since two electrons are required for the process. However, deep level Auger processes involving pairs of electrons bound to the same defect do not exhibit this characteristic concentration dependence. Riddoch and Jaros (1980) showed that the capture cross section for Auger effects is very sensitive to the relationship between the impurity wave function and the relevant distorted Bloch states participating in the transition. The authors showed that the Auger capture cross section is approximately $10^{-16} - 10^{-18}$ cm^2, which is comparable to the capture cross section of a multiphonon capture.

https://doi.org/10.1017/CBO9780511599828.005 Published online by Cambridge University Press

Radiative recombination of a conduction band electron with a deep level is shown in Fig. 2.7(c). Transitions between deep levels and continuum states are characterized in terms of an optical capture cross section σ_c^{opt}, and a photoionization cross section, σ^{opt}. The optical photoionization cross section is defined as the probability that an absorption transition takes place between a deep level state and the continuum states for one incoming photon per unit area per second (Jaros, 1982). The optical capture cross section is defined as the probability that a photon is emitted due to the transition involving a continuum level and the deep level for one incoming electron per unit area and per unit time. Under thermal equilibrium conditions, the photon capture and photoionization rates balance in detail. The theoretical framework of radiative deep level transitions is based on the well-known theory of radiative band-to-band transitions developed by van Roosbroeck and Shockley (1954). This approach was modified by Solar and Burstein (1955) and by Blakemore (1967) to radiative extrinsic processes. There are only few reports on optical transitions involving deep levels. Banks et al. (1980) reported photoionization cross sections of deep level transitions in GaAs and GaP. Line shape analysis of the deep level transitions was also performed.

Multiphonon recombination processes usually have a very small capture cross section, e.g. $<10^{-20}$ cm^2 (Lax, 1959, 1960). The small capture cross section is due to the fact that several phonons are required for a single recombination process. However, if the deep center is neutral and does not have excited states, then cascade recombination is less likely to occur. Multiphonon recombination may then be the dominating recombination mechanism. Henry and Lang (1977) argued that multiphonon emission during recombination processes with deep centers in GaAs and GaP is the prevailing process. The authors further showed that large lattice relaxation (see Sect. 2.3.2) is required for multiphonon recombination. This result is in agreement with Kovarskii (1962), who pointed out that multiphonon processes are likely in centers with large lattice relaxation.

2.3.2 Large lattice relaxation

The model of large lattice relaxation of deep centers can satisfactorily explain a host of experimental characteristics frequently observed with deep centers. In the model of lattice relaxation, the generalizing assumption is made that the atomic configuration of the center and of the surrounding host lattice depends on the charge state of the deep center. For example, the position, bonding structure, and electron configuration of the center can change with the charge state of the center. Based on this assumption, many features of deep levels can be explained

https://doi.org/10.1017/CBO9780511599828.005 Published online by Cambridge University Press

including (i) a thermally activated capture cross section, (ii) a thermally activated emission energy that exceeds the thermal ionization energy, and (iii) a optical ionization energy that is much larger than the thermal activation energy of the center.

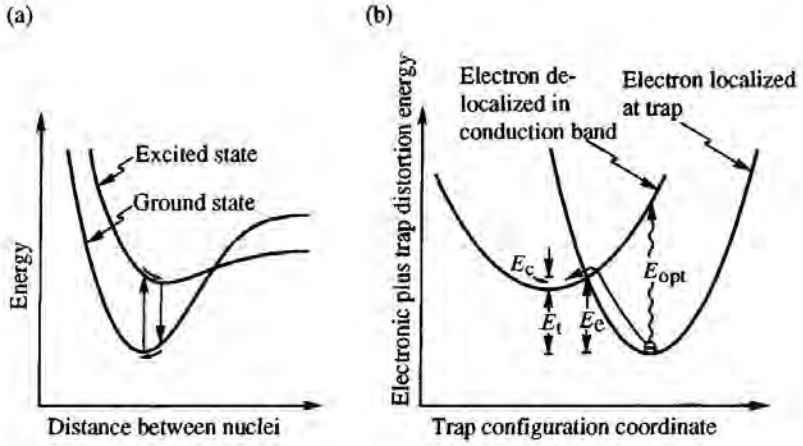


Fig. 2.8. Schematic illustration of the configuration coordinate diagram for (a) a molecule exhibiting a Franck–Condon shift (small relaxation) and for (b) a deep center with large lattice relaxation. Thermally activated capture with energy E_c is required for the case of large lattice relaxation, while thermal activation is not required for the electron transition to the ground state in the case of small lattice relaxation.

To appreciate the effect of large lattice relaxation, it is useful to recall the Franck–Condon principle (Davydov, 1976). The Franck–Condon principle applies to optical transitions between two electronic states of a molecule. The total energy of the molecule is shown in Fig. 2.8(a) for two different electronic configurations as a function of the distance between the nuclei. In the first configuration, the electron of interest occupies the ground state. In the second configuration, the electron is in an excited state. Other electrons are not considered here. The distance between the nuclei assumes a value to minimize the total energy of the molecule. For the dependences shown in Fig. 2.8(a), the distance between the nuclei is smaller if the electron occupies the ground state orbital. The two atoms are more loosely bound, i.e. the distance between the radii is larger, if the electron occupies the excited state. Franck argued that a change in the *electronic* configuration can occur much more rapidly than a change in the *atomic* configuration, due to the much smaller masses of electrons as compared to those of the nuclei (Davydov, 1976). Therefore, the transitions, indicated by

https://doi.org/10.1017/CBO9780511599828.005 Published online by Cambridge University Press

arrows in Fig. 2.8(a), can only be 'vertical', i.e. without a change in the atomic configuration. As a result, the emission energy is always smaller than the absorption energy. The difference in energy is referred to as the ***Franck–Condon shift*** or ***Franck–Condon energy***. The excess energy, i.e. the difference between absorption and emission energy, usually results in vibrational energy of the molecule (Finkelnburg, 1958).

The Franck–Condon principle can be applied to deep centers in semiconductors by replacing the distance between nuclei by a ***generalized configuration coordinate***. Movement along the configuration coordinate implies the continuous change of the atomic configuration of the deep trap as well as of its surrounding lattice. A change in the configuration coordinate can represent the displacement of an impurity out of its equilibrium position, the breakage of an electronic bond, the distortion of the lattice, or a similar effect. It is further assumed that the configuration of the center is changed upon changing the charge state of the center. In the following, we consider a trap with two charge states, namely with an electron localized at the center (charged) and with the electron delocalized in the conduction band. Furthermore, we assume that the elastic energy changes quadratically upon changing the configuration coordinate (Hooke's law). Such a situation is illustrated in Fig. 2.8(b) which shows the total energy, i.e. electronic energy, plus trap distortion energy as a function of the configuration coordinate.

It is useful to differentiate between different magnitudes of lattice relaxation. Lattice relaxation is termed *small*, if vertical absorption or emission transitions can occur from the minimum of each parabola to the other parabola. Such *small lattice relaxation* is shown in Fig. 2.8(a). Lattice relaxation is called *large*, if vertical absorption or emission transitions are not possible from the minimum of each parabola to the other parabola. Such a *large lattice relaxation* is illustrated in the configuration coordinate diagram shown in Fig. 2.8(b). The figure shows an absorption transition from the lower to the upper parabola. However, a vertical transition from the minimum of the upper parabola to the lower parabola is not possible. As a consequence, systems with large lattice relaxation exhibit metastability, i.e. cannot return to the ground state via a vertical transition.

Recombination of an electron with the deep level exhibiting large lattice relaxation cannot occur via a *vertical* transition but requires the thermal excitation of the center to the crossover point of the two parabola. Once excited to the crossover point, the center can relax to its ground state. The capture cross section of the deep center can be written as (Henry and Lang, 1977)

https://doi.org/10.1017/CBO9780511599828.005 Published online by Cambridge University Press

$$\sigma = \sigma^{\infty} e^{-E_c/kT} \tag{2.35}$$

where E_c is the thermal activation energy for capture and σ^{∞} is the capture cross section for $T \to \infty$. The thermally activated capture cross section of Eq. (2.35) has a fundamentally different functional dependence on temperature as compared to ordinary deep levels with a capture cross section independent of temperature. A second possibility for the relaxation of the deep trap is optical excitation. 'Vertical' excitation from the parabola representing the empty deep center to the parabola representing the (filled) center results in a finite probability that the trap relaxes to its filled ground state. The photon energy required for this transition exceeds the thermal capture energy, E_c, for the configuration coordinate diagram shown in Fig. 2.8(b). Large lattice relaxation of deep centers can also account for a large difference between the thermal ionization energy and the transient emission energy of deep levels. As illustrated in Fig. 2.8(b), the emission of an electron from the deep level into the conduction band requires an energy of E_e. This energy is relevant in transient emission experiments such as deep level transient spectroscopy (Lang and Logan, 1977; Lang et al., 1979). However, under (non-transient) thermal equilibrium conditions, the distribution of electrons between the deep level ground state and conduction band states is given by the energy difference between the two states, E_t. This thermal ionization energy is relevant in Hall measurements carried out at a given temperature. Note that the activation energy for transient emission is the sum of the thermal activation energy and the capture activation energy, i.e.

$$E_e = E_t + E_c \tag{2.36}$$

as shown in Fig. 2.8(b).

The optical excitation energy of a electron occupying the deep level can be much larger than the thermal excitation energy for levels exhibiting large lattice relaxation. As with the Franck–Condon effect in molecules, it is reasonable to assume that changes in the electronic configuration of the center occur much more rapidly than changes in the structural (atomic) configuration. Therefore, the optical excitation of an electron to the conduction band must be a 'vertical' transition. The optical excitation energy, E_{opt}, can therefore be much larger than the thermal excitation energy, as shown in the example of Fig. 2.8(b). Much larger optical excitation energies have indeed been observed in centers exhibiting large lattice relaxation (Lang et al., 1979).

Large lattice relaxation can also explain the ***persistent photoconductivity*** effect which is observed at low temperatures in several III–V semiconductors (see e.g. Nelson, 1977). After photoexcitation of the deep centers at low temperatures,

https://doi.org/10.1017/CBO9780511599828.005 Published online by Cambridge University Press

337-TA-1366 RX-0255

the conductivity increases. Most of the increase in photoconductivity persists even after the exciting illumination is turned off (Schubert and Ploog, 1984). The effect is referred to as persistent photoconductivity or persistent photocurrent. To explain the phenomenon of persistent photoconductivity, we refer to Fig. 2.8(b). After photoexcitation, the electron occupies conduction band states and contributes to the conductivity of the semiconductor. Recombination with the deep center requires that the thermal barrier of energy E_c be overcome. For $E_c >> kT$, recombination can be very slow and can last for more than 24 hours, since carriers cannot overcome the thermal capture barrier. The persistent photoconductivity effect can be quenched by heating the sample to higher temperatures at which carriers have sufficient thermal energy to overcome the capture barrier.

Centers with large lattice relaxation are frequently characterized by the ***Huang–Rhys parameters*** S (Huang and Rhys, 1950). This parameter describes the shape of the parabola in configuration space (see Fig. 2.8). The parameter is given by $S = E_e/\hbar\omega$, where E_e is the energy required for electron emission from the ground state of the center, and $\hbar\omega$ is the vibrational energy of the center. For $S >> 1$, i.e. $E_e >> \hbar\omega$, a thermally activated emission process is expected.

We finally note that deep centers exhibiting lattice relaxation can have different symmetries for the two different configurations, i.e. in their occupied and empty states. Therefore, experimental procedures sensitive to the symmetry of the center can contribute to the structural identification of the two configurations (for examples see Chap. 9).

2.3.3 The Poole–Frenkel effect

The emission characteristics of deep centers depend on the electric field applied to the semiconductor. For electric fields with a magnitude of $\sim 10^5$ V/cm, the change in emission characteristics can be quite significant (see, for example, Lang, 1974a, 1974b; Vincent et al., 1979). The change in emission characteristics is attributed to the Poole–Frenkel effect, which implies a lowering of the effective emission barrier due to an electric field (Frenkel, 1938a and 1938b). The effect is also named after Poole (1914) who observed that the conductivity of a dielectric increases exponentially with the electric field in the pre-breakdown regime (see also Bethe and Salpeter, 1957).

The schematic potential of a deep impurity is shown schematically in Fig. 2.9 without (dash-dotted line) and with (solid line) a superimposed electric field. The emission of an electron from the deep level can be either thermionic emission or thermally assisted tunneling as indicated by the two arrows (a) and (b), respectively. Below, we first consider the effect of thermionic emission and

https://doi.org/10.1017/CBO9780511599828.005 Published online by Cambridge University Press

RX-0255.98

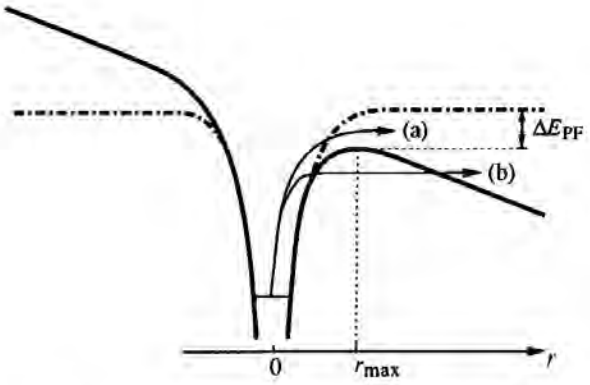


Fig. 2.9. Deep center potential with (solid line) and without (dash-dotted line) an electric field. Electron emission to the conduction band in the presence of an electric field can occur (a) via thermal emission (Poole–Frenkel emission) or (b) via thermally assisted tunneling.

subsequently the effect of tunneling through the barrier.

The potential of a deep level can be separated into a short-range potential and a long-range Coulomb potential. Since the short-range potential extends over only a few unit cells, the effect of an electric field on the short-range potential is small. The long-range Coulomb potential and the superimposed electric field are given by

$$V(r) = \frac{e}{4\pi\varepsilon r} + \mathcal{E} r \tag{2.37}$$

where $\mathcal{E}$ is the magnitude of the electric field and $r = 0$ is the location of the impurity core. The potential has an extremum ($dV/dr = 0$) at $r_{max} = (e/4\pi\varepsilon\mathcal{E})^{1/2}$. The lowering of the emission barrier is then obtained by inserting r_{max} into Eq. (2.37), i.e.

$$\Delta E_{PF} = eV(r_{max}) = e\sqrt{\frac{e\mathcal{E}}{\pi\varepsilon}} \tag{2.38}$$

The emission probability for a deep center in an electric field is obtained from Eq. (2.9) by including the Poole–Frenkel energy, i.e.

https://doi.org/10.1017/CBO9780511599828.005 Published online by Cambridge University Press

$$e_n^{PF} = \nu\sigma N_c \, e^{-(E_t - \Delta E_{PF})/kT} \tag{2.39}$$

where $E_t = E_C - E_T$, and the ground-state degeneracy of the trap is assumed to be unity. Equation (2.39) is based on purely thermionic emission and tunneling through the barrier is neglected. Furthermore, Eq. (2.39) is a one-dimensional approximation. The three-dimensional case was developed by Hartke (1968) and by Milnes (1973).

In addition to thermal emission, tunneling is an alternative mechanism for electron emission from a deep level. This mechanism is indicated by the arrow (b) in Fig. 2.9. As a simple approximation, the tunnel barrier can be considered to be triangular. Tunneling through triangular barriers is known as Fowler–Nordheim tunneling which has the classical tunneling probability of

$$p(E) = e^{-K} \tag{2.40a}$$

with

$$K = \frac{4}{3} \frac{\sqrt{2m^*}}{e\hbar\mathcal{E}} E_t \, . \tag{2.40b}$$

The emission probability due to tunneling is then given by (Korol, 1977)

$$e_n^{tunnel} = \gamma \frac{E_t}{eK} e^{-K} \, . \tag{2.41}$$

The prefactor γ depends on the short-range potential of the deep level. Korol (1977) obtained a value of $\gamma = e/3\hbar$ for the model potential used. Subsequently, thermally assisted tunneling was considered by Vincent et al. (1979) and Pons and Makram-Ebeid (1979). At high temperatures, the emission probability can be approximated by

$$e_n \cong e_{n_0} \, e^{-2SkT/\hbar\omega} \tag{2.42}$$

where the Huang–Rhys parameter S represents the strength of the electron–phonon coupling.

https://doi.org/10.1017/CBO9780511599828.005 Published online by Cambridge University Press

3

Semiconductor statistics

The concentration of neutral impurities, ionized impurities, and free carriers in a doped semiconductor depends on a large number of parameters such as the impurity atom concentration, the free carrier mass, the bandgap energy, and the dielectric constant. The interdependences of the free majority and minority carrier concentration, the impurity concentration, impurity ionization energy as well as some other constants and materials parameters are given by ***semiconductor statistics***. Semiconductor statistics describes the probabilities that a set of electronic states are either vacant or populated.

Electronic states include localized impurity states as well as delocalized conduction and valence band states. In the simplest case, an impurity has a single state with no degeneracy ($g_0 = 1$). However, an impurity may have a degenerate ground state ($g_0 > 1$) as well as excited levels which may need to be considered. The states in the bands and their dependence on energy are described by the ***density of states***. In semiconductor heterostructures, the free motion of carriers is restricted to two, one, or zero spatial dimensions. In order to apply semiconductor statistics to such systems of reduced dimensions, the density of states in quantum wells (two dimensions), quantum wires (one dimension), and quantum dots (zero dimensions), must be known. The density of states in such systems will also be calculated in this chapter.

Classical Maxwell–Boltzmann statistics and quantum mechanical Fermi–Dirac statistics are introduced to calculate the occupancy of states. Special attention is given to analytic approximations of the Fermi–Dirac integral and to its approximate solutions in the non-degenerate and the highly degenerate regime. In addition, some numerical approximations to the Fermi–Dirac integral are summarized.

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

The *activation energy* of impurities will be frequently used in this chapter. It is useful to recall the interdependence of free energy, internal energy, enthalpy, entropy, and activation energy. To do so, consider the electronic ionization of an impurity, for example a donor

$$D^0 \leftrightarrow D^+ + e\,. \tag{3.1}$$

The effective work necessary to accomplish the ionization process at a constant temperature and pressure equals to the change of ***Gibbs free energy*** of the system (Kittel and Kroemer, 1980; Reif, 1965). In thermodynamics, Gibbs free energy G is defined as

$$G = H - TS\,, \tag{3.2}$$

$$H = E + PV \tag{3.3}$$

where H is the reaction enthalpy, S the entropy, E the internal energy, and PV the product of pressure and volume of the system. The change in Gibbs free energy occurring during the donor ionization process of Eq. (3.1) at a constant temperature T is then given by

$$\Delta G = \Delta H - T\Delta S\,, \tag{3.4}$$

$$\Delta H = \Delta E + P\Delta V \tag{3.5}$$

where constant temperature and constant pressure is assumed. Gibbs free energy is the proper energy to be used in a Boltzmann factor or Fermi function (see Sect. 3.3). The change in volume of the system occurring during chemical reactions can be quite significant. However, the change in volume during the electronic reaction of Eq. (3.1) is very small since the valence electron configuration does not change. The change in volume can therefore be neglected. In this chapter, the change in entropy as well as the mechanical work ($P\Delta V$) are neglected. In this case, $\Delta G \cong \Delta H \cong \Delta E$. The energy required for the ionization reaction of Eq. (3.1) is the difference in internal energy, i.e. the difference in energy of states occupied by the electron before and after the ionization process. The change in free energy for donors can then be written as $\Delta G \cong \Delta H \cong \Delta E = E_C - E_D = E_d$, that is, the ionization energy equals the donor level energy relative to the bottom of the conduction band. The enthalpy and the entropy of ionization of centers in semiconductors were further considered by Thurmond (1975) and by Van Vechten and Thurmond (1976a, 1976b). The authors made simple estimates of the entropy of ionization of coulombic, isoelectronic, and vacancy-type defects in semiconductors by

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

considering the effect of localized and free-carrier charge distributions upon the lattice modes. The empirical values of these entropies are observed as the temperature variation of the corresponding ionization levels (i.e. the term $T\Delta S$ in Eq. (3.4)). The change in entropy during the ionization reaction of Au-related levels in Si was considered by Lang et al. (1980), who differentiated between the entropy change due to electronic degeneracy and due to atomic vibrational changes. The authors showed that the change in entropy can be a small fraction (10%) of the ionization enthalpy.

3.1 Density of continuum states

Carriers occupy either localized impurity states or delocalized continuum states in the conduction band or valence band. In the simplest case, each impurity has a single, non-degenerate state. Thus, the density of impurity states equals the concentration of impurities. The energy of the impurity states is the same for all impurities (of the same species) as long as the impurities are sufficiently far apart and do not couple. The density of continuum states is more complicated and will be calculated in the following sections. Several cases will be considered including (i) a spherical, single-valley band, (ii) an anisotropic band, (iii) a band with multiple valleys, and (iv) the density of states in a semiconductor with reduced degrees of freedom such as quantum wells, quantum wires, and quantum boxes. Finally the *effective* density of states will be calculated.

3.1.1 Single-valley, spherical, and parabolic band

The simplest band structure of a semiconductor consists of a single valley with an isotropic (i.e. spherical), parabolic dispersion relation. This situation is closely approximated by, for example, the conduction band of GaAs. The electronic density of states is defined as the number of electron states per unit volume and per unit energy. The finiteness of the density of states is a result of the ***Pauli principle***, which states that only two electrons of opposite spin can occupy one volume element in phase space. The ***phase space*** is defined as a six-dimensional space composed of real space and momentum space. We now define a 'volume' element in phase space to consist of a range of positions and momenta of a particle, such that the position and momentum of the particle are distinguishable from the positions and momenta of other particles. In order to be distinguishable, the range of positions and momenta must be equal or exceed the range given by the ***uncertainty relation***. The volume element in phase space is then given by

$$\Delta x \Delta y \Delta z \Delta p_x \Delta p_y \Delta p_z = (2\pi\hbar)^3 . \tag{3.6}$$

The 'volume' element in phase space is $(2\pi\hbar)^3$. For systems with only one

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

degree of freedom, Eq. (3.6) reduces to the one-dimensional Heisenberg uncertainty principle $\Delta x \Delta p_x = 2\pi\hbar$. The Pauli principle states that two electrons of opposite spin occupy a 'volume' of $(2\pi\hbar)^3$ in phase space. Using the de Broglie relation ($p = \hbar k$) the 'volume' of phase space can be written as

$$\Delta x \Delta y \Delta z \Delta k_x \Delta k_y \Delta k_z = (2\pi)^3 \,. \tag{3.7}$$

The ***density of states*** per unit energy and per unit volume, which is denoted by $\rho_{DOS}(E)$, allows us to determine the total number of states per unit volume in an energy band with energies E_1 and E_2 according to

$$N = \int_{E_1}^{E_2} \rho_{DOS}(E)\, dE \,. \tag{3.8}$$

Note that N is the total number of states per unit volume, and $\rho_{DOS}(E)$ is the density of states per unit energy *and* per unit volume. To obtain the density of states per unit energy dE, we have to determine how much unit-volumes of k-space is contained in the energy interval E and $E + dE$, since we already know that one unit volume of k-space can contain two electrons of opposite spin.

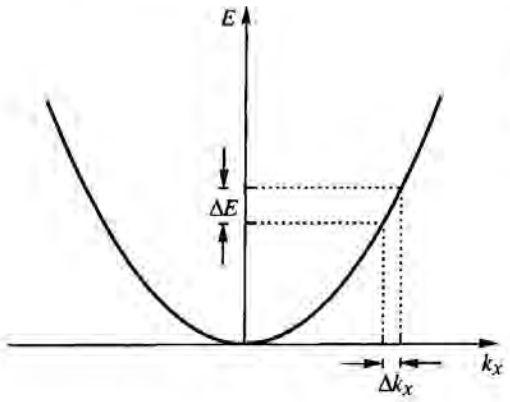


Fig. 3.1. Parabolic dispersion relation with a k-space interval Δk_x and a corresponding energy interval $\Delta E = (\partial E/\partial k_x)\Delta k_x$.

In order to obtain the volume of k-space included between two energies, the *dispersion relation* will be employed. A one-dimensional, parabolic dispersion relation $E = E(k_x)$ is shown in Fig. 3.1. For a given dE one can easily determine the corresponding length in k-space, as illustrated in Fig. 3.1. The k-space length associated with an energy interval dE is simply given by the slope of the dispersion relation. While the one-dimensional dispersion relations can be illustrated easily, the three-dimensional dispersion relation cannot be illustrated in three-dimensional space. To circumvent this difficulty, *surfaces of constant*

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

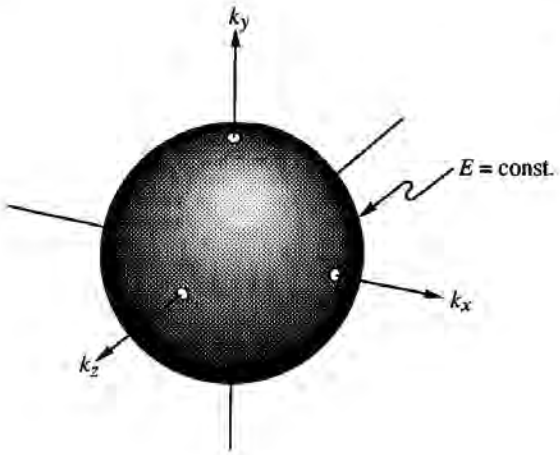


Fig. 3.2. Constant energy surface for a single valley, isotropic band.

energy in k-space are frequently used to illustrate a three-dimensional dispersion relation. As an example, the constant energy surface in k-space is illustrated in Fig. 3.2 for a spherical, single-valley band. A large separation of the constant energy surfaces, i.e. a large Δk for a given ΔE, indicates a weakly curved dispersion and a large effective mass.

In order to obtain the volume of k-space enclosed between two constant energy surfaces, which correspond to energies E and $E + \mathrm{d}E$, we (first) determine $\mathrm{d}k$ associated with $\mathrm{d}E$ and (second) integrate over the entire constant energy surface. The 'volume' of k-space enclosed between the two constant energy surface is thus given by

$$V_{k\text{-space}}(E) = \mathrm{d}E \int_{\text{Surface}} \frac{\partial k}{\partial E(k)}\,\mathrm{d}s \tag{3.9}$$

where $\mathrm{d}s$ is an area element of the constant energy surface. In a three-dimensional k-space we use $\nabla_k = (\partial/\partial k_x, \partial/\mathrm{d}k_y, \partial/\mathrm{d}k_z)$ and obtain

$$V_{k\text{-space}}(E) = \mathrm{d}E \int_{\text{Surface}} \frac{\mathrm{d}s}{\nabla_k \mathrm{E}(k)}\,. \tag{3.10}$$

Since an electron requires a volume of $4\pi^3$ in phase space, the number of states per unit volume is given by

$$N(E) = \frac{1}{4\pi^3}\,\mathrm{d}E \int_{\text{Surface}} \frac{\mathrm{d}s}{\nabla_k E(k)}\,. \tag{3.11}$$

Finally, we obtain the density of states per unit energy and unit volume according to

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

$$\rho_{\mathrm{DOS}}(E) = \frac{1}{4\pi^3} \int_{\mathrm{Surface}} \frac{\mathrm{d}s}{\nabla_k E(k)} . \tag{3.12}$$

In this equation, the surface element $\mathrm{d}s$ is always perpendicular to the vector $\nabla_k E(k)$. Note that the surface element $\mathrm{d}s$ is in k-space and that $\mathrm{d}s$ has the dimension m^{-2}.

Next we apply the expression for the density of states to *isotropic parabolic* dispersion relations of a three-dimensional semiconductor. In this case the surface of constant energy is a sphere of area $4\pi k^2$ and the parabolic dispersion is $E = \hbar^2 k^2/2m^* + E_{\mathrm{pot}}$ where k is the wave vector. Insertion of the dispersion in Eq. (3.12) yields the density of states in a semiconductor with a single-valley, isotropic, and parabolic band

$$\boxed{\rho_{\mathrm{DOS}}^{\mathrm{3D}}(E) = \frac{1}{2\pi^2} \left(\frac{2m^*}{\hbar^2}\right)^{3/2} \sqrt{E - E_{\mathrm{pot}}}} \tag{3.13}$$

where E_{pot} is a potential energy such as the conduction band edge or the valence band edge energy, E_C or E_V, respectively.

3.1.2 Single-valley, anisotropic, parabolic band

In an anisotropic single-valley band, the dispersion relation depends on the spatial direction. Such an anisotropic dispersion is found in III–V semiconductors in which the L- or X- point of the Brillouin zone is the lowest minimum, for example in GaP or AlAs. The surface of constant energy is then no longer a sphere, but an ellipsoid, as shown in Fig. 3.3. The three main axes of the ellipsoid may have different lengths, and thus the three dispersion relations are curved differently. If the main axes of the ellipsoid align with a cartesian coordinate system, the dispersion relation is

$$E = \frac{\hbar^2 k_x^2}{2m_x^*} + \frac{\hbar^2 k_y^2}{2m_y^*} + \frac{\hbar^2 k_z^2}{2m_z^*} . \tag{3.14}$$

The vector $\nabla_k E$ is given by $\nabla_k E = (\hbar^2 k_x/m_x^*, \hbar^2 k_y/m_y^*, \hbar^2 k_z/m_z^*)$. Since the vector $\nabla_k E$ is perpendicular on the surface element, the *absolute* values of $\mathrm{d}s$ and $\nabla_k E$ can be taken for the integration. Integration of Eq. (3.12) with the dispersion relation of Eq. (3.14) yields the density of states in an anisotropic semiconductor with parabolic dispersion relations, i.e.

$$\rho_{\mathrm{DOS}}(E) = \frac{\sqrt{2}}{\pi^2 \hbar^3} \sqrt{m_x^* m_y^* m_z^*} \sqrt{E - E_{\mathrm{pot}}} . \tag{3.15}$$

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

If the main axes of the constant-energy ellipsoid do not align with the k_x, k_y, and k_z axes of the coordinate system then m_x^*, m_y^*, and m_z^* can be formally replaced by m_1^*, m_2^*, and m_3^*.

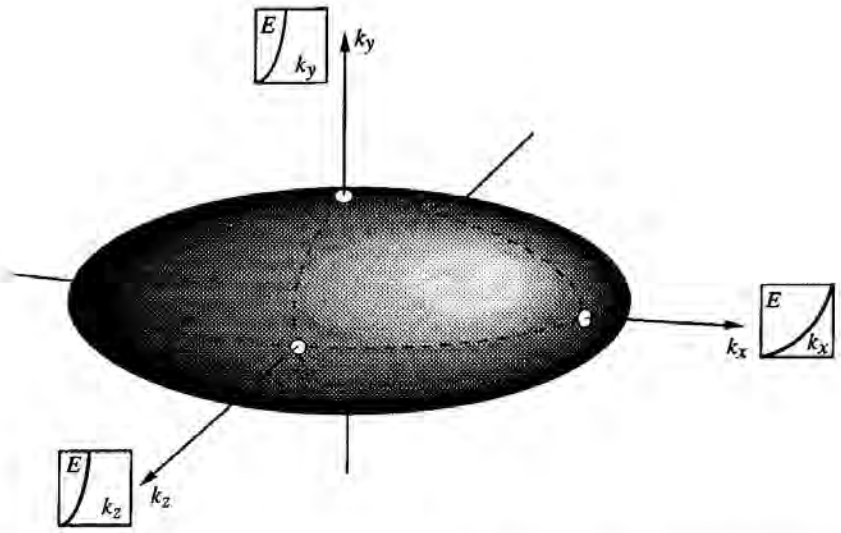


Fig. 3.3. Ellipsoidal constant energy surface with a weakly curved dispersion relation along the k_x axis and strongly curved dispersion along the k_y and k_z axis.

Frequently, the constant energy surfaces are rotational ellipsoids, that is, two of the main axes of the ellipsoid are identical. The axes are then denoted as the transversal and the longitudinal axes for the short and long axes, respectively. Such a rotational ellipsoid is schematically shown in Fig. 3.3. A relatively light mass is associated with the (short) transversal axis, while a relatively heavy mass is associated with the (long) longitudinal axis. If the masses are denoted as m_t^* and m_l^* for the transversal and the longitudinal mass, respectively, Eq. (3.15) can be modified according to

$$\rho_{\mathrm{DOS}}(E) = \frac{\sqrt{2}}{\pi^2 \hbar^3} \sqrt{m_l^* m_t^{*2}} \sqrt{E - E_{\mathrm{pot}}} \,. \tag{3.16}$$

The anisotropic masses m_x^*, m_y^*, m_z^*, m_l^*, and m_t^* are frequently used to define a ***density-of-states effective mass***. This mass is given by

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

337-TA-1366 RX-0255

$$m^*_{DOS} = (m^*_x m^*_y m^*_z)^{1/3} \tag{3.17a}$$

$$m^*_{DOS} = (m^2_t m_l)^{1/3} \tag{3.17b}$$

The density of states is then given by

$$\rho_{DOS}(E) = \frac{\sqrt{2}}{\pi^2 \hbar^3} (m^*_{DOS})^{3/2} \sqrt{E - E_{pot}} \,. \tag{3.18}$$

Note that for isotropic semiconductors the effective mass coincides with the density-of-states effective mass.

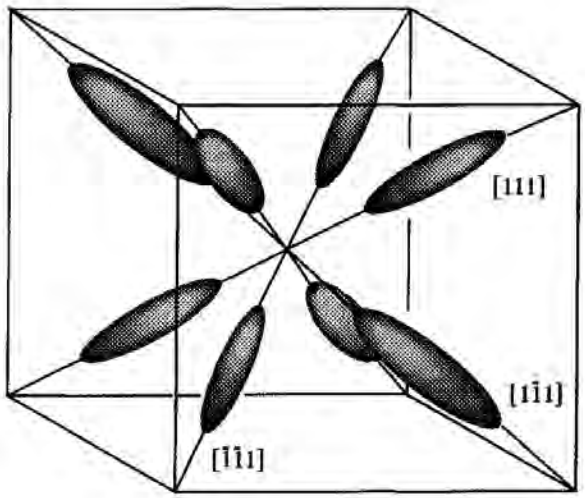


Fig. 3.4. Constant energy surface for the L-point of the Brillouin zone. The band structure consists of eight equivalent rotational ellipsoids.

3.1.3 Multiple valleys

At several points of the Brillouin zone, several equivalent minima occur. For example, eight equivalent minima occur at the L-point as schematically shown in Fig. 3.4. Each of the valleys can accommodate carriers, since the minima occur at different k_x, k_y, and k_z values, i.e. the Pauli principle is not violated. The density of states is thus obtained by multiplication with the number of equivalent minima, that is

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

RX-0255.108

$$\rho_{DOS}(E) = \frac{M_c\sqrt{2}}{\pi^2\hbar^3}\sqrt{m_1^* m_2^* m_3^*}\sqrt{E - E_{pot}} \tag{3.19}$$

where M_c is the number of equivalent minima and m_1^*, m_2^*, and m_3^* are the effective masses for motion along the three main axes of the ellipsoid.

3.1.4 Reduced spatial dimensions

Semiconductor heterostructure allows one to change the band energies in a controlled way and confine charge carriers to two (2D), one (1D), or zero (0D) spatial dimensions. Due to the confinement of carriers, the dispersion relation along the confinement direction is changed. The change in dispersion relation results in a change in the density of states.

Confinement of a carrier in one spatial dimension, e.g. the z-direction results in the formation of quantum states for motion along this direction. Consider the ground state in a quantum well of width L_z with infinitely high walls. The ground-state energy is obtained from the solution of Schrödinger's equation and is given by

$$E_0 = \frac{\hbar^2}{2m^*}\left(\frac{\pi}{L_z}\right)^2 . \tag{3.20}$$

The particle in the quantum well can assume a range of momenta in the z-direction; the range is given by the uncertainty principle, i.e.

$$\Delta k_z = \Delta p_z/\hbar = 2\pi/L_z . \tag{3.21}$$

The dispersion relation for motion along the confinement (z-) direction is thus given by

$$E = E_0 \quad \text{for the entire range of } k_z . \tag{3.22}$$

The dispersion is flat, i.e. constant for all values of k_z. The z-component of the vector $\nabla_k E$ (see Eq. (3.12)) is therefore zero and need not be considered.

We next consider the x- and y- direction and recall that the Schrödinger equation is separable for the three spatial dimensions. Thus, the kinetic energy in the xy-plane is given by

$$E = \frac{\hbar^2}{2m^*}(k_x^2 + k_y^2) \tag{3.23}$$

for a parabolic dispersion.

The surface of constant energy for the dispersion relation given by Eq. (3.23) is shown in Fig. 3.5, and is a circle around $k_x = k_y = 0$. The density of states of such a 2D electron system is obtained by similar considerations as for the 3D

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

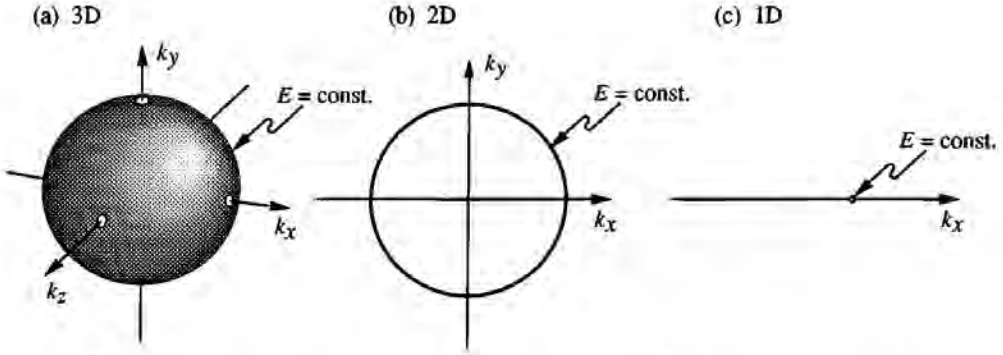


Fig. 3.5. Constant energy surfaces of a (a) 3-dimensional, (b) 2-dimensional, and (c) 1-dimensional system. The surfaces are a sphere, a circle, and a point for 3D, 2D, and 1D systems, respectively.

case. The reduced phase space now consists only of the xy-plane and the k_x and k_y coordinates. Correspondingly, the two-dimensional density of states is the number of states per *unit-area* and unit-energy. The volume of k-space between the circles of constant energy is given by Eq. (3.10). The equation is evaluated most conveniently in polar coordinates in which $k_r = (k_x^2 + k_y^2)^{1/2}$ is the radial component of the k-vector. The surface integral reduces to a line integral and the total length of the circular line is $2\pi k_r$. The volume of k-space then obtained is

$$V_{k\text{-space}}^{2D}(E) = dE \int_{\text{Surface}} \frac{ds}{\nabla_k E(k)} = \frac{2\pi m^*}{\hbar^2}. \tag{3.24}$$

Since two (one) electrons of opposite spin require a volume element of $(2\pi)^2$ in phase space, the density of states of a 2D electron system is given by

$$\boxed{\rho_{DOS}^{2D}(E) = \frac{m^*}{\pi\hbar^2} \quad \text{for } E \geq E_0} \tag{3.25}$$

where E_0 is the ground state of the quantum well system. For energies $E \geq E_0$, the 2D density of states is a constant and does not depend on energy. If the 2D semiconductor has more than one quantum state, each quantum state has a state density of Eq. (3.25). The total density of states can be written as

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

$$\rho_{\mathrm{DOS}}^{\mathrm{2D}}(E) = \frac{m^*}{\pi\hbar^2} \sum_n \sigma(E - E_n) \tag{3.26}$$

where E_n are the energies of quantized states and $\sigma(E - E_n)$ is the step function.

We next consider a one-dimensional (1D) system, the quantum wire, in which only one direction of motion is allowed, e.g. along the x-direction. The dispersion relation is then given by $E = \hbar^2 k_x^2 / 2m^*$. The 'volume' (i.e. length-unit) in k-space is obtained in analogy to the three-dimensional and two-dimensional case according to Eq. (3.10). The 'surface' integral reduces to a single point in k-space, i.e. the point $k = k_x$. Thus, the volume of k-space is given by

$$V_{k\text{-space}}^{\mathrm{1D}}(E) = \int_{\text{Surface}} \frac{\delta(k_x - k_{x0})\,\mathrm{ds}}{\nabla_{\mathrm{k}} E(k_x)}$$

$$= \sqrt{\frac{m^*}{2\hbar^2(E - E_0)}} \quad \text{for } E \geq E_0 \,. \tag{3.27}$$

The volume in phase space of two electrons with opposite spin is given by 2π and thus the 1D density of states is given by

$$\boxed{\rho_{\mathrm{DOS}}^{\mathrm{1D}}(E) = \frac{1}{\pi\hbar} \sqrt{\frac{m^*}{2(E - E_0)}} \quad \text{for } E \geq E_0} \tag{3.28}$$

Note that the density of states in a 3-, 2- and 1-dimensional system has a functional dependence on energy according to $E^{1/2}$, E^0, and $E^{-1/2}$, respectively. For more than one quantized state, the 1D density of states is given by

$$\rho_{\mathrm{DOS}}^{\mathrm{1D}}(E) = \frac{1}{\pi\hbar} \sum_n \sqrt{\frac{m^*}{2\sigma(E - E_n)}} \tag{3.29}$$

where E_n are the energies of the quantized states of the wire.

Finally, we consider the density of states in a zero-dimensional (0D) system, the quantum box. No free motion is possible in such a quantum box, since the electron is confined in all three spatial dimensions. Consequently, there is no k-space available which could be filled up with electrons. Each quantum state of a 0D system can therefore be occupied by only two electrons. The density of states is therefore described by a δ-function.

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

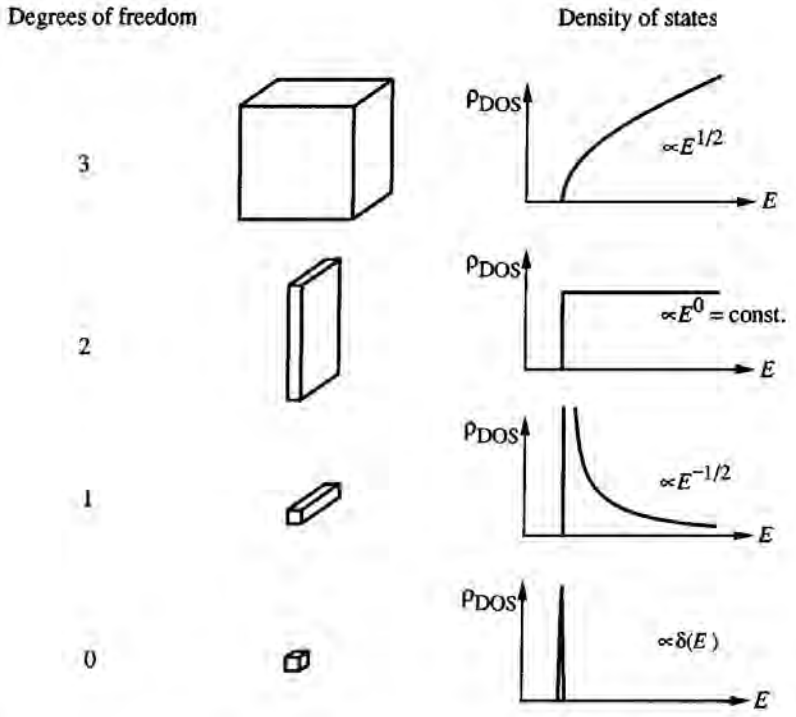


Fig. 3.6. Electronic density of states of semiconductors with 3, 2, 1, and 0 degrees of freedom for electron propagation. Systems with 2, 1, and 0 degrees of freedom are referred to as quantum wells, quantum wires, and quantum boxes, respectively.

$$\rho_{DOS}^{0D}(E) = 2\delta(E - E_0) \tag{3.30}$$

For more than one quantum state, the density of states is given by

$$\rho_{DOS}^{0D}(E) = \sum_n 2\delta(E - E_n) . \tag{3.31}$$

The densities of states for one quantized level for a 3D, 2D, 1D, and 0D electron system are schematically illustrated in Fig. 3.6.

3.1.5 Effective density of states

The ***effective density of states*** is introduced in order to simplify the calculation of the population of the conduction and valence band. The basic simplification made is that all band states are assumed to be located directly at the band edge. This situation is illustrated in Fig. 3.7 for the conduction band. The 3D density of

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

states has square-root dependence on energy. The effective density of states is δ-function-like and occurs at the bottom of the conduction band.

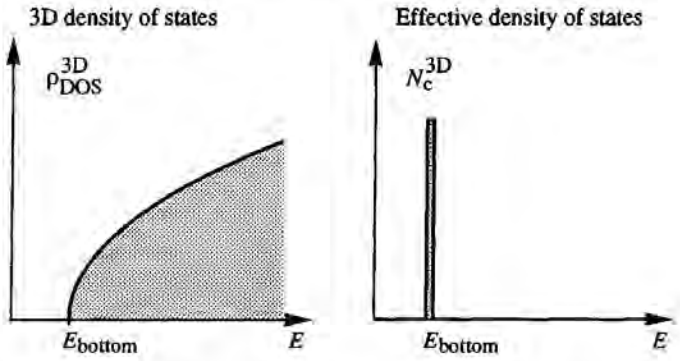


Fig. 3.7. Energy dependent density of states, ρ_{DOS}^{3D}, and *effective* density of states, N_c^{3D}, at the bottom of the conduction band.

An electronic state can be either occupied by an electron or unoccupied. Quantum mechanics allows us to attribute to the state a probability of occupation. The total electron concentration in a band is then obtained by integration over the product of state density and the probability that the state is occupied, that is

$$n = \int_{E_{bottom}}^{E_{top}} \rho_{DOS}(E) f(E)\, dE \tag{3.32}$$

where $f(E)$ is the (dimensionless) probability that a state of energy E is populated (see Sect. 3.2). The limits of the integration are the bottom and the top energy of the band, since the electron concentration in the entire band is of interest.

As will be seen in the next section, the probability of occupation, $f(E)$, is given by the Maxwell–Boltzmann distribution (the Maxwell–Boltzmann distribution applies only to non-degenerate situations which will be defined in the next section). The Maxwell–Boltzmann distribution, also frequently referred to as the Boltzmann distribution, is given by

$$f_B(E) = \exp\left[-\frac{E - E_F}{kT}\right] \tag{3.33}$$

where E_F is the Fermi energy (for a definition of the Fermi energy the reader is again referred to the next section). Using Eq. (3.32), the electron concentration can be determined by evaluating the integral.

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

The *effective* density of states at the bottom of the conduction band is now defined as the density of states which yields, with the Boltzmann distribution, the *same* electron concentration as the true density of states, that is

$$n = \int_{E_{\text{bottom}}}^{E_{\text{top}}} \rho_{\text{DOS}}(E)\, f_{\text{B}}(E)\, \mathrm{d}E = N_{\text{c}} f_{\text{B}}(E = E_{\text{C}}) \tag{3.34}$$

where N_{c} is the effective density of states at the bottom of the conduction band and E_{C} is the energy of the bottom of this band. Strictly speaking, the effective density of states has no physical meaning but is simply a mathematical tool to facilitate calculations. For completeness, Eqs. (3.32) and (3.34) are now given explicitly using the Boltzmann distribution and the density of states of an isotropic three-dimensional semiconductor:

$$n = \int_{E_{\text{C}}}^{\infty} \frac{1}{2\pi^2} \left(\frac{2m^*}{\hbar^2} \right)^{3/2} \sqrt{E - E_{\text{C}}}\; \mathrm{e}^{-(E - E_{\text{F}})/kT}\, \mathrm{d}E\,, \tag{3.35}$$

$$n = N_{\text{c}}\, \mathrm{e}^{-(E_{\text{C}} - E_{\text{F}})/kT}\,. \tag{3.36}$$

The upper limit of the integration can be taken to be infinity without loss of accuracy due to the strongly converging Boltzmann factor. Evaluation of the integral in Eq. (3.35) and comparison with Eq. (3.36) yields the effective density of states

$$\boxed{N_{\text{c}} = \frac{1}{\sqrt{2}} \left(\frac{m^* kT}{\pi \hbar^2} \right)^{3/2}} \tag{3.37}$$

Note that the effective density of states given by Eq. (3.37) applies to one minimum in the conduction band. If there are a number of M_{c} equivalent minima in the conduction band, the corresponding density of states must be multiplied by M_{c}. Furthermore, if the band structure is anisotropic, the effective mass m^* must be replaced by the density-of-states effective mass m^*_{DOS}. For a degenerate valence band with heavy and light holes, the effective density of states is the sum of both effective state densities, that is

$$N_{\text{v}} = \frac{1}{\sqrt{2}} \left(\frac{m^*_{\text{hh}} kT}{\pi \hbar^2} \right)^{3/2} + \frac{1}{\sqrt{2}} \left(\frac{m^*_{\text{lh}} kT}{\pi \hbar^2} \right)^{3/2}. \tag{3.38}$$

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

The effective density of states in a two-dimensional system (i.e. a system with two degrees of freedom) is obtained by the identical procedure as the three-dimensional effective density of states. The equations analogue to Eqs. (3.35) and (3.36) then read

$$n^{2D} = \int_{E_C}^{\infty} \frac{m^*}{\pi\hbar^2} e^{-(E - E_F)/kT} \, dE \,, \tag{3.39}$$

$$n^{2D} = N_c^{2D} e^{-(E_C - E_F)/kT} \tag{3.40}$$

where N_c^{2D} is the two-dimensional effective density of states. The carrier concentration n^{2D} represents the number of electrons per unit-area and is also referred to as the 2D density. Evaluation of the integral yields

$$\boxed{N_c^{2D} = \frac{m^*}{\pi\hbar^2} kT} \tag{3.41}$$

Finally, the effective density of states of a one-dimensional (1D) system is obtained in a similar way. The 1D density, i.e. the number of carriers per unit length is given by

$$n^{1D} = \int_{E_C}^{\infty} \frac{1}{\pi\hbar} \sqrt{\frac{m^*}{2(E - E_C)}} e^{-(E - E_F)/kT} \, dE \,, \tag{3.42}$$

$$n^{1D} = N_c^{1D} e^{-(E_C - E_F)/kT} \,. \tag{3.43}$$

The one-dimensional effective density of states is obtained as

$$\boxed{N_c^{1D} = \sqrt{\frac{m^* kT}{2\pi\hbar^2}}} \tag{3.44}$$

The evaluation of a zero-dimensional density of states does not yield a simplification of the carrier-density calculation, since the zero-dimensional density of states is δ-function like. Table 3.1 summarizes the dispersion relation, the density of states, and the effective density of states of semiconductors with various degrees of freedom.

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

Table 3.1. *Density of states for semiconductor with 3, 2, 1, and 0 degrees of freedom for propagation of electrons. The dispersion relations are assumed to be parabolic. The formulas can be applied to anisotropic semiconductors if the effective mass m^* is replaced by the density-of-states effective mass m^*_{DOS}. If the semiconductor has a number of M_c equivalent minima, the corresponding density of states must be multiplied by M_c. The bottom of the band is denoted as E_C and $\sigma(E)$ is the step-function.*

Degrees of freedom	Dispersion (kinetic energy)	Density of states	Effective density of states
3(bulk)	$E = \frac{\hbar^2}{2m^*}(k_x^2 + k_y^2 + k_z^2)$	$\rho_{DOS}^{3D} = \frac{1}{2\pi^2}\left[\frac{2m^*}{\hbar^2}\right]^{3/2}\sqrt{E - E_C}$	$N_c^{3D} = \frac{1}{\sqrt{2}}\left[\frac{m^* kT}{\pi\hbar^2}\right]^{3/2}$
2 (slab)	$E = \frac{\hbar^2}{2m^*}(k_x^2 + k_y^2)$	$\rho_{DOS}^{2D} = \frac{m^*}{\pi\hbar^2}\sigma(E - E_C)$	$N_c^{2D} = \frac{m^*}{\pi\hbar^2}kT$
1 (wire)	$E = \frac{\hbar^2}{2m^*}k_x^2$	$\rho_{DOS}^{1D} = \frac{1}{\pi\hbar}\sqrt{\frac{m^*}{2(E - E_C)}}$	$N_c^{1D} = \sqrt{\frac{m^* kT}{2\pi\hbar^2}}$
0 (box)	—	$\rho_{DOS}^{0D} = 2\delta(E - E_C)$	$N_c^{0D} = 2$

Innoscience1366-00138789

3.2 Classical and quantum statistics

Typical densities of free carriers in semiconductors range from 10^{15} cm^{-3} to 10^{20} cm^{-3}. It is impossible to describe the energies or velocities of those carriers individually. An alternative to the individual characterization of particles is the *statistical* description of a carrier system. The statistical description uses *probabilities* of velocities or energies rather than knowing these quantities for all individual carriers. Thus, the statistical treatment represents a simplification. The derivation of the energy distribution function treats the carrier system as an *ideal gas*, for example a gas of oxygen molecules. The ideal gas is assumed to have only *elastic* collisions between atoms or molecules. Furthermore, the energy of the gas molecules is assumed to be purely translational *kinetic*. Since these properties are applied to the electron or hole system, those systems are frequently referred to as *electron-gases* or *hole-gases*.

Semiconductor statistics includes both classical statistics and quantum statistics. Classical or Maxwell–Boltzmann statistics is derived on the basis of purely classical physics arguments. In contrast, quantum statistics takes into account two results of quantum mechanics, namely (i) the Pauli exclusion principle which limits the number of electrons occupying a state of energy E and (ii) the finiteness of the number of states in an energy interval E and $E + dE$. The finiteness of states is a result of the Schrödinger equation. In this section, the basic concepts of classical statistics and quantum statistics are derived. The fundamentals of ideal gases and statistical distributions are summarized as well since they are the basis of semiconductor statistics.

3.2.1 Probability and distribution functions

Consider a large number N of free classical particles such as atoms, molecules or electrons which are kept at a constant temperature T, and which interact only weakly with one another. The energy of a single particle consists of *kinetic energy* due to translatory motion and an internal energy for example due to rotations, vibrations, or orbital motions of the particle. In the following we consider particles with only kinetic energy due to translatory motion. The particles of the system can assume an energy E, where E can be either a discrete or a continuous variable. If N_i particles out of N particles have an energy between E_i and $E_i + dE$, the probability of any particle having any energy within the interval E_i and $E_i + dE$ is given by

$$f(E_i)\, dE = \frac{N_i}{N} \tag{3.45}$$

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

where $f(E)$ is the ***energy distribution function*** of a particle system. In statistics, $f(E)$ is frequently called the *probability density function.* The total number of particles is given by

$$\sum_i N_i = N \tag{3.46}$$

where the sum is over all possible energy intervals. Thus, the integral over the energy distribution function is

$$\int_0^\infty f(E)\, dE = \sum_i \frac{N_i}{N} = 1 . \tag{3.47}$$

In other words, the probability of any particle having an energy between zero and infinity is unity. Distribution functions which obey

$$\int_0^\infty f(E)\, dE = 1 \tag{3.48}$$

are called ***normalized*** distribution functions.

The ***average energy*** or ***mean energy*** $\bar{E}$ of a single particle is obtained by calculating the total energy and dividing by the number of particles, that is

$$\bar{E} = \frac{1}{N} \sum_i N_i E = \int_0^\infty E f(E)\, dE . \tag{3.49}$$

In addition to energy distribution functions, velocity distribution functions are valuable. Since only the kinetic translatory motion (no rotational motion) is considered, the velocity and energy are related by

$$E = \tfrac{1}{2} m v^2 . \tag{3.50a}$$

The average velocity and the average energy are related by

$$\bar{E} = \tfrac{1}{2} m \overline{v^2} \tag{3.50b}$$

where $\overline{v^2}$ is called the *mean square velocity.* The *root-mean-square* (rms) velocity is defined as

$$v_{\text{rms}} = \sqrt{\overline{v^2}} \tag{3.50c}$$

and is the velocity corresponding to the average energy

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

$$\bar{E} = \tfrac{1}{2} m v_{\text{rms}}^2 \,. \tag{3.50d}$$

In analogy to the energy distribution we assume that N_i particles have a velocity within the interval v_i and $v_i + \mathrm{d}v$. Thus,

$$f(v)\,\mathrm{d}v = \frac{N_i}{N} \tag{3.51}$$

where $f(v)$ is the normalized velocity distribution. Knowing $f(v)$, the following relations allow one to calculate the mean velocity, the mean square velocity, and the root-mean-square velocity

$$\bar{v} = \int_0^\infty v f(v)\ \mathrm{d}v \,, \tag{3.52a}$$

$$\overline{v^2} = \int_0^\infty v^2 f(v)\ \mathrm{d}v \,, \tag{3.52b}$$

$$v_{\text{rms}} = \sqrt{\overline{v^2}} = \left[\int_0^\infty v^2 f(v)\ \mathrm{d}v\right]^{1/2} \,. \tag{3.52c}$$

Up to now we have considered the velocity as a scalar. A more specific description of the velocity distribution is obtained by considering each component of the velocity $v = (v_x, v_y, v_z)$. If N_i particles out of N particles have a velocity in the 'volume' element $v_x + \mathrm{d}v_x$, $v_y + \mathrm{d}v_y$, and $v_z + \mathrm{d}v_z$, the distribution function is given by

$$f(v_x, v_y, v_z)\,\mathrm{d}v_x\mathrm{d}v_y\mathrm{d}v_z = \frac{N_i}{N} \,. \tag{3.53}$$

Since $\sum_i N_i = N$, the velocity distribution function is normalized, i.e.

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(v_x, v_y, v_z)\,\mathrm{d}v_x\mathrm{d}v_y\mathrm{d}v_z = 1 \,. \tag{3.54}$$

The average of a specific propagation direction, for example v_x is evaluated in analogy to Eqs. (3.52a–c). One obtains

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

$$\bar{v}_x = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} v_x f(v_x, v_y, v_z)\,dv_x dv_y dv_z\,, \tag{3.55a}$$

$$\overline{v_x^2} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} v_x^2 f(v_x, v_y, v_z)\,dv_x dv_y dv_z\,, \tag{3.55b}$$

$$v_{x,\mathrm{rms}} = \sqrt{\overline{v_x^2}} = \left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} v_x^2 f(v_x, v_y, v_z)\,dv_x dv_y dv_z\right]^{1/2}. \tag{3.55c}$$

In a closed system the mean velocities are zero, that is $\bar{v}_x = \bar{v}_y = \bar{v}_z = 0$. However, the mean square velocities are, just as the energy, not equal to zero.

3.2.2 Ideal gases of atoms and electrons

The basis of classical semiconductor statistics is ideal gas theory. It is therefore necessary to make a small excursion into this theory. The individual particles in such ideal gases are assumed to interact weakly, that is collisions between atoms or molecules are a relatively seldom event. It is further assumed that there is no interaction between the particles of the gas (such as electrostatic interaction), unless the particles collide. The collisions are assumed to be (i) *elastic* (i.e. total energy and momentum of the two particles involved in a collision are preserved) and (ii) of very short duration.

Ideal gases follow the universal gas equation (see e.g. Kittel and Kroemer, 1980)

$$PV = RT \tag{3.56}$$

where P is the pressure, V the volume of the gas, T its temperature, and R is the universal gas constant. This constant is independent of the species of the gas particles and has a value of $R = 8.314\ \mathrm{J\ K^{-1}\ mol^{-1}}$.

Next, the pressure P and the kinetic energy of an individual particle of the gas will be calculated. For the calculation it is assumed that the gas is confined to a cube of volume V, as shown in Fig. 3.8. The quantity of the gas is assumed to be 1 mole, that is the number of atoms or molecules is given by Avogadro's number, $N_{\mathrm{Avo}} = 6.023 \times 10^{23}$ particles per mole. Each side of the cube is assumed to have an area $A = V^{2/3}$. If a particle of mass m and momentum mv_x (along the x-direction) is elastically reflected from the wall, it provides a momentum $2mv_x$ to reverse the particle momentum. If the duration of the collision with the wall is dt, then the force acting on the wall during the time dt is given by

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

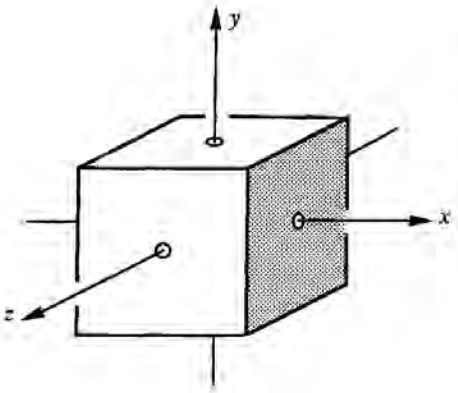


Fig. 3.8. Cubic volume confining one mole ($N_{Avo} = 6.023 \times 10^{23}$ atoms/mole) of an ideal gas. The pressure of the ideal gas exerted on one side of the cube (shaded area) is given by Eq. (3.59).

$$F = \frac{dp}{dt} \tag{3.57a}$$

where the momentum change is $dp = 2mv_x$. The pressure P on the wall during the collision with one particle is given by

$$dP = \frac{F}{A} = \frac{1}{A}\frac{dp}{dt} \tag{3.57b}$$

where A is the area of the cube's walls. Next we calculate the total pressure P experienced by the wall if a number of N_{Avo} particles are within the volume V. For this purpose we first determine the number of collisions with the wall during the time dt. If the particles have a velocity v_x, then the number of particles hitting the wall during dt is $(N_{Avo}/V)Av_x dt$. The fraction of particles having a velocity v_x is obtained from the velocity distribution function and is given by $f(v_x, v_y, v_z)\,dv_x\,dv_y\,dv_z$. Consequently, the total pressure is obtained by integration over all positive velocities in the x-direction

$$P = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty} \frac{N_{Avo}}{V} Av_x\,dt\,f(v_x,v_y,v_z)\,dv_x\,dv_y\,dv_z\,\frac{2mv_x}{A\,dt}. \tag{3.58a}$$

Since the velocity distribution is symmetric with respect to positive and negative x-direction, the integration can be expanded from $-\infty$ to $+\infty$

$$P = \frac{N_{Avo}}{V} m \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} v_x^2 f(v_x, v_y, v_z)\,dv_x\,dv_y\,dv_z = \frac{N_{Avo}}{V} m\,\overline{v_x^2}. \tag{3.58b}$$

Since the velocity distribution is isotropic, the mean square velocity is given by

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

$$\overline{v^2} = \overline{v_x^2} + \overline{v_y^2} + \overline{v_z^2} \quad \text{or} \quad \overline{v_x^2} = \frac{1}{3}\overline{v^2} . \tag{3.58c}$$

The pressure on the wall is then given by

$$P = \frac{1}{3}\overline{v^2}\frac{N_{\text{Avo}}}{V} m . \tag{3.59}$$

Using the universal gas equation, Eq. (3.56), one obtains

$$RT = \frac{2}{3} N_{\text{Avo}} \frac{1}{2} m\overline{v^2} . \tag{3.60}$$

The average kinetic energy of one mole of the ideal gas can then be written as

$$\bar{E} = \overline{E_{\text{kin}}} = \frac{3}{2} RT . \tag{3.61}$$

The average kinetic energy of one single particle is obtained by division by the number of particles, i.e.

$$\boxed{\bar{E} = \bar{E}_{\text{kin}} = \frac{3}{2} kT} \tag{3.62}$$

where $k = R/N_{\text{Avo}}$ is the Boltzmann constant. The preceding calculation has been carried out for a three-dimensional space. In a one-dimensional space (one degree of freedom), the average velocity is $\overline{v^2} = \overline{v_x^2}$ and the resulting kinetic energy is given by

$$\overline{E_{\text{kin}}} = \tfrac{1}{2}kT \text{ (per degree of freedom)} . \tag{3.63}$$

Thus the kinetic energy of an atom or molecule is given by $\frac{1}{2}kT$. Equation (3.63) is called the ***equipartition law***, which states that each 'degree of freedom' contributes $\frac{1}{2}kT$ to the total kinetic energy.

Next we will focus on the energetic distribution of electrons. The properties which have been derived in this section for atomic or molecular gases will be applied to free electrons of effective mass m^* in a crystal. To do so, the interaction between the electrons and the lattice must be negligible and electron–electron collisions must be a relatively seldom event. Under these circumstances we can treat the electron system as a classical ideal gas.

3.2.3 Maxwell velocity distribution

The Maxwell velocity distribution describes the distribution of velocities of the particles of an ideal gas. It will be shown that the Maxwell velocity distribution

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

is of the form

$$f_M(v) = A \exp\left[-\frac{\frac{1}{2}mv^2}{kT}\right] \tag{3.64a}$$

where $\frac{1}{2} mv^2$ is the kinetic energy of the particles. If the energy of the particles is purely kinetic, the Maxwell distribution can be written as

$$f_M(E) = A \exp\left[-\frac{E}{kT}\right] . \tag{3.64b}$$

The proof of the Maxwell distribution of Eq. (3.64) is conveniently done in two steps. In the first step, the exponential factor is demonstrated, i.e. $f_M(E) = A \exp(-\alpha E)$. In the second step it is shown that $\alpha = 1/kT$.

In the theory of ideal gases it is assumed that collisions between particles are elastic. The total energy of two electrons before and after a collision remains the same, that is

$$E_1 + E_2 = E_1' + E_2' \tag{3.65}$$

where E_1 and E_2 are the electron energies before the collision and E_1' and E_2' are the energies after the collision. The probability of a collision of an electron with energy E_1 and of an electron with energy E_2 is proportional to the probability that there is an electron of energy E_1 and a second electron with energy E_2. If the probability of such a collision is p, then

$$p = B f_M(E_1) f_M(E_2) \tag{3.66}$$

where B is a constant. The same consideration is valid for particles with energies E_1' and E_2'. Thus, the probability that two electrons with energies E_1' and E_2' collide is given by

$$p' = B f_M(E_1') f_M(E_2') . \tag{3.67}$$

If the change in energy before and after the collision is ΔE, then $\Delta E = E_1' - E_1$ and $\Delta E = E_2 - E_2'$. Furthermore, if the electron gas is in equilibrium, then $p = p'$ and one obtains

$$f_M(E_1) f_M(E_2) = f_M(E_1 + \Delta E) f_M(E_2 - \Delta E) . \tag{3.68}$$

Only the exponential function satisfies this condition, that is

$$f_M(E) = A \exp(-\alpha E) \tag{3.69}$$

where α is a positive yet undetermined constant. The exponent is chosen negative

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

to assure that the occupation probability decreases with higher energies. It will become obvious that α is a universal constant and applies to all carrier systems such as electron-, heavy- or light-hole systems.

Next, the constant α will be determined. It will be shown that $\alpha = 1/kT$ using the results of the ideal gas theory. The energy of an electron in an ideal gas is given by

$$E = \tfrac{1}{2}\, mv^2 = \tfrac{1}{2}m(v_x^2 + v_y^2 + v_z^2)\,. \tag{3.70}$$

The exponential energy distribution of Eq. (3.69) and the normalization condition of Eq. (3.54) yield the normalized velocity distribution

$$f(v_x, v_y, v_z) = \left(\frac{m\alpha}{2\pi}\right)^{3/2} \exp\left[-\tfrac{1}{2}\, m\alpha(v_x^2 + v_y^2 + v_z^2)\right]\,. \tag{3.71}$$

The average energy of an electron is obtained by (first) calculating the mean square velocities, $\overline{v_x^2}, \overline{v_y^2}, \overline{v_z^2}$ from the distribution and (second) using Eq. (3.70) to calculate E from the mean square velocities. One obtains

$$E = (3/2)\ \alpha^{-1}\,. \tag{3.72}$$

We now use the result from classic gas theory which states according to Eq. (3.62) that the kinetic energy equals $E = (3/2)kT$. Comparison with Eq. (3.72) yields

$$\alpha = (kT)^{-1} \tag{3.73}$$

which concludes the proof of the Maxwell distribution of Eq. (3.64).

Having determined the value of α, the explicit form of the normalized ***maxwellian velocity distribution*** in cartesian coordinates is

$$\boxed{f_M(v_x, v_y, v_z) = \left(\frac{m}{2\pi kT}\right)^{3/2} \exp\left[-\frac{\tfrac{1}{2}m(v_x^2 + v_y^2 + v_z^2)}{kT}\right]} \tag{3.74a}$$

Due to the spherical symmetry of the maxwellian velocity distribution, it is useful to express the distribution in spherical coordinates. For the coordinate transformation we note that $f_M(v_x, v_y, v_z)\,dv_x dv_y dv_z = f_M(v)\,dv$, and that a volume element $dv_x dv_y dv_z$ is given by $4\pi v^2 dv$ in spherical coordinates. The maxwellian velocity distribution in spherical coordinates is then given by

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

337-TA-1366 RX-0255

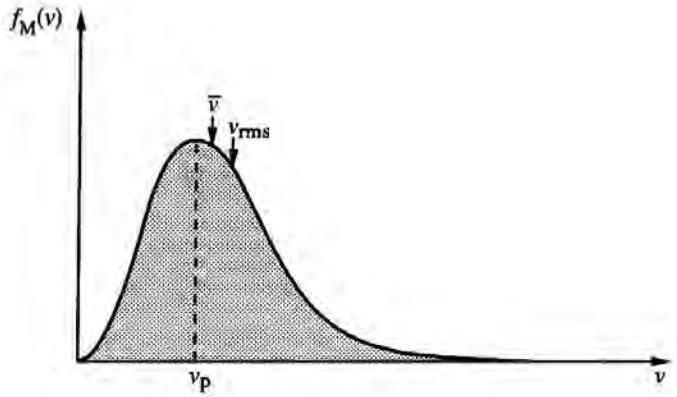


Fig. 3.9. Schematic maxwellian velocity distribution $f_M(v)$ of an ideal elecron gas. The velocity with the highest probability, v_p, is lower than the mean velocity, $\bar{v}$, and the root-mean-square velocity, v_{rms}.

$$f_M(v) = \left(\frac{m}{2\pi kT}\right)^{3/2} (4\pi v^2) \exp\left[-\frac{\frac{1}{2}mv^2}{kT}\right]. \tag{3.74b}$$

The maxwellian velocity distribution is shown in Fig. 3.9. The peak of the distribution, that is the most likely velocity, is $v_p = (2kT/m)^{1/2}$. The mean velocity is given by $\bar{v} = (8kT)/(\pi m)^{1/2}$. The root-mean-square velocity can only be obtained by numerical integration.

3.2.4 The Boltzmann factor

The maxwellian velocity distribution can be changed to an energy distribution by using the substitution $E = \frac{1}{2}mv^2$. Noting that the energy interval and the velocity interval are related by $dE = mvdv$ and that the number of electrons in the velocity interval, $f_M(v)\,dv$, is the same as the number of electrons in the energy interval, $f_{MB}(E)\,dE$, then the energy distribution is given by

$$\boxed{f_{MB}(E) = \frac{2}{\sqrt{\pi}} \frac{\sqrt{E}}{(kT)^{3/2}} e^{-E/kT}} \tag{3.75}$$

which is the ***Maxwell–Boltzmann distribution***.

For large energies, the exponential term in the Maxwell–Boltzmann distribution essentially determines the energy dependence. Therefore, the high-

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

RX-0255.125

337-TA-1366 RX-0255

energy approximation of the Maxwell–Boltzmann distribution is

$$f_B(E) = A\, e^{-E/kT} \tag{3.76}$$

which is the ***Boltzmann distribution***. The exponential factor of the distribution, $\exp(-E/kT)$, is called the ***Boltzmann factor*** or ***Boltzmann tail***. The Boltzmann distribution does not take into account the quantum mechanical properties of an electron gas. The applicability of the distribution is therefore limited to the classical regime, i.e. for $E >> kT$.

3.2.5 The Fermi–Dirac distribution

In contrast to classical Boltzmann statistics, the quantum mechanical characteristics of an electron gas are taken into account in Fermi–Dirac statistics. The quantum properties which are explicitly taken into account are

1. The *wave character* of electrons. Due to the wave character of electrons the Schrödinger equation has only a *finite number of solutions* in the energy interval E and $E + dE$.
2. The *Pauli principle* which states that an eigenstate can be occupied by only two electrons of opposite spin.

Since the Pauli principle strongly restricts the number of carriers per energy level, higher states are populated even at zero temperature. This situation is illustrated in Fig. 3.10, where two electron distributions are illustrated at zero temperature. The distribution in Fig. 3.10(a) does not take into account the Pauli principle while that in Fig. 3.10(b) does.

The first restriction imposed by quantum mechanics is the *finiteness of states* within an energy interval E and $E + dE$. The finiteness of states played a role in the derivation of the density of states, which was determined in Sect. 3.1. The density of states in an isotropic semiconductor was shown to be

$$\rho_{DOS}(E) = \frac{1}{2\pi^2}\left[\frac{2m^*}{\hbar^2}\right]^{3/2}\sqrt{E} \tag{3.77}$$

where E is the kinetic energy. Note that for the derivation of the density of states the Pauli principle has been taken into account. Therefore, the states given by Eq. (3.77) can be occupied only by *one* electron. Since the number of states per velocity-interval will be of interest, Eq. (3.77) is modified using $E = \frac{1}{2}mv^2$ and $dE = mvdv$. Note that the number of states per energy interval dE is the same as

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

RX-0255.126

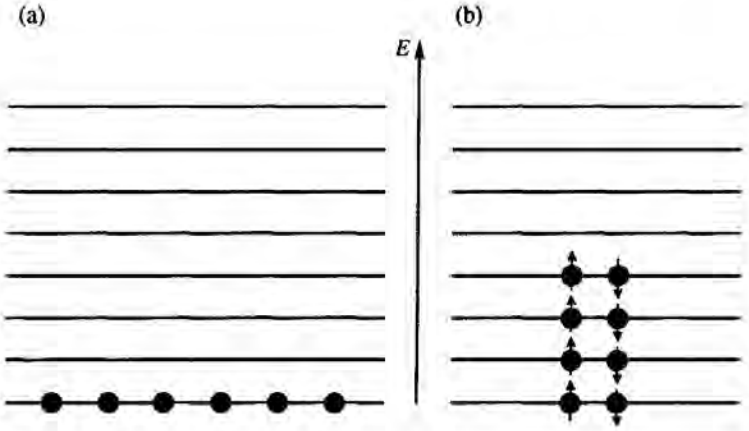


Fig. 3.10. Distribution of electrons at zero temperature among discrete energy levels (a) without Pauli principle and (b) with Pauli principle and spin taken into account. Spin 'up' and 'down' is illustrated by arrows.

the number of states per velocity interval dv, i.e. $\rho_{DOS}(E)\,dE = \rho_{DOS}(v)\,dv$. The number of states per velocity interval (and per unit volume) is then given by

$$\rho_{DOS}(v) = \frac{m^3}{\pi^2\hbar^3}v^2 \tag{3.78}$$

for an isotropic semiconductor.

The *Fermi–Dirac distribution*, also called the Fermi distribution, gives the probability that a state of energy E is occupied. Since the Pauli principle has been taken into account in the density of states given by Eq. (3.77), each state can be occupied by at most one electron. The Fermi distribution is given by

$$f_F(E) = \left[1 + \exp\left(\frac{E - E_F}{kT}\right)\right]^{-1} \tag{3.79}$$

where E_F is called the Fermi energy. At $E = E_F$ the Fermi distribution has a value of ½. For small energies the Fermi distribution approaches 1; thus low-energy states are very likely to be populated by electrons. For high energies the Fermi distribution decreases exponentially; states of high energy are less likely to the populated. Particles which follow a Fermi distribution are called ***fermions***. Electrons and holes in semiconductors are such fermions. A system of particles which obey ***Fermi*** statistics are called a ***Fermi gas***. Electrons and holes constitute such Fermi gases.

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

337-TA-1366 RX-0255

An approximate formula for the Fermi distribution can be obtained for high energies. One obtains for $E >> E_F$

$$f_F(E) \cong \exp\left[-\frac{E - E_F}{kT}\right] = f_B(E) . \tag{3.80}$$

This distribution coincides with the Boltzmann distribution. Thus the (quantum-mechanical) Fermi distribution and the (classical) Boltzmann distribution coincide for high energies, i.e. in the classical regime.

Next we prove the Fermi distribution of Eq. (3.79) by considering a collision between two electrons. For simplification we assume that one of the electrons has such a high energy that it belongs to the classical regime of semiconductor statistics. Quantum statistics applies to the other low-energy electron. During the collision of the two electrons, the energy is conserved

$$E_1 + E_2 = E_1' + E_2' \tag{3.81}$$

where, as before (Eq. (3.65)), E_1 and E_2 are electron energies before the collision and E_1' and E_2' are the energies after the collision.

The probability for the transition $(E_1, E_2) \rightarrow (E_1', E_2')$ is given by

$$p = f_F(E_1) f_B(E_2)[1 - f_F(E_1')][1 - f_B(E_2')] \tag{3.82a}$$

where it is assumed that E_2 and E_2' are relatively large energies and the corresponding electron can be properly described by the Boltzmann distribution. The terms $[1 - f_F(E_1')]$ and $[1 - f_B(E_2')]$ describe the probability that the states of energies E_1' and E_2' are empty, and are available for the electron after the collision. Further simplification is obtained by considering that E_2' is large and therefore $[1 - f_B(E_2')] \cong 1$. Equation (3.82a) then simplifies to

$$p = f_F(E_1) f_B(E_2)[1 - f_F(E_1')] . \tag{3.82b}$$

The same considerations are valid for the transition $(E_1', E_2') \rightarrow (E_1, E_2)$. The probability of this transition is given by

$$p' = f_F(E_1') f_B(E_2')[1 - f_F(E_1)] . \tag{3.83}$$

Under equilibrium conditions both transition probabilities are the same, i.e. $p = p'$. Equating Eqs. (3.82b) and (3.83), inserting the Boltzmann distribution for $f_B(E)$, and dividing by $f_F(E_1) f_F(E_1') f_B(E_2)$ yields

$$\frac{1}{f_F(E_1')} - 1 = \left[\frac{1}{f_F(E_1)} - 1\right] \exp\left[\frac{E_2 - E_2'}{kT}\right] \tag{3.84}$$

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

RX-0255.128

which must hold for all E_1 and E_1'. This condition requires that

$$\frac{1}{f_F(E)} - 1 = A \exp \frac{E}{kT} \tag{3.85}$$

where A is a constant. If the value of the constant is taken to be $A = \exp(-E_F/kT)$ one obtains the ***Fermi–Dirac distribution***

$$\boxed{f_F(E) = \left[1 + \exp\left(\frac{E - E_F}{kT}\right)\right]^{-1}} \tag{3.86}$$

which proves Eq. (3.79).

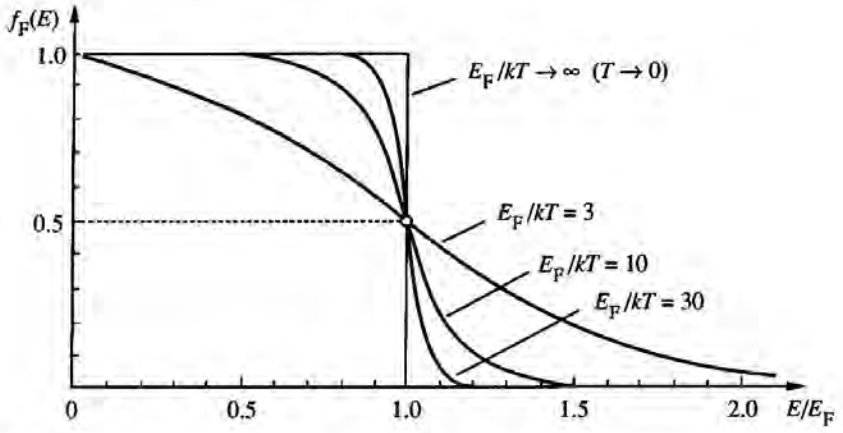


Fig. 3.11. Fermi–Dirac distribution as a function of E/E_F for different temperatures.

The Fermi–Dirac distribution is shown for different temperatures in Fig. 3.11. At the energy $E = E_F$ the probability of a state being populated has always a value of ½ independent of temperature. At higher temperatures, states of higher energies become populated. Note that the Fermi–Dirac distribution is symmetric with respect to E_F, that is

$$f_F(E_F + \Delta E) = 1 - f_F(E_F - \Delta E) \tag{3.87}$$

where ΔE is any energy measured with respect to the Fermi energy.

The Fermi–Dirac velocity distribution of the particles in a Fermi gas is obtained by multiplication of Eq. (3.78) with Eq. (3.79)

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

$$g(v) = \rho_{DOS}(v) f_F(v)$$

$$= \frac{m^3 v^2}{\pi^2 \hbar^3} \left[1 + \exp\left[\frac{\frac{1}{2}mv^2 - E_F}{kT}\right]\right]^{-1} \tag{3.88}$$

where we have used the fact that the energy of the Fermi gas is purely kinetic, i.e. $E = \frac{1}{2}mv^2$. Note that $g(v)$ is the number of carriers per velocity interval v and $v + dv$ and per unit volume. If the velocity v is expressed in terms of its components, then the spherical volume element, $4\pi v^2 dv$, is modified to a volume element in rectangular coordinates, $dv_x dv_y dv_z$. Thus, using $g(v)dv = g(v_x, v_y, v_z)\, dv_x dv_y dv_z$, one obtains

$$g(v_x, v_z, v_z) = \frac{m^3}{\pi^2 \hbar^3} \frac{1}{4\pi} \left\{1 + \exp\left[\frac{\frac{1}{2}m(v_x^2 + v_y^2 + v_z^2) - E_F}{kT}\right]\right\}^{-1} \tag{3.89}$$

which is the Fermi velocity distribution (per unit volume) in cartesian coordinates.

The Fermi distribution of energies of an ideal gas is obtained by multiplication of Eq. (3.77) with Eq. (3.79) and is given by

$$g(E) = \frac{1}{2\pi^2} \left(2\frac{m}{\hbar^2}\right)^{3/2} \sqrt{E} \left[1 + \exp\left[\frac{E - E_F}{kT}\right]\right]^{-1} \tag{3.90}$$

when $g(E)$ is the number of particles in the energy interval E and $E + dE$ and per unit volume.

3.2.6 The Fermi–Dirac integral of order $j = +\frac{1}{2}$

The Fermi–Dirac integral of order $j = +\frac{1}{2}$ allows one to calculate the free carrier concentration in a three-dimensional (3D) semiconductor. The free carrier concentration in one band, e.g. the conduction band, of a semiconductor is obtained from the product of density of states and the state occupation probability, i.e.

$$n = \int_{E_C}^{E_{top}} \rho_{DOS}(E)\, f_F(E)\, dE \,. \tag{3.91}$$

Integration over all conduction band states is required to obtain the total concentration. The upper limit of integration is the top of the conduction band and can be extended to infinity. This extension of $E_C^{top} \to \infty$ can be done without

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

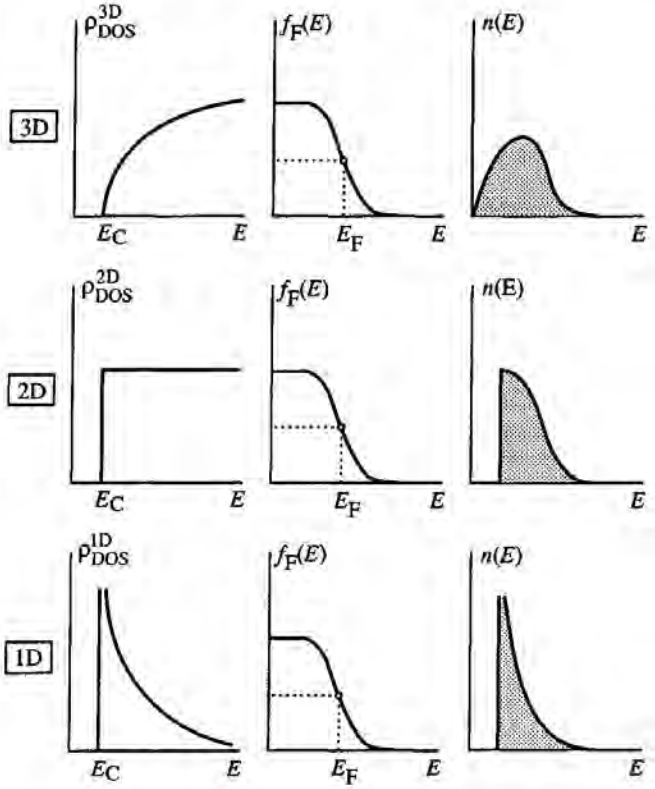


Fig. 3.12. Density of states (ρ_{DOS}), Fermi–Dirac distribution function (f_F) and carrier concentration (n) as a function of energy for a 3D, 2D, and 1D system. The shaded areas represent the total carrier concentration in the conduction band.

losing accuracy, since $f_F(E)$ converges strongly at high energies. The two functions, $\rho_{DOS}(E)$, $f_F(E)$, and their product are schematically shown in Fig. 3.12 for a semiconductor with three, two, and one, spatial degrees of freedom. The concentration per unit energy $n(E)$ is the product of state density and distribution function.

Equation (3.91) is evaluated by inserting the explicit expressions for the state density (Eq. (3.13)) and the Fermi–Dirac distribution (Eq. (3.79)). One obtains

$$n = \frac{1}{2\pi^2}\left(\frac{2m^* kT}{\hbar^2}\right)^{3/2} \int_0^\infty \frac{\eta^{1/2}}{1 + \exp(\eta - \eta_F)}\, d\eta \qquad (3.92)$$

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

where $\eta = E/kT$ and $\eta_F = -(E_C - E_F)/kT$ is the reduced Fermi energy. η_F is positive when E_F is inside the conduction band. The equation can be written in a more convenient way by using the Fermi–Dirac integral of order j, which is defined by (Sommerfeld, 1928; Sommerfeld and Frank, 1931)†

$$F_j(\eta_F) = \frac{1}{\Gamma(j+1)} \int_0^\infty \frac{\eta^j}{1 + \exp(\eta - \eta_F)} \, d\eta \tag{3.93}$$

where $\Gamma(j+1)$ is the Gamma-function. With $j = ½$ and $\Gamma(3/2) = \sqrt{\pi}/2$ one obtains

$$n = N_c F_{½}(\eta_F) \tag{3.94}$$

where N_c is the effective state density at the bottom of the conduction band (see Table 3.1). The Fermi–Dirac integral $F_{½}$ is shown in Fig. 3.13 along with several approximations which will be discussed later. For $j = ½$ the Fermi–Dirac integral is

$$F_{½}(\eta_F) = \frac{1}{\Gamma(3/2)} \int_0^\infty \frac{\eta^{½}}{1 + \exp(\eta - \eta_F)} \, d\eta \, . \tag{3.95}$$

The evaluation of the integral cannot be done analytically. Even though the numerical calculation of the Fermi–Dirac integral is straightforward, it proves frequently convenient to use *approximate* analytic solutions of $F_{½}(\eta_F)$.

For analytic approximations of the Fermi-Dirac integral $n/N_c = F_{½}(\eta_F)$, the inverse function is frequently used, that is the reduced Fermi energy η_F is expressed as a function of n/N_c. A number of analytic approximations developed prior to 1982 have been reviewed by Blakemore (Blakemore, 1982). To classify various approximations, we differentiate between non-degeneracy and degeneracy. In the *non-degenerate* regime, the Fermi energy is below the bottom of the conduction band, $E_F << E_C$. In the *degenerate* regime the Fermi energy is at or above the bottom of the conduction band.

(1) ***Extreme non-degeneracy (3D)***

In the case of extreme non-degeneracy (i.e. $E_C - E_F >> kT$ or $F_{½} << 1$

† Sommerfeld's original definition of the Fermi–Dirac integral omitted the term of the Gamma–function, i.e. $F_j^s(\eta_F) = \int_0^\infty \eta^j/[1 + \exp(\eta - \eta_F)] \, d\eta$. The modern definition of the Fermi–Dirac integral of Eq. (3.93) has the following advantages: (i) Unlike F_j^s, the functions F_j exist for negative orders of j, e.g. $j = -½, -1, -3/2$ etc. (ii) In the non-degenerate limit in which $\eta_F << 0$, all members of the $F_j(\eta_F)$ family reduce to $F_j(\eta_F) \to \exp\eta_F$ for all j. (iii) The *derivative* of the Fermi–Dirac integral of integer order j can be expressed as a Fermi–Dirac integral of order $(j-1)$, i.e. $(\partial/\partial\eta_F)F_j(\eta_F) = F_{j-1}(\eta_F)$.

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

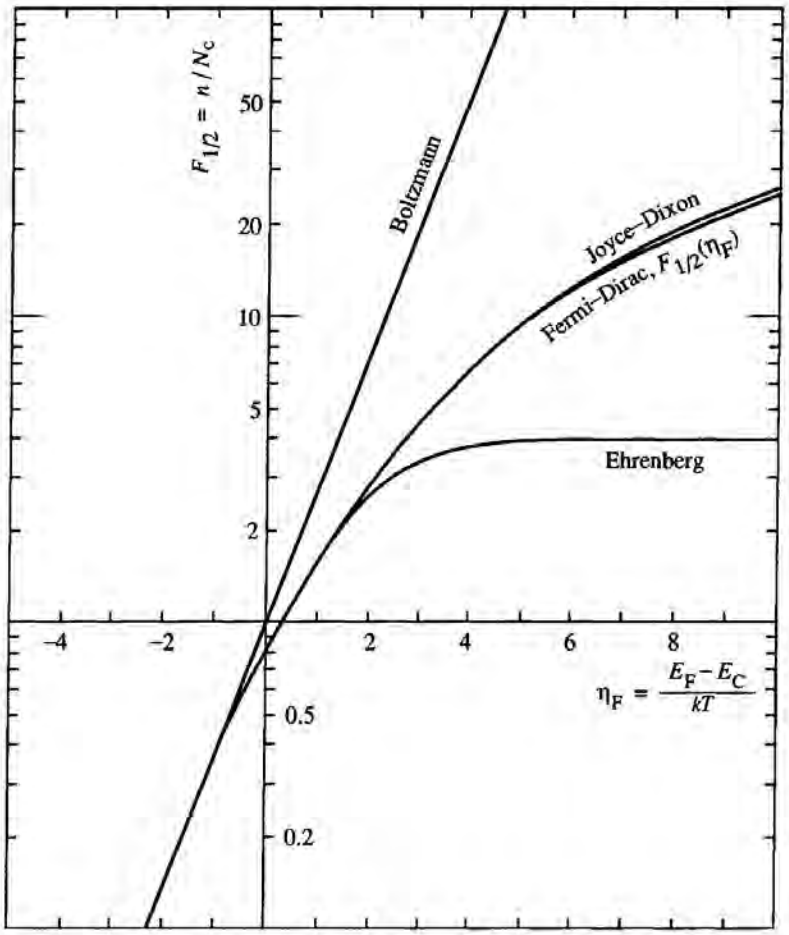


Fig. 3.13. Fermi–Dirac integral of order 1/2 as a function of the reduced Fermi energy η_F. Also shown are the Boltzmann distribution, the Joyce–Dixon approximation and the Ehrenberg approximation.

or $n << N_c$) the Fermi–Dirac distribution approaches the Boltzmann distribution. One obtains

$$\eta_F = -\frac{E_C - E_F}{kT} \cong \ln (n/N_c) \tag{3.96a}$$

which is shown in Fig. 3.14. This approximation is good when the Fermi

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

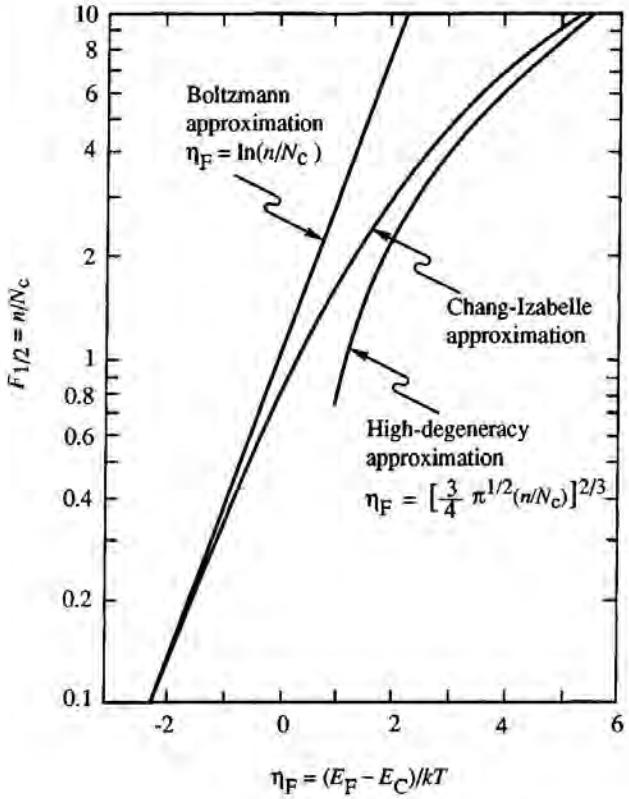


Fig. 3.14. Approximations for the Fermi–Dirac integral $F_{1/2}(\eta_F)$. Shown are the low-density approximation (Boltzman distribution), the high-degeneracy approximation and the full-range Chang–Izabelle approximation.

energy is $2kT$ or more below the bottom of the conduction band. Rearrangement of the equation yields the carrier concentration as a function of the Fermi energy in the non-degenerate limit

$$n = N_c \exp\left[-\frac{E_C - E_F}{kT}\right] \tag{3.96b}$$

(2) ***Extreme degeneracy (3D)***

In the case of extreme degeneracy (i.e. $(E_F - E_C) >> kT$ or $F_{½} >> 1$ or $n >> N_c$) the Fermi–Dirac integral reduces to

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

$$\eta_F = -\frac{E_C - E_F}{kT} \cong \left[\frac{3}{2}\Gamma\left(\frac{3}{2}\right)\frac{n}{N_c}\right]^{2/3} \cong \left(\frac{3}{4}\sqrt{\pi}\,\frac{n}{N_c}\right)^{2/3} \tag{3.97a}$$

which is shown in Fig. 3.14. The range of validity for this approximation is $E_F - E_C > 10\,kT$, i.e. when the Fermi energy is well within the conduction band. Rearrangement of the equation and using Eq. (3.37) for the effective density of states yields the carrier concentration as a function of the Fermi energy in the degenerate limit

$$n = \frac{1}{3\pi^2}\left[\frac{2m^*(E_F - E_C}{\hbar^2}\right]^{3/2} \tag{3.97b}$$

(3) ***The Ehrenberg Approximation (3D)***

This approximation (Ehrenberg, 1950) was developed for weak degeneracy and is shown in Fig. 3.13. The approximation is given by

$$\eta_F = -\frac{E_C - E_F}{kT} \cong \ln\frac{n}{N_c} - \ln\left[1 - \frac{1}{4}\frac{n}{N_c}\right]. \tag{3.98}$$

For small n the second logarithm term approaches zero, that is the Boltzmann distribution is recovered. The range of validity of the approximation is limited to $E_F - E_C \leq 2kT$, i.e. to weak degeneracy.

(4) ***The Joyce–Dixon approximation (3D)***

An approximation valid for a wider range of degeneracy was developed by Joyce and Dixon (Joyce and Dixon, 1977; Joyce, 1978). This approximation expresses the reduced Fermi energy as a sum of the Boltzmann term and a polynomial, i.e.

$$\eta_F = -\frac{E_C - E_F}{kT} \cong \ln\frac{n}{N_c} + \sum_{m=1}^{4} A_m \left[\frac{n}{N_c}\right]^m \tag{3.99a}$$

where the first four coefficients A_m are given by

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

$$A_1 = \frac{\sqrt{2}}{4} = 3.535\,53 \times 10^{-1} ,$$

$$A_2 = -4.950\,09 \times 10^{-3} , \tag{3.99b}$$

$$A_3 = 1.483\,86 \times 10^{-4} ,$$

$$A_4 = -4.425\,63 \times 10^{-6} .$$

The Joyce–Dixon approximation given here is shown in Fig. 3.13 and can be used for degeneracies of $E_F - E_C \leq 8kT$. Inclusion of higher terms in the power series ($m > 4$) allows one to extend the Joyce-Dixon approximation to higher degrees of degeneracy.

(5) ***The Chang–Izabelle Approximation (3D)***

The Chang–Izabelle approximation (Chang and Izabelle, 1989) is a full-range approximation which is valid for non-degenerate as well as degenerate semiconductors. The approximation is motivated by the fact that low-density and high-density approximations are available (see Eqs. (3.96) and (3.97)) which are the exact solutions in the two extremes. The Chang–Izabelle approximation represents the construction of a function, which approaches the low-density solution and the high-density solution of the Fermi–Dirac integral as shown in Fig. 3.14. The reduced Fermi energy is then given by

$$\eta_F = \frac{E_C - E_F}{kT} \cong \ln \frac{n}{N_c + n} + \left[\frac{3}{2}\Gamma\left(\frac{3}{2}\right)\right]^{2/3} \frac{n/N_c}{(A + n/N_c)^{1/3}} \tag{3.100a}$$

where

$$n/N_c = F_{1/2} , \tag{3.100b}$$

$$\frac{3}{2}\Gamma\left(\frac{3}{2}\right) = \frac{3}{2}\,\frac{\sqrt{\pi}}{2} = 1.329\,34 , \tag{3.100c}$$

$$A = \frac{[(3/2)\Gamma(3/2)]^2\,(n_0/N_c)^3}{[\ln(1 + N_c/n_0)]^3} , \tag{3.100d}$$

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

$$\frac{n_0}{N_c} = \frac{n(E_F = E_C)}{N_c} = F_{1/2}(\eta_F = 0) = 0.765\ 15\ . \qquad (3.100e)$$

One can easily verify that Eq. (3.100a) recovers the low-density approximation and the high-density approximation for $n << N_c$ and $n >> N_c$, respectively. Furthermore, the approximation yields an exact solution for $\eta_F = 0$, i.e. when the Fermi energy touches the bottom of the conduction band. The largest relative error of η_F is 1% in the Chang–Izabelle approximation. Chang and Izabelle (1989) showed that the relative error can be further reduced by a weighting function and a polynomial function. Using these functions, the maximum relative error is reduced to 0.033%.

(6) ***The Nilsson approximation (3D)***

The Nilsson approximation (Nilsson, 1973) is valid for the entire range of Fermi energies. It is given by

$$\eta_F = -\frac{\ln(n/N_c)}{n/N_c - 1} + \left[\frac{3}{4}\sqrt{\pi}\frac{n}{N_c}\right]^{2/3} + \frac{(3/2)\sqrt{\pi}(n/N_c)}{[3 + (3/4)\sqrt{\pi}(n/N_c)]^2}\ . \qquad (3.101)$$

The maximum relative error of the approximation is 1.1%.

3.2.7 The Fermi–Dirac integral of order $j = 0$

The Fermi–Dirac integral of order $j = 0$ allows one to calculate the free carrier density in a two-dimensional (2D) semiconductor. For semiconductor structures with only two degrees of spatial freedom, the Fermi–Dirac integral is obtained from Eq. (3.91) by insertion of the two-dimensional density of states. One obtains for the 2D carrier density

$$n^{2D} = \int_{E_C}^{E_{top}} \rho_{DOS}^{2D}(E)\, f_F(E)\, dE = \frac{m^* kT}{\pi\hbar^2} \int_0^\infty [1 + \exp(\eta - \eta_F)]^{-1}\, d\eta \qquad (3.102a)$$

where $\eta = E/kT$ and $\eta_F = (E_F - E_C)/kT$ are reduced energies. The integral can be written as the Fermi–Dirac integral of zero ($j = 0$) order

$$F_{j=0}(\eta_F) = \frac{1}{\Gamma(1)} \int_0^\infty [1 + \exp(\eta - \eta_F)]^{-1}\, d\eta \qquad (3.102b)$$

where $\Gamma(1) = 1$ is the Gamma-function. The two-dimensional carrier density can be written by using the effective density of states of a 2D system given by

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

Eq. (3.41). One obtains

$$n^{2D} = N_c^{2D} F_0(\eta_F) \tag{3.103}$$

which is formally similar to the corresponding equation in three dimensions (Eq. (3.94)). The Fermi–Dirac integral of zero order ($j = 0$) can be solved analytically. Using the integral formula

$$\int \frac{dx}{1 + e^x} = -\ln(1 + e^{-x}) , \tag{3.104}$$

one obtains

$$F_0(\eta_F) = \ln(1 + e^{\eta_F}) . \tag{3.105}$$

Thus, the two-dimensional carrier density depends on the reduced Fermi energy according to

$$\frac{n^{2D}}{N_c^{2D}} = \ln(1 + e^{\eta_F}) . \tag{3.106}$$

Rearrangement of the equation yields the Fermi energy as a function of the carrier density

$$\boxed{\eta_F = -\frac{E_C - E_F}{kT} = \ln\left[\exp\left[\frac{n^{2D}}{N_c^{2D}}\right] - 1\right]} \tag{3.107}$$

(1) ***Extreme non-degeneracy (2D)***

Approximation for the low-density regime ($n^{2D} << N_c^{2D}$) and the high-density regime ($n^{2D} >> N_c^{2D}$) can be easily obtained from Eqs. (3.106) and (3.107). In the low-density regime one obtains

$$\boxed{\eta_F = -\frac{E_C - E_F}{kT} = \ln\frac{n^{2D}}{N_c^{2D}}} \tag{3.108a}$$

which is valid if the Fermi energy is much below the bottom of the conduction subband. Rearrangement of the equation yields the two-dimensional carrier density as a function of the Fermi energy in the non-degenerate limit

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

$$n^{2D} = N_c^{2D} \exp\left[-\frac{E_C - E_F}{kT}\right]. \tag{3.108b}$$

(2) ***Extreme degeneracy (2D)***

In the high-density regime one obtains

$$\boxed{\eta_F = -\frac{E_C - E_F}{kT} = \frac{n^{2D}}{N_c^{2D}}} \tag{3.109a}$$

Rearrangement of the equation and insertion of the explicit expression for N_c^{2D} given in Eq. (3.41) yields

$$n^{2D} = \frac{m^*}{\pi\hbar^2}(E_F - E_C) \quad \text{(for high degeneracy)} \tag{3.109b}$$

that is the Fermi energy and the two-dimensional density follow a linear relation for two-dimensional structures.

3.2.8 The Fermi–Dirac integral of order $j = -½$

The Fermi-Dirac integral of order $j = -½$ allows one to calculate the free carrier density (per unit length) in a one-dimensional (1D) semiconductor. In semiconductor structures with only one degree of spacial freedom, the Fermi–Dirac integral is obtained from Eq. (3.91) by insertion of the 1D density of states of Eq. (3.28). One obtains

$$n^{1D} = \int_{E_c}^{E_{top}} \rho_{DOS}^{1D}(E)\, f_F(E)\, dE$$

$$= \frac{kT}{\pi\hbar}\sqrt{\frac{m^*}{2kT}} \int_0^\infty \frac{\eta^{-½}}{1 + \exp(\eta - \eta_F)}\, d\eta \tag{3.110}$$

where $\eta = E/kT$ and $\eta_F = -(E_C - E_F)/kT$ are reduced energies. The integral can be written as the Fermi–Dirac integral of order $j = -½$

$$F_{-½}(\eta_F) = \frac{1}{\Gamma(½)} \int_0^\infty \frac{\eta^{-½}}{1 + \exp(\eta - \eta_F)}\, d\eta \tag{3.111}$$

where $\Gamma(½) = \sqrt{\pi}$ is the Gamma-function. Using the effective density of states of a one-dimensional system (see Eq. (3.44)), the one-dimensional density can be

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

written as

$$n^{1D} = N_c^{1D} F_{-½}(\eta_F) \tag{3.112}$$

which is similar to the corresponding equations in three (Eq. (3.94)) and two (Eq. (3.103)) dimensions. The Fermi–Dirac integral of order $j = -½$ can only be obtained by numerical integration or by approximate solutions which will be discussed in the following sections. The $j = -½$ Fermi–Dirac integral has asymptotic solutions for the regimes of non-degeneracy and high degeneracy.

(1) ***Extreme non-degeneracy (1D)***

In the regime of extreme non-degeneracy ($n^{1D} << N_c^{1D}$) the Fermi–Dirac integral of order $j = -½$ approaches the Boltzmann distribution. One obtains

$$\boxed{\eta_F = -\frac{E_C - E_F}{kT} = \ln \frac{n^{1D}}{N_c^{1D}}} \tag{3.113a}$$

This approximation is good for $E_F - E_C \leq 2kT$, i.e. when the Fermi energy is at least $2kT$ below the bottom of the conduction subband. Rearrangement of the equation yields the one-dimensional carrier density as a function of the Fermi energy in the non-degenerate limit

$$n^{1D} = N_c^{1D} \exp\left[-\frac{E_C - E_F}{kT}\right]. \tag{3.113b}$$

(2) ***Extreme degeneracy (1D)***

In the case of extreme degeneracy ($n^{1D} >> N_C^{1D}$) the Fermi–Dirac integral of order $j = -½$ reduces to

$$\boxed{\eta_F = -\frac{E_C - E_F}{kT} \cong \frac{1}{4}[\Gamma(½)]^2 \left[\frac{n^{1D}}{N_c^{1D}}\right]^2 \cong \frac{\pi}{4}\left[\frac{n^{1D}}{N_c^{1D}}\right]^2} \tag{3.114}$$

The range of validity of the approximation is $E_F - E_C > 10\ kT$, i.e. when the Fermi energy is well within the conduction band.

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

3.3 Carrier concentrations

The free carrier concentration in semiconductors depends on a number of parameters such as the doping concentration, impurity activation energy, temperature, and other parameters. Given the results of the previous sections on the density of states and the distribution functions, the carrier concentration can now be calculated. In the calculation intrinsic, extrinsic, and compensated semiconductors will be considered.

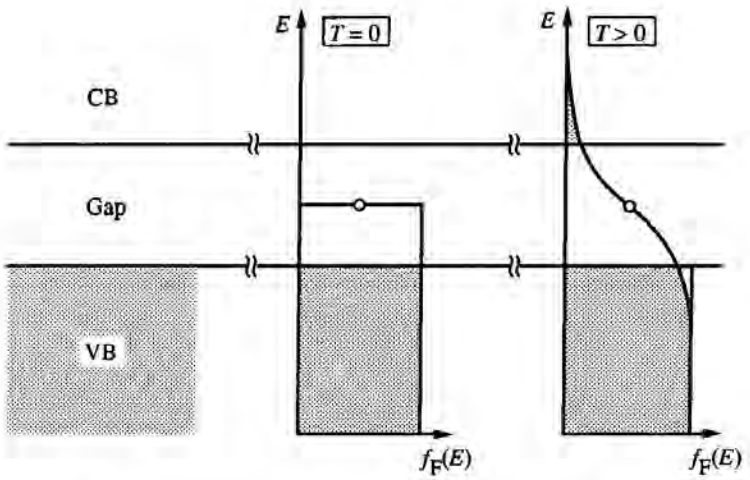


Fig. 3.15. Carrier distributions in the conduction band (CB) and valence band (VB) and Fermi distribution function in an intrinsic semiconductor at $T = 0$ K and at finite temperatures $T > 0$ K.

3.3.1 Intrinsic semiconductors

The carrier concentration of pure, undoped semiconductors is determined by thermal excitation of electrons from the valence band to states in the conduction band. An intrinsic semiconductor has a filled valence band and an empty conduction band at zero temperature. This property is the very definition of semiconductors. The band diagram along with the $T = 0$ K Fermi distribution function is shown in Fig. 3.15.

As the temperature increases, a small fraction of electrons in the valence band is excited into the conduction band. Thus the number of holes (unoccupied states) p in the valence band coincides with the number of electrons n in the conduction band. Semiconductors for which $n = p$ are called *intrinsic*. The condition that the concentration of electrons coincides with the concentration of holes requires

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

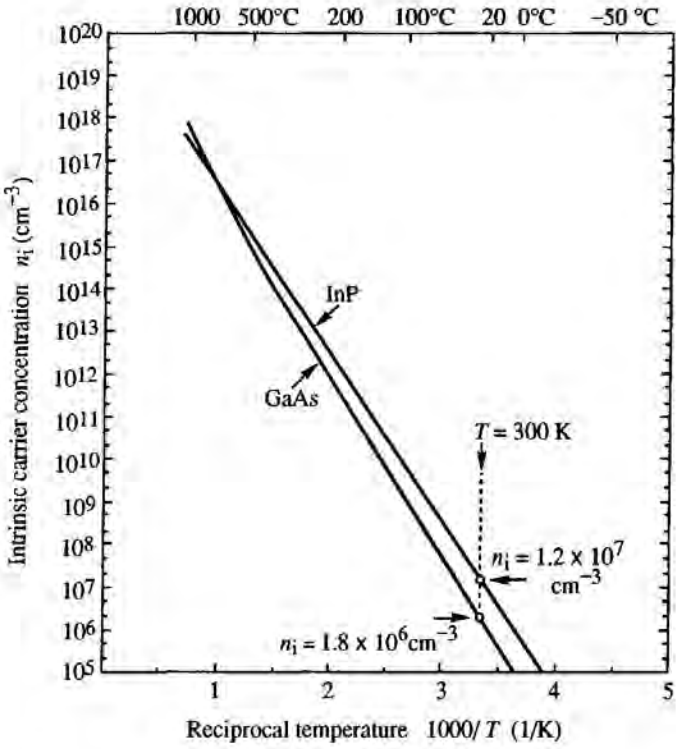


Fig. 3.16. Intrinsic carrier concentrations of GaAs and InP as a function of temperature. The slopes of the curves are proportional to the gap energy (after Thurmond, 1975; Laufer et al., 1980).

that the Fermi energy be within the forbidden gap. The position of the Fermi energy in the gap is visualized in Fig. 3.15. The electron and hole concentrations are given by

$$n = p\,, \tag{3.115a}$$

$$N_c F_{1/2}(\eta_F) = N_v F_{1/2}(\eta_F)\,. \tag{3.115b}$$

Since the Fermi energy is within the forbidden gap, i.e. many values of kT below the conduction band and many values of kT above the valence band, simpler Boltzmann statistics can be used instead of Fermi–Dirac statistics. Equation (3.115b) then simplifies to

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

$$N_c \exp\left[-\frac{E_C - E_F}{kT}\right] = N_v \exp\left[-\frac{E_F - E_V}{kT}\right]. \tag{3.116}$$

Rearrangement of the equation and the definition of the gap energy $E_g = E_C - E_V$ yields for the Fermi energy of an *intrinsic* semiconductor

$$E_F = E_V + \frac{1}{2}E_g + \frac{kT}{2}\ln\frac{N_v}{N_c}. \tag{3.117}$$

In this equation, $E_V + \frac{1}{2} E_g$ represents the mid-gap energy. Since the logarithmic function changes weakly with N_v/N_c, the Fermi energy of an intrinsic semiconductor is approximately at mid-gap. The temperature dependence of the intrinsic Fermi energy is weak due to the (weak) logarithmic dependence of the Fermi energy on the temperature. Using Boltzmann statistics the Fermi energy allows us to determine the ***intrinsic carrier concentration***, n_i, of electrons and holes in an undoped semiconductor.

$$\boxed{n_i = \sqrt{N_v N_c}\exp\left[-\frac{E_g}{2kT}\right]} \tag{3.118}$$

According to this equation the intrinsic carrier concentration increases exponentially with temperature. In addition, the effective density of states have the comparatively weak temperature dependence of $N_{c,v} \propto T^{3/2}$. The intrinsic carrier concentration is of special importance. Calculating the product of electron and hole concentration for *any* (non-degenerate) Fermi level using Boltzmann statistics yields

$$np = n_i^2 = N_v N_c \exp\left[-\frac{E_g}{kT}\right]. \tag{3.119}$$

Thus the product np is a constant at a given temperature and, since the result does not depend on the Fermi level, is independent of the doping concentration. The intrinsic carrier concentrations of GaAs and InP are shown as a function of temperature in Fig. 3.16.

3.3.2 Extrinsic semiconductors (single donor species)

Substitutional donors and acceptors have an excess or a deficit electron in their outer electron shell, respectively, as compared to the replaced lattice atom.

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

Donors have one excess electron which can be *donated* to the conduction band. ***Acceptors*** have one less electron than the replaced lattice atom and can *accept* an electron from the filled valence band of the semiconductor, thereby creating a *hole*. The detailed physical properties of shallow impurities were considered in Chap. 1. Here we consider donors and acceptors being represented by an energy state close to the conduction band edge (donor) or close to the valence band edge (acceptor), as shown in Fig. 3.17. In the following, the free carrier concentration as a function of temperature is investigated in a semiconductor with donor impurities of concentration N_D of one chemical species.

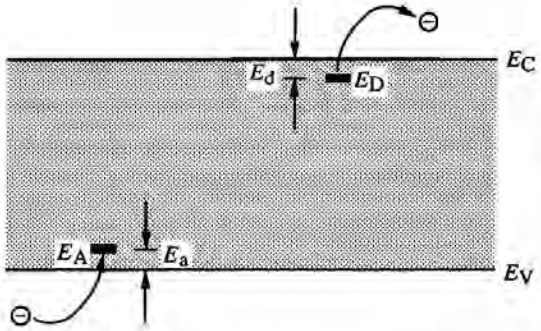


Fig. 3.17. Energy levels of acceptors and donors in the semiconductor band diagram. A donor (acceptor) level at energy E_D (E_A) has an ionization energy E_d (E_a).

The charge state of donors is *neutral* when occupied by an electron and positively charged if the electron is excited to the conduction band. The total concentration of donors is the sum of neutral donor concentration and ionized donor concentration, i.e.

$$N_D = N_D^0 + N_D^+ . \tag{3.120}$$

The energy of the donor impurity state is denoted as E_D. The donor energy is frequently given with respect to the conduction band edge, that is

$$E_d = E_C - E_D . \tag{3.121}$$

The probability of occupation of an acceptor or donor follows Fermi–Dirac statistics. Consequently, the concentration of neutral donors, i.e. donors occupied by an electron is

$$N_D^0 = N_D \, f_F(E_D) \tag{3.122}$$

where $f(E_D)$ is the value of the Fermi–Dirac distribution (Eq. (3.79)) at the energy of the donor. With $N_D^+ = N_D - N_D^0$ one obtains the concentration of

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

ionized donors

$$N_D^+ = N_D[1 - f_F(E_D)] = N_D\left[1 - \frac{1}{1 + \frac{1}{g}\exp\left(\frac{E_D - E_F}{kT}\right)}\right]$$
$$= N_D\left[1 + g\exp\left(\frac{E_F - E_D}{kT}\right)\right]^{-1} \quad (3.123)$$

where g is the ground-state degeneracy of the donor. The value of the ground-state degeneracy in GaAs is $g = 2$ for hydrogen-like donors since the donor can donate one electron of either spin (see Chap. 1). The ground-state degeneracy of acceptors in GaAs is $g = 4$, since the acceptor can accept electrons of either spin from the heavy-hole and the light-hole valence band (see Chap. 1). Note that Eq. (3.123) is limited to concentrations below the Mott transition (see Chap. 1). Above the Mott transition, impurities cannot bind charge carriers, i.e. donors and acceptors cannot be in the neutral charge state.

If a semiconductor has one carrier type dominating due to doping, the other carrier type has an extremely small equilibrium concentration. If, for example, GaAs is doped with $N_D = 10^{17}$ cm^{-3} donors and $n \cong 10^{17}$ cm^{-3}, the hole-concentration inferred from Eq. (3.118) at 300 K is $p = n_i^2/n = 3.2 \times 10^{-3}$ cm^{-3}. Thus, there are approximately 3 holes in 1000 cm^3 of this n-type semiconductor. The very small concentration of the ***minority carrier*** allows us to completely neglect minority carriers in many semiconductor structures. Such semiconductor devices are called ***majority carrier*** devices.

Charge neutrality is maintained in a doped semiconductor and has to be taken into account in addition to Fermi–Dirac statistics. Since minority carriers can be neglected, the free carrier concentration coincides with the ionized dopant concentration. If we restrict ourselves to n-type semiconductors, then

$$n = N_D^+ . \quad (3.124)$$

We now consider the semiconductor at low temperatures, when most electrons occupy donor states. Then Boltzmann statistics can be used for the occupation of conduction band states according to

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

$$n = N_c \exp\left[-\frac{E_C - E_F}{kT}\right]. \tag{3.125}$$

If Fermi–Dirac statistics are used for the occupation of the donor level according to Eq. (3.123), one obtains a quadratic equation for the free carrier concentration

$$n^2 - \frac{1}{g} N_D N_c \, e^{-E_d/kT} + \frac{1}{g} n N_c \, e^{-E_d/kT} = 0. \tag{3.126}$$

At low temperatures the free carrier concentration, n, is much smaller than the donor concentration, N_D. Thus, the third term of the quadratic equation is much smaller than the second term. The free carrier concentration is given by

$$\boxed{n \equiv \left(\frac{1}{g} N_D N_c\right)^{1/2} \exp\left(-\frac{E_d}{2kT}\right)} \tag{3.127}$$

where the ground state degeneracy for donors is $g = 2$.

Equation (3.127) was first obtained by de Boer and van Geel (1935) by the method described here. The formulas can also be obtained by minimizing the free energy change due to thermal excitation of electrons from donor states to conduction band states (Mott and Gurney, 1940). At higher temperatures all donors become ionized. The carrier concentration is then constant $n = N_D^+ = N_D$ and independent of temperature. This temperature regime is called the ***saturation regime.***

As the temperature is increased even further, the *intrinsic* carrier concentration n_i increases and at sufficiently high temperatures assumes values comparable or higher than the dopant concentration. For most technologically useful semiconductors, the crossover from the saturation to the ***intrinsic regime*** occurs at temperatures much higher than room temperature. The three temperature regimes (i) thermal ionization regime (ii) saturation regime and (iii) intrinsic regime are shown schematically in Fig. 3.18 along with the associated activation energies.

The thermal ionization energy of a donor can be obtained from the slope of n versus reciprocal temperature. Rearrangement of Eq. (3.127) yields

$$E_d = -2k \frac{d(\ln n)}{d(1/T)} \tag{3.128}$$

which allows one to determine E_d directly from the temperature dependent carrier concentration. The change in carrier concentration with increasing temperatures

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

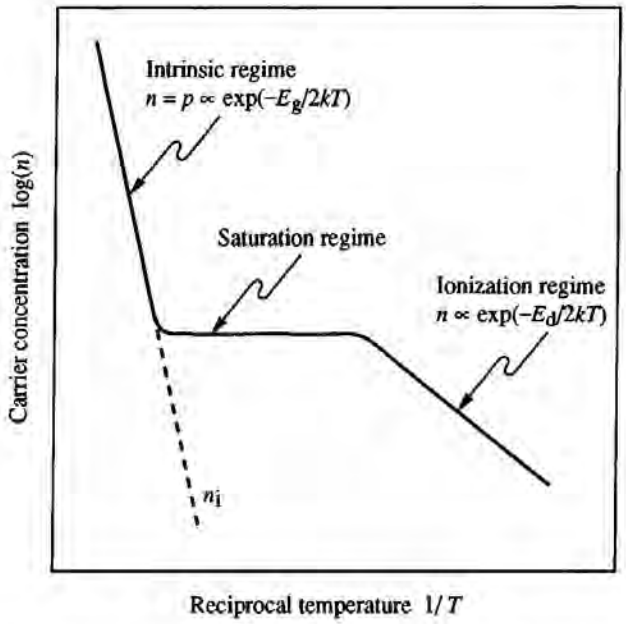


Fig. 3.18. Carrier concentration as a function of reciprocal temperature for an uncompensated n-type semiconductor. The donor is assumed to form one level at energy E_d below the conduction band edge. Three regimes, namely (i) the ionization regime, (ii) the saturation regime, and (iii) the intrinsic regime and their corresponding activation energies are indicated (after Smith, 1986).

also implies a continuously changing Fermi level in the semiconductor. Consider an n-type semiconductor. At low temperatures the donor levels are filled, while the conduction band is empty. Thus, the Fermi level must be slightly above the donor level. As the temperature increases, the Fermi distribution becomes smeared out, as conduction band states become filled and donor states become unoccupied. Simultaneously the Fermi level moves deeper into the forbidden gap. At still higher temperatures, the Fermi level approaches the (near) mid-gap level and the semiconductor becomes intrinsic.

3.3.3 Extrinsic semiconductors (two donor species)

In the following, the free carrier concentration as a function of temperature is investigated in a semiconductor with two different species of donor impurities. It is assumed that the two donors form two different energy levels in the gap of the semiconductor. The two types of donor levels can originate from two different chemical species (e.g. Sn and Te donors in GaAs).

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

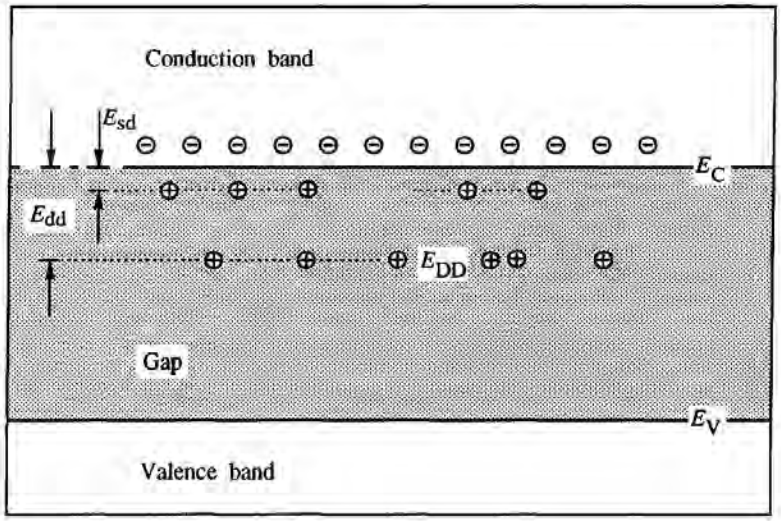


Fig. 3.19. Band diagram of a semiconductor with two species of donors namely a shallow donor with energy E_{SD} (ionization energy E_{sd}) and a deep donor with energy E_{DD} (ionization energy E_{dd}).

If both donor types have very similar thermal ionization energies, there is no necessity to differentiate between the two types of donors. It is therefore assumed that the two types of donors have markedly different ionization energies. In particular, we assume that one of the donors is relatively shallow and the other one is relatively deep. The band diagram of a semiconductor with two types of donors with energies E_{SD} and E_{DD} is shown in Fig. 3.19. If the shallow and deep donor concentrations are given by N_{SD} and N_{DD}, respectively, then the free carrier concentration is given by

$$n = N_{SD}^{+} + N_{DD}^{+} \,. \tag{3.129}$$

The carrier concentration in the conduction band is given by the Boltzmann distribution

$$n = N_c \exp\left[-\frac{E_C - E_F}{kT}\right] \tag{3.130}$$

while Fermi–Dirac statistics is assumed for the donor levels (see Eq. (3.123)).

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

$$N_{SD}^{+} = N_{SD} - \frac{N_{SD}}{1 + \frac{1}{g}\exp\left[\frac{E_{SD} - E_F}{kT}\right]}, \tag{3.131a}$$

$$N_{DD}^{+} = N_{DD} - \frac{N_{DD}}{1 + \frac{1}{g}\exp\left[\frac{E_{DD} - E_F}{kT}\right]} \tag{3.131b}$$

where E_{SD} and E_{DD} are the energies of the donor states. We assume that $E_{sd} \ll E_{dd}$ and $E_{dd} \ll E_g$, where $E_{sd} = E_C - E_{SD}$ and $E_{dd} = E_C - E_{DD}$.

All donors are neutral at very low temperatures. As temperature increases, shallow donors will donate their electrons to the conduction band, until all shallow donors are ionized. As the temperature is further increased, the deep donors start to become ionized until all deep donors are ionized. At even higher temperatures the intrinsic carrier concentration exceeds the dopant concentration and the semiconductor becomes intrinsic. In the following, the ionization regimes of the shallow and the deep donor are investigated.

At low temperatures when both types of donors are neutral, the Fermi energy is higher than the shallow donor energy. Then the energy difference between the Fermi energy and the deep donor energy is relatively large and according to Eq. (3.131b), the deep donor can be considered as neutral, i.e. $N_{DD}^{+} = 0$. The carrier concentration is then given by

$$n = N_{SD}^{+} = N_{SD} - \frac{N_{SD}}{1 + \frac{1}{g}\exp\left[\frac{E_{SD} - E_F}{kT}\right]}. \tag{3.132}$$

Using Boltzmann statistics for the conduction band one obtains the quadratic equation

$$n^2 - \frac{1}{g} N_{SD} N_c \exp(-E_{sd}/kT) + \frac{1}{g} n N_c \exp(-E_{sd}/kT) = 0 \tag{3.133}$$

which is identical to the single donor equation Eq. (3.126). Thus, the low temperature solution is

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

$$n \cong \left(\frac{1}{2} N_{SD} N_c\right)^{1/2} \exp\left(-\frac{E_{sd}}{2kT}\right) \tag{3.134}$$

where the ground-state degeneracy is assumed to be $g = 2$. Ionization of the shallow donor continues until all shallow donors are ionized, i.e. $n = N_{SD}^{+} = N_{SD}$.

As the temperature is increased further, deep donors become ionized. Using Boltmann statistics for the conduction band (Eq. (3.130)) and Fermi–Dirac statistics for the deep donor (Eq. 131b) one obtains the quadratic equation

$$(N_{DD}^{+})^2 + N_{DD}^{+}\left[N_{SD}^{+} + \frac{1}{g} N_c \exp(-E_{dd}/kT)\right] - \frac{1}{g} N_{DD} N_c \exp(-E_{dd}/kT) = 0. \tag{3.135}$$

For $N_{SD}^{+} >> (g^{-1}) N_c \exp(-E_{dd}/kT)$ one obtains

$$(N_{DD}^{+})^2 + N_{SD}^{+} N_{DD}^{+} \cong \frac{1}{g} N_{DD} N_c \exp(-E_{dd}/kT) . \tag{3.136a}$$

Since the free carrier concentration is the sum of ionized deep and shallow donor concentration ($n = N_{SD}^{+} + N_{DD}^{+}$) the equation can be written as

$$n(n - N_{SD}^{+}) \cong \frac{1}{2} N_{DD} N_c \exp(-E_{dd}/kT) \tag{3.136b}$$

where the donor ground-state degeneracy is assumed to be $g = 2$.

At the elevated temperatures considered here, the shallow donor is ionized ($N_{SD}^{+} = N_{SD}$) and therefore the slope of the carrier density with respect to temperature follows the proportionality

$$n^2 - nN_{SD} \propto \exp(-E_{dd}/kT) \tag{3.136c}$$

Note that this equation is significantly different from the simple relation $n \propto \exp(-E_d/kT)$, which would lead to incorrect results if applied to a semiconductor with two types of donors. Equation (3.136c) was applied to shallow and deep Si donors in $Al_x Ga_{1-x} As$ (Schubert and Ploog, 1984).

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

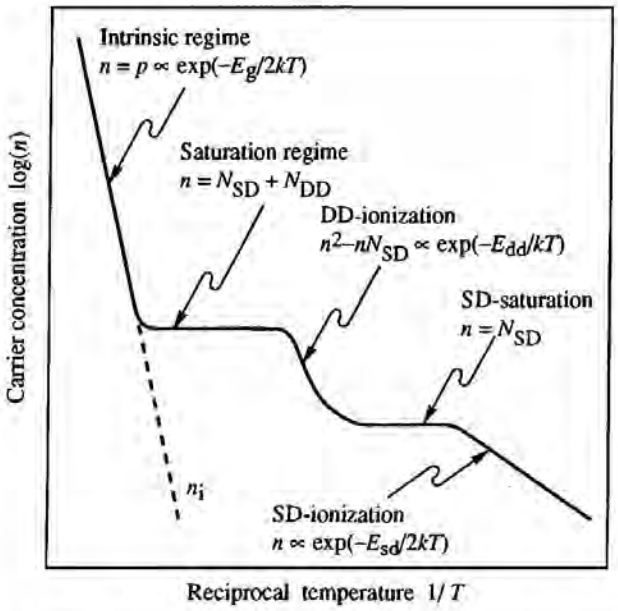


Fig. 3.20. Carrier concentration as a function of reciprocal temperature for a semiconductor with two different dopant species, namely a shallow donor (SD) and a deep donor (DD). At low temperatures the shallow donor becomes ionized and saturates when $n = N_{SD}$. At higher temperatures the deep donor becomes ionized and saturates ($n = N_{SD} + N_{DD}$). At even higher temperatures the semiconductor becomes intrinsic ($n = n_i$). The activation energies of each regime are indicated.

The ionization of the deep donor continues until shallow and deep donors are ionized, which corresponds to the saturation regime. At even higher temperatures the intrinsic carrier concentration increases above the dopant concentration and the semiconductor becomes intrinsic. The carrier concentration is shown in Fig. 3.20 for a semiconductor containing two different donor species as a function of temperature. The different saturation and ionization regimes along with their activation energies are indicated in the figure. Note that the different ionization regimes discussed above may not be as clearly distinguishable if the difference between E_{sd} and E_{dd} is small.

3.3.4 Compensated semiconductors

A partially compensated semiconductor contains dopant atoms of one type (n- or p-type) and, in addition, a smaller number of dopants of the other type. The band diagram of a partially compensated n-type semiconductor with a small number of acceptors is shown in Fig. 3.21. The free carrier concentration is given by

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

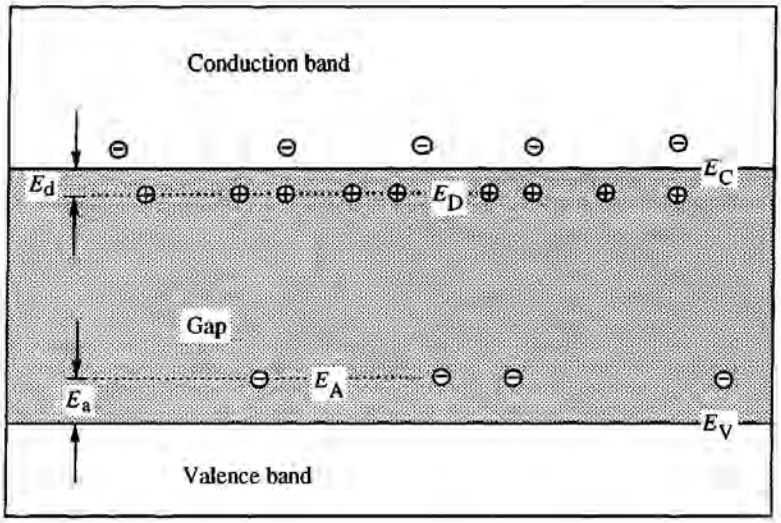


Fig. 3.21. Band diagram of a lightly compensated n-type semiconductor.

$$n = N_D^+ - N_A^- \,. \tag{3.137}$$

Electrons from donors prefer to occupy lower-energy acceptor states at all temperatures. Thus, even for low temperatures ($T \to 0$ K) *all* acceptors and *some* donors are ionized. Fermi–Dirac statistics for the occupation of the donor state (Eq. (3.122)) and Eq. (3.137) yields

$$n + N_A^- = N_D^+ = N_D - \frac{N_D}{1 + \frac{1}{g} \exp\left(\frac{E_D - E_F}{kT}\right)} \,. \tag{3.138}$$

Using Boltzmann statistics for the population of the conduction band (Eq. (3.96b)) allows one to eliminate the Fermi energy. One obtains the quadratic equation

$$n^2 + n\left[N_A + \frac{1}{g} N_c \, e^{-E_d/kT}\right] - \frac{1}{g} N_c \, e^{-E_d/kT} (N_D - N_A) = 0 \tag{3.139}$$

where acceptors are assumed to be ionized at all temperatures, i.e. $N_A = N_A^-$. The solution of the quadratic equation is, for $g = 2$:

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

$$n = -\tfrac{1}{2}\left[N_A + \tfrac{1}{2} N_c\, e^{-E_d/kT}\right] + \tfrac{1}{2}\left[\left(N_A + \tfrac{1}{2} N_c\, e^{-E_d/kT}\right)^2 + 2N_c\, e^{-E_d/kT}(N_D - N_A)\right]^{1/2} \quad (3.140)$$

At low temperatures when $\frac{1}{2}N_c \exp(-E_d/kT) \ll N_A$ the equation simplifies to

$$n = -\tfrac{1}{2} N_A + \tfrac{1}{2} N_A \left[1 + 2N_c\, e^{-E_d/kT}\left(\frac{N_D - N_A}{N_A^2}\right)\right]^{1/2}. \quad (3.141)$$

An approximate solution of the equation can be found by applying $\sqrt{1+x} \cong 1 + \frac{1}{2}x$ (for $x \ll 1$) to the square root term of the equation.

$$\boxed{n \cong \tfrac{1}{2} N_c \frac{N_D - N_A}{N_A} e^{-E_d/kT}} \quad (3.142)$$

Note that the temperature dependence of the carrier concentration as a function of temperature has a different slope as compared with the uncompensated semiconductor (see Eq. (3.127)). The slopes are different by a factor of two for the compensated and uncompensated case.

As the temperature is further increased, $N_D \gg \frac{1}{2} N_c \exp(-E_d/kT) \gg N_A$ and Eq. (3.140) simplifies to

$$n \cong \left(\tfrac{1}{2} N_D N_c\right)^{1/2} \exp\left(-\frac{E_d}{2kT}\right) \quad (3.143)$$

which is identical to the uncompensated case given by Eq. (3.127).

Even further increase of the temperature results in fully ionized donors. The free carrier concentration is then given by $n = N_D^+ - N_A = N_D - N_A$. The two different slopes for the carrier concentration vs. temperature have indeed been observed experimentally (Morin, 1959).

3.4 Law of mass action

The law of mass action is frequently applied to chemical reactions and allows one to calculate the concentration of reagents as a function of temperature. Consider the two starting reagents A and B and the products C and D. The chemical

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

reaction between the reagents can be written as

$$A + B \overset{\Delta G}{\leftrightarrow} C + D \tag{3.144}$$

where ΔG is the change in Gibbs free energy of the reaction. We refer to the introduction of this chapter and assume that $\Delta G \cong \Delta H \cong \Delta E$, i.e. the reaction enthalpy is a good approximation of the Gibbs free energy. ΔE is defined to be positive and negative for *endothermic* and *exothermic* reactions, respectively. Under thermal equilibrium conditions, the reaction of Eq. (3.144) follows the *principle of detailed balance* which states that at a given temperature the reaction rate of a process and the reaction rate of the reverse process balance in detail independent of other processes. Therefore, the *concentrations* of the reagents, [A], [B], [C], and [D] are *constant*. The ***law of mass action*** states that the ratio of the products of the concentrations of the final and initial reagents of the reaction is a constant which depends only on (i) temperature and (ii) the energy of the reaction, i.e.

$$\boxed{\frac{[C][D]}{[A][B]} = K = c \exp\left[-\frac{\Delta E}{kT}\right]} \tag{3.145}$$

where K and c are constants. At a given temperature the ratio of the concentrations is a constant. For low temperatures ($T \to 0$) and positive ΔE the right-hand side of Eq. (3.145) approaches zero and consequently $[C][D] \to 0$. Thus, the equilibrium of the endothermic chemical reaction is strongly shifted to the left-hand side in Eq. (3.144) at low temperatures.

The law of mass action is also applicable to *electronic reactions* in semiconductors. Such electronic processes are, for example, the ionization process of an impurity, the band-to-band excitation of an electron–hole pair, and other electronic equilibrium reactions. The application of the law of mass action to such processes will be discussed below.

Note that the law of mass action does not tell us anything about reaction *rates*. Such rates depend on the activation energy, which is illustrated by a potential hill in Fig. 3.22. Low activation energies result in fast reaction rates. However, the ratio of the concentrations of the reagents under equilibrium conditions (see Eq. (3.145)) depends only on ΔE. Note further that the law of mass action applies to *diluted* chemical solutions. Application of the law of mass action to electronic processes in semiconductors is therefore limited to *non-degenerate* carrier

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

337-TA-1366 RX-0255

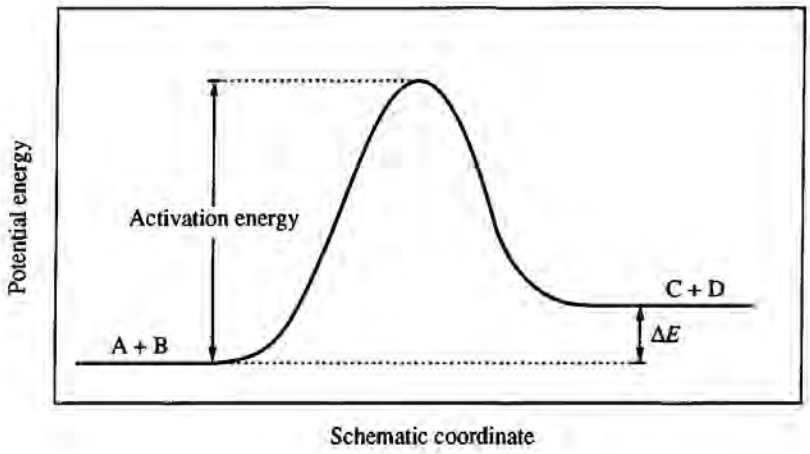


Fig. 3.22. Schematic potential energy diagram of a chemical reaction with the initial reagents A and B and the final reagents C and D. The equilibrium constant is determined by ΔE. The reaction rate is determined by the activation energy.

concentrations (Madelung, 1970). In such non-degenerate semiconductors, the (effective) density of states (in either the conduction or the valence band) is much larger than the free carrier density. Thus, the free carrier concentration is not limited by the finite density of states.

3.4.1 Intrinsic semiconductors

In an undoped semiconductor, free carriers are solely generated by excitation of an electron from the (filled) valence band to the (empty) conduction band. Excitation of an electron from the top of the valence band to the bottom of the conduction band requires the gap energy, E_g. (For a discussion on how the excitation energy relates to free energy, entropy, and enthalpy, we refer to the introduction of this chapter). The band-to-band excitation process can be expressed as

$$\mathrm{VE} \overset{E_g}{\leftrightarrow} \mathrm{e} + \mathrm{h} \tag{3.146}$$

where VE is a valence electron, and h and e are a free hole in the valence band and a free electron in the conduction band, respectively. The concentration of h and e are p and n, respectively. The concentration of valence electrons is very large and is not significantly changed by the process of Eq. (3.146). It is therefore reasonable to assume that [VE] = const. The law of mass action is then given in

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

RX-0255.155

337-TA-1366 RX-0255

analogy to Eqs. (3.144) and (3.145) by

$$np = c\, e^{-E_g/kT} \tag{3.147a}$$

where c is a constant. This result has the same functional dependence as Eq. (3.119), which was obtained from semiconductor statistics and the charge neutrality condition. The as yet unknown constant c is determined by comparison with Eq. (3.119) to be $c = N_c N_v$. With the definition of the intrinsic carrier concentration $n_i = (np)^{1/2}$ one obtains

$$\boxed{np = n_i^2 = N_c N_v e^{-E_g/kT}} \tag{3.147b}$$

The exponential dependence of the intrinsic carrier concentration has thus been derived using a second independent method in terms of the law of mass action.

3.4.2 Extrinsic semiconductors (single donor species)

For the calculation of the temperature dependence of extrinsic semiconductors we consider first an n-type semiconductor with donor energy E_d. The ionization process can be written as

$$D^0 \overset{E_d}{\leftrightarrow} D^+ + e \tag{3.148}$$

where D^0 and D^+ are neutral and ionized donors, respectively. If the concentration of D^0, D^+ and e are given by N_D^0, N_D^+ and n, respectively, then the law of mass action is

$$\frac{n\, N_D^+}{N_D^0} = c\, e^{-E_d/kT}\,. \tag{3.149a}$$

The equation can be simplified by considering $n \ll N_D$. Recalling that $n = N_D^+$ and with $N_D^0 = N_D - N_D^+ \cong N_D$ one obtains

$$n^2 \cong c\, N_D\, e^{-E_d/kT} \tag{3.149b}$$

which is the same functional dependence as Eq. (3.127). Comparison of the two equations yields the value of the constant $c = \frac{1}{2}N_c$. The electron concentration is then given by

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

RX-0255.156

$$\boxed{n \cong (\tfrac{1}{2} N_D N_c)^{1/2} e^{-E_d/2kT}} \quad (3.149c)$$

Note that the factor of 2 arises from the twofold spin-degeneracy of the donor.

3.4.3 Extrinsic semiconductors (two donor species)

Consider a semiconductor with two types of dopants with different thermal ionization energies. For simplicity we assume a shallow and a deep donor with ionization energies E_{sd} and E_{dd}, respectively. At low temperature, the deep donor is assumed to be neutral and the reaction is given by

$$SD^0 \overset{E_{sd}}{\leftrightarrow} SD^+ + e \quad (3.150)$$

where SD^0 and SD^+ refer to the neutral and ionized shallow donor, respectively. As shown in the preceding section, this reaction leads to the concentration

$$n \cong (\tfrac{1}{2} N_{SD} N_c) \exp(-E_{sd}/2kT) . \quad (3.151)$$

As the temperature is further increased all shallow donors are ionized, i.e. $N_{SD}^+ = N_{SD}$ and deep donors start to become ionized. The corresponding reaction is

$$DD^0 \overset{E_{dd}}{\leftrightarrow} DD^+ + e \quad (3.152)$$

where DD^0 and DD^+ are the neutral and ionized donor, respectively. Application of the law of mass action yields

$$\frac{N_{DD}^+ n}{N_{DD}^0} = c \exp(-E_{dd}/kT) . \quad (3.153a)$$

This equation can be simplified by considering the temperature range in which only a few deep donors are ionized, i.e. $N_{DD}^+ \ll N_{DD}$ and $N_{DD} \cong N_{DD}^0$. Since the free electron density is the sum of shallow and ionized deep donor concentrations, i.e. $n = N_{SD} + N_{DD}^+$ one obtains

$$n(n - N_{SD}) \cong cN_{DD} \exp(-E_{dd}/kT) \quad (3.153b)$$

which has the same functional dependence as Eq. (3.136b). A comparison of the two equation allows one to determine the constant $c = \frac{1}{2} N_c$. The electron concentration is then given by

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

$$n(n - N_{SD}) \cong \tfrac{1}{2} N_{DD} N_c \exp(-E_{dd}/kT) \,. \tag{3.153c}$$

3.4.4 Compensated semiconductors

In n-type semiconductors, donors are the dominant dopant species; however some residual acceptors are always present as well. Even at very low temperatures, some donors are ionized, because of the availability of lower energy acceptor states. For $T \to 0$ the free carrier concentration tends to zero, $n \to 0$, but $N_D^+ = N_A^- = N_A$. As the temperature is increased, additional donors become ionized. The corresponding process is

$$D^0 \overset{E_d}{\leftrightarrow} D^+ + e \tag{3.154}$$

where E_d is the thermal ionization energy of the donor. If the concentrations of D^0, D^+ and e are N_D^0, N_D^+, and n, respectively, then the law of mass action is

$$\frac{n N_D^+}{N_D^0} = c\, e^{-E_d/kT} \,. \tag{3.155a}$$

The equations can be simplified by considering low temperatures; then $n \to 0$, $N_D^+ \cong N_A$, and $N_D^0 \cong N_D - N_A$ which yields

$$n \cong c \frac{N_D - N_A}{N_A} e^{-E_d/kT} \tag{3.155b}$$

which has the same functional dependence as Eq. (3.143) which in turn was obtained by statistical considerations. The constant c is obtained by comparison with Eq. (3.143) and has a value of $c = \tfrac{1}{2}N_c$. The law of mass action for a partially compensated semiconductor then is given by

$$\boxed{n \cong \tfrac{1}{2} N_c \frac{N_D - N_A}{N_A} e^{-E_d/kT}} \tag{3.155c}$$

Note that the slope of n versus reciprocal temperature is described by the activation energy E_d.

If the temperature is increased further and $N_A \ll n \ll N_D$ then $n = N_D^+ - N_A \cong N_D^+$ and $N_D^0 = N_D - N_D^+ \cong N_D$. Insertion of these conditions into Eq. (3.155a) yields

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

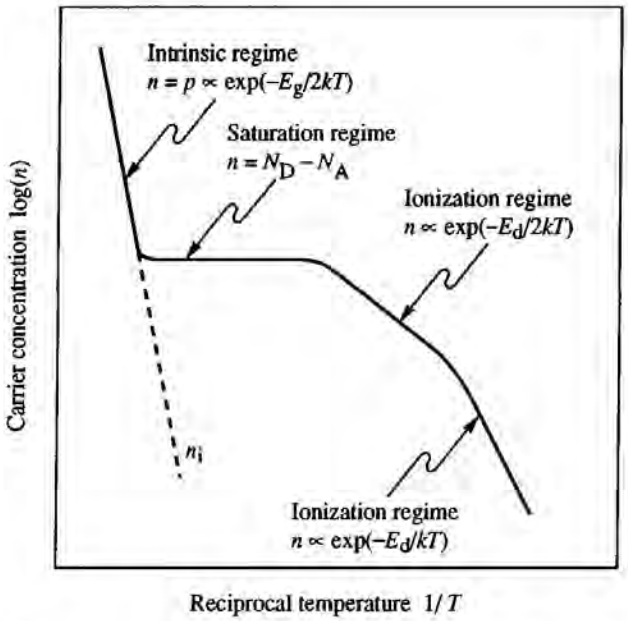


Fig. 3.23. Carrier concentration as a function of reciprocal temperature for a partially compensated n-type semiconductor. The ionization regime is characterized by two different activation energies (E_d and $E_d/2$).

$$n \cong \left(\tfrac{1}{2} N_c N_D\right)^{1/2} e^{-E_d/2kT} \quad (3.156)$$

Note that the slope of n versus reciprocal temperature is described by the activation energy $\frac{1}{2}E_d$, which is a factor of two smaller than the activation energy of the compensated semiconductor of Eq. (3.155c). The carrier concentration versus temperature of a compensated semiconductor is schematically illustrated in Fig. 3.23, where the different regimes are indicated.

https://doi.org/10.1017/CBO9780511599828.006 Published online by Cambridge University Press

4

Growth technologies

The two growth technologies molecular-beam epitaxy (MBE) and vapor-phase epitaxy (VPE) are frequently employed for epitaxial growth, i.e. the crystalline growth of a thin layer on a crystalline substrate. High-quality epitaxial semiconductors with well controlled doping, composition, and thickness can be grown by MBE and VPE. Furthermore, the two growth methods MBE and VPE are capable of growing atomically abrupt doping and compositional profiles. This chapter covers the basic concepts of MBE, VPE, and related growth technologies.

4.1 Molecular-beam epitaxy

Molecular-beam epitaxy (MBE) is a powerful epitaxial growth technique whose historical origin dates back to the 1950s when Günther (1958) developed the 'three temperature method'. For the epitaxial growth of a compound semiconductor, e.g. GaAs, by coevaporation, the temperatures of the effusion cells (Ga and As) and the GaAs substrate must conform with the three temperature rule. Günther showed that the substrate temperature must be *between* the two effusion cell temperatures for stoichiometric growth. In the experiments of Günther, the temperature of the group-V oven (As, Sb) was much lower than the temperature of group-III (In) oven. The substrate temperature was kept at a temperature between the two oven temperatures. Stoichiometric, polycrystalline growth of InAs and InSb was demonstrated on glass substrates using the three temperature method. The choice of the group-V to group-III flux ratio is of crucial importance. The group-V element flux is selected to be larger than the group-III flux. Excess group-V atoms on the substrate reevaporate from the substrate (due to its higher temperature) except for the group-V atoms that have reacted with group-III atoms on the substrate. Stoichiometric growth of

https://doi.org/10.1017/CBO9780511599828.007 Published online by Cambridge University Press

compound semiconductors thus requires excess group-V pressure.

The first epitaxial growth of III–V semiconductors was achieved by Davey and Pankey (1968). They used single-crystal GaAs substrates and a simple glass tube as a growth chamber.

The technology of molecular-beam epitaxy, as it is known today, is mainly due to Cho and coworkers (Cho, 1971; Cho and Casey, 1974; Cho and Arthur, 1975; Cho, 1985). Among the innovations introduced by Cho are a large area liquid N_2 cryoshroud encircling effusion cells and substrate, improved beam flux controllability, separate effusion cells for the group-III and group-V elements, and increased cell capacity. He demonstrated n-type and p-type conductivity in GaAs using Si, Sn, and Mg as doping elements. He made the first use of RHEED in an MBE system. Furthermore, he demonstrated the suitability of high-quality epitaxial layers for microwave and optoelectronic devices. In the 1980s the technology of MBE was further advanced by a large number of improvements including a load-lock sample introductory chamber, rotatable substrate holder, and better vacuum conditions.

The epitaxial growth of GaAs was studied in detail by several groups. Arthur (1967, 1968) investigated the kinetic behavior of Ga and As_2 beams impinging onto GaAs surfaces. The atomic and molecular beams, were obtained by evaporation of GaAs under ultra-high vacuum (UHV) conditions. The chopping of the atomic and molecular beams made possible the study of the growth kinetics. The growth kinetics of GaAs growth were studied in further detail by Foxon et al. (1974) and by Foxon and Joyce (1975, 1977) using a modulated molecular beam and reflection high-energy electron diffraction (RHEED). The thermal accommodation coefficients, surface lifetimes, sticking coefficients, desorption energies, and the reaction order were determined in their studies. The present understanding of GaAs growth from Ga and As_4/As_2 beams emerged mainly from their experiments.

The basic components of an MBE system which are inside the ultra-high vacuum chamber of the growth apparatus are schematically shown in Fig. 4.1. The constituents of the semiconductor are evaporated from effusion cells which are equipped with mechanical shutters. The shutters allow one to initiate and to terminate either the epitaxial growth or the doping. The substrate holder is rotated during growth for better uniformity. Several instruments are included within typical MBE systems which allow *in situ* analysis. These include a reflection high-energy electron diffraction (RHEED) electron gun with fluorescent screen, a residual gas analyzer (RGA), an Auger electron spectrometer (AES), and a mass spectrometer. In the following, the essential components of a MBE system and the materials used for the system will be discussed.

https://doi.org/10.1017/CBO9780511599828.007 Published online by Cambridge University Press

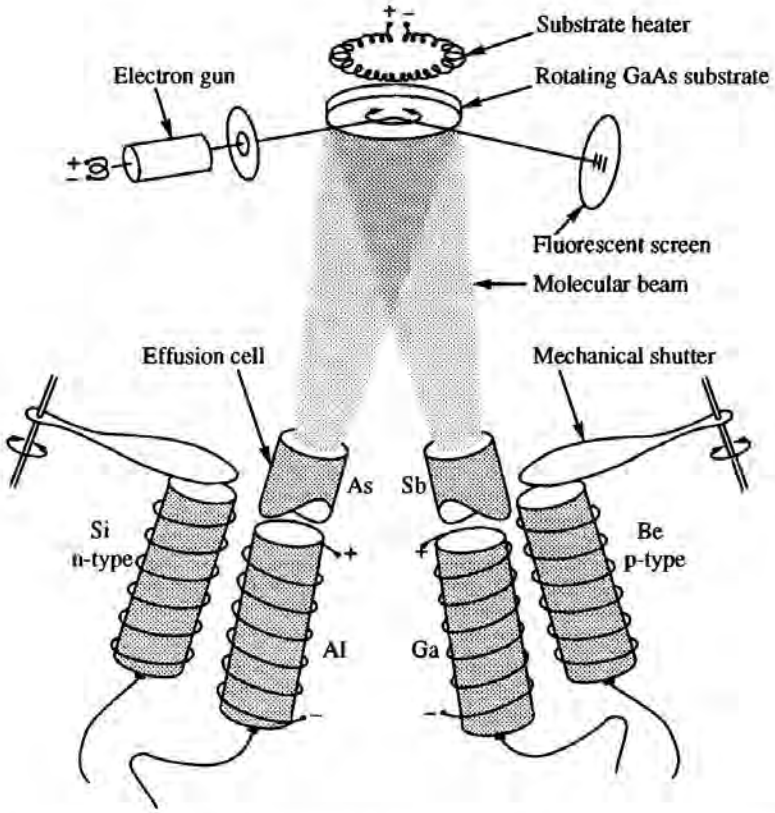


Fig. 4.1. Sketch showing the essential components of a molecular-beam epitaxy (MBE) system consisting of resistively heated Knudsen cells with shutters and the substrate which is rotating for better uniformity. A reflection high-energy electron diffraction (RHEED) gun and fluorescent screen are used to determine the reconstruction of the semiconductor surface. The components shown are enclosed in ultra-high vacuum (UHV). A cryoshroud (not shown) filled with liquid nitrogen encircles the growth area and the effusion cells.

4.1.1 Materials

The materials used in an MBE system need to fulfill several requirements due to the large temperature differences and the ultra-high vacuum conditions within the MBE apparatus. The temperatures range from $T = 77$ K for the cryoshroud filled with liquid N_2 to temperatures exceeding 1400 °C occurring in high temperature effusion cells. Furthermore, the low pressures of 10^{-10} Torr require that the materials do not outgass even at high temperatures. Thus, the requirements for materials used in the MBE apparatus are (i) low vapor pressure at high temperatures, (ii) low outgassing at high temperatures, (iii) small thermal expansion coefficient, and (iv) good mechanical properties.

https://doi.org/10.1017/CBO9780511599828.007 Published online by Cambridge University Press

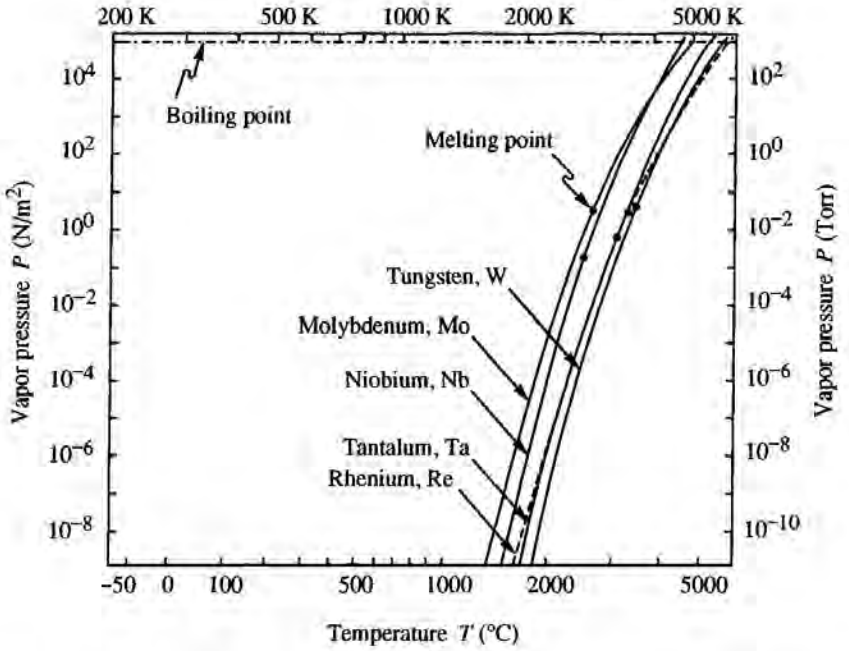


Fig. 4.2. Vapor pressure data of metals frequently used for high-temperature components of MBE systems.

The vapor pressure curves of metals with very low vapour pressure are shown in Fig. 4.2. Tungsten (W) is the metal with the lowest vapor pressure. Other metals with slightly higher vapor pressure are tantalum (Ta), molybdenum (Mo), rhenium (Re), and niobium (Nb). These metals are frequently used for the high temperature components of the MBE system due to the requirement of low vapor pressure. Tungsten is frequently used for heating wires; Ta for heat shields and heater ribbons; Mo for the substrate mounting block; Re and W in thermocouples ($W_{0.95}Re_{0.05}/W_{0.74}Re_{0.26}$). Stainless steel is used for parts not heated to very high temperatures but which require high mechanical stability. For example, stainless steel is used for the vacuum chamber walls and for the cryoshroud filled with liquid N_2.

Dielectric materials used in an MBE chamber are required to have low vapor pressure, low outgassing, small thermal expansion, and good mechanical properties. Boron nitride (BN), graphite (C), and quartz (SiO_2) fulfill these requirements. Very pure BN can be made by pyrolysis. Such pyrolytic BN is used for the crucibles of effusion cells and for the tubes of cracker cells. Graphite is used for some crucibles of effusion cells.

https://doi.org/10.1017/CBO9780511599828.007 Published online by Cambridge University Press

4.1.2 Effusion cells

A critical part of an MBE system are the effusion cells which are used to evaporate the constituent elements of the semiconductor as well as the dopants. The effusion cells are of Knudsen-type geometry. In this geometry, collisions between evaporated atoms (or molecules) are much less frequent than collisions between atoms (or molecules) and the crucible wall. As a result, the flux distribution from a cylindrical effusion cell with Knudsen-type geometry is jet-like and can be expressed by a cosine function (Knudsen, 1909). The distribution of atoms or molecules arriving at the substrate is cosine-like as well.

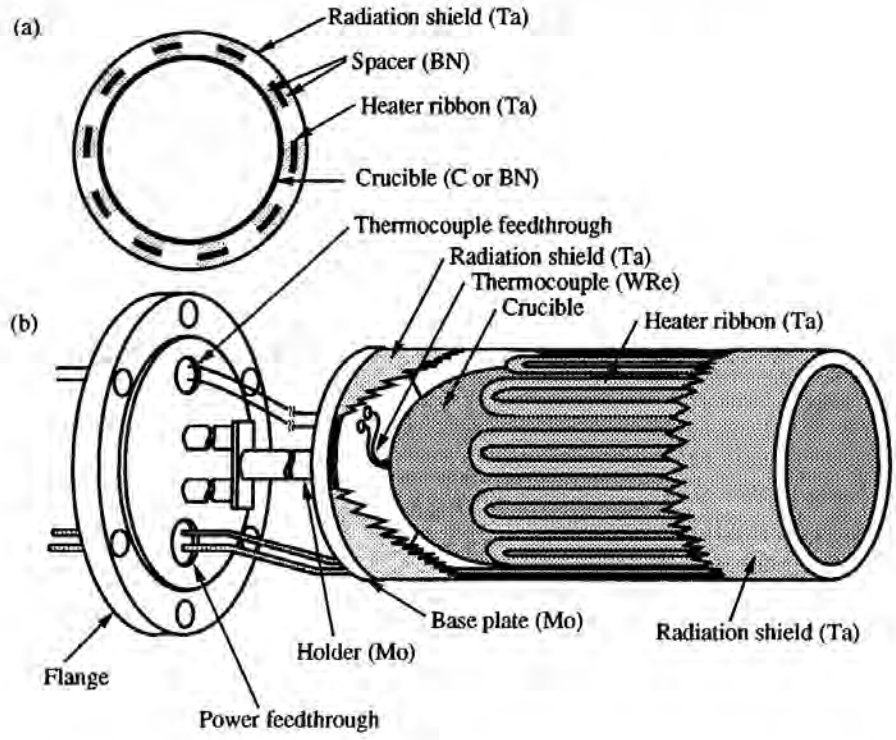


Fig. 4.3. (a) Top-view and (b) side-view of a Knudsen-type effusion cell with a cylindrical crucible. The crucible is heated by low-induction Ta ribbons and heat-shielded by Ta foil. A thermocouple is used to measure the temperature of the crucible. The effusion cell is mounted on a conflat flange with feedthroughs for the heater and the thermocouple.

A sketch of an effusion cell mounted on a conflat flange is shown in Fig. 4.3. The essential components of an effusion cell are the crucible, heating wire or ribbon, heat shield, and thermocouple. The thermocouple touches the bottom of the crucible for good thermal contact and temperature accuracy. The crucible is

https://doi.org/10.1017/CBO9780511599828.007 Published online by Cambridge University Press

337-TA-1366 RX-0255

typically made out of BN or C. Cylindrically shaped as well as conically shaped crucibles are employed. Conically shaped crucibles have better flux homogeneity as compared to cylindrically shaped ones.

The heater of the effusion cell typically consists of W wire or Ta ribbon. The heater filament is mechanically kept in place by BN spacers. It is desirable for the heater filament to have a low inductance. The low-inductance wiring makes possible the employment of alternating current (ac) supplies for the heater and is also aimed to reduce magnetic fields which would perturb the beam of the RHEED electron gun. Therefore, a meander-type rather than the spiral-type filament pattern is used in effusion cells.

The extreme temperature differences occurring in and around effusion cells require effective shielding. Effusion cells are surrounded by liquid N_2 of $T = 77$ K. On the other side, temperatures exceeding 1400 °C occur at the heater of high-temperature effusion cells. To avoid radiation losses, effusion cells are shielded by thin metal foils mounted on a base plate. Tantalum is frequently used as a shielding foil. Molybdenum is used for the base plate supporting the Ta shield. Some effusion cells use water cooling to accommodate the high temperatures of the heater and the crucible. However, the provision of sufficient shielding allows one to avoid water cooling. Uncooled effusion cells are available for temperatures up to 1400 °C, which is sufficient for most applications.

The symmetry axis of an effusion cell points towards the center of the substrate. This geometry is used for all MBE systems with a non-rotating substrate to provide best thickness uniformity of the epitaxial films. In MBE systems with rotating substrates, the symmetry axis of the effusion cell points off the center of the wafer. The off-center focal point provides uniformity for the largest possible area.

4.1.3 Mechanical shutter

The mechanical shutters allow one to rapidly stop or initiate the molecular fluxes of effusion cells. The shutters can thus be used to stop and initiate growth and doping of the semiconductor. The time required to open and to close a shutter must be much shorter than the time required to grow a monolayer of the semiconductor. Such fast shutters are required to grow atomically abrupt interfaces. Since typical growth rates are 1 monolayer/s ≡ 2.8 Å/s, opening and closing times of 0.1 s are desirable.

Several shutter constructions are employed in MBE systems including rotating shutters and sliding shutters. An example of a sliding shutter is shown in Fig. 4.4. The shutter is magnetically coupled from the UHV side to the

https://doi.org/10.1017/CBO9780511599828.007 Published online by Cambridge University Press

RX-0255.165

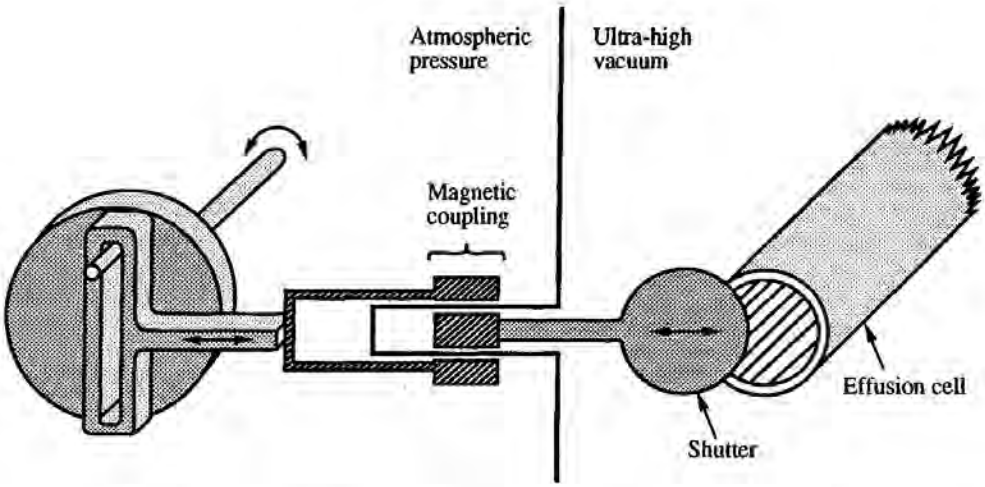


Fig. 4.4. Schematic illustration of a mechanical shutter with translational movement. The coupling between the UHV and the atmospheric pressure environment is provided by magnets. The shutter has 'soft stops', i.e. the translational velocity of the shutter continuously decreases to zero velocity at the end points of the shutter movement.

atmospheric pressure side. The magnets on the atmospheric pressure side move back and forth by means of a rod with an elongated opening. A disk driven by a motor provides the translational motion of the actuator rod. The shutter design shown in Fig. 4.4 has proven to be exceptionally reliable and can be used for thousands of opening and closing cycles. A particular advantage of the shutter is its soft start and stop. That is, the velocity of the actuator rod goes continuously to zero at its starting and at its stopping point. This 'soft' action is due to the conversion of the rotational movement of the disk into the translational movement of the shutter. Thus, the shutter velocity is a sinusoidal function of time. The two end points of the shutter movement are the zeros of the sinusoidal function.

4.1.4 Substrate holder and temperature

The substrate holder provides a mechanical means of holding and transporting the semiconductor substrate in and out of the MBE chamber. The mounting of the semiconductor substrate is either done by soldering with liquid indium or by indium-free clamping of the substrate onto the holder. The material used for substrate holders is Mo which is characterized by a low vapor pressure and little outgassing.

https://doi.org/10.1017/CBO9780511599828.007 Published online by Cambridge University Press

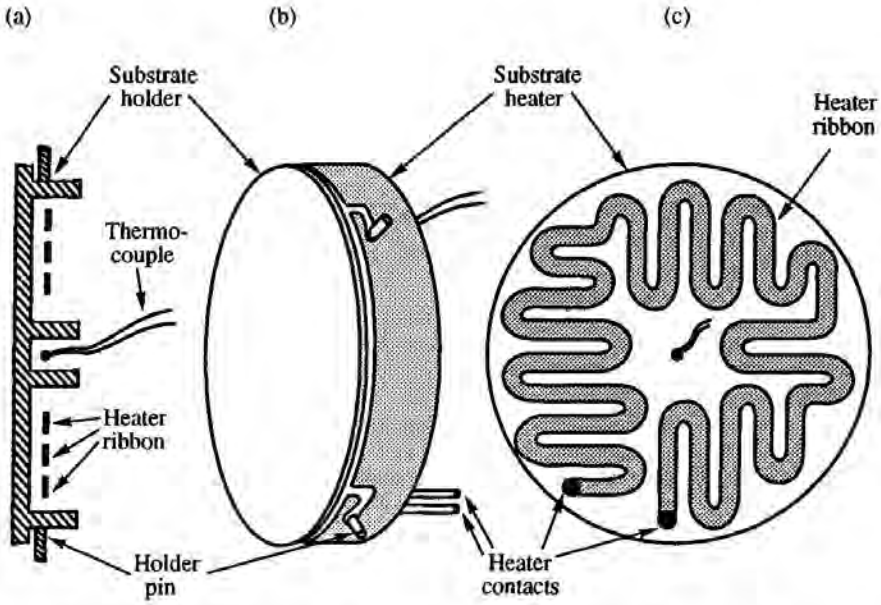


Fig. 4.5. Schematic illustration of a substrate holder (a,b) and a substrate heater (b, c) of an MBE system. The thermocouple is surrounded by the substrate holder to provide good thermal contact. The thermocouple does not 'see' the heating ribbon which is low-inductance meander-shaped in order not to perturb the beam of the RHEED electron gun.

An example of a substrate holder is shown in Fig. 4.5, which also shows the heater ribbon and the thermocouple. The heater ribbon is arranged in a low-inductive pattern. Such a pattern reduces magnetic fields which deflect the beam of the RHEED electron gun. Either tantalum (Ta) heater ribbons or tungsten (W) heater wire are used. However, to date, Ta heater ribbons are more popular since they are less brittle than W wire. Thermocouples are used to measure the substrate temperature.

The calibration and measurement of the substrate temperature can be done using several methods. At the congruent sublimation temperature of the substrate (e.g. 640 °C for GaAs) the surface reconstruction changes from a 2×4 reconstruction to a 3×4 reconstruction. The reconstruction can be monitored by means of the RHEED pattern. The congruent sublimation temperature can then be used as a calibration point for the thermocouple measurement. Another method of measuring the substrate temperature makes use of the InSb melting point at 525 °C. The solid-to-liquid transition can be visually observed through a viewport. The congruent sublimation temperature and the InSb melting temperature can then be used as calibration temperatures for the thermocouple.

https://doi.org/10.1017/CBO9780511599828.007 Published online by Cambridge University Press

337-TA-1366 RX-0255

Another convenient method of measuring substrate temperatures makes use of a pyrometer which analyzes the radiant spectrum of the substrate holder. Comparison with the known black-body radiation spectrum allows one to estimate the substrate temperature. The pyrometer method can be used if the emissivity of the substrate is known. Furthermore, the windows of the MBE apparatus must be clean and uncoated to provide the true radiant spectrum of the substrate.

Another method of measuring the GaAs substrate temperature was proposed by Hellman and Harris (1987). The GaAs fundamental absorption edge shifts by approximately 50 meV during heating from 300K to typical MBE growth temperatures and can be measured by using the infra-red light emitted by the substrate heater filament. The light transmitted through the GaAs substrate can be measured and the fundamental absorption edge can be determined. The authors claimed an accuracy of ±10 °C and a precision of ±2 °C of this method.

4.1.5 Pumps

The ultra-high vacuum of an MBE system require high-efficiency pumping systems. A number of different pumping systems are used in MBE systems which are all based on different physical principles. ***Ion getter pumps*** have an *ionization stage* in which residual gas atoms are ionized. The ionized atoms (or molecules) are then accelerated toward a metal foil (Ta) by means of a magnetic and electric field. The gas atoms impinge with high impact onto the metal foil where they are adsorbed. ***Cryopumps*** consist of a cold stage which is cooled to approximately 10 K by means of a closed cycle He refrigeration system. At the low temperature of 10 K most gases (e.g. O_2, N_2, CO_2) condense at the cold stage. The residual pressures achievable with cryopumps are very low and can be in the 10^{-11} Torr range. ***Titanium sublimation pumps*** consist of a Ti coil that is periodically heated for several seconds. The metal evaporates from the coil and is deposited on the inside walls of the MBE chamber. The freshly deposited Ti layer is very reactive and getters gas molecules from the vacuum. The Ti getter pump is frequently used during bake out times when the evaporating Ti does not interfere with the growing crystal. ***Turbo pumps*** are turbines with high rotation speed. Turbo pumps are capable of high pumping volumes which are required in gas-source MBE and chemical-beam epitaxy systems. ***Liquid N_2 sorption pumps*** are used to rough-pump UHV systems after venting. The sorption pumps are activated by cooling to liquid N_2 temperature, where the pumping action is due to the highly absorptive molecular sieve in the pump. Sorption pumps are preferred over mechanical roughing pumps due to their cleanliness.

https://doi.org/10.1017/CBO9780511599828.007 Published online by Cambridge University Press

RX-0255.168

4.2 Gas-source molecular-beam epitaxy

The solid sources of group-III and group-V elements of a molecular-beam epitaxy (MBE) growth system can be replaced by gaseous sources. These sources are gaseous compounds containing the group-III, group-V, or doping elements. Compounds used are hydrides (e.g. AsH_3, PH_3) for group-V elements and organometallics [e.g. $Ga(CH_3)_3$, $Al(C_2H_6)_3$] for group-III elements. In this section we restrict our discussion to the replacement of group-V elements with gas sources, i.e. *hydride* sources. The group-III elements are supplied by solid sources as in conventional MBE. For brevity we refer to hydride gas-source MBE as GSMBE.

The motivation for the usage of hydrides for the group-V elements is due to the difficulty of conventional MBE to grow compounds containing phosphorus. Phosphorus-containing compounds such as $Ga_xIn_{1-x}P_yAs_{1-y}$ lattice matched to InP substrates are of paramount importance for long-wavelength (λ = 1.3 – 1.6 μm) optoelectronic components such as light-emitting and light-detecting devices. However, the evaporation of red phosphorus from a solid source results in tetramers (P_4) in the gas phase. Phosphorous tetramers have an extremely low sticking coefficient ($<10^{-3}$) on InP substrates (Tsang, 1985b). Therefore, phosphorus is hardly incorporated during growth of $Ga_xIn_{1-x}P_yAs_{1-y}$. Furthermore, it is difficult to control the composition 'y' of the quaternary compound due to the low sticking coefficient of P_4. To date it is well accepted that high quality $Ga_xIn_{1-x}P_yAs_{1-y}$ epitaxial films with good compositional control cannot be grown by solid-source MBE.

The group-V solid sources of As and P were replaced with arsine (AsH_3) and phosphine (PH_3) by Panish (1980). The replacement was intended to achieve a well-controlled flux of phosphorus and phosphorus plus arsenic for the MBE growth of heterostructures of $Ga_xIn_{1-x}P_yAs_{1-y}$ and InP on InP substrates. A schematic illustration of a GSMBE system is shown in Fig. 4.6, which has solid, elemental effusion cells for the use of group-III elements and dopants. The hydrides arsine and phosphine are thermally cracked into atoms and dimers (i.e. P, P_2, As, and As_2) and enter the MBE chamber through a single port. Many other components of the GSMBE system such as substrate holder, cryoshroud etc. are identical or similar to components in conventional MBE systems.

The flux of the hydrides into the hydride cracker cells is controlled outside the high-vacuum growth chamber. The flow of hydrides can be controlled by either high precision *flow* or high precision *pressure* controllers. Ventilation lines are frequently employed to avoid flow transients. Flow transients typically occur as a gas flow is turned on or off until a steady-state flow rate is achieved. Such flow

https://doi.org/10.1017/CBO9780511599828.007 Published online by Cambridge University Press

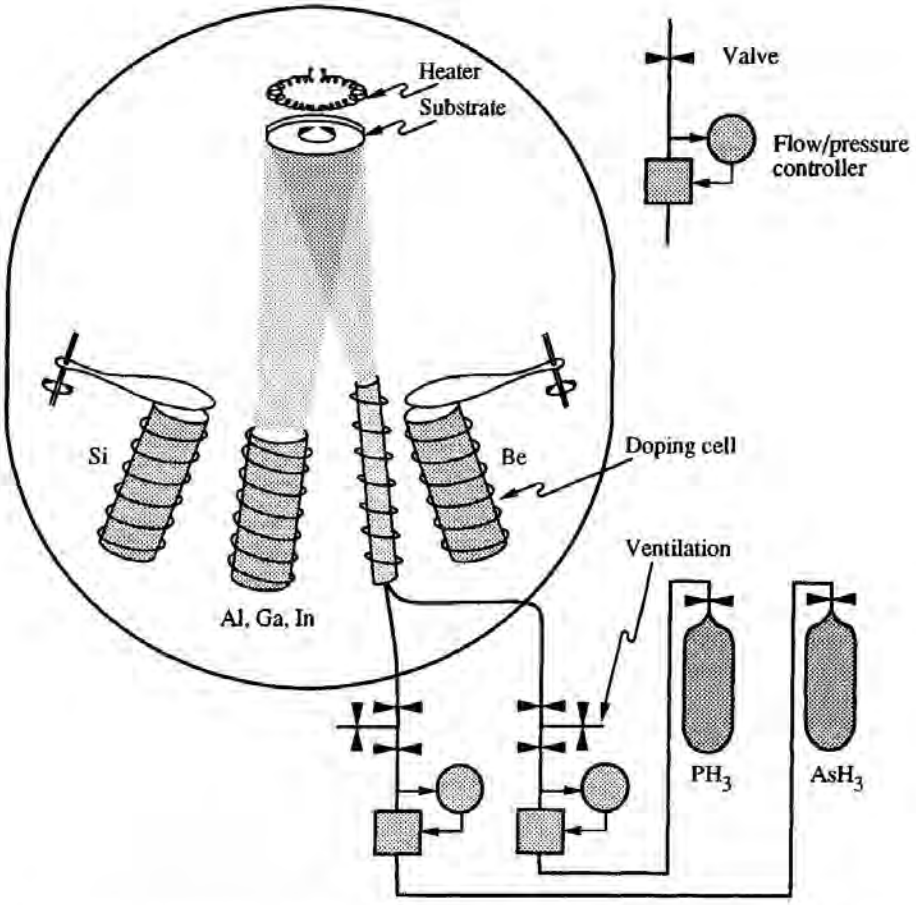


Fig. 4.6. Schematic illustration of a gas-source molecular-beam epitaxy (GSMBE) system. Group-V elements are provided as hydrides (AsH_3 and PH_3) which are thermally cracked upon entering the high vacuum chamber. Group-III elements and doping elements are evaporated from effusion cells.

transients can be avoided by using ventilation lines. The gas flow is not switched on or off but rather directed into the MBE chamber or to the ventilation line using three valves as illustrated in Fig. 4.6.

It is desirable to thermally crack the hydrides into their constituent elements using a cracker cell, since uncracked hydrides result in insufficient incorporation of group-V elements into the III–V film. The use of uncracked hydrides relies on the reaction and decomposition at the heated substrate surface which was shown to be insufficiently small (Panish, 1980; Veuhoff et al., 1981). Thus, hydrides need to be cracked upon entering the MBE chamber.

A schematic illustration of a hydride cracker cell is shown in Fig. 4.7. The

https://doi.org/10.1017/CBO9780511599828.007 Published online by Cambridge University Press

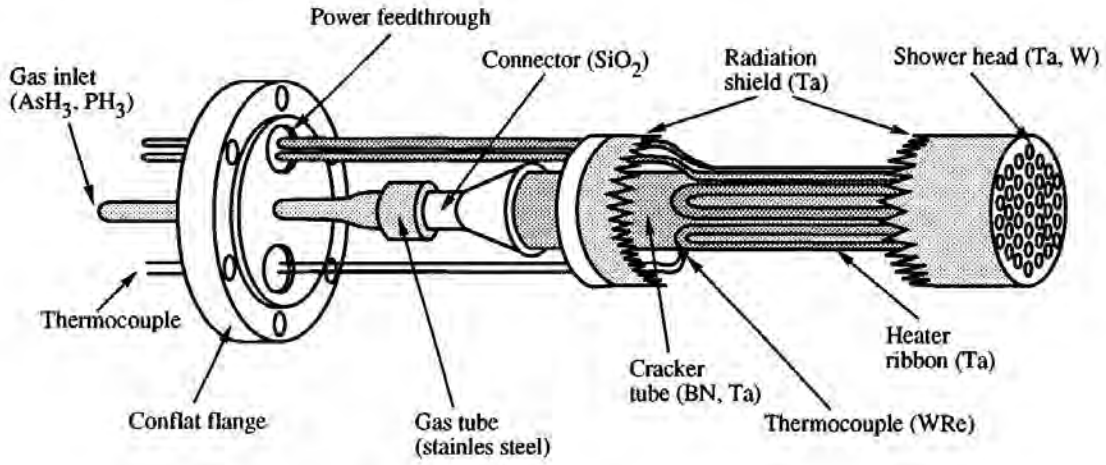


Fig. 4.7. Schematic illustration of a cracker cell used to thermally decompose arsine and phosphine. The cracker tube may be made of a material (e.g. Ta) with cathalytic characteristics to faciliate cracking. The cracker tube is heated by a Ta ribbon and is monitored by a thermocouple. A 'shower head' is provided at the exit of the cracker cell to provide a homogeneous flux of group-V elements. The cracker cell is mounted on a conflat flange with feedthroughs for the heater and the thermocouple.

cell is mounted on a conflat flange for compatibility with conventional MBE vacuum ports. Thermal cracking is achieved in the cracker tube which is heated by a Ta heater ribbon to temperatures of 1000 °C. The cracker tube can be made from several materials including Ta and pyrolytic BN. Tantalum was shown to have catalytic properties for the cracking process and therefore has a higher cracking efficiency at a given temperature as compared to the inert BN (Chow and Chaoi, 1983; Panish et al., 1985). Another method of thermally cracking hydrides was demonstrated by Calawa (1981) who used heated Ta wire for thermal decomposition. As illustrated in Fig. 4.7, the temperature of the cracker tube is monitored by a thermocouple. A Ta foil surrounds the cracker tube as a radiation shield. The exit end of the cracker cell is equipped with a 'shower head' in order to diffuse and homogenize the group-V element beam. The 'shower head' consists of many small holes through which the group-V elements exit the cracker cell.

All group-V elements enter the MBE chamber through a single port. The single port provides good homogeneity if two group-V elements are used simultaneously. That is, the growth of $Ga_xIn_{1-x}P_yAs_{1-y}$ is expected to have good P/As homogeneity and little variation of the composition 'y' across the

https://doi.org/10.1017/CBO9780511599828.007 Published online by Cambridge University Press

337-TA-1366 RX-0255

wafer.

The species created by cracking of hydrides are As and P atoms, As_2 and P_2 dimers, as well as As_4 and P_4 tetramers. Although As and P atoms may be obtained at very high cracker temperatures (>> 1000 °C) and low hydride pressures, the predominant species obtained by cracking are dimers. These dimers are desirable due to their large sticking coefficient on the substrate. A cracker temperature of or in excess of 1000 °C insures that dimer species dominate (Panish and Temkin, 1989).

The range of operating pressures in GSMBE is higher compared to conventional solid-source MBE. The higher pressure is mainly due to hydrogen which is generated by the cracking of arsine and phosphine. The typical operating pressure during GSMBE is in the 10^{-5} Torr range. Appropriate pumping systems for this pressure range are high-capacity cryopumps, turbo molecular pumps, and diffusion pumps. Diffusion pumps are considered unacceptable for solid-source MBE due to possible contamination of the UHV environment of MBE. In contrast, diffusion pumps are occasionally employed in GSMBE systems with no drastic degradation of the material quality.

High-quality epitaxial materials have been grown by GSMBE including $Al_xGa_{1-x}As/GaAs$, $Ga_{0.47}In_{0.53}As/InP$, $Al_{0.48}In_{0.52}As/InP$, and $Ga_xIn_{1-x}P_yAs_{1-y}/InP$. A review on quantum well structures, superlattice *pin* photodetectors, and bipolar transistors grown by GSMBE was given by Panish (1987). He demonstrated $Ga_{0.47}In_{0.55}As$ and $Ga_xIn_{1-x}P_yAs_{1-y}$ quantum wells of thickness 5 Å and near atomic monolayer abruptness. Excitonic enhancement of the photoluminescence and absorption spectra were also demonstrated. In the $Al_xGa_{1-x}As/GaAs$ material system, high-quality selectively doped heterostructures were demonstrated (Schubert et al., 1987). Thus, it is well established that GSMBE is capable of producing device quality epitaxial material in the GaAs as well as in the InP material systems. Even though most of the highest purity material and the highest electron mobilities were reported for growth by MBE, the epitaxial layers grown by GSMBE are of near comparable quality.

A drawback of GSMBE with hydrides is the toxicity of arsine and phosphine. Arsine is an extremely toxic compound. The usage of hydrides necessitates continuous gas analysis and monitoring of the growth facility.

4.3 Chemical-beam epitaxy

The growth of III–V semiconductors by chemical-beam epitaxy (CBE) employs gas sources for both the group-III as well as the group-V elements. Ultra-high vacuum chambers as used for solid-source MBE and hydride GSMBE are also

https://doi.org/10.1017/CBO9780511599828.007 Published online by Cambridge University Press

RX-0255.172

employed for CBE. The motivation for chemical beam epitaxy is to combine the advantages of MBE and organometallic vapor-phase epitaxy (Sect. 4.4).

The sources used for group-V elements are the hydrides of arsenic and phosphorus which are thermally cracked upon entering the reactor chamber. The reagents providing the group-V elements are thus the same as in gas-source MBE. The precursors used for group-III sources are organometallics, i.e. the group-III element is attached to three alkyl molecules. The alkyl groups most frequently used are derived from methane (CH_4) and ethane (C_2H_6). The corresponding organometallic molecules are trimethylaluminum [$Al(CH_3)_3$, TMAl], trimethylgallium [$Ga(CH_3)_3$, TMGa], trimethylindium [$In(CH_3)_3$, TMIn], triethylaluminum [$Al(C_2H_5)_3$, TEAl], triethylgallium [$Ga(C_2H_5)_3$, TEGa], and triethylindium [$In(C_2H_5)_3$, TEIn]. The employment of organometallic compounds for the group-III precursors has given rise to the name of organometallic molecular-beam epitaxy (OMMBE). Frequently, CBE and OMMBE are used as synonyms.

Organometallic group-III sources and hydride group-V sources in an MBE chamber were first used at the beginning of the 1980s (Veuhoff et al., 1981). The authors reported the use of TMGa and arsine for the growth of GaAs. This initial work was not encouraging due to a high carbon incorporation obtained in GaAs and, as was shown later, especially in $Al_xGa_{1-x}As$. High p-type conductivity was obtained in these materials as a result of the C incorporation. Tsang (1984) reported the CBE growth of GaAs and InP using organometallic group-III compounds and thermally cracked arsine and phosphine for the group-V compounds. Low p-type background doping in GaAs grown by CBE of 1.2×10^{15} cm^{-3} was reported by Horiguchi et al. (1986).

A schematic illustration of a CBE system is shown in Fig. 4.8. The components related to the hydrides are identical to the components of the GSMBE system discussed in the previous section. It will not be further discussed here. Both group-III and group-V chemicals are introduced into the high-vacuum reactor in which the molecular beams are directed to the substrate. Many components of CBE systems not related to gas handling are identical to corresponding components of MBE systems. Among these components are the substrate holder and heater, in-situ analysis instruments, and the UHV recipient.

The gas handling of the organometallic precursors is schematically shown in Fig. 4.8. Although many organometallic compounds, such as TMGa or TMAl, are liquid at room temperature, some (e.g. TMGa) have a sufficiently high vapor pressure to enter the reaction chamber at a high enough rate to enable growth. Others, however, (e.g. TEGa) have a lower vapor pressure that is insufficient to drive the compound into the reaction chamber. In such cases of low pressure,

https://doi.org/10.1017/CBO9780511599828.007 Published online by Cambridge University Press

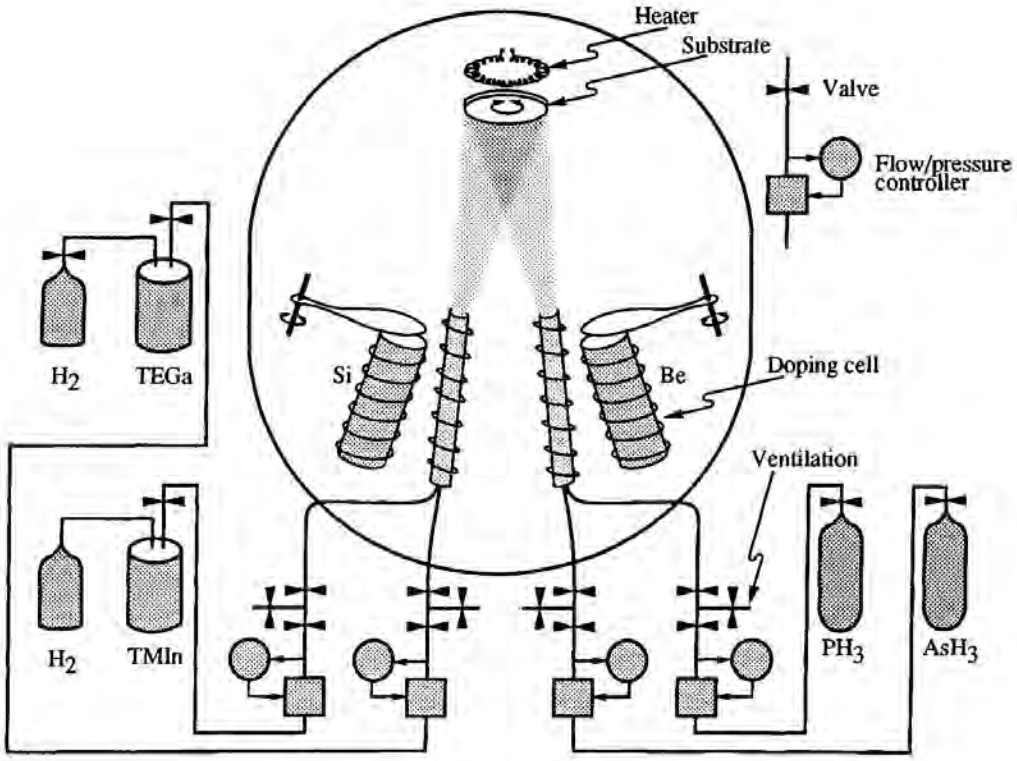


Fig. 4.8. Schematic illustration of a chemical-beam epitaxy (CBE) system which uses organometallic group-III compounds and cracked hydrides as group-V compounds.

organometallic bubblers are used to increase the group-III flux. In such bubblers, an inert gas (e.g. H_2 or N_2) bubbles through the liquid organometallic compound and serves as a carrier gas for the group-III compound. The flow of the carrier gas plus the organometallic compound is regulated by flow or pressure controllers. The flow is then switched between the reaction chamber and a ventilation line, while the flow through the controller is maintained constant. Such a constant flow is provided in order to avoid flux transients (see Sect. 4.2).

The group-III organometallics and their carrier gas are introduced into the high-vacuum recipient through a group-III cell. Figure 4.9 shows a gas introduction cell for organometallics. The two purposes of the cell are to introduce the organometallics into the growth chamber and to provide a homogeneous distribution of the molecular beam inside the reaction chamber. Thermal cracking of the organometallics is *not* provided by the cell, since thermal pyrolysis of the organometallics occurs readily at the heated substrate at

https://doi.org/10.1017/CBO9780511599828.007 Published online by Cambridge University Press

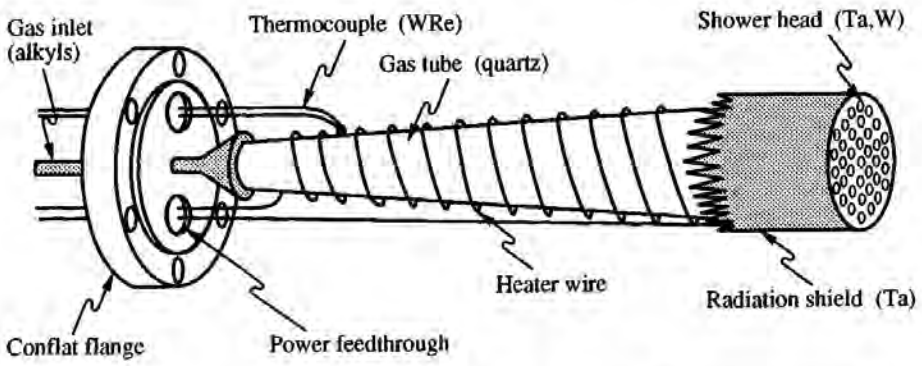


Fig. 4.9. Schematic illustration of a group-III cell for organometallic compounds. The cell is heated to moderate temperatures of 50 °C to avoid condensation of organometallics on the quartz walls of the cell. A 'shower head' provides a uniform flux distribution.

temperatures of 500–700 °C. The group-III cell is heated to a typical temperature of 50 °C. This moderate temperature is selected in order to avoid condensation of organometallics in the cell. This would occur in unheated cells due to the cryoshroud encircling the cell. The cell temperature is measured by a thermocouple. The material used for the gas tube is typically quartz, but the material is not critical since the tube is not heated to high temperatures. The cell is equipped with a 'shower head' at the gas exit in order to provide a homogeneous flux distribution of the organometallics. For heat shielding, the cell is surrounded by a Ta radiation shield which thermally insulates the cell from the liquid N_2 cryoshroud. Finally, the group-III cell is mounted on a conflat flange to provide compatibility with the UHV recipient.

Among the advantages of chemical-beam epitaxy systems are the simplicity and flexibility of source changes without breaking the vacuum of the reaction chamber and good homogeneity over large areas of the substrate. In comparison with OMVPE systems, chemical-beam epitaxy (i) uses less than 1% of AsH_3 and PH_3, (ii) allows in-situ surface analysis instruments to be used, and (iii) has a smaller memory effect. A comprehensive review on the properties of chemical-beam epitaxy was given by Tsang (1989).

The pressure regime of chemical-beam epitaxy is in the 10^{-4} Torr range which is higher than MBE and GSMBE. The higher pressure is due to the unused alkyl radicals as well as to hydrogen. The high gas-load necessitates large-volume pumping systems in order to keep the pressure in the 10^{-4} to 10^{-5} Torr range. Turbo molecular pumps as well as cryopumps are well-suited pumping

https://doi.org/10.1017/CBO9780511599828.007 Published online by Cambridge University Press

systems for CBE systems.

High quality $Ga_{0.47}In_{0.53}As/InP$ and quaternary $Ga_xIn_{1-x}P_yAs_{1-y}/InP$ grown by CBE was reported in the literature. Characterization of CBE-grown material includes optical and transport properties of $Ga_{0.47}In_{0.53}As$ (Tsang et al., 1986), optical properties of $Ga_{0.47}In_{0.53}As/InP$ quantum wells (Tsang and Schubert, 1986), and optical properties of quaternary $Ga_xIn_{1-x}P_yAs_{1-y}$ (Tsang et al., 1987). These studies demonstrated the capability of CBE to grow high-quality epitaxial films with low background impurity concentration, high carrier mobility, and intense photoluminescence spectra with narrow linewidths. Furthermore, high-performance electronic devices processed from CBE grown material were reported including heterobipolar transistors (Tsang et al., 1988) and heterostructure field effect transistors (Schubert et al., 1988).

A problem area of CBE is the growth of high-quality $Al_xGa_{1-x}As$ with sufficiently low background impurity concentration. Although GaAs growth with low background concentrations was reported (Pütz et al., 1986; Chiu et al., 1987), the impurity background concentration of $Al_xGa_{1-x}As$ was found to be very high, typically in the $10^{17} - 10^{19}$ cm^{-3} range. Such high background impurity concentrations are unacceptable for device applications. The high C background concentration in $Al_xGa_{1-x}As$ is due to the high reactivity of Al which tends to form strong Al–C bonds (which are stronger than the Ga–C bonds). Carbon is thus incorporated on group-V sites and acts as an acceptor. The employment of chemicals other than TMAl and TEAl have reduced the C background concentration to lower levels (Chiu et al., 1991). However, the $Al_xGa_{1-x}As$ quality grown by CBE is still inferior to MBE grown $Al_xGa_{1-x}As$.

4.4 Organometallic vapor-phase epitaxy

Among the different epitaxial growth techniques, organometallic vapor-phase epitaxy (OMVPE) has proven the ability to grow state of the art heterostructures, quantum wells and superlattices. Low-threshold current-injection lasers and efficient photodetectors were also demonstrated by the OMVPE growth technique. OMVPE is used in manufacturing environments for the production of optoelectronic devices. Multiwafer systems are frequently used in such manufacturing systems. OMVPE is a promising candidate for future high-throughput epitaxial growth systems.

Historically, the development of OMVPE started with the work of Manasevit (1968) and of Manasevit and Simpson (1969). The authors grew single crystal GaAs initially on insulating, lattice mismatched substrates such as Al_2O_3. Subsequently, the authors grew binary (GaAs, GaP) and ternary (GaP_xAs_{1-x}, $GaAs_xSb_{1-x}$) semiconductors on GaAs. The epitaxial growth was achieved by

https://doi.org/10.1017/CBO9780511599828.007 Published online by Cambridge University Press

using organometallics and hydrides for the supply of group-III and group-V precursors, respectively. Device-quality materials were grown with OMVPE at the end of the 1970s. Dupuis and Dapkus (1978) reported the first continuous wave operation of $Al_xGa_{1-x}As/GaAs$ double-heterostructure lasers at room temperature grown by OMVPE. This publication stimulated further work in the field and is considered as one of the milestones in OMVPE research.

GROWTH METHOD		PRESSURE	GAS FLOW
OMVPE	Atmosph. Pressure	760 Torr	Viscous
	Low Pressure	$>10^{-2}$ Torr	Viscous
GSMBE	CBE/MOMBE	$\cong 10^{-5}$ Torr	Molecular
	Hydride GSMBE	$\cong 10^{-6}$ Torr	Molecular
MBE		$\cong 10^{-9}$ Torr	Molecular

Fig. 4.10. Operating pressures and gas flow characteristics for different epitaxial growth techniques.

The gas flow in VPE reactors is fundamentally different from the gas transport in GSMBE and MBE (Tsang, 1989). The gas flow and the operating pressures of the three growth technologies are compared in Fig. 4.10. Growth techniques which operate under high vacuum ($< 10^{-5}$ Torr) and ultra-high vacuum conditions ($< 10^{-7}$ Torr) such as GSMBE and MBE have gas transport in terms of a *molecular beam*. That is, the source materials evaporated from effusion cells or injected into the reaction chamber have the nature of a *beam*. The molecules forming the beam have one preferential direction and furthermore molecules forming the beam do not interact. In contrast, OMVPE operating at pressures $> 10^{-2}$ Torr has a *viscous* gas flow. That is, the molecules forming the gas flow are strongly interacting. The flow pattern follows the laws of gas dynamics which is in clear contrast to the beam nature found in high-vacuum systems.

The viscous flow of gases in OMVPE systems necessitates that the reaction chamber have a smooth shape, i.e. no sharp corners and dead volumes. Such dead volumes are defined as having a zero or a very small gas velocity. A typical

https://doi.org/10.1017/CBO9780511599828.007 Published online by Cambridge University Press

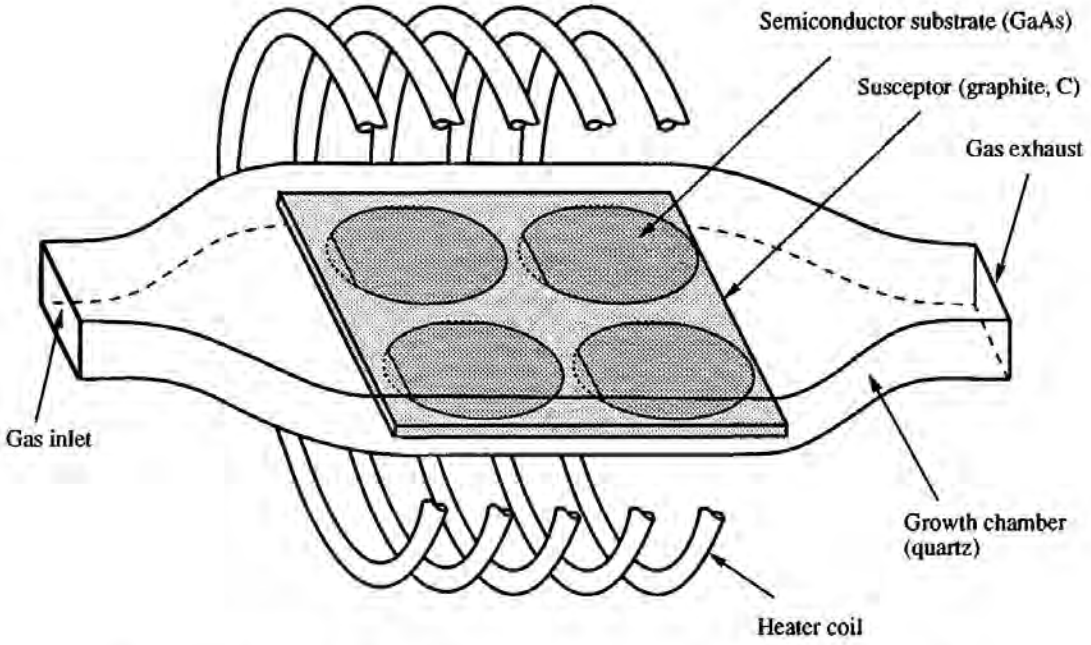


Fig. 4.11. Schematic illustration of the growth chamber of an organometallic vapor-phase epitaxy (OMVPE) system.The wafers are on a conductive susceptor which is heated by an induction coil. The growth chamber has no sharp corners, thus making possible a laminar, non-turbulent gas flow.

reaction chamber of an OMVPE system is shown in Fig. 4.11. The reaction chamber walls are typically made of quartz and have smooth shapes to enable a laminar, non-turbulent gas flow. The reaction chamber shown has a multiwafer susceptor for high-volume throughput. The requirements for the susceptor are (i) low vapor pressure in order to avoid incorporation of the susceptor material into the growing semiconductor and (ii) high electrical and thermal conductivity. Graphite is frequently used as a susceptor material. The desirable high thermal conductivity of the susceptor provides a homogeneous heat distribution across the entire susceptor area.

The susceptor is heated by radio-frequency (rf) induction as indicated in Fig. 4.11. In order to induce currents in the susceptor, it must be conductive. The inductive heating is provided by a coil around the reaction chamber. The coil generates a high-frequency field inside the reaction chamber and generates an induced current in the susceptor which in turn resistively heats the substrate.

A laminar flow of the gases inside the reaction chamber is necessary in order to avoid the ***memory effect***. The memory effect manifests itself by a lack of

https://doi.org/10.1017/CBO9780511599828.007 Published online by Cambridge University Press

abruptness of compositional and doping profiles. *Dead volumes* inside the gas manifold and inside the reaction chamber gradually release gases over a period of time and result in graded (rather than abrupt) interfaces and in smoothly changing (rather than abrupt) doping profiles. A reduction of the memory effect is obtained by evacuating the reaction chamber, i.e. by low-pressure OMVPE. ***Low-pressure OMVPE*** operates at a pressure of 10^{-2} Torr to 100 Torr. As a result of the lower pressure, the gas velocity inside the reactor increases. The increased gas velocity is advantageous in several respects. First, as already mentioned, the memory effect is less pronounced. Second, the homogeneity of composition, thickness, and doping across the growing wafer is improved. Third, interfaces between two different semiconductors are more abrupt in low-pressure OMVPE.

Vertical design and ***horizontal design*** reaction chambers have been realized (Ludowise, 1985). In the vertical design, the gas flow enters the reaction chamber from the top and is deflected at the wafer surface. In horizontal design reactors (as shown in Fig. 4.11) the gas flow is parallel to the surface of the wafer. The wafer is exposed to a more uniform gas stream than in vertical design chambers. To date, horizontal design reactors are the preferred solution due to better homogeneity of doping and thickness of the epitaxial layers. More recently, the horizontal design was further improved by using rapidly rotating substrate holders (see, for example, Tanaka et al., 1987). The rapid rotation of the substrates provides significantly improved homogeneity of the thickness, composition and doping across the wafer.

The chemical starting materials used in OMVPE are organometallics for the group-III elements and hydrides for the group-V elements. The starting chemicals must satisfy several requirements for OMVPE growth. First, the chemicals must thermally decompose at the growth temperature which is typically in the 500–800 °C range for III–V semiconductor growth. If the chemicals are stable and do not decompose at the growth temperature, incorporation at the wafer surface does not occur. Second, the starting chemicals must be either gaseous or have a relatively high vapour pressure at room temperature. Gaseous chemicals are easily transported through tubes from the reservoir to the reaction chamber. Liquid and solid compounds are not suited for such transport.

The vapor pressure of organometallics varies over several orders of magnitude at room temperature. The vapor pressures of the most popular organometallics are shown in Fig. 4.12. As a general tendency, molecules with larger alkyl groups have lower vapor pressures than molecules with small alkyl groups. For example, the vapor pressure of TMGa is higher than the vapor pressure of TEGa. The same trend applies to TMAl, TEAl, TMIn, and TEIn. If the vapor pressure of a

https://doi.org/10.1017/CBO9780511599828.007 Published online by Cambridge University Press

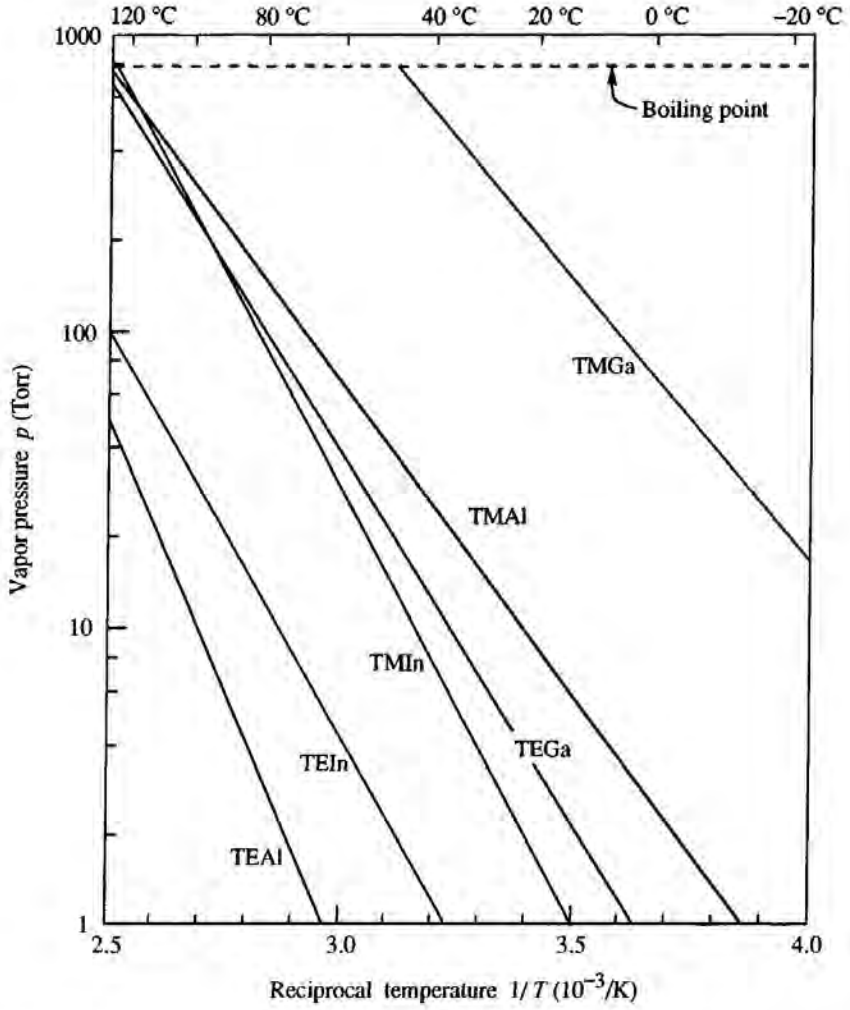


Fig. 4.12. Vapor pressure of organometallic group-III compounds as a function of reciprocal temperature. A carrier gas such as H_2 is needed to transport a compound into the reaction chamber if the vapor pressure of the compound (e.g. TEAl) is very low.

chemical is low, a small flow rate results. The supply rate of such low-pressure chemicals can be enhanced by means of hydrogen which is bubbled through the liquid organometallic compound. Thus, H_2 can serve as a carrier gas for low vapor pressure organometallics.

The starting material containers and the gas manifold system of an OMVPE system is schematically shown in Fig. 4.13. The starting chemicals are either directly fed into the gas manifold or are equipped with an inert gas (H_2) bubbler. Furthermore, valves are used to completely open and close the flow of reagents.

https://doi.org/10.1017/CBO9780511599828.007 Published online by Cambridge University Press

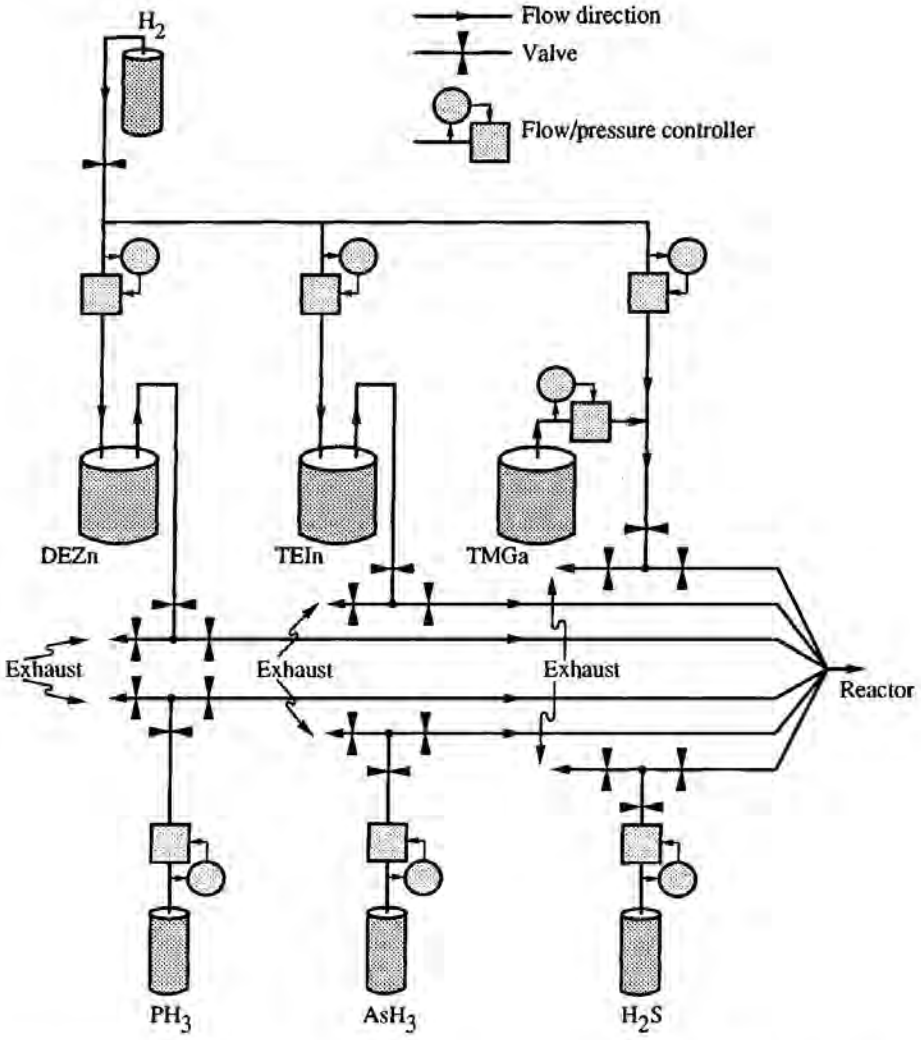


Fig. 4.13. Gas manifold of an organometallic vapor-phase epitaxy (OMVPE) system. The gas flows are switched between the reaction chamber and the ventilation line in order to provide a constant flow through the flow/pressure controllers.

The flow of reagents and of the carrier gas is regulated by either pressure or flow controllers. All chemical reagents are mixed before they enter the growth chamber. The pre-mixing of chemicals results in a homogeneous gas composition which in turn results in a homogeneous composition of the epitaxially grown layers.

The switching of source reagents occurs with the use of exhaust or ventilation valves (see Fig. 4.13). The gas flow is switched from the reaction chamber to the

https://doi.org/10.1017/CBO9780511599828.007 Published online by Cambridge University Press

Innoscience1366-00138854

exhaust or vice versa. It is important to note that the gas flow is never shut off, but is instead directed to either the growth chamber or to the exhaust at a *constant* flow rate. The purpose of this procedure is to minimize flow transients. Flow transients usually occur at transitions between different materials during growth and, as a result, the compositions of the growing semiconductor cannot be changed abruptly. The use of exhaust valves reduces the magnitude of such flow transients.

The precursors of the OMVPE technique are thermally cracked at the hot susceptor and substrate surface. As an example, we consider TMGa.

$$(CH_3)_3Ga + \frac{3}{2}H_2 \leftrightarrow 3CH_4 + Ga\,. \tag{4.1}$$

The TMGa molecule is cracked into alkyl radicals (CH_3) which form stable methane (CH_4) with the excess hydrogen in the growth chamber. The cracking process of the hydrides is straightforward. Considering arsine as an example, the pyrolysis can be written as

$$2AsH_3 + \text{Energy} \leftrightarrow As_2 + 3H_2 \tag{4.2a}$$

and

$$4AsH_3 + \text{Energy} \leftrightarrow As_4 + 6H_2\,. \tag{4.2b}$$

The species of arsenic formed by the pyrolysis are As_2 as well as As_4. The formation of arsenic monomers is unlikely since they are not stable at the pressure and temperature range of interest (Panish and Temkin, 1989).

The combination of the above two chemical reactions yields the complete reaction occurring during OMVPE growth. The complete reaction for the starting reagents TMGa and AsH_3 is thus given by

$$(CH_3)_3Ga + AsH_3 \leftrightarrow GaAs + 3CH_4\,. \tag{4.3}$$

The relative fractions of reactions that are pyrolytic (Eq. (4.1) and (4.2)) and of the reactions that are complete (Eq. (4.3)) are not known. However it is clear that pyrolysis does occur and is a necessary condition for the growth of the semiconductor.

A vertical cut through an OMVPE reactor is shown in Fig. 4.14. The reactor shown is of the horizontal type. Pyrolysis (as defined in Eqs. (4.1) and (4.2)) occurs in the vicinity of the susceptor and the semiconductor substrate. A ***stagnant gas layer*** develops on top of the susceptor, the substrate, and at the reactor walls. This stagnant boundary layer is a gas layer that has a very small or zero velocity. Such low flow velocities are due to mechanical friction of the gas

https://doi.org/10.1017/CBO9780511599828.007 Published online by Cambridge University Press

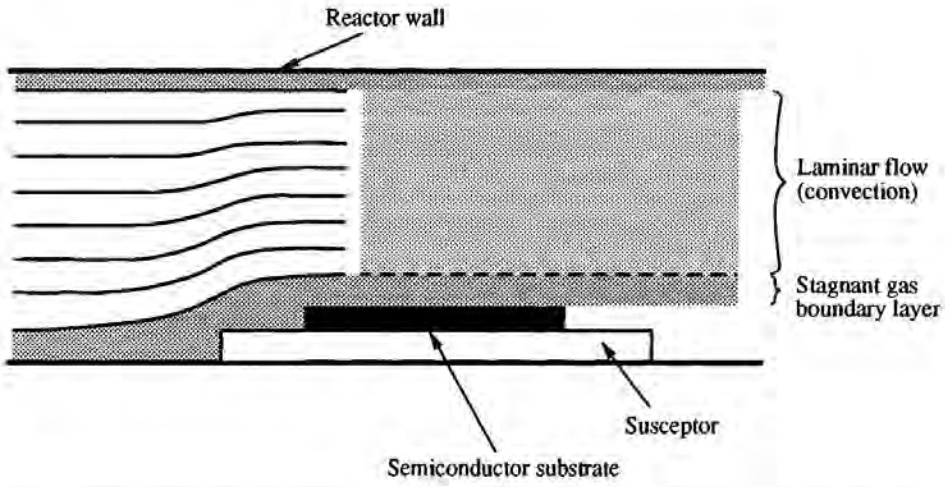


Fig. 4.14. Schematic illustration of the gas flow in a vapor-phase epitaxy (VPE) reaction chamber. Good mixing of gases is achieved in the laminar flow region. A stagnant gas layer develops in the vicinity of the substrate. Gas transport through the stagnant layer occurs through diffusion rather than through convection.

at the substrate surface. The spatial extent of the stagnant layer is limited to the vicinity of the susceptor, substrate, and reactor walls. Laminar gas flow occurs sufficiently far away from the walls and the susceptor. As a result of the stagnant layer, reagents are not supplied directly from the laminar flow region (see Fig. 4.14 for illustration) to the substrate surface by *convection*. Instead, the stagnant layer becomes depleted by the growth reaction and new reagents are supplied to the semiconductor surface by means of *diffusion* of reagents from the laminar flow region to the growing epitaxial layer surface. The occurrence of the stagnant layer and the transport of reagents through this layer by diffusion is one of the major differences between OMVPE and CBE (Tsang, 1989).

One of the drawbacks of OMVPE is the toxicity of its source reagents, especially AsH_3, and the amount needed during growth. Arsine is a highly toxic gas and can only be used with many elaborate and expensive safety precautions. Furthermore, the amount of reagents required for OMVPE is much higher (typically a factor of 10–100, Tsang, 1990) as compared to the CBE growth technique. The use of such large amounts of dangerous chemicals represents a potential safety hazard.

Excellent parameter control is desirable for epitaxial growth. Many parameters of OMVPE growth are interdependent. For example, the substrate

https://doi.org/10.1017/CBO9780511599828.007 Published online by Cambridge University Press

temperature influences the epitaxial growth rate. At low growth temperatures, the pyrolysis is incomplete and the growth rate decreases with OMVPE growth. This characteristic of OMVPE is in contrast to MBE, where growth rate and substrate temperature are independent over a large range of growth temperatures. Due to the interdependence of so many parameters in OMVPE, all the parameters must be controlled very accurately in order to achieve small variations of epitaxial layer thickness, composition, and doping. The requirement of accurate parameter control and the interdependence of many parameters makes OMVPE a growth technology which is less flexible than MBE.

On the other hand, OMVPE is very versatile and allows one to grow virtually all III–V semiconductors. Excellent heterostructures, quantum well structures, and superlattice structures have been grown by OMVPE. In addition, electronic as well as optoelectronic devices with impressive performance have been grown by OMVPE. Recently extensive reviews were published on OMVPE (Razeghi, 1989; Stringfellow, 1989; Swaminathan and Macrander, 1991), which provide an overview on the growth, physics, and chemistry of the OMVPE technique as well as device characteristics of OMVPE-grown structures.

4.5 Vapor-phase epitaxy

Conventional vapor phase-epitaxy (VPE) which preceded the now prominent OMVPE, does not employ organometallic compounds for the group-III elements. Instead, conventional VPE uses the chlorides of the group-III elements as source gases. VPE was first demonstrated by Tietjen and Amick (1966). The authors used gaseous arsine and phosphine as group-V precursors and gallium chloride for the growth of GaP_xAs_{1-x}. An alternative to hydrides for the group-V precursors are chlorides (e.g. $AsCl_3$ or PCl_3). VPE is frequently referred to as ***hydride VPE*** and ***chloride VPE*** for growth with group-V hydride precursors and group-V chloride precursors, respectively.

The growth of III-V semiconductors by VPE is performed in fused silica reactors consisting of three zones. Such a three zone VPE reactor is shown schematically in Fig. 4.15. The zones of the reactor are the *source zone*, the *baffle zone*, and the *growth* zone. The flow of gases to the reaction chamber is controlled by either flow or pressure regulators. The gas manifold system is typically operated in a growth or ventilation mode, which allows one to maintain a constant flow or pressure. The switching between growth and ventilation mode reduces flow and pressure transients that always occur upon changing the nominal flow or pressure.

The first zone of the reaction chamber is called the ***source zone***. The purpose of the source zone is to generate gaseous compounds containing the group-III

https://doi.org/10.1017/CBO9780511599828.007 Published online by Cambridge University Press

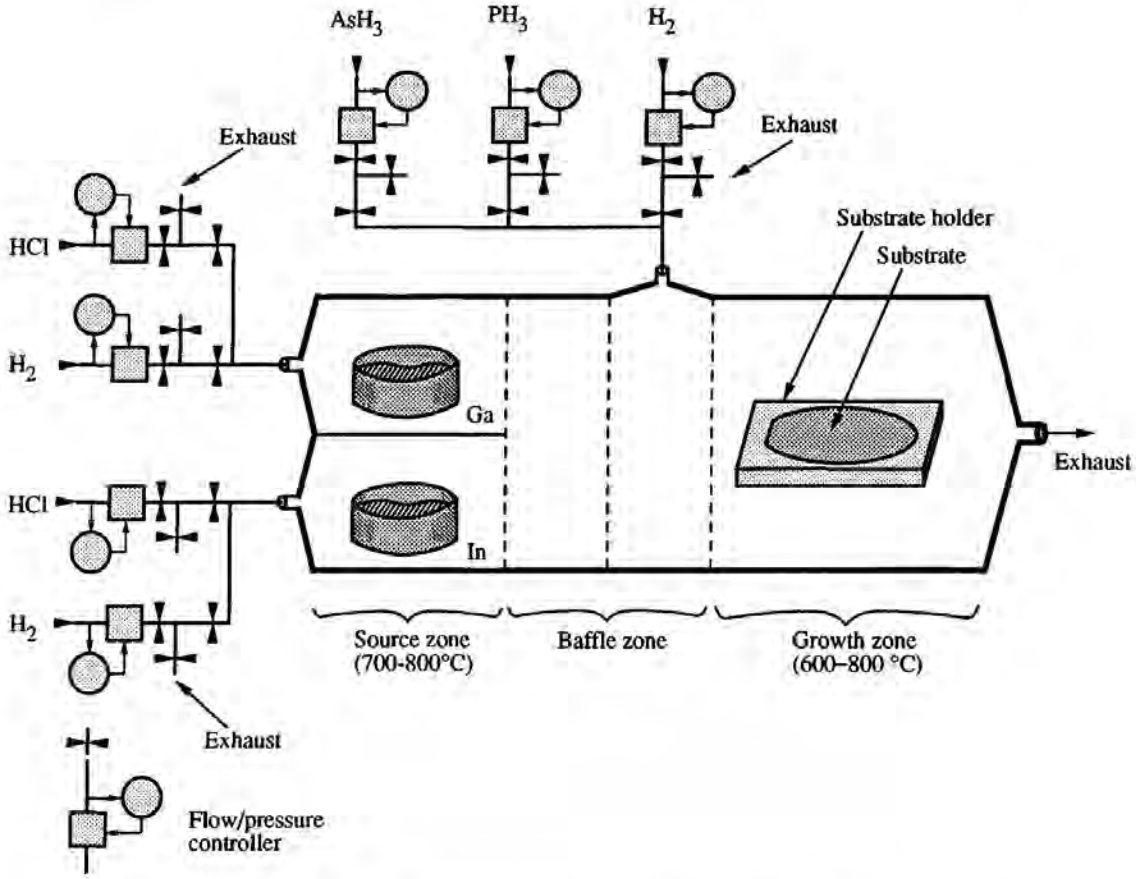


Fig. 4.15. Schematic illustration of a hydride vapor-phase epitaxy (VPE) system. The group-V elements are provided by means of hydrides (AsH_3 and PH_3). Group-III elements are transported by means of their chlorides using hydrochloric acid as transport reagent.

elements. Chlorides are the most popular inorganic group-III precursors. They are synthesized by hydrogen chloride gas flowing over and reacting with the elemental group-III sources. The chemical reaction for a gallium source is given by

$$HCl + Ga \leftrightarrow GaCl + \tfrac{1}{2} H_2 , \quad (4.4a)$$

$$HCl + GaCl \leftrightarrow GaCl_2 + \tfrac{1}{2} H_2 , \quad (4.4b)$$

$$HCl + GaCl_2 \leftrightarrow GaCl_3 + \tfrac{1}{2} H_2 . \quad (4.4c)$$

All three of the above chlorides, i.e. GaCl, $GaCl_2$, and $GaCl_3$ are present in the

https://doi.org/10.1017/CBO9780511599828.007 Published online by Cambridge University Press

337-TA-1366 RX-0255

gas phase of a VPE reactor. The source zone of the VPE reactor is heated to temperatures of typically 700–800 °C to accelerate the chemical reaction.

The second zone of the reactor is called the ***baffle zone*** and has the purpose of mixing the group-III with the group-V precursor. Such a mixing is required to obtain epitaxial growth with good homogeneity of composition and uniformity of layer thickness. The baffle zone consists typically of several perforated walls which diffuse the gas flow. Tietjen and Amick (1966) used a so-called ***reaction zone*** instead of the baffle zone. The reaction zone was heated to a temperature of 850 °C to enhance the chemical reaction between the group-III and group-V constituents.

The third zone of the VPE reactor is the ***growth zone***. This zone includes the semiconductor substrate which is placed on top of the substrate holder (substrate holders are termed *susceptors* if heated by radio-frequency induction). The heating of all the three zones of the VPE system is usually provided by a furnace surrounding the reactor. Such reactors are called hot-wall reactors. It is desirable that the furnace has several zones in order to adjust the temperature in each zone independently. Arsine and phosphine are thermally cracked in the hot growth zone of the reactor. The pyrolysis reactions can be written as

$$2AsH_3 + \text{Energy} \leftrightarrow As_2 + 3H_2 , \tag{4.5a}$$

$$4AsH_3 + \text{Energy} \leftrightarrow As_4 + 6H_2 . \tag{4.5b}$$

The corresponding pyrolysis reaction for phosphine is given by

$$2PH_3 + \text{Energy} \leftrightarrow P_2 + 3H_2 , \tag{4.5c}$$

$$4PH_3 + \text{Energy} \leftrightarrow P_4 + 6H_2 . \tag{4.5d}$$

The chemical species of arsenic and phosphorus formed are thus dimers and tetramers. The formation of arsenic and phosphorus monomers is unlikely since the monomers are thermodynamically unstable at the pressure and temperature range of interest (Panish and Temkin, 1989).

For the complete chemical reaction of chloride VPE growth, we consider GaAs as an example. The reaction is given by

$$GaCl + AsH_3 \leftrightarrow GaAs + HCl + H_2 . \tag{4.6}$$

Hence, gaseous hydrochloric acid is generated during VPE growth. Since HCl is toxic, it needs to be neutralized by a base before being released into the exhaust system.

The VPE growth reaction (Eq. (4.6)) is an equilibrium reaction which depends strongly on the growth temperature. In other words, the growth rate can be

https://doi.org/10.1017/CBO9780511599828.007 Published online by Cambridge University Press

RX-0255.186

positive (growth) as well as negative (etching). As a result, the VPE growth technique is well suited for growth on previously processed structures, for selective area growth, and for growth on patterned substrates (Johnston et al., 1989). On the other hand, the strong dependence of the growth rate on the substrate temperature can be disadvantageous due to the interdependence of parameters. It is generally desirable for the growth parameters to be independent.

A further disadvantage of chloride/hydride VPE is the difficulty of growing thin quantum well structures. The growth of thin, homogeneous quantum well structures by VPE has not hitherto been demonstrated. This lack is partly due to the high growth rates of VPE of typically 100 Å/s = 36 μm/hr at moderate growth temperatures. Furthermore, polycrystalline growth occurs on the fused silica walls of the reactor in hot-wall reactors used for VPE. The deposition of material on the reactor walls changes the heat-radiation pattern of the VPE reactor and results in changing growth rates, which complicates the growth rate control.

https://doi.org/10.1017/CBO9780511599828.007 Published online by Cambridge University Press

5

Doping with elemental sources

Intentional impurities (dopants) can be introduced during epitaxial crystal growth in terms of either elemental doping sources or chemical compounds. This chapter is devoted to aspects of elemental doping sources and to the characteristics of these dopants. The following chapter discusses non-elemental, gaseous doping precursors. The incorporation of elemental doping impurities is mediated through thermal evaporation of the doping element from an effusion cell. Such elemental solid evaporation sources are commonly used for molecular-beam epitaxy (MBE) and related techniques discussed in Chap. 4.

5.1 Doping techniques and calibration

The incorporation of doping atoms originating from solid elemental doping sources is mediated through thermal evaporation within a vacuum environment. The solid source is heated to a temperature T at which the desired amount of the doping element evaporates. The rate of evaporation, the source and substrate geometry, and the sticking probability of impurity atoms on the crystal surface determine the incorporation rate of the impurities into the growing epitaxial crystal. These parameters are discussed below.

5.1.1 Doping impurity flux

The concentration of doping impurities in an epitaxial layer is usually controlled by the impurity-cell temperature. Consider an epitaxial crystal of GaAs that grows at a rate of 1 monolayer per second, i.e. at a rate of 6×10^{14} GaAs molecules per cm^2 per second. Assume further that doping with concentration of 1×10^{18} cm^{-3} is desired. Using a thickness of one monolayer of 2.83 Å, an impurity atom flux of 2.83×10^{10} cm^{-2} s^{-1} is required (unity sticking of

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

impurity atom on the substrate is assumed). This example shows the simple relationship between the required dopant flux and the doping concentration in the crystal. However, the relationship between the desired dopant *flux* and the effusion cell *temperature* is important as well and is calculated below.

The vapor pressure of a doping source can be calculated from the ***ideal gas*** law which is given by

$$PV = N_{\text{Avo}}\, kT \tag{5.1}$$

where P is the pressure of $N_{\text{Avo}} = 6.022 \times 10^{23}$ atoms (or molecules) residing within a volume V at the temperature T. The vapor pressure of a solid can be expressed as $P = P_0 \exp(-H_v/kT)$, where P_0 is a constant and H_v is the latent heat of vaporization. Furthermore, the thermal velocity, v_{th}, of the atoms (molecules) is given by the ***equipartition law***, i.e.

$$\frac{1}{2}\, mv_{\text{th}}^2 = \frac{3}{2}\, kT \tag{5.2}$$

where m is the mass of one atom (molecule). The average flux of atoms (per cm^2 per second) is then given by

$$F = \frac{N}{V}\, v_{\text{th}} = (P/kT)\sqrt{3kT/m} \tag{5.3}$$

which can be used to calculate the required impurity atom flux at the substrate surface. Consider further that the impurity source has an effective area A and is located at a distance r from the substrate. Assuming that the flux is isotropically distributed over a half-sphere, the impurity atom flux at the evaporation source is given by

$$F_{\text{source}} = \frac{\frac{1}{2}\pi r^2}{A}\, F = \frac{\frac{1}{2}\pi r^2}{A}\, P\sqrt{3/kTm}\ . \tag{5.4}$$

For a given doping flux F_{source}, the equation gives us one condition for the pressure P and the temperature T. Of course, P and T are also related by the vapour pressure curves (Marton and Marton, 1979) which represent a second condition for P and T.

As an example, consider Si doping of GaAs at a concentration of $5 \times 10^{18}\ \text{cm}^{-3}$ with a solid Si source with an area of $1\ \text{cm}^2$ located at a distance of 50 cm from the semiconductor substrate. At a growth rate v_g, the required doping atom flux is given by

$$F = N_{\text{Si}}\, v_g\ . \tag{5.5}$$

At a growth rate of 1 monolayer per second ($v_g \equiv 2.8$ Å/s) the required doping

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

flux at the growing layer is obtained as $F = 1.4 \times 10^{11}$ cm^{-2}s^{-1}. The flux at the Si source is $F_{source} = \frac{1}{2} F\pi r^2/A = 5.5 \times 10^{14}$ cm^{-2}. Using Eq. (5.4) and the vapor pressure versus temperature curve of Si (Marton and Marton, 1979) allows one to determine the required Si-source vapor pressure and temperature. One obtains, for a Si mass of $m_{Si} = 14 \times 10^{-3}$ kg/$6.022 \times 10^{23} = 2.3 \times 10^{-26}$ kg, a Si vapor pressure of $P_{Si} \cong 1.5 \times 10^{-8}$ Torr and a Si-source temperature of approximately $T_{Si} \cong 1000$ °C.

The above calculation is insufficient in two respects. First, the impurity gas atoms have a maxwellian velocity distribution rather than the average velocity used in the calculation. Second, the impurity atom flux is assumed to be isotropic; this assumption is not justified for Knudsen-type effusion cells. Even for dopant evaporation from filaments an isotropic distribution of the impurity-atom flux is hardly fulfilled. Nevertheless, the calculation provides a fast and simple means of estimating the temperature and pressure of a thermal doping source.

5.1.2 Effusion cells

Doping impurities are frequently evaporated from ***effusion cells***. Effusion cells employ filaments for resistively heating crucibles which contain the impurity element. The material used for the crucibles is typically a chemically inert material with a low vapor pressure even at high temperatures. Pyrolytic boron nitride (BN) and graphite (C) are suited for this purpose. In ***Knudsen-type*** effusion cells (Knudsen, 1909), evaporated atoms or molecules interact weakly with each other and mostly with the walls of the effusion cell. Such effusion cells have a preferred direction of evaporation which is the symmetry axis of the effusion cell. The radiant pattern of the cell is given by a cosine distribution. Such a cosine distribution of the flux is schematically shown in Fig. 5.1(a). In the following the cosine distribution will be derived.

Consider an effusion cell heated to a temperature T. The area of the effusion cell is denoted as ΔA. Atoms (or molecules) evaporate from the source and leave the effusion cell through its opening. The velocity distribution function of the particles is, under equilibrium conditions, given by the maxwellian velocity distribution (see Chap. 3). In cartesian and spherical coordinates, the distribution is given by

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

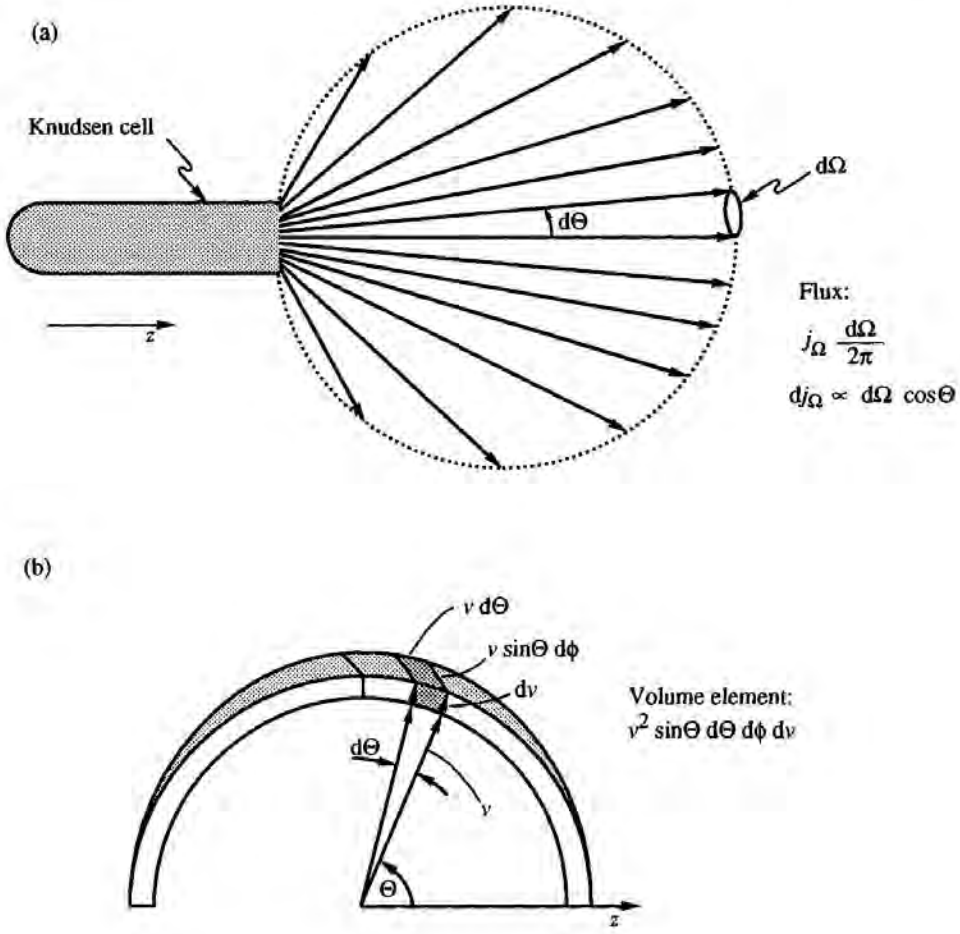


Fig. 5.1. (a) Schematic illustration of an effusion cell and the cosine-shaped distribution of the effusion flux. (b) Velocity 'volume element' in spherical coordinates (velocity v, polar angle Θ, and azimuthal angle ϕ) used to calculate the spatial flux distribution of a Knudsen cell.

$$f(v_x, v_y, v_z)\,dv_x\,dv_y\,dv_z = \left(\frac{m}{2\pi kT}\right)^{3/2} \exp\left[-\frac{\frac{1}{2}m(v_x^2 + v_y^2 + v_z^2)}{kT}\right] dv_x\,dv_y\,dv_z \quad (5.6a)$$

$$f(v)\,v^2 \sin\theta\,d\theta\,d\phi\,dv = \left(\frac{m}{2\pi kT}\right)^{3/2} \exp\left[-\frac{\frac{1}{2}mv^2}{kT}\right] v^2 \sin\theta\,d\theta\,d\phi\,dv \quad (5.6b)$$

where m is the mass of one atom (or molecule). It is convenient to transform the velocity distribution into spherical coordinates using the spherical coordinates v,

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

ϕ, and θ. The 'volume' elements are then given by $dv_x dv_y dv_z = v^2 \sin\theta\, d\theta\, d\phi\, dv$, as illustrated in Fig. 5.1(b). Assuming that the symmetry axis of the effusion cell is the z-axis, all atoms with a positive z-component of the velocity can leave the effusion cell. The number of particles which exit the effusion cell through the area ΔA during the time Δt with a velocity between v and dv is given by

$$dn_{\Delta A}(v) = dn(v)\,\frac{\Delta A \Delta t\, v_z}{V} = N f(v_x, v_y, v_z)\, dv_x dv_y dv_z \frac{\Delta A \Delta t\, v_z}{V}$$

$$= \frac{\Delta A \Delta t\, N}{V}\left[\frac{m}{2\pi kT}\right]^{3/2} e^{-mv^2/2kT}\, v\cos\theta\, v^2 \sin\theta\, d\theta\, d\phi\, dv$$

$$= \frac{\Delta A \Delta t\, N}{V}\left[\frac{m}{2\pi kT}\right]^{3/2} e^{-mv^2/2kT}\, v^3 \cos\theta\, d\Omega\, dv \qquad (5.7)$$

where $v_z = v\cos\theta$ is the z-component of the velocity, $d\Omega = \sin\theta\, d\theta\, d\phi$ is the solid angle, N is the total number of gas atoms in the cell, and V is the volume of the cell. Integration over all velocities and division by ΔA and Δt yields the flux of atoms out of the effusion cell into the solid angle $d\Omega$

$$dj_\Omega = \int_0^\infty \frac{N}{V}\left[\frac{m}{2\pi kT}\right]^{3/2} e^{-mv^2/2kT}\, v^3 dv \cos\theta d\Omega = \frac{1}{4\pi}\,\frac{N}{V}\,\bar{v}\cos\theta d\Omega\,. \qquad (5.8)$$

where we used $\bar{v} = 4\pi \int_0^\infty v f(v) v^2 dv$. Defining the total flux out of the effusion cell as $j_{2\pi} = (N/V)\bar{v}/4$, the flux distribution is given by

$$\boxed{dj_\Omega = j_{2\pi}(d\Omega/\pi)\cos\theta} \qquad (5.9)$$

The result shows that the flux of doping atoms out of the effusion is not isotropic. The flux has its preferential direction along the z-direction, i.e. along the symmetry axis of the effusion cell. The angle-dependence of the effusion follows a cosine law. As a check we integrate the flux over the entire half-space and obtain

$$\int_{2\pi} dj_\Omega = j_{2\pi} \int_0^{2\pi} d\phi \int_0^{\pi/2} \frac{1}{\pi}\cos\theta\sin\theta d\theta = j_{2\pi} \qquad (5.10)$$

which is the expected result.

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

5.1.3 Effusion cell calibration

Doping concentrations in III–V semiconductors need to be controlled over a wide range, for example from 10^{15} to 10^{20} cm^{-3}. Doping control over 5 orders of magnitude requires the ***calibration*** of the doping process. In a typical calibration process several samples are grown at the same growth rate but different doping effusion cell temperatures. The thicknesses of such calibration epitaxial layers is typically on the order of 1 μm. The total thickness of the epitaxial layer must be much thicker than the width of the surface depletion layer in order to avoid large corrections of free carrier concentrations, e.g. measured by the Hall effect (see below). The surface depletion layer thickness, W_D, depends on the doping concentration according to

$$W_D = \sqrt{\frac{2\varepsilon}{eN_D}\phi_B} \tag{5.11}$$

where ε is the permittivity, e is the electronic charge, and ϕ_B is the barrier height (in volts) of the semiconductor surface.

Typical calibration measurements for Be (p-type) and Si (n-type) doping of MBE-grown GaAs is shown in Figs. 5.2 and 5.3, respectively. The Be effusion cell operates at lower temperatures due to the higher vapor pressure of Be as compared to Si. The growth rate should be constant for all doping cell temperatures. The doping concentration has a quasi-linear relationship if plotted versus reciprocal temperature which reflects the exponential increase of the vapor pressure with temperature.

The doping concentration is frequently measured by the Hall effect. The Hall effect allows one to measure the free electron or hole concentration. The free carrier concentration does not always coincide with the doping concentration. In order to obtain a reliable result (i) all dopants must be electrically active, (ii) no compensation of dopants must occur, and (iii) surface depletion effects must be negligible. Furthermore, the Hall carrier concentration differs from the true free carrier concentration by the so-called Hall factor. The Hall factor causes the measured Hall carrier concentration to deviate from the true free carrier concentration by about 20% depending on the dominant carrier scattering mechanism (Putley, 1960).

5.1.4 Filament doping

Evaporation of doping impurities from filaments was pioneered by Malik (Malik et al., 1988) for C doping of GaAs. The C doping was achieved by resistively heating a graphite filament which was cut in serpentine shape to yield a resistance of 1 Ω and an evaporation area of 1 cm^2. The filament was clamped on both ends

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

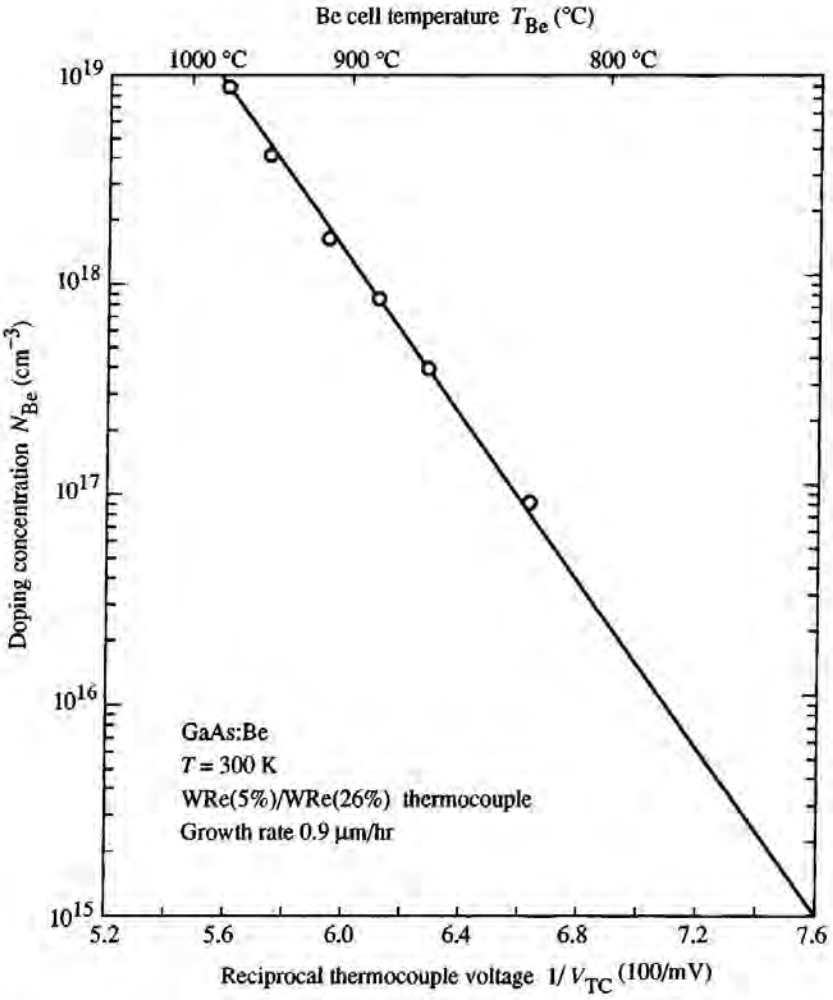


Fig. 5.2. Calibration of Be doping concentration as a function of the effusion cell temperature (thermocouple voltage) for MBE growth. The Be doping concentrations are measured by the Hall effect (courtesy of R. F. Kopf).

using flat tantalum (Ta) tabs attached to Ta rods which in turn were fastened to a dual current vacuum feed through. Malik demonstrated C doping of GaAs up to a concentration of 10^{20} cm^{-3} using C filament doping.

A schematic illustration of a graphite filament is shown in Fig. 5.4. The filament is connected to two electrical conductors which are in turn connected to the contact rods. The entire filament arrangement is S-shaped in order to accommodate the significant thermal expansion of the C filament which may be heated to temperatures as high as 2500 °C.

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

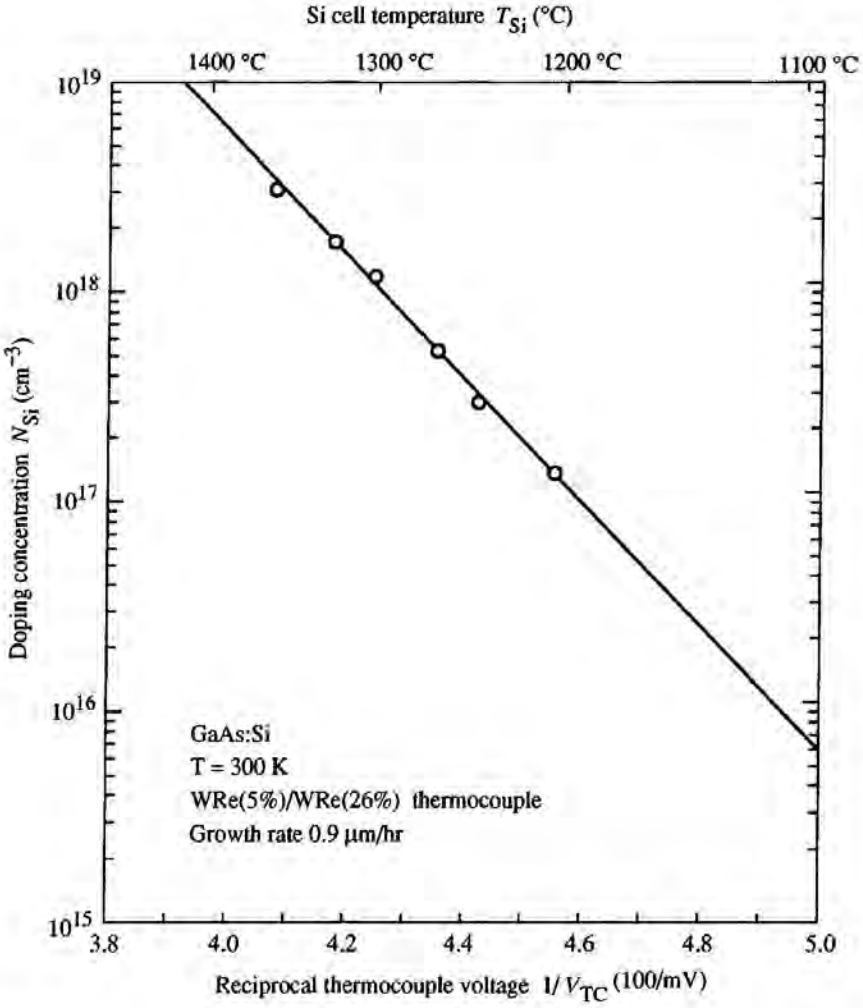


Fig. 5.3. Calibration of Si doping concentration as a function of the effusion cell temperature (thermocouple voltage) for MBE growth. The Si doping concentration is measured by the Hall effect (courtesy of R. F. Kopf).

Doping filaments need to meet several requirements. First, the filament made of the doping element needs to be electrically conductive. Isolating elements such as sulfur (S) are therefore not suited for filament doping. Second, if temperatures >2000 °C are required for evaporation, filament doping is the only choice. Effusion cells are currently available for temperatures as high as 1400 °C (e.g. from EPI Div. Chorus Corp., St. Paul, Minnesota). Graphite, which requires temperatures as high as 2500 °C, cannot be evaporated from effusion cells.

The longevity of filaments is generally shorter compared to effusion cell doping. After many evaporations, the filament becomes thinner and requires an

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

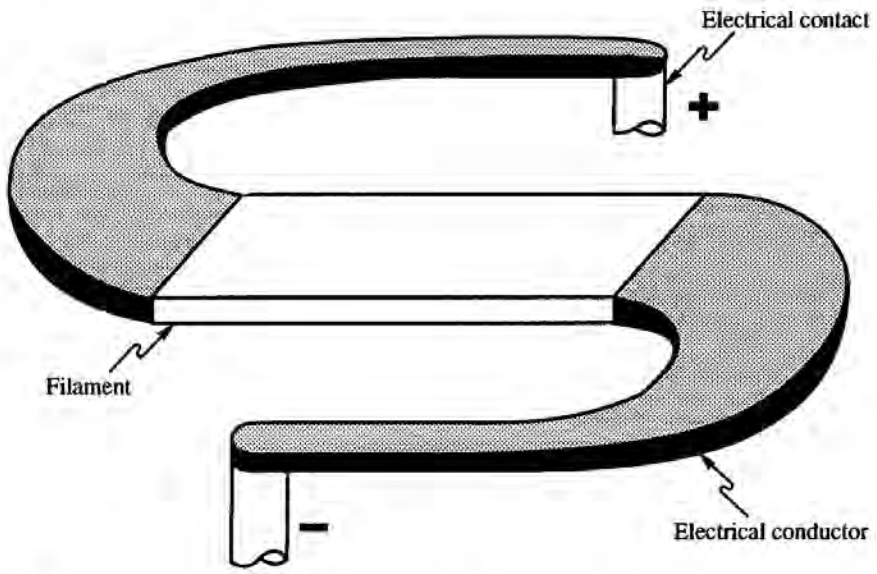


Fig. 5.4. Schematic illustration of a graphite filament for C-doping. The filament is S-shaped in order to accommodate thermal expansion effects.

adjustment of the doping calibration. The purity, homogeneity, and microscopic structure of graphite filaments strongly influence their longevity. Nevertheless, Malik et al. (1988) demonstrated that the relative thickness change of the C filament per typical growth run is relatively small ($< 10^{-3}$). Thus, the graphite filament yields reproducible doping levels over many growth experiments without the need of frequent recalibrations.

Several other groups have used graphite filaments for C doping (Ito and Ishibashi, 1991; Hoke et al. 1991). These reports confirm the feasibility of well-controlled C doping up to doping levels of 10^{20} cm^{-3} using graphite filaments. Hoke et al. (1991) further investigated the influence of the III/V ratio during growth of GaAs and the nature of the As species (As_2 and As_4) on C doping. The authors found no significant change in C doping characteristics for different growth conditions.

Silicon filaments have been employed for n-type doping of III–V semiconductors (Pfeiffer and West, 1987). Single crystal Si wafers of appropriate resistivity cleaved to a stripe-type geometry were used as filament material. The Si filaments have excellent purity, homogeneity and stability resulting in a long lifetime of the filaments.

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

5.1.5 Filament calibration

The doping calibration of graphite filaments is similar to effusion cell calibrations. Epitaxial films are grown at different graphite filament temperatures (or power) which results in different doping concentrations. The doping calibration curve for graphite filament doping is shown in Fig. 5.5 (Nagle et al., 1991). The figure shows the measured hole concentration (by the Hall effect) as a function of the reciprocal graphite filament temperature. The temperature was measured by a hot-wire pyrometer with an accuracy of ±25 K. Within the limits of the accuracy, the measured temperature was found to be exactly proportional to the fourth root of the electrical input power of the filament, i.e. $T \propto P^{1/4}$ for temperatures between 1600 K and 2400 K.

The temperature dependence $T \propto P^{1/4}$ is the dependence predicted by the ***Stefan–Boltzmann law of radiation***. The Stefan–Boltzmann law applies to radiant black bodies and predicts that the radiant energy is proportional to the fourth power of the temperature according to

$$\boxed{E/V = \pi^2 (kT)^4 / (15\hbar^3 c^3)} \tag{5.12}$$

where E/V is the radiant heat per unit volume and c is the velocity of light. From the exact $T \propto P^{1/4}$ dependence shown in Fig. 5.5, Nagle et al. (1991) concluded that the heat losses of the filament are dominated by radiation loss. Nagle et al. (1991) used an emissivity factor of 0.8 (Reynolds, 1968) and calculated a radiant area which was very close to the geometrical area of the graphite filament.

5.1.6 Calibration using electrochemical profiling

The doping calibration of epitaxial layers can be achieved efficiently, accurately, and conveniently using electrochemical capacitance–voltage (CV) profiling. This method allows one to calibrate the entire doping range (e.g. 10^{16} cm^{-3} to 10^{19} cm^{-3}) in a *single* growth run. The electrochemical profiling calibration method is therefore less tedious than calibration via the Hall measurements discussed previously.

The electrochemical CV-profiling method consists of alternate capacitance–voltage measurements and electrochemical etching. The CV-technique is a well-known method to measure the net donor ($N_D^+ - N_A^-$) or the net acceptor ($N_A^- - N_D^+$) concentration in semiconductors. The CV-technique using electrochemical etches was proposed by Ambridge and Faktor (1975a, 1975b). The authors demonstrated the use of an electrolyte as a barrier contact (Schottky contact) and also the electrolytical etching of the semiconductor. Both processes are carried out alternately in order to measure a doping *profile*. An

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

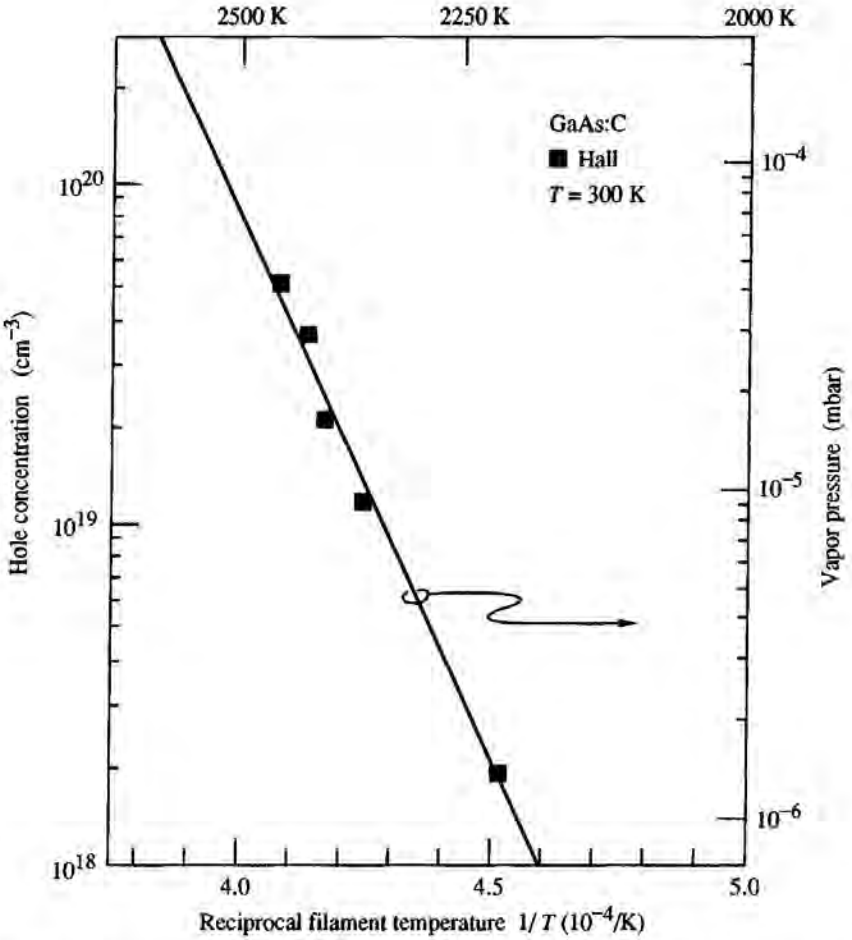


Fig. 5.5. Measured hole concentration at $T = 300$ K in GaAs: C as a function of temperature. The C filament temperature is measured with a pyrometer. The solid line is the calculated vapor pressure of graphite (after Nagle et al., 1991, courtesy of R. J. Malik).

accurate measure of the etching depth is obtained by integrating the etching current and by application of Faraday's law. Recently, the electrochemical CV-technique was reviewed by Blood (1986). This review covers many aspects of the technique including its limitations.

Two electrochemical CV-profiles on p-type and n-type GaAs are shown in Figs. 5.6 and 5.7, respectively. The measurement is carried out at room temperature. Six different effusion cell temperatures are chosen for the calibration which results in six different doping concentrations. The lowest and highest doping concentration are in the 10^{16} cm^{-3} and 10^{18} cm^{-3} range,

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

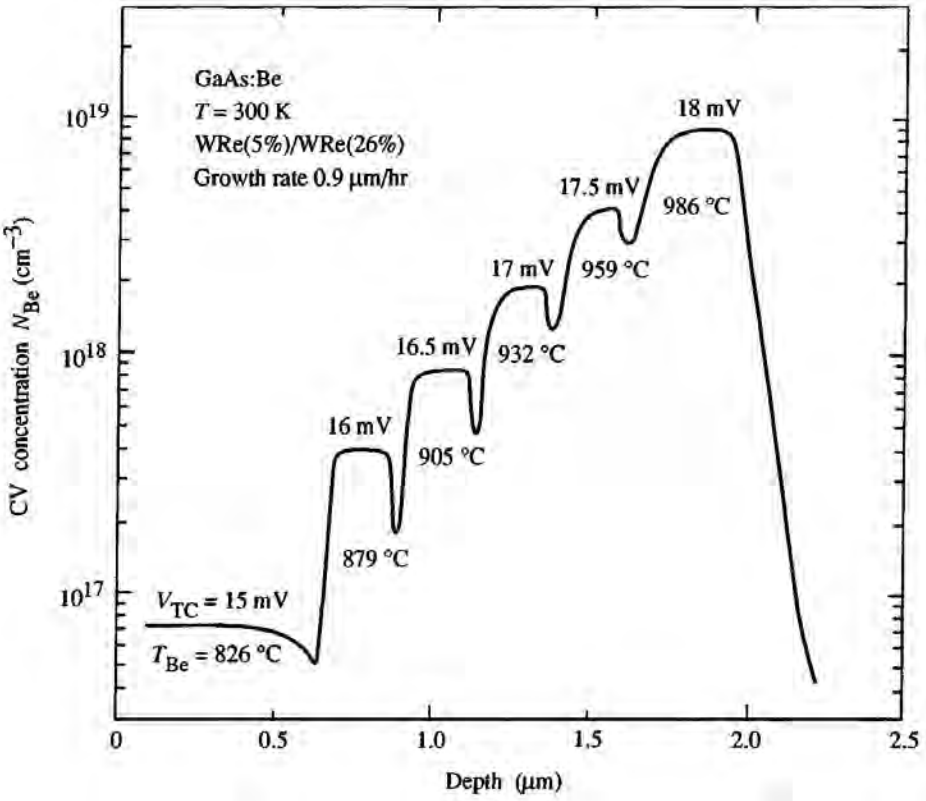


Fig. 5.6. Electrochemical capacitance–voltage depth profile on p-type GaAs:Be with different doping levels grown for calibration purposes. The thermocouple voltages and the corresponding temperatures are indicated for each doping level (courtesy of R. F. Kopf).

respectively. Figures 5.6 and 5.7 demonstrate that the entire doping calibration for one doping cell can be achieved with a single epitaxial growth run.

The accuracy of the CV-measurement is approximately ±5%. Two major factors limit the accuracy of the measurement. First, the dielectric constants of III–V semiconductors are frequently not known very accurately. Second, the contact diameter of the electrolyte Schottky contact may deviate from its nominal value. Since the dielectric constant and the contact diameter enter the calculated concentration, a small systematic error results. The accuracy of the capacitance measurement is usually excellent for state of the art capacitance bridges. Furthermore, the precision (repeatability) of the measurement is excellent, typically much better than 1%. Electrochemical CV-profiling equipment is offered commercially by several companies (e.g. by Polaron Equipment Ltd., Watford, UK).

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

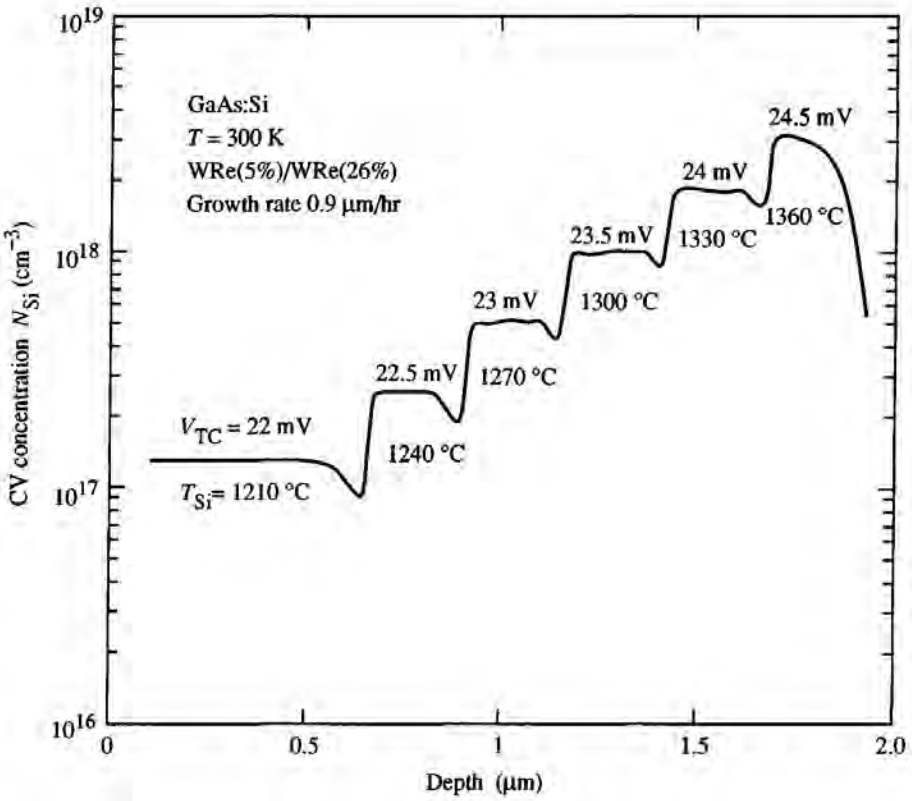


Fig. 5.7. Electrochemical capacitance–voltage depth profile on n-type GaAs:Si with different doping concentrations grown for calibration purposes. The thermocouple voltages and the corresponding temperatures are indicated for each doping level (courtesy of R. F. Kopf).

5.2 Doping homogeneity and uniformity

The production of III–V semiconductor devices from epitaxially grown wafers requires knowledge of the doping homogeneity along the growth axis as well as the uniformity across the wafer. Questions of production yield and manufacturing tolerances are closely related to the homogeneity and uniformity of layer thicknesses and doping. This section summarizes the result of a doping homogeneity study on Si-doped GaAs. The study includes depth-homogeneity, lateral uniformity, and a distribution test. The wafers were grown in a Varian Gen II MBE system with 5 cm rotating In-free substrate mounting.

Doping inhomogeneities are the result of three major causes. First, doping inhomogeneities along the radial direction of a wafer can be the result of non-uniform impurity flux. As shown in the previous section, the flux distribution is

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

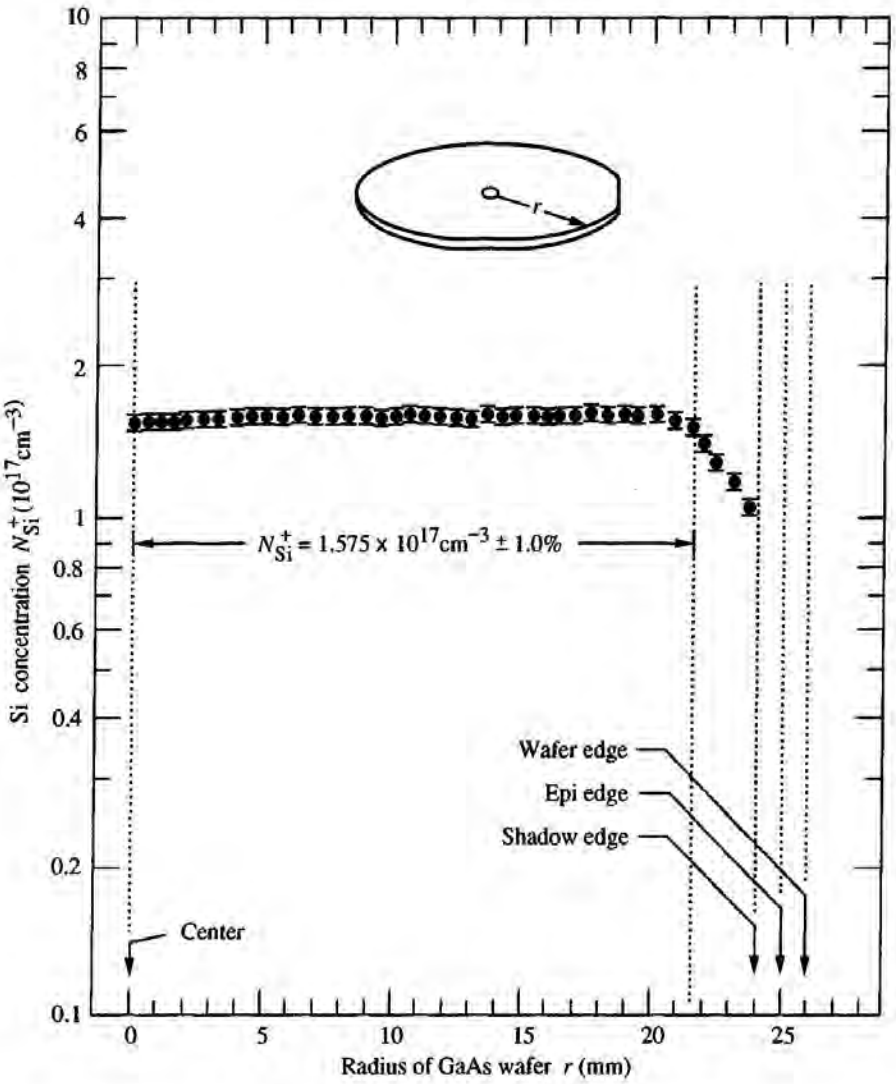


Fig. 5.8. Doping concentration of Si-doped GaAs grown by MBE as a function of distance from the center of the wafer.

cosine-shaped which results in a decrease of the doping concentration at the edges of the wafer. Second, thickness non-uniformities result in doping variation even if the dopant flux is perfectly uniform. Third, variations of the doping concentration along the growth axis can be caused by temporal variations of effusion cell or filament fluxes. If the doping cell (or other growth cells) temperature varies with time, the dopant flux and the doping concentration vary along the growth axis. Such temperature variations may also be caused by temperature changes in the cell environment (e.g. the liquid N_2 cryoshroud).

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

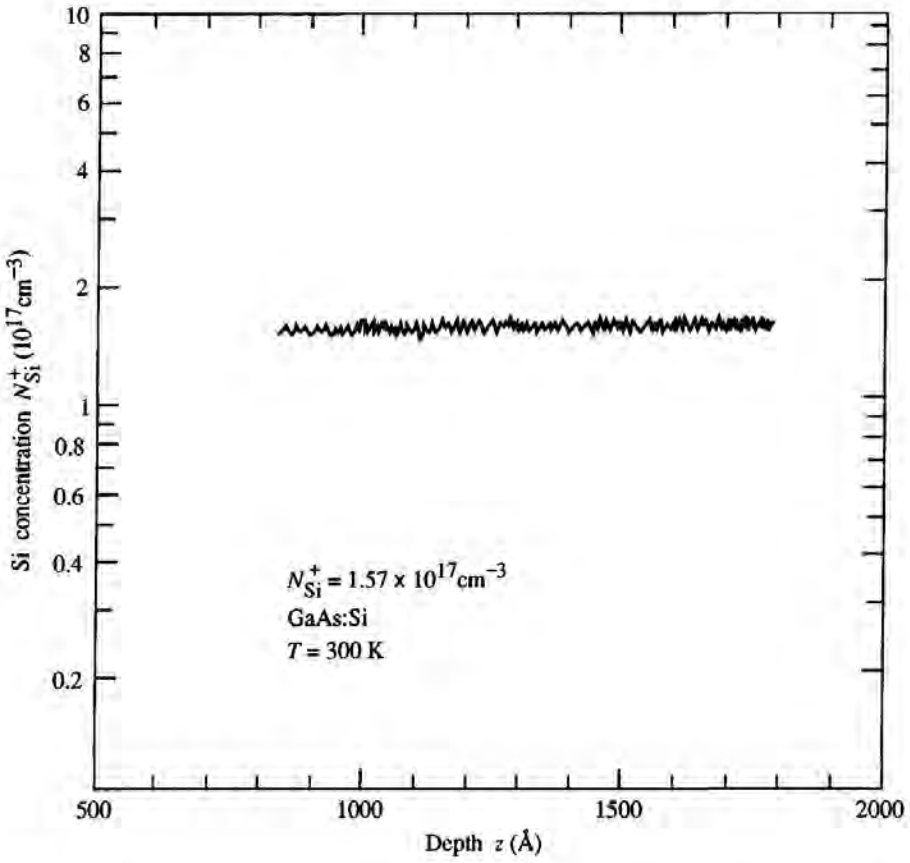


Fig. 5.9. Doping concentration of Si-doped GaAs grown by MBE as a function of depth.

The doping concentration of a 5 cm GaAs:Si wafer is shown in Fig. 5.8 as a function of distance from the center of the wafer. The concentration is measured by the capacitance–voltage technique using 500 μm diameter Ti/Au Schottky contacts. The accuracy of the technique is ±5% limited mostly by the uncertainty of the dielectric constant. The precision of the measurement is typically better than 1%. Since the wafer was rotated during growth, the doping pattern has rotational symmetry. The average doping concentration is 1.57×10^{17} cm^{-3}. The standard deviation of the measured points within a radius < 22 mm is better than 1%. This is a typical result for the MBE system used for this study. Comparable results can be expected for modern MBE systems from other manufactures.

The doping concentration decreases near the edge of the wafer. As indicated in Fig. 5.8, the physical edge of the *substrate* is 26 mm from the center of the

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

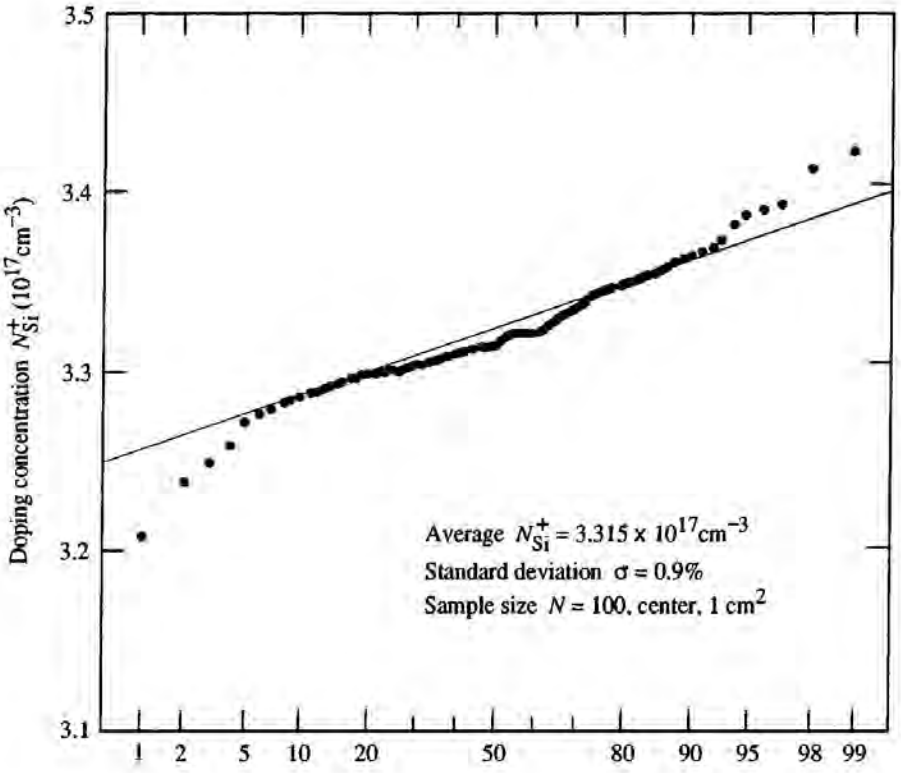


Fig. 5.10. Probability plot of the doping concentration of a Si-doped GaAs wafer. The 100 probing points are located in the center 1 cm^2 area of the 5 cm diameter wafer. The relative standard deviation of N_{Si}^{+} for the 100 points is 0.9%.

wafer. The physical edge of the *epitaxial layer* is 25 mm from the center of the wafer. The outer ring with 1 mm width is covered by mounting clamps. The outer 1 mm of the epitaxial layer has rough surface morphology when inspected under an optical microscope and is due to a shadow effect. At the edge of the epilayer the wafer holder used for In-free mounting shadows the epitaxial growth, resulting in non-stoichiometric and polycrystalline growth which results in incomplete doping activation and a general deterioration of the electronic properties. The decrease of the doping concentration at the edge of the wafer is therefore attributed to such edge effects, and cannot be explained by the cosine-shaped flux distribution of the effusion cell.

The doping concentration of Si-doped GaAs as a function of depth is shown in Fig. 5.9. The Si concentration is determined by a C-V measurement using Ti/Au Schottky contacts. The measurement temperature is 300 K. The range of depths accessible to the measurement ranges from 800 Å to 1800 Å at this doping

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

337-TA-1366 RX-0255

concentration. The doping concentration shown in Fig. 5.9 does not exhibit any depth dependence. The Si concentration is 1.57×10^{17} cm^{-3} for the entire depth range. The constant concentration demonstrates the excellent thermal stability of the doping effusion cell, i.e. the lack of any temperature excursions. It further demonstrates that the growth rate is constant. Any change in growth rate would inevitably result in a change in doping concentration.

The doping concentration of the Si-doped GaAs wafer is shown in a probability plot in Fig. 5.10. The measurement consists of 100 sampling points located in the center 1 cm^2 of a 5 cm GaAs wafer. The average doping concentration is 3.315×10^{17} cm^{-3}, and the lowest and highest concentration are 3.22×10^{17} cm^{-3} and 3.42×10^{17} cm^{-3}, respectively. The standard deviation of the measurement sample is 0.9%. The variation of the doping concentration is attributed to fluctuations in the contact area of the mechanical shadow mask used to define the Ti/Au Schottky contacts. The solid line drawn in Fig. 5.10 represents a fit to the experimental points. If the distribution were a normal (gaussian) distribution, experimental points would follow a straight line. Any non-gaussian distribution would significantly deviate from the linear line in the probability plot. Since the experimental points approximately follow a straight line the distribution of doping concentrations is either gaussian or close to a gaussian distribution.

5.3 Elemental dopants

This section covers the doping elements used in III–V semiconductors and has the sub-sections: 5.3.1 Group-IIa dopants, 5.3.2 Group-IIb dopants, 5.3.3 Group-IV dopants, and 5.3.4 Group-VI dopants. The sections cover the doping impurities used for epitaxial growth, bulk growth, diffusion, and ion implantation. Doping impurities that are introduced as chemical compounds (precursors), which is typical in vapor-phase epitaxy, will be discussed in the next chapter.

The present chapter also covers the specific characteristics of individual impurity elements including amphotericity, saturation, redistribution mechanisms, and incorporation conditions. The properties specific to a certain impurity will be emphasized, for example the autocompensation of Si or the segregation of Sn. However, many questions go beyond the individual characteristics of impurities. For example, *why* does Si autocompensate in GaAs but Be in GaAs does not? *Why* is the diffusion coefficient of Zn in GaAs concentration dependent? *Why* does Be strongly redistribute at concentrations $> 10^{20}$ cm^{-3}? Such questions which go beyond the experimental facts will not be discussed in this chapter. Instead the general characteristics of shallow impurity elements including the underlying physics, chemistry, and material science will be discussed in Chaps. 7

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

RX-0255.204

and 8.

5.3.1 Group-IIa dopants

The group-IIa impurities include Be, Mg, Ca, Ba, and Ra. The elements close to the bottom of the IIa column have large atomic radii. They are therefore not expected to incorporate easily in III–V semiconductors. The impurity elements near the top of the IIa column incorporate on the cation sites of III–V semiconductors. The group-IIa impurity atoms have two electrons in their outer shell (oxidation state 2), i.e. one electron less than the cation they substitute for. The group-IIa elements therefore are shallow acceptors in III–V semiconductors.

Beryllium, Be

Summary: Be is a shallow acceptor in III–V semiconductors with high electrical activity. Be diffuses slowly at all except the highest concentrations. Strong redistribution of Be impurities occurs in the very high concentration regime. Be is used in MBE-growth, ion implantation, and as AuBe alloy for p-type ohmic contact metallization.

Be substitutes for cations and is shallow acceptor in all III-V semiconductors. The ionization energy of Be is 28 meV in GaAs and 41 meV in InP as determined by photoluminescence measurements (Ashen et al., 1975; Skromme et al., 1984). However, lower activation energies of 19 meV for Be in GaAs were also reported (Wood, 1985a and 1985b).

Be exhibits near complete activation in III–V semiconductors without compensation effects. The high activation of Be in GaAs is illustrated in Fig. 5.11 which shows the two-dimensional hole density as a function of the Be density (Schubert et al., 1990a). Be impurities were deposited during interruption of epitaxial growth by MBE. A 1000 Å thick GaAs cap layer was grown after impurity deposition on the non-growing surface. The experimental result of Fig. 5.11 demonstrates that the free hole concentration follows the Be impurity concentration in a linear fashion. Furthermore, the experimental free carrier concentration is in agreement with the Be concentration inferred from the deposition rate (which is obtained from an independent secondary ion mass spectrometry calibration) shown by the solid line. Therefore, all Be impurities incorporated into the GaAs crystal are electrically active. The fraction of electrically inactive Be impurities is insignificantly small.

However, the high activation of Be impurities depends on the host material and the incorporation conditions. Experimental results of Hamm et al. (1989) demonstrate that a fraction of Be impurities in $Ga_{0.47}In_{0.53}As$ are inactive especially at high growth temperatures. The lack of activity can be caused by Be

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

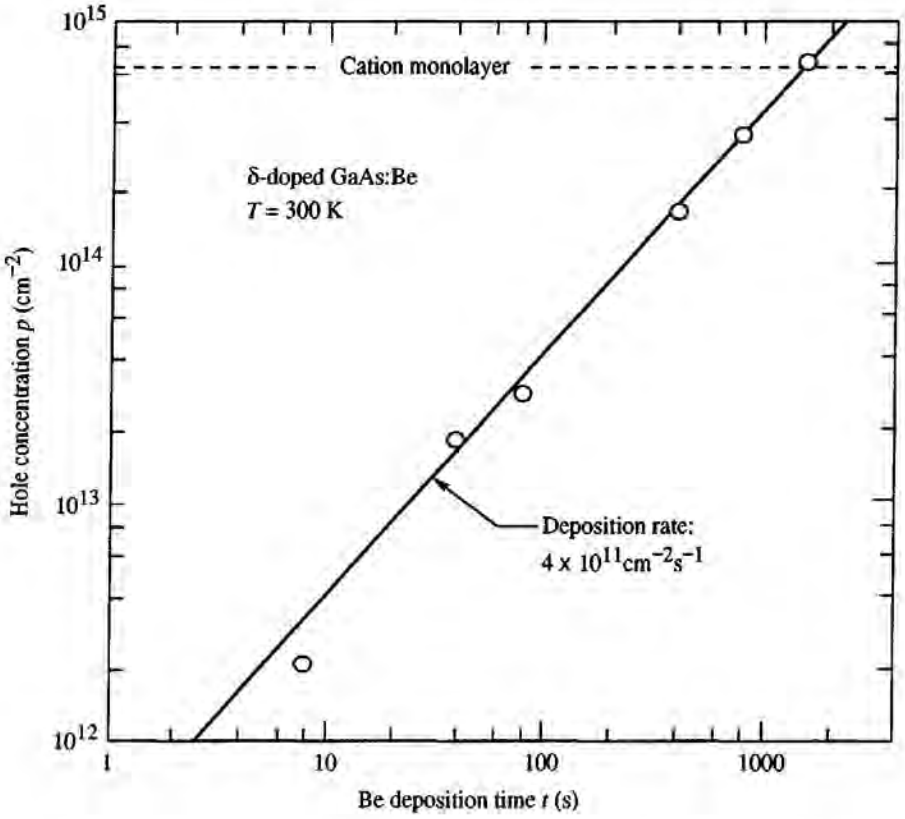


Fig. 5.11. Two-dimensional hole concentration of Be-doped GaAs as a function of the Be deposition time (Be density). The dopants are deposited during the suspension of MBE growth at a temperature of 500 °C. The hole density follows the acceptor density without any indication of saturation effects.

precipitates (Hamm et al., 1989) or by inactive Be atoms on the semiconductor surface. Inactive Be impurities in GaAs were concluded by Duhamel et al. (1981) to be due to Be interstitials. The authors found an increasing density of stacking faults with increasing Be concentration which they attributed to an increasing concentration of Be complexes. A degradation of surface morphology in highly Be doped GaAs was reported Ilegems (1977) which further indicates the dependence of Be incorporation on the growth conditions.

Be in GaAs is a slowly diffusing impurity for all except the highest Be concentrations. Ilegems (1977) and McLevige et al. (1978) found a small diffusion coefficient of $D = 10^{-13} - 10^{-14}$ cm^2/s for $T = 900$ °C and $N_{Be} = 10^{19}$ cm^{-3}. The temperature dependence of the Be diffusion coefficient was determined to be $D = D_0 \exp(-E_a/kT)$ with $E_a = 1.95$ eV and

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

$D_0 = 2 \times 10^{-5}$ cm^2 (Schubert et al. 1990a) for samples with $N_{Be} \cong 10^{19}$ cm^{-3}. Both results establish that Be is a very slow diffuser in GaAs. The diffusion of Be in GaP was studied by Ilegems and O'Mara (1972). The authors found a small diffusion length for $N_{Be} < 10^{19}$ cm^{-3} but a strong redistribution of Be at concentrations $> 10^{19}$ cm^{-3}.

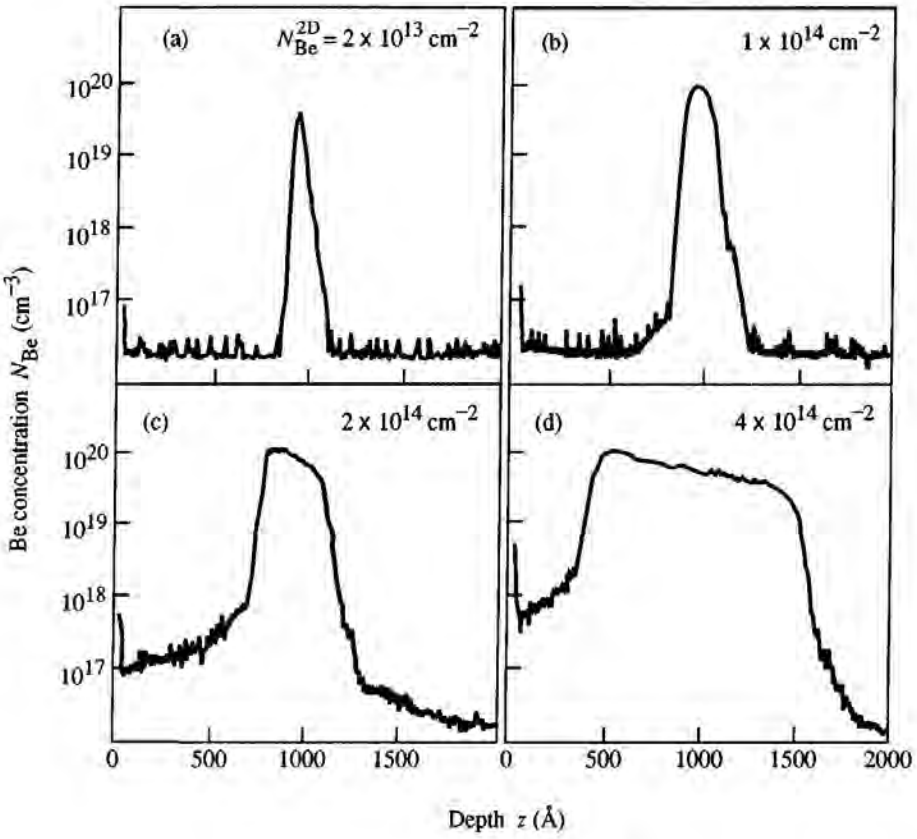


Fig. 5.12. Secondary ion mass spectrometry (SIMS) profiles of Be-doped GaAs with different densities. The dopants are deposited during the suspension of MBE growth at 500 °C. Strong spreading of Be occurs at high densities. The maximum attainable Be concentration is 10^{20} cm^{-3}.

Strong redistribution of Be in GaAs occurs at very high doping concentrations of about 10^{20} cm^{-3}. The occurrence of strong impurity redistribution is illustrated in Fig. 5.12 which shows the SIMS (secondary ion mass spectrometry) profiles of different Be densities in GaAs doped during growth interruption (Schubert et al., 1990a). A 1000 Å thick top layer is grown after the growth interruption. Figure 5.12 illustrates that little Be redistribution occurs at two-

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

dimensional Be densities of $\leq 2 \times 10^{13}$ cm^{-2}. However, for high Be densities, strong redistribution of Be occurs in such a way that peak concentration of Be does not exceed 10^{20} cm^{-3}.

It was proposed by Schubert et al. (1992) that the strong redistribution of Be in GaAs at concentrations $N_{Be} \cong 10^{20}$ cm^{-3} is due to coulombic repulsion of ionized substitutional impurities. The authors showed that high concentration of ionized impurities results in substantial coulombic interaction energies which limit the incorporation of impurities to a maximum concentration which is on the order of $10^{19} - 10^{21}$ cm^{-3} depending on the diffusion coefficient and the growth rate. The authors further showed that the flat top-hat distribution of Be shown in Fig. 5.12 can be fully explained by repulsive coulombic interaction. A further discussion of the impurity concentration limit caused by coulombic interaction is given in Chaps. 7 and 8.

Beryllium is also used as dopant for ohmic contact metallization to p-type semiconductors. A typical contact metal is a AuBe alloy with a composition of 99.1% Au and 0.9% Be. Low contact resistances are obtained by typical alloying cycles of 20 seconds at 380 °C.

Beryllium easily getters oxygen and forms the oxide BeO. Both Be and BeO, but especially BeO, are highly toxic and carcinogenic. Precautions are required for their handling.

Magnesium, Mg

Summary: Mg is a shallow acceptor impurity in III–V semiconductors. The Mg incorporation coefficient is low at typical MBE growth temperatures. Only a small fraction of Mg is incorporated during MBE at high substrate temperatures.

Magnesium occupies cation sites and is a shallow acceptor in III–V semiconductors with an ionization energy of 28 mV in GaAs and 41 meV in InP (Ashen et al., 1975; Skromme et al., 1984). Mg doping of GaAs and $Al_xG_{1-x}As$ was first reported in 1972 by Cho and Panish. The authors reported a low sticking coefficient of Mg in GaAs grown at 560 °C. However, they found a much larger sticking coefficient for Mg-doped $Al_xGa_{1-x}As$ compared to GaAs at the same growth temperature of 560 °C. The authors attributed the increased incorporation of Mg into $Al_xGa_{1-x}As$ to the enhanced concentration of oxygen in $Al_xGa_{1-x}As$ which forms strong MgO bonds and thus prevents Mg from desorbing. Hole concentrations of $p = 10^{19}$ cm^{-3} in Mg-doped $Al_{0.20}Ga_{0.80}As$ were reported. These results and their interpretation are somewhat surprising since an impurity-oxygen complex is not expected to act as a simple acceptor.

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

In subsequent publications Joyce and Foxon (1977) reported a high incorporation coefficient during MBE growth but a low electrical activity of the incorporated Mg impurities. Furthermore, Kirchner et al. (1981) showed that even if Mg is incorporated into GaAs it can be electrically inactive due to the formation of MgO. The authors showed that the formation of MgO can be reduced by adding a nonvolatile oxygen-gettering element (such as Al) to the Ga effusion cell, or by growing $Al_xGa_{1-x}As$ in order to form AlO complexes rather than electrically inactive MgO complexes. Finally, Wood et al. (1982) carefully analyzed the dependence of Mg incorporation into GaAs as a function of the growth temperature. A low growth temperature of 500 °C results in a Mg impurity incorporation coefficient of 0.3. High growth temperatures of 600 °C reduce the incorporation coefficient by three orders of magnitude to 0.3×10^{-3} probably due to the high vapor pressure of Mg. The reduced incorporation coefficient leads to a reduction of the free hole concentration from the 10^{18} cm^{-3} to the 10^{16} cm^{-3} range which occurs concomitantly with the increase of growth temperature from 500 °C to 600 °C.

The current understanding of Mg impurities in GaAs and $Al_xGa_{1-x}As$ is summarized as follows. The incorporation coefficient of Mg is close to unity only for temperatures <500 °C and decreases rapidly at higher temperatures. Magnesium has a strong affinity to oxygen and the formation of MgO leads to electrically inactive complexes. These properties restrict the use of Mg to materials grown at low temperatures <500 °C.

Calcium, Ca

Doping of GaAs with Ca revealed no electrical activity (Wood et al., 1982). An effusion cell charged with Ca_3As_2 was used as the Ca doping source in this study.

5.3.2 Group-IIb dopants

The group-IIb impurities include Zn, Cd, and Hg. Mercury has a high vapor pressure and is therefore very volatile. A low incorporation coefficient is therefore expected for Hg doping. Both zinc and cadmium were employed for doping of III–V semiconductors.

Zinc, Zn

Summary: Zn is a shallow acceptor in III–V semiconductors with high electrical activity. Zn diffuses rapidly at high concentrations. Zn doping is used in OMVPE growth, solid-source and vapor-phase diffusion, and in a AuZn alloy for p-type ohmic contact metallization.

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

337-TA-1366 RX-0255

Zinc substitutes for cations in III–V semiconductors and is a shallow acceptor. The Zn ionization energy is 31 meV in GaAs and 47 meV in InP as determined by photoluminescence measurements (Asher et al., 1975; White et al., 1973). Zn-acceptors exhibit high electrical activity and show no evidence of autocompensation.

Doping of GaAs with Zn by MBE was first reported by Arthur (1973) who showed that the bonding of Zn to the GaAs surface is weak and that Zn desorbs rapidly from the surface at temperatures of 600 °C. The activation energy for desorption was shown to equal the enthalpy for vaporization of pure Zn. The high vapor pressure of Zn thus makes Zn unsuited for doping during MBE growth.

Zinc is also diffused from a Zn layer into III–V semiconductors. For example, pn-junctions can be formed by diffusing Zn into n-type semiconductors. Evaporated Zn films on the semiconductors serve as a reservoir. It is useful to provide As overpressure during diffusion of Zn into GaAs to reduce As outdiffusion at high temperatures. The As overpressure is usually achieved by an additional, elemental As source enclosed with the semiconductor in an ampoule.

Zinc in GaAs was shown to have a strongly concentration-dependent diffusion coefficient. It was shown that the Zn diffusion profiles cannot be explained by the well-known Fickian (i.e. concentration-independent) diffusion (Tuck and Kadhim, 1972; Kadhim and Tuck, 1972; Tuck, 1988). In order to obtain the concentration-dependent diffusion coefficient Kadhim and Tuck (1972) employed the Boltzmann–Matano method (see for example Tuck, 1988). This method allows one to determine the concentration dependent diffusion coefficient. The Zn diffusion coefficient in GaAs at $T = 1000$ °C as a function of concentration is shown in Fig. 5.13 (Kadhim and Tuck, 1972). The authors found a diffusion coefficient as high as 10^{-7} cm^2/s at the highest Zn concentrations. The dependence of the diffusion coefficient is well described by the power law $D \propto N_{Zn}^{1.8}$. An exponent of two (2) in the power law (i.e. $D \propto N_{Zn}^{2}$) can be explained by the substitutional-interstitial diffusion model discussed in Chap. 8. Zinc is considered to be a model impurity for the substitutional-interstitial diffusion mechanism.

Zinc is a popular dopant in VPE and CBE growth. Diffusion coefficients of Zn in OMVPE-grown GaAs are several orders of magnitudes smaller as compared to the 'infinite' source diffusion experiments (i.e. diffusion from a Zn film). At concentration of 1×10^{18} cm^{-3} the Zn diffusion coefficients reported at 900 °C are $5 \times 10^{-12} - 10 \times 10^{-12}$ cm^2/s (Hobson et al., 1989; Enquist et al., 1988). At the lower growth temperature of 625 °C the authors reported a diffusion coefficient as low as 7×10^{-17} cm^2/s. Such low diffusion coefficients make feasible the abrupt Zn doping profiles which are required in heterostructure

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

RX-0255.210

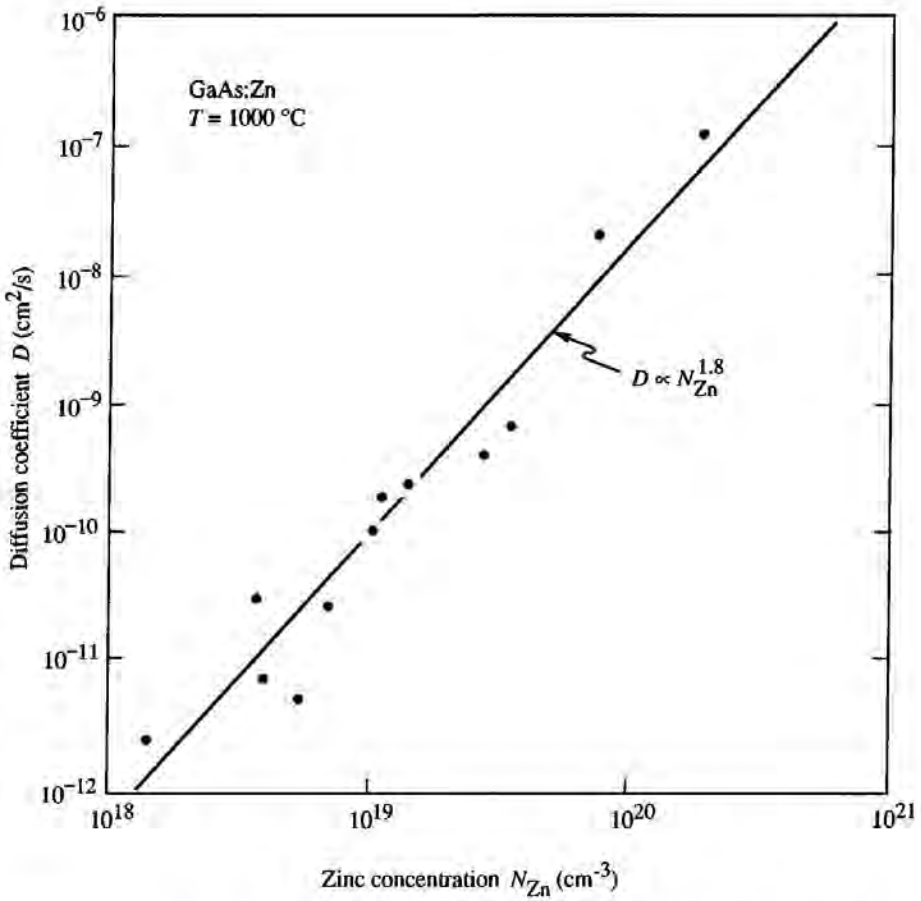


Fig. 5.13. Diffusion coefficient of Zn in GaAs at $T = 1000$ °C as a function of the Zn concentration. The diffusion experiments were performed at the As dissociation pressure (after Kadhim and Tuck, 1972).

devices.

Zinc is also used for ohmic contact formation with p-type III–V semiconductors. A AuZn alloy (95% Au, 5% Zn) is frequently used for contact metallization. A typical alloying cycle for AuZn contacts on GaAs is 20 s at 380 °C.

Cadmium, Cd

Cadmium is a shallow acceptor in GaAs and InP with an activation energy of 35 meV and 56 meV respectively (Ashen et al., 1975; White et al., 1973). It has a vapor pressure approximately 1–2 orders of magnitude higher than Zn. Therefore, Cd is quite volatile which precludes it as a viable doping impurity for MBE growth.

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

Cadmium has been implanted in $Ga_{0.47}In_{0.53}As$ in order to obtain p-type conductivity (Vescan et al., 1984). The authors attempted to activate the Cd impurities by annealing at a temperature of 800 °C. Despite annealing, only a small fraction of Cd impurities were electrically activated. A further drawback of Cd is its toxicity.

Mercury, Hg

Mercury was reported to form an acceptor level in InP with an ionization energy of 95 meV (Williams et al., 1973). Mercury has a high vapor pressure which makes it very volatile. This unfavorable property precludes Hg as a viable acceptor in III–V semiconductors.

5.3.3 Group-IV dopants

The group-IV impurities include C, Si, Ge, Sn, and Pb. The latter element has a large atomic radius and is therefore not expected to incorporate easily in III–V semiconductors. The remaining group-IV impurities, i.e. C, Si, Ge, and Sn share ***four key characteristics*** which are summarized as follows. First, all column IV impurities are ***amphoteric***, i.e. they can occupy either the group-III or the group-V (cation or anion) site of the zincblende crystal. As a result, group-IV impurities are donors or acceptors for cation site or anion site occupancy, respectively. Second, column-IV impurities ***autocompensate***, which is a result of their amphotericity. Consider a column-IV impurity that predominantly occupies the cation site, i.e. is a donor. If some of the impurity atoms occupy anion sites, i.e. acceptor sites, then these acceptors partially compensate the donors which results in a reduction of the free carrier concentration. Since the same species of impurities cause the compensation, the phenomenon is called autocompensation. Third, the electrical activity ***saturates*** at high impurity concentrations. With increasing doping concentration the tendency of group-IV impurities to autocompensate increases and ultimately leads to the saturation of the free carrier concentration. Fourth, the free carrier concentration depends on ***growth temperature*** and ***V/III flux ratio*** during epitaxial growth by MBE, GSMBE and OMVPE. Generally, low growth temperatures result in the highest free carrier concentration. Furthermore, a V/III flux ratio that makes the desired crystal site available (e.g. a high V/III flux ratio for Si doping in order for Si to predominantly occupy group-III sites) results in the highest free carrier concentrations.

What is the materials physics behind these very general properties? There are, of course, several answers and models that explain the specific properties of group-IV impurities. The general models of impurity characteristics will be discussed in Chaps. 7 and 8. In the following, the experimental part of impurity

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

characteristics will be emphasized and the specific properties of group-IV impurities will be illustrated and summarized.

Carbon, C

Summary: Carbon is a shallow acceptor in GaAs and a shallow donor in InAs. It strongly compensates in $Ga_{0.47}In_{0.53}As$. C is a stable impurity and diffuses very slowly. Very high acceptor concentrations $>10^{19}$ cm^{-3} can be achieved by filament doping during MBE and $>10^{20}$ cm^{-3} in CBE.

Carbon substitutes for anions in GaAs and forms a shallow acceptor level of 26 meV (Ashen et al., 1975). Although C rarely forms acceptors in InP (Skromme et al., 1984), a C acceptor level of 43 meV in InP has been reported (Hess et al., 1974). The ionization energies were determined by photoluminescence experiments.

Carbon is a very stable impurity in GaAs grown by MBE. Malik et al. (1988) showed for C-filament doping of GaAs that very little diffusion of C impurities occurs at 800 °C. The negligible diffusion found by Malik et al. agrees with earlier reports on C doping by flow-rate modulation epitaxy (Kobayashi et al., 1987). Kobayashi et al. reported a diffusion coefficient of 2×10^{-16} cm^2/s at a temperature of 800 °C and a C concentration of $\cong 10^{18}$ cm^{-3}. Subsequently, the temperature dependence of C diffusion in GaAs was determined and expressed as $D = D_0 \exp(-E_a/kT)$ with $E_a = 1.75$ eV and $D_0 = 5 \times 10^{-8}$ cm^2/s (Schubert, 1990) which yields $D = 3 \times 10^{-16}$ cm^2/s at 800 °C. The diffusion coefficient of C is about two orders of magnitude lower than that of Be in GaAs which further illustrates the stability of C.

Carbon acceptor concentrations of 5×10^{19} cm^{-3} (Malik et al., 1988, Hoke et al., 1991a and 1991b) were achieved in GaAs using C (graphite) filament sources. Even higher concentrations $>10^{20}$ cm^{-3} were achieved by using alkyl sources as C doping precursors (see Chap. 6).

Carbon incorporation during MBE growth depends on the V/III flux ratio. The measured hole concentration of C-doped GaAs as a function of the V/III flux ratio is shown in Fig. 5.14. (Hoke et al., 1991a and 1991b). The As_4/Ga flux ratio was varied between 10 and 34. With increasing flux ratio a decreasing hole concentration is measured. The relative change in the free hole concentration is approximately 20%. The authors also compared C-doped GaAs grown with As_4 and As_2. They found no difference in carrier concentration and mobility for As_4 and As_2 grown samples that exceeded their experimental uncertainty of 10%. The experimental results shown in Fig. 5.14 can be understood in terms of increased C autocompensation at high V/III flux ratios. Carbon predominantly

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

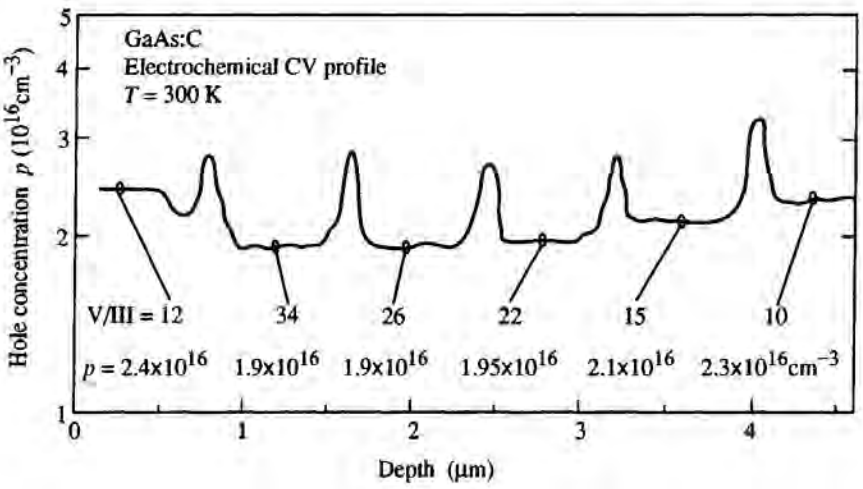


Fig. 5.14. Hole concentration versus depth for epitaxial C-doped GaAs containing six 0.8 μm thick layers grown with different As_4/Ga flux ratios. The layers are delineated by doping spikes (after Hoke et al., 1991a).

occupies group-V sites. However, as the As_4 flux increases, As sites become rapidly occupied and are less likely to be occupied by C. As a result, C is more likely to occupy Ga sites on which C is a donor. The experimental result of Fig. 5.14 can thus be explained by autocompensation of C at high V/III flux ratios. The difference in bond energy of the Ga–C bond (2.65 eV/bond) and the As–C bond (2.47 eV/bond) (Hartley and Patai, 1982) is sufficiently large to make As the preferred C site. However, the difference in bond energy of ≅ 700 meV (four bonds) is approximately nine times larger than kT at the growth temperature which does not rule out the occupation of the Ga site by a fraction of C atoms. Furthermore, caution is required when Ga–C and As–C bond energies are applied to GaAs:C crystals.

Room-temperature Hall mobilities of C- and Be-doped GaAs and $Al_{0.30}Ga_{0.70}As$ grown by MBE are shown in Fig. 5.15 (Hoke et al., 1991b). The plot shows very similar mobilities for C and Be doping. Furthermore, the hole mobilities agree with values compiled by Blakemore (1982). The hole mobilities in $Al_xGa_{1-x}As$ are lower by a factor of 2–2.5 than the mobilities in GaAs. The reduction in mobility is due to alloy scattering caused by potential fluctuation (due to random cation distribution) in the ternary $Al_xGa_{1-x}As$.

While C is an acceptor in GaAs, it is a donor in InAs. In $Ga_{0.47}In_{0.53}As$ carbon strongly compensates but still shows residual p-type conductivity (Ito and Ishibashi, 1991). It is likely that the conductivity type of C-doped $Ga_{0.47}In_{0.53}As$

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

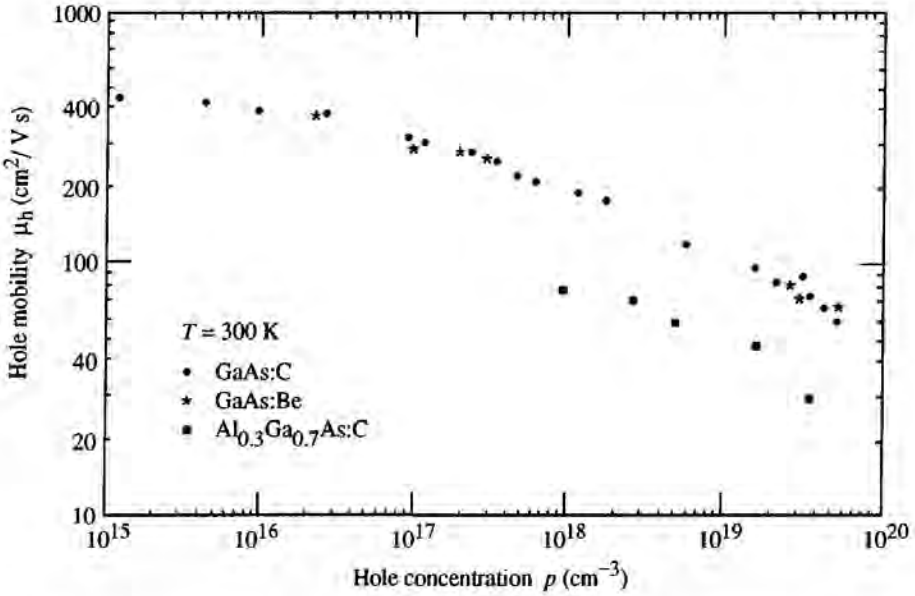


Fig. 5.15. Room temperature Hall mobility as a function of hole concentration for MBE-grown C- and Be-doped GaAs and for C-doped $Al_{0.30}Ga_{0.70}As$ (after Hoke et al., 1991b).

depends strongly on the V/III flux ratio. Ito and Ishibashi showed that under their growth conditions a conversion from p-type to n-type conductivity occurs at an In mole fraction of $x = 0.60$ in $Ga_{1-x}In_xAs$.

High non-radiative recombination rates have been found in MBE-grown highly C-doped GaAs using photoluminescence studies (Calderon et al., 1992). The room-temperature photoluminescence intensity was found to decrease strongly at C doping concentrations exceeding 10^{19} cm^{-3}. The non-radiative rate was larger than the expected Auger recombination rate. The authors therefore concluded that deep centers associated with the high C concentrations are the origin of the high non-radiative recombination.

Silicon, Si

Summary: Si is a shallow donor in all III–V semiconductors. Si auto compensates little at low concentrations, but strongly in the high concentration regime. It diffuses slowly and slightly segregates along with the growing surface during MBE growth. Si is a popular n-type dopant for MBE growth and ion implantation.

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

337-TA-1366 RX-0255

Silicon predominantly occupies cation sites in III–V semiconductors and is a hydrogenic shallow donor with 4–6 meV ionization energy in GaAs and InP. A small fraction of Si incorporates as an acceptor. The Si acceptor ionization energy is 35 meV in GaAs (Ashen et al., 1975). Even though Si is predominantly a donor, it can, under specific growth conditions, also be incorporated as an acceptor for example in liquid-phase epitaxial growth (Rupprecht et al., 1966; Spitzer and Panish, 1969).

Silicon is a stable impurity in III–V semiconductors with only little diffusion occurring during GaAs growth at moderate temperatures. The diffusion coefficient was expressed in terms of the equation $D = D_0 \exp(-E_a/kT)$ with an activation energy $E_a = 2.45$ eV and $D_0 = 4 \times 10^{-4}$ cm^2/s for diffusion of atomic Si in GaAs (Schubert et al., 1988). Greiner and Gibbons (1984) investigated the diffusion of Si pairs in GaAs and found a higher diffusion coefficient for pair diffusion. The diffusion coefficient of Si in $Al_xGa_{1-x}As$ was found to be larger than in GaAs (Lanzilloto et al., 1989). The authors found that Si spreads significantly further in $Al_{0.25}Ga_{0.75}As$ as compared to GaAs. The determination of the Si diffusion coefficient in $Al_{0.30}Ga_{0.70}As$ yielded indeed higher values, namely an activation energy of $E_a = 1.3$ eV and $D_0 = 4 \times 10^{-8}$ cm^2/s (Schubert, et al., 1989).

In addition to the pure diffusive motion of Si in GaAs, Si was also found to redistribute predominantly to the surface during growth by MBE (Beall, et al., 1988). The authors found marked preferential migration towards the surface especially at high growth temperatures. The authors proposed that conventional surface segregation can account for the Si redistribution (Harris et al., 1984). A second model for surface migration is based on the drift of impurities in the surface field (Schubert et al., 1990b). The electric field is the result of the pinning of the Fermi level at the semiconductor surface. The two segregation models will be discussed further in Chap 8.

The free electron concentration of Si-doped III–V semiconducting material exhibits a peculiar saturation in the high concentration regime. The measured free electron concentration of Si-doped GaAs grown by MBE is shown in Fig. 5.16 as a function of reciprocal Si effusion cell temperature (Chai et al., 1981). For effusion cell temperatures ≤ 1250 °C, the free electron concentration does not increase further but saturates at approximately 5×10^{18} cm^{-3}. The exact concentration at which the free carrier concentration saturates depends on the growth conditions. Values of $n = 1.1 \times 10^{19}$ cm^{-3} (Heiblum et al., 1983), 1.6×10^{19} cm^{-3} (Malik et al., 1992), and 1.8×10^{19} cm^{-3} (Ogawa and Baba, 1985) have been reported for the maximum electrically active Si concentration in GaAs. Such high values for the free carrier concentration are obtained under

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

Innoscience1366-00138889

RX-0255.216

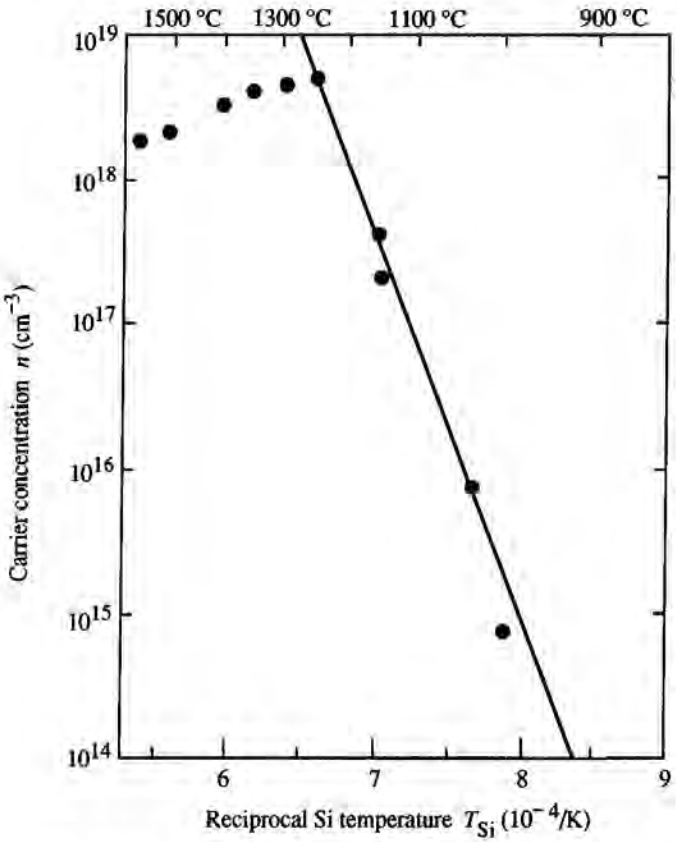


Fig. 5.16. Free electron concentration of Si-doped GaAs as a function of the reciprcal Si effusion cell temperature. For Si cell temperatures > 1250 °C, the free electron concentration saturates (after Chai et al., 1981).

optimized growth conditions, i.e. for low growth temperatures (e.g. 500 °C), high V/III flux ratios (e.g. V/III = 20), or growth with As_2. Under fairly standard growth conditions (T = 580 °C, V/III = 4) the onset of free carrier saturation in MBE grown Si-doped GaAs occurs at a Si concentration of $4 \times 10^{18} - 5 \times 10^{18}$ cm^{-3} (Kopf and Schubert, 1990).

The physical origin of the saturation of the free electron concentration in highly Si-doped GaAs was shown to be electrically inactive Si, Si pairs, and Si acceptors, that drastically increase in concentration for nominal doping concentrations above 10^{19} cm^{-3} (Maguire et al., 1987), as demonstrated by infra-red local vibrational mode absorption spectroscopy on highly Si-doped GaAs grown by MBE. Two local vibrational mode spectra are shown in Fig. 5.17 for nominal Si concentrations of 5×10^{18} cm^{-3} and 3×10^{19} cm^{-3}. The spectrum obtained from the sample doped at 5×10^{18} cm^{-3} reveals that ^{28}Si on

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

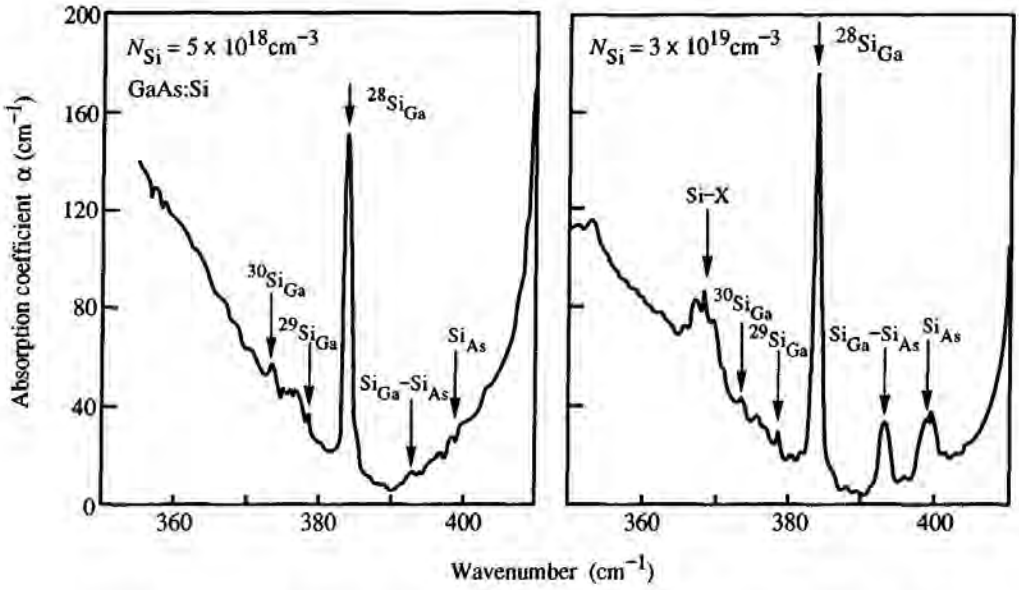


Fig. 5.17. Local vibrational mode (LVM) spectra of heavily Si-doped GaAs with Si concentrations of $5 \times 10^{18} cm^{-3}$ and $3 \times 10^{19} cm^{-3}$. At the higher concentration, Si_{As}, Si_{Ga}–Si_{As} pairs, and Si complexes are detected which compensate for Si_{Ga} donors, and thus reduce the free carrier concentration (after Maguire et al., 1987).

Ga sites is the dominant species. Furthermore, vibrational resonances of the isotopes $^{29}Si_{Ga}$ and $^{30}Si_{Ga}$ are found. Thus Si occupies predominantly substitutional group-III sites at this doping concentration. The concentration of Si acceptors, i.e. Si on As sites, as well as neutral donor–acceptor pairs is comparatively small.

The local vibrational mode spectrum changes markedly at a doping concentration of $N_{Si} = 3 \times 10^{19}$ cm^{-3}. Three new lines appear which the authors attributed to (i) Si acceptors, i.e. Si on arsenic sites, (ii) neutral Si donor–acceptor pairs, and (iii) a Si-complex (Si–X) possibly being a Ga-vacancy-Si complex. However, an unambiguous identification of the Si–X complex is not provided. The authors assume that all of the three configurations, that is Si_{As}, $Si_{As} - Si_{Ga}$, and $Si - X$, contribute to the reduction of the free carrier concentration. The free carrier concentration measured is only 8×10^{17} cm^{-3} at the nominal Si doping concentration of 3×10^{19} cm^{-3}. Maguire et al. concluded that 40% of the Si exists as Si_{Ga}, 19% as Si_{As}, 14% as $Si_{Ga} - Si_{As}$, and 27% as Si–X. The work clearly shows that electrically inactive Si causes the saturation in highly Si-doped GaAs.

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

An alternative model for the free carrier saturation in n-type GaAs was given by Theis et al. (1988). The authors proposed that at sufficiently high doping concentrations the Fermi level crosses a second, non-hydrogenic Si level which is resonant with the conduction band. As a result, this resonant level becomes occupied and inhibits a further increase of the free electron concentration.

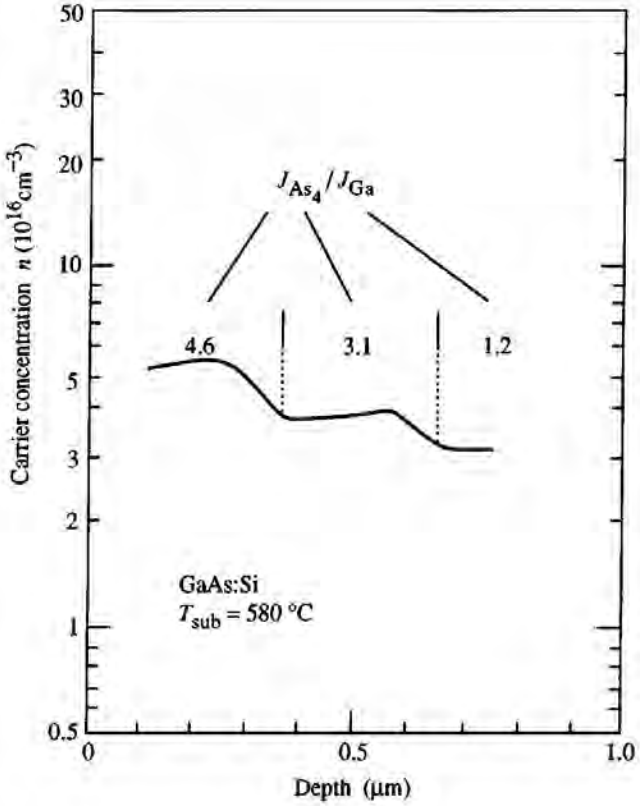


Fig. 5.18. Effect of V/III flux ratio on the free carrier concentration of MBE-grown Si-doped GaAs (after Chai et al., 1981).

Further experiments clarifying the Si incorporation into GaAs were performed by Chai et al. (1981). The authors investigated the dependence of the free carrier concentration of Si-doped GaAs as a function of the MBE growth conditions. Chai et al. varied the As_4/Ga flux ratio and the growth temperature. The experimental free carrier concentration versus depth for different As_4/Ga flux ratios is shown in Fig. 5.18. The flux ratio is varied between V/III = 1.2 – 4.6, i.e. epitaxial growth was under As stabilized conditions. The free carrier concentration increases with increasing V/III flux ratio from

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

$n = 3.2 \times 10^{16}$ cm^{-3} (V/III = 1.2) to $n = 5.3 \times 10^{16}$ cm^{-3} (V/III = 4.6). The increase of the free carrier concentration with V/III flux can be understood in terms of the availability of sites and the amphoteric nature of Si in GaAs. Consider the chemical reaction

$$Si_{Ga} + V_{As} \leftrightarrow Si_{As} + V_{Ga} \tag{5.13}$$

where V_{As} and V_{Ga} are As and Ga vacancies, respectively. In order to increase the concentration of Si_{Ga} (Si donors) it is, according to the *law of mass action*, advantageous to decrease the concentration V_{As}. The reduction of As vacancies on the surface of the growing semiconductor as well as in the bulk is achieved by higher As partial pressures, i.e. higher V/III flux ratios. As the V/III ratio becomes larger the Si donor concentration increases along with the free carrier concentration. The increase of the free carrier concentration appears as steps in Fig. 5.18. The incorporation of amphoteric impurities is further discussed in Chap. 7.

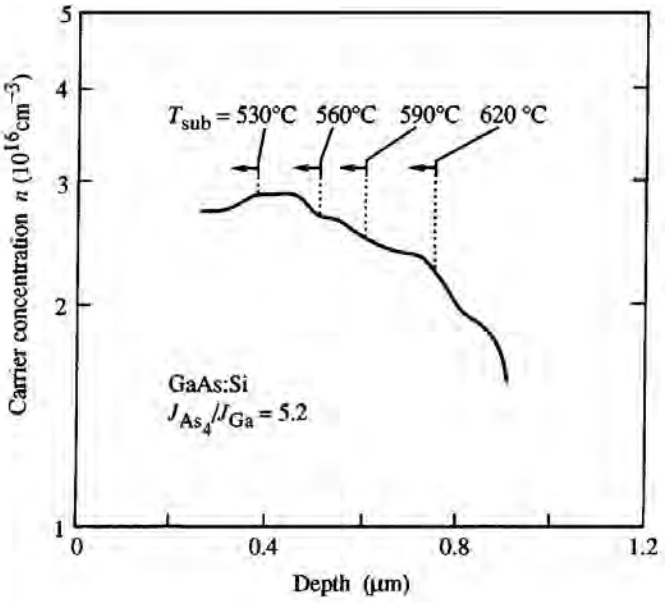


Fig. 5.19. Effect of substrate temperatute T_{sub} on the free electron concentration of Si-doped GaAs grown by MBE at a constant V/III flux ratio (after Chai et al., 1981).

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

The free carrier concentration of Si-doped GaAs as a function of depth for different growth temperatures is shown in Fig. 5.19. (Chai et al., 1981). The substrate temperature was lowered during growth from 620 °C to 530 °C in several temperature steps. Figure 5.19 shows that the free carrier concentration increases as the growth temperature is lowered. The authors argue that the increase is expected for amphoteric impurities. A high substrate temperature causes an increase in the As_4 desorption rate from the surface resulting in an increased V_{As} concentration. The increase in V_{As} concentration shifts the reaction equilibrium of the chemical reaction Eq. (5.13) to the right-hand side of the equation, which increases the concentration of Si acceptors, Si_{As}. As a result, the carrier concentration decreases at high substrate temperatures in agreement with the experimental result shown in Fig. 5.19.

Another explanation of the decrease in free carrier concentration at high growth temperatures is based on the different formation energies of Si_{Ga} and Si_{As}. Obviously Si_{Ga} has a larger formation energy than Si_{As} since Si is predominantly a donor. Denoting the difference in formation energy as ΔE ($\Delta E > 0$) the ratio of Si acceptor to Si donor concentration is given by

$$[Si_{As}]/[Si_{Ga}] \propto \exp(-\Delta E/kT) \tag{5.14}$$

where ΔE is the difference in formation energy of Si_{Ga} and Si_{As} (and the entropy term is neglected). At high temperatures the exponential term in Eq. (5.14) approaches unity, i.e. the Si acceptor concentration Si_{As} increases which leads to a decrease in the free carrier concentration. For a more detailed discussion of compensation the reader is referred to Chap. 7. Chai et al. pointed out that the decrease of the free carrier concentration (see Fig. 5.19) at a substrate temperature of 530 °C is inconsistent with the expected further increase. They attributed the decrease to a lower crystalline quality obtained at this low growth temperature and to enhanced C incorporation. C acceptors tend to compensate the Si donors which agrees with the experimental finding. Finally, the decrease of the free carrier concentration at high growth temperatures can be due to reevaporation of Si from the growing surface. However, Si has a relatively small vapor pressure and reevaporation of Si is therefore a minor effect.

Germanium, Ge

Summary: Ge is a strongly amphoteric impurity in III–V semiconductors. Both p- and n-type conductivity can be obtained depending on substrate temperature and V/III flux ratio during MBE growth. AuGe alloys are used for ohmic contact metallization to n-type semiconductors.

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

Germanium incorporates on either the cation site or anion site of a III–V semiconductor. It is an acceptor and a donor on the anion and cation site, respectively. The Ge acceptor ionization energy in GaAs was determined to be 40 meV by low-temperature photoluminescence measurements (Ashen et al., 1975). Ge is a shallow hydrogenic donor on the cation site and has an ionization energy of 4–6 meV in GaAs and InP.

Germanium is a strongly amphoteric impurity in GaAs. It was shown that MBE-grown GaAs doped with Ga can be either p-type or n-type (Cho and Hayashi, 1971). The authors further showed that the incorporation of Ge strongly depends on the surface chemistry, i.e. on the (i) V/III flux ratio and (ii) the growth temperatures. Under As-rich growth conditions, Ge incorporates as a donor; correspondingly, under Ga-stabilized growth conditions, i.e. a V/III flux ratio close to unity, Ge incorporates as an acceptor.

Similarly, the variation of the growth temperature allows one to change the conductivity type of Ge-doped GaAs. At growth temperatures below the congruent sublimation temperature (which is 630 °C for GaAs), the epitaxial surface is As stabilized as long as V/III > 1. As a result, Ge incorporates as a donor. However, for growth temperatures that exceed the congruent evaporation temperature, Ga-stabilized surfaces are obtained. As a result, Ge incorporates as an acceptor on anion sites. Cho and Hayashi (1971) demonstrated that pn junctions can be grown using Ge as the single impurity species. The conversion from p- to n-type was achieved by an increase of the V/III flux ratio.

Künzel et al. (1982) investigated the effect of the two arsenic vapor species As_2 and As_4 as well as of the substrate temperature on the incorporation of Ge in GaAs. The experimental result of this study is shown in Fig. 5.20 which depicts the compensation $[Ge_{As}]/[Ge_{Ga}]$ as a function of reciprocal substrate temperatures for the two species As_2 and As_4. Compensation increases strongly at higher growth temperatures. The high growth temperatures result in increased As reevaporation from the epitaxial surface and thus a reduced As coverage of the surface. As a result, Ge increasingly occupies As sites. The authors further showed different compensation for As_4 and As_2 evaporation. As_2 species were generated by thermally cracking As_4. Künzel et al. (1982) argued that crystals grown with the more reactive As_2 have higher As surface coverage and therefore less compensation. It is interesting to note that the two solid lines intersect unity compensation, i.e. $[Ge_{As}]/[Ge_{As}] = 1$, at a temperature of 600–620 °C which is slightly below the congruent sublimation temperature for GaAs of 630 °C. The compensation of Ge and its dependence on the growth conditions preclude the use of Ge as a donor or acceptor for MBE-grown III–V semiconductor devices.

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

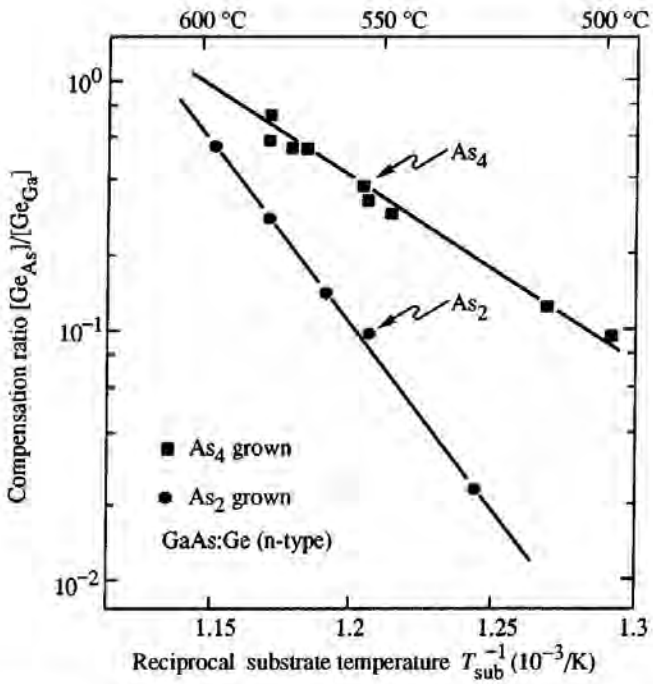


Fig. 5.20. Compensation of MBE-grown n-type Ge-doped GaAs as a function of substrate temperature. The use of As_2 dimers results in a reduction of auto-compensation compared to growth with As_4 tetramers (after Künzel et al., 1982).

Eutectic AuGe (88% Au, 12% Ge) is frequently used in ohmic contact metallization for III–V semiconductors. Alloying cycles of 420 °C for 20 s are sufficient to diffuse Ge into GaAs where it acts as donor. The chemistry and technology of alloyed AuGe ohmic contacts was reviewed by Braslau (1981). The author showed that in addition to Ge diffusion, the formation of domains with different chemical composition occurs in AuGe contacts.

Tin, Sn

Summary: Tin is a weakly compensating shallow donor in III–V semiconductors. Sn strongly segregates and diffuses at growth temperatures exceeding 500 °C.

Tin occupies cation sites in III–V semiconductors and is a hydrogenic shallow donor with 4–6 meV ionization energy in GaAs and InP. However, Sn was reported to incorporate as an acceptor in GaSb (Chang et al., 1977). Even though there are no reports on Sn acceptor incorporation in GaAs it is quite likely that a small fraction of Sn does indeed incorporate as acceptor in all III–V semiconductors.

The major drawback of Sn impurities in GaAs is the tendency to strongly

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

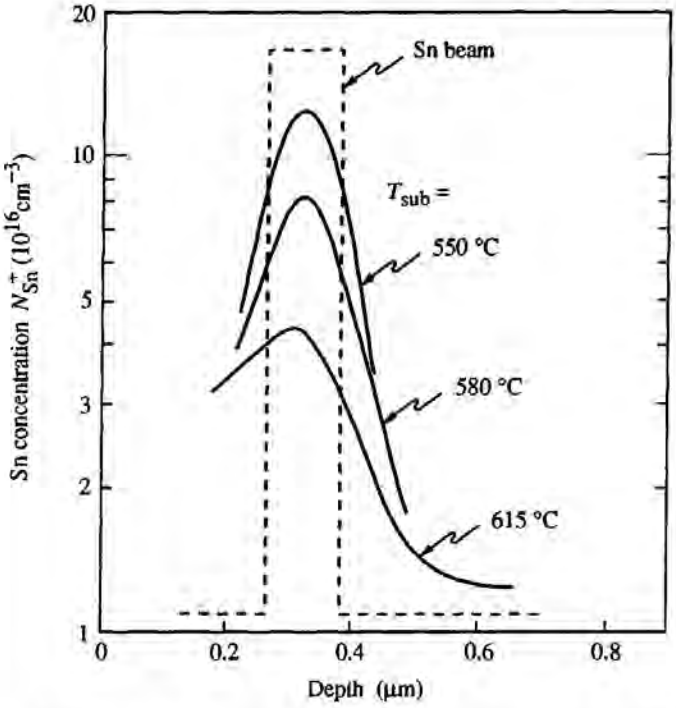


Fig. 5.21. Measured doping profiles of Sn-doped GaAs grown by MBE at different substrate temperatures. The dashed line is the ideal doping profile (without any dopant redistribution) calculated from the Sn beam intensity (after Cho, 1975).

redistribute during growth. Diffusion as well as surface segregation contribute to the redistribution, which was first discovered by Cho (1975). Cho analyzed Sn doping profiles in GaAs grown by MBE at different substrate temperatures. The measured Sn profiles are shown in Fig. 5.21 along with the calculated Sn flux profile indicated by the dashed line. The calculated Sn profile was obtained from the Sn beam intensity. It was assumed that all Sn impurity atoms impinging on the growing surface would incorporate into the crystal, i.e. a unity sticking coefficient was assumed.

The experimental result of Fig. 5.21 shows that Sn redistribution occurs for MBE growth temperatures of 550–615 °C. The impurity profile obtained from the 550 °C growth exhibits symmetric broadening towards the substrate side as well as towards the epitaxial surface. The symmetric broadening indicates that diffusion is the dominant redistribution mechanism. The impurity profile obtained from the 615 °C growth is clearly skewed towards the surface, i.e. the slope of the trailing edge of the profile is much steeper than the slope of the

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

leading edge indicating surface segregation of Sn impurities. Further experiments of Cho (1975) with multiple Sn doping pulses demonstrated that Sn surface segregation (rather than diffusion) is the dominant redistribution mechanism at high growth temperatures.

Ploog and Fischer (1978) used Auger electron spectroscopy (AES) and secondary ion mass spectrometry (SIMS) to study Sn surface segregation. The study revealed that a significant amount of Sn, about three orders of magnitude larger than the portion being incorporated into the growing film, was detected on the surface and the outermost atomic layers (<20 Å) of the epitaxial film. The segregation of Sn occurred at substrate temperatures as low as 490 °C. Ploog and Fischer (1978) also found that the segregation increases with Sn concentration.

Even though the strong redistribution rules out Sn as a useful donor for epitaxially grown III–V semiconductors devices, Sn doping of GaAs provides insight into and understanding of impurity surface segregation in semiconductors. Wood and Joyce (1978) proposed a surface rate limiting process in which the Sn is incorporated from a surface layer, not directly from the molecular beam. Harris et al. (1984) modeled the Sn surface segregation quantitatively assuming a lower energy of Sn atoms on the surface as compared to the bulk. At high growth temperatures, the Sn atoms segregate to the low-energy surface. Thermal equilibrium distribution of Sn is achieved for temperatures exceeding 600 °C. Growth at lower temperatures imposes a kinetic limitation on the Sn movement, and as a consequence, Sn atoms are trapped in the bulk. A further discussion of surface segregation of impurities in semiconductors is found in Chap. 8.

Lead, Pb

Lead has a large covalent atomic radius and is not expected to be very soluble in III–V semiconductors. There is only one report of Pb doping in GaAs (Wood, 1985a and 1985b). Lead doping of GaAs did not result in an increase of the free carrier concentration.

5.3.4 Group-VI dopants

The group-VI impurities include O, S, Se, Te, and Po and are called the chalcogens. Polonium has a comparatively large atomic radius and is therefore not expected to incorporate easily in typical III–V semiconductors. Oxygen on the other hand forms deep donor levels in GaAs. Incorporation of oxygen into any III–V semiconductors should therefore be avoided. The remaining useful group-VI impurities are S, Se, and Te. They occupy anion sites in III–V semiconductors and are shallow donors. Group-VI donors are not amphoteric. However, group-VI impurities can be electrically inactive in III–V semiconductors. All group-VI impurities have comparatively high vapor

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

337-TA-1366 RX-0255

pressures. The vapor pressure decreases with increasing atomic number. That is, S has the highest and Te the lowest vapor pressure of the three elements S, Se, and Te. As a consequence of the high vapor pressure, group-VI impurities tend to reevaporate from the epitaxial surface. Furthermore, the incorporation of chalcogens depends strongly on the growth temperature.

Oxygen, O

Oxygen forms deep donor levels in III–V semiconductors. In GaAs two donor levels 400 meV and 670 meV below the conduction band edge were reported (see Chap. 9). Oxygen has a strong affinity to Al-containing compounds such as $Al_xGa_{1-x}As$ and $Al_xIn_{1-x}As$. It is desirable to avoid any oxygen incorporation.

Sulfur, S

Summary: Sulfur is a shallow donor in III–V semiconductors. The impurity has a high vapor pressure and the incorporation depends strongly on the epitaxial growth temperature.

Sulfur occupies anion sites in III–V semiconductors and is a shallow donor with 6 meV ionization energy in GaAs. It is used frequently for III–V semiconductors in which Si is a strongly amphoteric donor, e.g. in GaSb. Poole et al. (1988) studied S doping of GaSb and demonstrated that incorporated S in epitaxial GaSb has near 100% electrical activity. Free carrier concentrations exceeding 10^{18} cm^{-3} were achieved.

Sulfur has a high vapor pressure even at relatively low temperatures. This characteristic imposes serious restrictions on the maximum bake-out temperatures of MBE systems. Typical bake-out temperatures are in the range of 150 °C to 250 °C; at such temperatures S evaporates and would deplete an elemental S effusion cell during bake-out. In order to avoid this depletion, Davies et al. (1981) developed an electrochemical Ag_2S cell located within a conventional Knudsen cell. The solid-state electrochemical cell contained Pt/Ag/AgI/Ag_2S/Pt where the two Pt metallization layers serve as electrodes. Upon application of a bias to the electrochemical cell, elemental sulfur is generated which is then evaporated in the heated effusion cell. This method of S generation is compatible with the ultra-high vacuum of an MBE system; the cell is bakeable to approximately 350 °C without depletion of S from the cell. Davies et al. (1981) demonstrated low (5×10^{15} cm^{-3}) as well as high (2×10^{18} cm^{-3}) fully active S donor concentrations in GaAs.

Other compounds used for S doping of GaAs during MBE include lead sulfide, PbS. Wood (1978) showed that, upon evaporation of PbS from an effusion cell, S incorporates into the GaAs epitaxial layer while Pb does not,

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

Innoscience1366-00138899

RX-0255.226

probably due to its large atomic radius. Doping levels ranging from 10^{16} cm^{-3} to 5×10^{18} cm^{-3} were demonstrated with the PbS cell. The major drawback of the PbS cell is its high vapor pressure which can result in a depletion of the effusion cell even during bake-out at 200 °C.

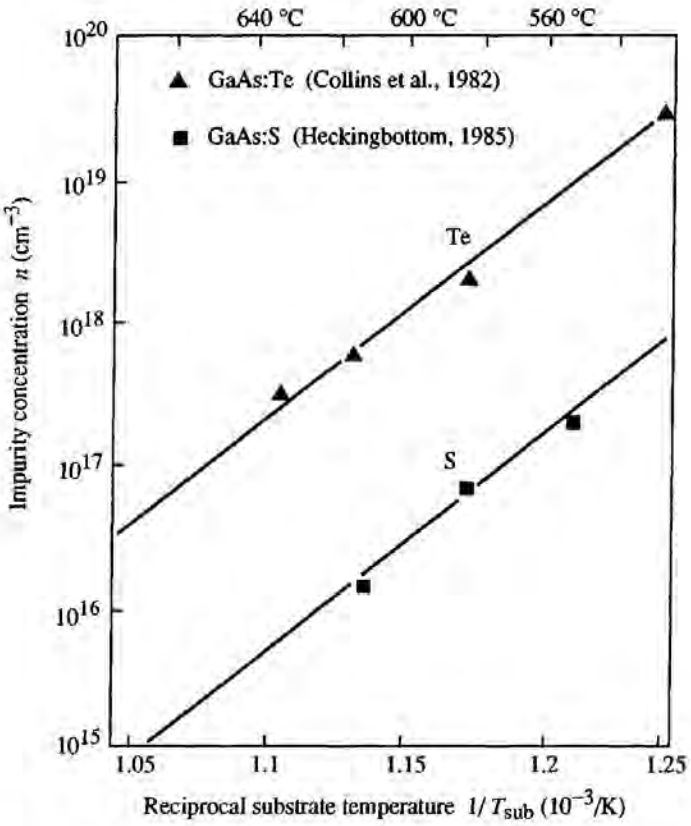


Fig. 5.22. Free electron concentration of S- and Te-doped GaAs grown by MBE as a function of reciprocal substrate temperature. The sulfur and tellurium doping impurities are obtained from S_2 and SnTe beams, respectively (after Collins et al., 1982 and Heckingbottom, 1985).

The high vapor pressure of S makes the incorporation of S strongly dependent upon the epitaxial growth temperature. Heckingbottom (1985) showed that the concentration of incorporated S in GaAs varies by one order of magnitude in a growth temperature interval of approximately 50 °C. The experimental carrier concentration of S-doped GaAs is shown in Fig. 5.22 as a function of substrate temperature. Heckingbottom attributed the reduction of S doping concentration at elevated substrate temperatures to the reevaporation of S from the growing surface. He also suggested that gallium sulfide is formed at higher growth

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

temperatures according to $2Ga + \frac{1}{2}S_2 \leftrightarrow Ga_2S$ where Ga_2S is a gaseous volatile compound.

The high S vapor pressure and the strong dependence of the S doping concentration on the substrate temperature make the accurate control of S doping difficult. These characteristics make S clearly inferior to Si in GaAs. For other materials, such as GaSb in which Si is strongly amphoteric, S seems to be the better choice.

Selenium, Se

Summary: Selenium is a high vapor pressure shallow donor in III–V semiconductors. Se has high electrical activity with no tendency to compensate.

Selenium occupies anion lattice sites in III–V semiconductors and forms shallow, hydrogenic donors. It has an ionization energy of 6 meV below the conduction band edge in GaAs. Selenium and other chalcogens have a high vapor pressure which makes them difficult to use in an ultra-high vacuum environment. Effusion cells with elemental Se baked at temperatures of 250 °C would release considerable amounts of Se. Therefore elemental Se effusion cells are not used in MBE growth systems. Selenides such as lead selenide (PbSe) have much lower vapor pressures as compared to elemental Se (Hirama, 1964). Wood (1978) used PbSe for n-type doping in GaAs using a conventional effusion cell. The author showed that only Se but not Pb is incorporated during MBE growth and that the Pb does not interfere with the electrical and structural properties of GaAs grown by MBE at substrate temperatures of 480 °C. Even the coverage of GaAs with multiple atomic layers of lead did not result in incorporation of Pb. The chemical incorporation reaction proposed is given by $PbSe + V_{As} \leftrightarrow Pb\uparrow + Se_{As}$ where V_{As} and $Pb\uparrow$ are an As vacancy and a volatile lead atom, respectively.

The electrical characteristics of Se-doped GaAs revealed no evidence of electrical neutrality or autocompensation of the Se impurities (Wood, 1978). Instead, the Se impurities were assumed to be highly efficient. Electrically active Se doping levels as high as 3×10^{18} cm^{-3} were demonstrated. The growth of multiple doping pulses did not reveal evidence of excessive impurity redistribution during growth at 490 °C. However, an increase of the growth temperature to 560 °C resulted in drastically decreased free carrier concentrations. The decrease in free carrier concentration was attributed to reevaporation of PbSe from the GaAs surface.

It appears that the major drawback of Se as a doping impurity is its high vapor pressure which makes its handling in a ultra-high vacuum environment difficult. Furthermore, the strong dependence of the free carrier concentration on the

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

substrate temperature makes control of the growth parameter complicated. Se doping is therefore used only if Si doping is not feasible. In compounds such as GaAsSb or GaSb, Si is strongly amphoteric. In these compounds, Se doping (along with other chalcogens) is a viable alternative to Si doping.

Tellurium, Te

Summary: Tellurium is a non–hydrogenic, shallow donor in III-V semiconductors. Te has both high electrical activity and a high vapor pressure and its incorporation depends strongly on growth conditions especially the growth temperature.

Tellurium occupies anion sites in III–V semiconductors with an ionization energy of 30 meV in GaAs. The ionization energy is much larger than expected for hydrogen-like donors in GaAs which have typical ionization energies of 5 meV. The larger ionization energy is probably due to a central cell correction of the big Te atom in the GaAs lattice.

Tellurium has not been evaporated from effusion cells containing elemental Te. Elemental Te is avoided in the ultra-high vacuum environment of an MBE system due to its high vapor pressure. Instead, Te is evaporated from compounds that have a lower vapor pressure such as SnTe (Collins et al., 1982), PbTe (Subbanna et al., 1988) and Sb_2Te_3 (Chiu et al., 1990). Tellurium doping was done predominantly in Sb containing III–V compounds such as GaSb and $GaAs_{1-x}Sb_x$ due to the strong amphoteric nature of Si in these compounds (Chang et al., 1977; Yano et al., 1978).

At doping concentrations $N_{Te} << 10^{18}$ cm^{-3} Te is highly active. Collins et al. (1982) and Chiu et al. (1990) showed that the free carrier concentration follows the Te concentration for $N_{Te} << 10^{18}$ cm^{-3}. However at a high doping concentration of $N_{Te} \geq 10^{18}$ cm^{-3} the electrical activity of Te is reduced. At a Te concentration of 5×10^{19} cm^{-3} in GaAs the free carrier concentration was only 4×10^{18} cm^{-3} indicating that less than 10% of the Te atoms were electrically active.

The high vapor pressure of Te makes the incorporation of Te strongly dependent on the growth conditions. The concentration of Te in GaAs is shown in Fig. 5.22 as a function of the growth temperature (Collins et al., 1982). High Te concentrations exceeding 10^{19} cm^{-3} are achieved at the low growth temperature of 530 °C. As the growth temperature is increased to 630 °C the Te concentration drops by two orders of magnitude. A strong temperature dependence of the Te incorporation was also observed in GaSb by Chiu et al. (1990). However, the authors found that Te incorporation is insensitive to the

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

growth temperature for $T < 540$ °C. Efficient incorporation of Te occurred for temperatures < 540 °C.

Polonium, Po

Polonium is an impurity with a large atomic radius. It is not expected to be very soluble in III–V semiconductors.

https://doi.org/10.1017/CBO9780511599828.008 Published online by Cambridge University Press

6

Gaseous doping sources

Elemental impurities cannot be used as doping sources in vapor-phase epitaxial growth systems because such impurities would assume the ambient temperature in the reactor and condense on the reactor walls. Vapor-phase epitaxial (VPE) growth techniques therefore require gaseous doping compounds, i.e. chemical gases transporting the doping impurity. Such gaseous doping compounds are called ***doping precursors*** and are mandatory for growth techniques with a viscous gas flow such as VPE. Growth techniques with molecular flow can employ either elemental doping sources or doping precursors. Chemical-beam epitaxy (CBE) frequently employs gaseous doping compounds. The use of precursors in CBE is motivated by parasitic chemical reactions between the organometallic sources used for group-III elements and elemental doping sources which result in a degradation in elemental source purity (see Sect. 6.1). Thus the two techniques, VPE and CBE, use doping precursors for different reasons.

The use of non-elemental chemical precursors allows a wide selection of chemicals. The most obvious requirement for these chemicals are (i) a sufficient vapor pressure at room temperature (ii) thermal decomposition of the chemical at the growth temperature, and (iii) no parasitic chemical reactions before and after thermal decomposition. These requirements can be satisfied by either *inorganic* precursors, e.g. silane or disilane, or *organometallic* precursors, e.g. dimethylzinc (DMZn).

Dopant incorporation from precursors can be extremely complicated and is controlled by a multitude of parameters. Chemistry, transport by convection and diffusion, thermodynamics, and fluid dynamics govern the incorporation of doping impurities. Even though considerable progress has been made in the understanding of doping incorporation, Razeghi (1989) concluded that the exact

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

chemical decomposition pathways in OMVPE are not yet clearly understood.

It is useful to keep in mind the general chemical characteristics of the doping precursors containing alkyl groups. Alkyls are hydrocarbons with varying length of the carbon chain. The most frequently used alkyls are methane, ethane, and propane, i.e. CH_4, C_2H_6, and C_3H_8, respectively. The first general characteristic of alkyls is that their stability decreases with the length of the carbon chain. For example, ethane is more stable against thermally activated dissociation than propane. The second general characteristic is that the vapour pressure decreases with increasing length of the carbon chain. Methane is thus the alkyl with the highest vapor pressure. This characteristic also holds for organometallic compounds. For example, trimethylgallium (TMGa) has a higher vapor pressure than triethylgallium (TEGa). The same is valid for other organometallic compounds such as TMAl/TEAl or TMIn/TEIn. Furthermore, the two characteristics mentioned above are valid not only for carbon hydrides (alkyls) but also for silicon hydrides. That is, disilane (Si_2H_6) has a lower vapor pressure than silane (SiH_4).

The vapor pressure of the group-II, -IV, and -VI doping elements also follows some general tendencies. Among the group-II elements, Be, Mg, Zn, and Cd are most important. Within this group the vapor pressure increases with atomic weight, i.e. the vapor pressure is lowest for Be and highest for Cd (see, for example, Shapira and Lichtman, 1979). Reevaporation of doping impurities from a growing semiconductor surface is strongest for Cd. The group-IV impurities C, Si, Ge, and Sn share the same trend for the vapor pressure. At a given temperature, Sn has the largest vapor pressure and C the smallest vapor pressure of the group-IV elements. Finally, the group-VI elements used for doping are S, Se, and Te. Among these elements, the vapor pressure at a given temperature decreases with atomic weight. Thus, S has the highest vapor pressure and has the highest rate of reevaporation from a growing crystal surface.

In this chapter the use of gaseous doping compounds for epitaxial III–V semiconductor growth will be discussed. After the analysis of the general requirements for gaseous precursors, the doping homogeneity of OMVPE will be assessed. Finally, the properties of different doping compounds for p- and n-type doping of III–V semiconductors will be summarized.

6.1 General considerations

Doping impurities in vapor-phase epitaxially (VPE) grown materials are introduced by means of gaseous chemical compounds, the doping ***precursors***. The use of chemical compounds rather than doping elements makes possible a wide selection of doping precursors. Among the most frequently used doping

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

precursors are the ***hydrides*** (e.g. SiH_4, H_2S) and ***organometallic*** compounds (e.g. diethylzinc, DEZn; tetramethyltin, TMSn).

There is a multitude of requirements for chemical doping precursors. Ideal doping precursors (i) do not react in the gas phase of the VPE system before reaching the substrate, (ii) fully decompose at or near the growing crystal surface, (iii) incorporate into the crystal with unity probability, (iv) are electrically active in the semiconductor, and (v) do not exhibit the memory effect. A sufficiently *high vapor pressure* is required at or near room temperature for all doping precursors. The typical mole fraction of the doping compound in the gas phase is $10^{-10} - 10^{-1}$. In order to achieve such mole fractions, a high vapor pressure is advantageous. However, the transport of the doping precursor can be enhanced, if necessary. The first possibility is to heat the precursor source above room temperature. Excessive heating ($>>$ 50 °C) should be avoided since condensation of the precursor may occur in the gas lines or at the reactor walls (in cold wall reactors) before reaching the epitaxial crystal surface. A second possibility to increase the flow of the doping precursor is the use of a carrier gas bubbler (see Sect. 4.3).

It is desirable that dopants not exhibit any ***memory effect*** (Sect. 4.4) which complicates the growth of abrupt doping profiles. A strong memory effect has been observed for the doping precursors hydrogen selenide (H_2Se) and diethyltellurium (DETe) by Lewis et al. (1984) and Houng and Low (1986), respectively. The authors found an unexpected high background concentration after doping GaAs with DETe. Houng and Low (1986) found that adsorption of DETe onto the cold stainless steel tube walls gives rise to the doping memory effect. The authors showed that a reduction of the doping line surface area and heating of the doping lines to 60 °C reduces the memory effect. Previously, Lewis et al. (1984) demonstrated that H_2Se exhibits memory effects in OMVPE-grown GaAs. As a consequence, graded doping profiles were obtained. The memory effect was attributed to adsorption of H_2Se onto the internal walls of the system and subsequent desorption after the doping source was shut off.

Complete ***thermal decomposition*** is a desirable property of doping precursors which should occur at all growth temperatures, i.e. in the range 600–800 °C for the growth of GaAs. To illustrate the requirement of pyrolysis, consider the following examples. First, methane (CH_4) could be considered as a carbon doping source. However, methane is a very stable molecule. It is not thermally decomposed in the entire temperature range of interest. Therefore, methane cannot be used as a doping precursor in OMVPE. Second, silane (SiH_4) is a fairly stable molecule. Only a fraction of SiH_4 molecules are decomposed at typical growth temperatures of 600–800 °C (Kuech et al., 1984). Since the

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

337-TA-1366 RX-0255

thermal decomposition efficiency depends exponentially on temperature, the incorporation of Si from a SiH_4 source is strongly growth temperature dependent. Such a strong temperature dependence is undesirable, since it does not allow the two parameters, i.e. doping and growth temperature, to be adjusted independently. Third, disilane (Si_2H_6) is a less stable molecule which decomposes into Si, H_2, and SiH_4 at the temperature range of interest ≤ 600 °C. The incorporation of Si from a disilane source is therefore approximately independent of growth temperature (Kuech et al., 1984). The above examples illustrate that complete pyrolysis is a desirable property for dopants. Other decomposition paths such as chemical reactions with other precursors or catalytic reaction are undesirable since they again lead to the interdependence of growth parameters. Such an interdependence should be avoided as much as possible to keep the growth and doping process well controlled.

Non-toxicity and non-flammability are desirable but not necessary conditions for doping precursors in OMVPE. In fact, many precursors are used despite their toxicity and flammability. Appropriate precautions, which are recommended by the manufacturers, are required when handling such doping precursors.

Doping precursors are required to not react in the gas phase of the growth system but only at the semiconductor surface. Premature, parasitic gas-phase reactions can alter the reaction pathways of doping precursors and complicate or even make impossible the detailed quantitative understanding of the doping process. It is finally desirable that doping incorporation is independent of reactor pressure and V/III flux ratio.

Doping precursors have also been used in chemical-beam epitaxy (CBE) which uses, as OMVPE, organometallics and hydrides for group-III and group-V elements, respectively. Elemental dopants were widely used in the 1980s for CBE growth (see e.g. Tsang, 1989). However, Skevington et al. (1990) found anomalous elemented source Si and Sn doping behaviour in InP grown by CBE. The authors used elemental source effusion cells for doping and found that the doping level decreases from growth run to growth run. The experimental result is depicted in Fig. 6.1 which shows the doping concentration of four consecutive growth runs of Si-doped InP. The effusion cell temperature, i.e. the nominal doping concentration during the four growth runs remained constant. Nevertheless, a clear decrease of the measured doping concentration from 5×10^{16} cm^{-3} to 2.8×10^{16} cm^{-3} is observed. The authors pointed out that the long-term aging process of the effusion cell caused major reproducibility problems, with the cell temperature requiring frequent adjustment to maintain a given doping level. Eventually, the required cell temperature became unacceptably high. When the effusion cell was removed, the Si charge was found

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

RX-0255.234

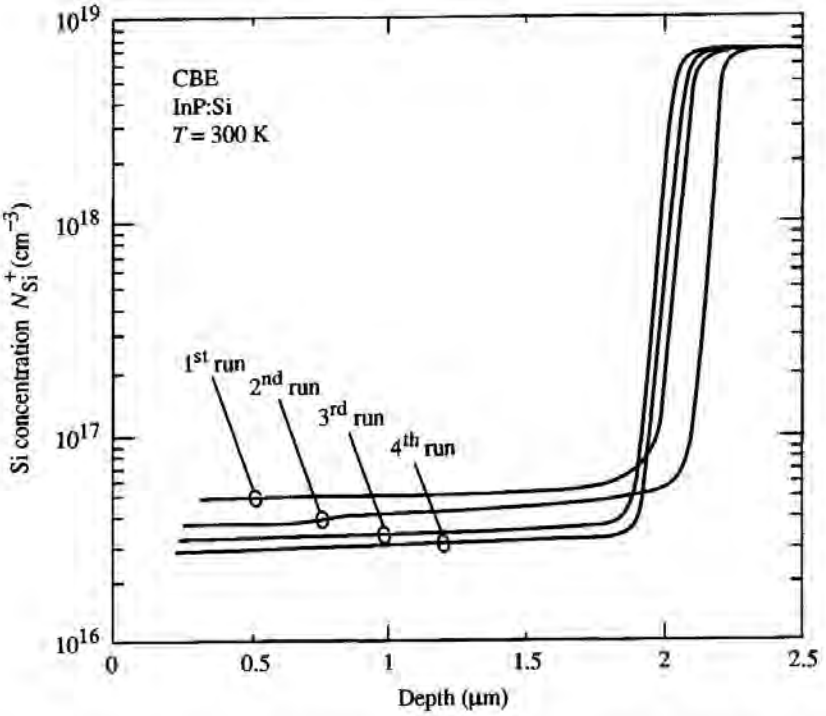


Fig. 6.1. Si doping profile in CBE-grown InP doped with an elemental Si effusion cell. The carrier concentration decreases monotonically in four consecutive growth runs (after Skevington et al., 1990).

to be covered by black and grey deposits. Skevington et al. (1990) also found reproducibility problems with Sn-doped InP. The Sn background concentrations began to vary, apparently unpredictably, up to an electron concentration in the 10^{18} cm^{-3} range. Even at low Sn effusion cell temperatures of 400 °C, the authors found a Sn background doping level in the 10^{17} cm^{-3} range. The Sn background doping level increased with the duration of the Sn effusion cell being heated to high temperatures.

The difficulties of elemental Si and Sn doping during CBE were attributed to chemical reactions between the organometallic group-III compounds with the charges of the heated doping effusion cells. Trimethylindium (TMIn) was used as the source material for cations. The authors proposed that silicon carbide (SiC) is formed as a result of the interaction of TMIn or alkyl radicals with the heated Si charge. Black carbon deposits were also found in the effusion cell. Furthermore, the authors proposed that an organometallic Sn compound was formed. Most likely is the formation of a methylated Sn compound (e.g. monomethyltin,

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

dimethyltin), which is more volatile than Sn and thus can be incorporated into the epitaxial layer. The formation of tetramethyltin is ruled out on the basis of its small incorporation probability. Thus, the study of Skevington et al. demonstrates the necessity of a non-elemental doping source for CBE.

It is likely that the organometallic compounds also react with p-type impurities. Berylliumcarbide (Be_2C) is probably formed as a result of a chemical reaction between group-III organometallics and a Be cell heated to operating temperatures. Organometallic compounds tend to coat the ceramic and metallic parts of elemental effusion cells. Chemical reactions and short-circuits can result.

6.2 Doping homogeneity and uniformity

The homogeneity of epitaxial layers grown by hydride VPE or organometallic VPE can be categorized into three groups which are (i) doping uniformity, (ii) thickness uniformity, and (iii) composition uniformity. These three types of uniformities are interdependent. For example, thickness non-uniformity across a wafer automatically translates into doping non-uniformity. The three-dimensional doping concentration is inversely proportional to the epitaxial layer thickness if the doping incorporation rate is perfectly uniform across the epitaxial layer surface. Thus, a non-uniform doping distribution can occur due to layer thickness variations despite perfectly uniform dopant incorporation.

Composition non-uniformity is known as a changing composition of a ternary or quaternary alloy semiconductor across an epitaxial layer. Composition uniformity should not be confused with the stoichiometry of an epitaxial layer. Composition non-uniformities in hydride VPE and organometallic VPE system are usually small. The constituent precursor compounds of a ternary or quaternary semiconductor alloy are premixed, i.e. mixed before entering the reaction chamber. The premixing process usually provides near perfect mixing and negligible inhomogeneity of precursors in the gas phase.

Composition inhomogeneity can result in doping inhomogeneities as well. Consider a doping impurity whose incorporation strongly depends on the alloy composition. Such a strong dependence is observed for carbon which is amphoteric in $Ga_xIn_{1-x}As$. Carbon occupies predominantly acceptor and donor sites in GaAs and InAs, respectively. Thus, a change in the alloy composition can result in doping inhomogeneity.

Finally, the doping non-uniformities can be a result of non-uniform doping incorporation. The physical origin for non-uniform doping incorporation can be threefold. First, the gas flow of the doping precursor can be non-uniform or even turbulent. Second, the doping gas can deplete along the gas flow direction. Such

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

a depletion results in a smaller doping concentration at the exit side on the wafer as compared to the gas entry side. Third, doping non-uniformities can be the result of a variation of the temperature at the growing surface. The incorporation of some doping precursors is strongly temperature dependent. For example, the Si doping concentration is known to be strongly temperature dependent if derived from silane. Thus, a temperature variation across the wafer can result in non-uniform doping distribution. Furthermore, if the temperature of the growing crystal changes with time, a change in doping concentration results. Temperature and, in general, process stability are therefore essential to achieve homogeneous doping.

Low-pressure VPE reactors improve doping, thickness, and compositional homogeneity of epitaxial crystals. Low-pressure VPE reactors operate in a pressure range < 1000 mbar (< 750 Torr), with typical values of approximately 10-100 mbar (7.5-75 Torr). The gas velocity is increased in low pressure reactors which leads to a more homogeneous and laminar gas flow. Nevertheless, depletion effects are not eliminated in low-pressure VPE reactors i.e. depletion effects persist as a cause for non-uniform doping.

Doping homogeneity and uniformity were studied on an Aixtron low-pressure OMVPE system with a horizontal reaction chamber by Zilko and Schubert (1991). The reaction chamber can accommodate two 5 cm or one 7.5 cm wafer. Two 5 cm wafers were used for the homogeneity study, one wafer being upstream (near the gas entry port of the reactor) and one downstream (near the exhaust of the reactor). The results obtained from the upstream wafer were found to have better uniformity. The following results are from the upstream wafer. The substrates were not rotated during growth. The growth temperature was 640 °C, the growth rate was 2.3 μm/hr, and the V/III ratio was 94. Silane (SiH_4) diluted to 1% in hydrogen was used as the Si-doping precursor. The growth experiment was conducted at a pressure of 20 mbar (15 Torr) which has proven favorable for uniformity purposes. The epitaxial GaAs layers have a nominal doping concentration of 1.5×10^{17} cm^{-3}. Capacitance–voltage measurements were carried out at 300 K on 500 μm diameter Ti/Au Schottky contacts to determine the doping concentration.

The result of the lateral uniformity on a 5 cm Si-doped GaAs epitaxial upstream wafer is shown in Fig. 6.2. One series of measurements was carried out along the direction of the gas flow. A second series of measurements was carried out normal to the gas flow direction. The average free electron concentration is 1.15×10^{17} cm^{-3}. The highest and lowest concentrations measured on the wafer are 1.37×10^{17} cm^{-3} and 1.02×10^{17} cm^{-3}, respectively. The standard deviation of the doping concentration is 10% and is calculated from 100

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

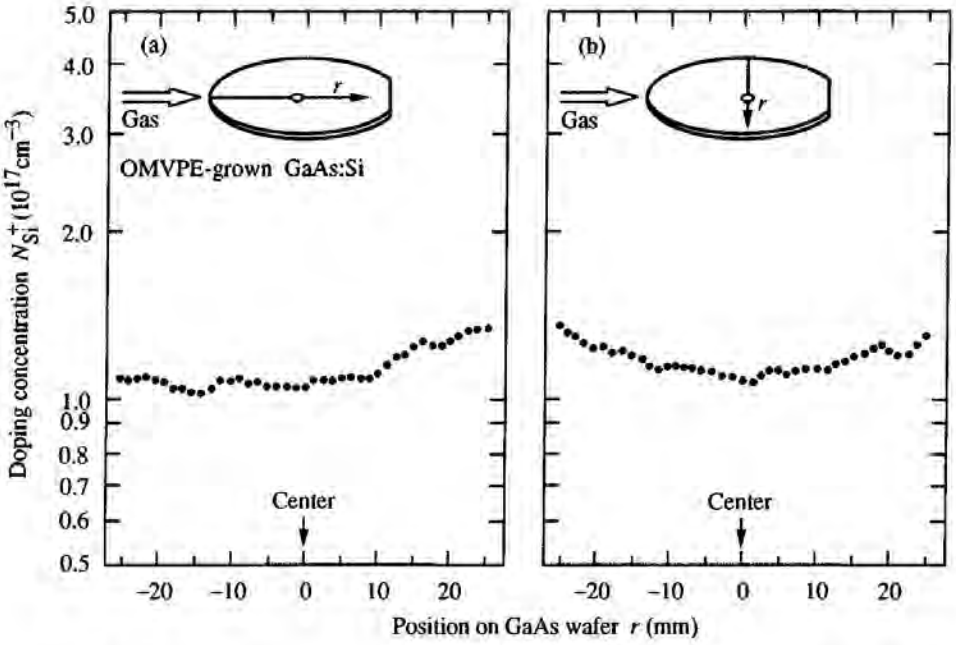


Fig. 6.2. Doping concentration of low-pressure OMVPE-grown n-type GaAs:Si along the diameter of a 5 cm wafer. The measurement points are (a) along the gas flow direction and (b) perpendicular to the gas flow direction.

measurement points.

The doping concentration measured along the gas flow direction has a clear upward trend near the exit side of the gas stream (see Fig. 6.2a). The increase is of a systematic nature which prevails from the center to the edge of the wafer. The increase is most likely due to depletion of the group-III precursor. The depletion of the group-III precursor results in a smaller growth rate near the downstream side of the wafer, since the group-III precursor determines the growth rate. Assuming a constant doping incorporation along the gas flow direction, the doping concentration increases due to the reduced thickness of the epitaxial layer. This interpretation is consistent with a thickness mapping of the epitaxial layer which shows a decrease of epitaxial layer thickness near the downstream side of the wafer. Note that a doping precursor depletion effect cannot explain the increase of the doping concentration. Doping precursor depletion is unlikely due to the small distribution coefficient of SiH_4, which is typically $\ll 10^{-2}$.

The doping concentration perpendicular to the gas stream is U-shaped. Higher doping concentrations are achieved at the edges than at the center of the

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

wafer. These higher concentrations could be the result of either a lower growth rate or a higher doping incorporation at the edges. A higher doping concentration results at higher temperature due to more efficient cracking of SiH_4. However, thickness mapping of the epitaxial layer revealed that the epitaxial layer is thinner at the two edges. The U-shaped concentration distribution shown in Fig. 6.2b is therefore most likely a result of a smaller growth rate at the epitaxial layer edges.

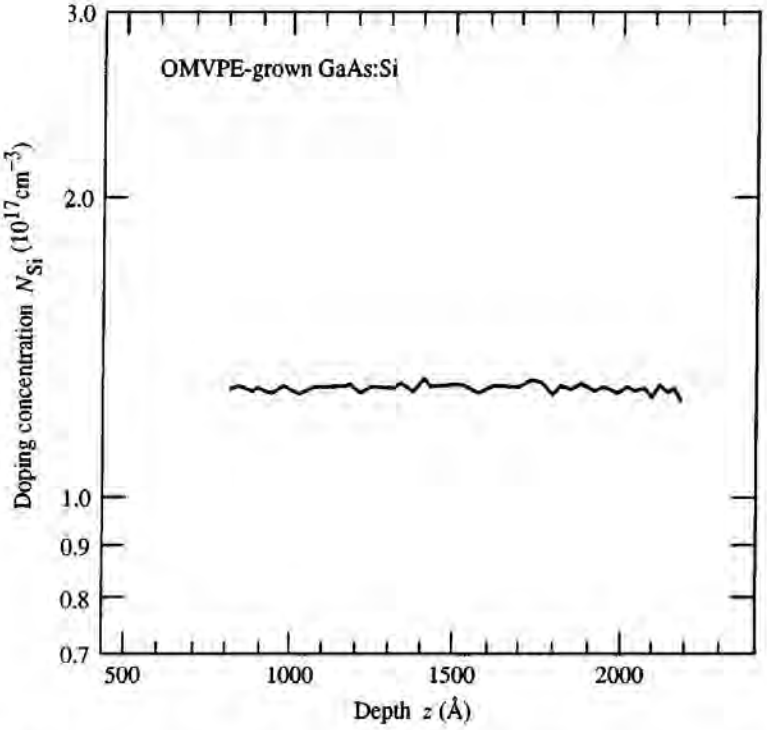


Fig. 6.3. Doping concentration of low-pressure OMVPE-grown n-type GaAs:Si versus depth.

The doping concentration of the OMVPE-grown GaAs is shown in Fig. 6.3 as a function of depth. Capacitance-voltage measurements were employed to measure the net active Si donor concentration ($N_{Si}^{+} - N_{Si}^{-}$) as a function of depth. Figure 6.3 reveals no systematic dependence of the doping concentration on depth. The average doping concentration is 1.29×10^{17} cm^{-3}. The variation of the doping concentration with depth is $\leq 1\%$ for all locations. The constant doping concentration with depth indicates excellent long-term stability and process control and no transient effects with long time constants.

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

Several reports in the literature indicate a doping uniformity similar to the results discussed above. Gersten et al. (1986) reported on (i) thickness, (ii) doping, and (iii) compositional uniformity of a large-scale OMVPE multiwafer "chimney" reactor with 90 wafers (4.6 cm × 7.6 cm). The authors found a doping uniformity of ±10% which was equal to the thickness uniformity. An Al-compositional uniformity in $Al_xGa_{1-x}As$ of ±3% was achieved. The authors emphasize the importance of accurate growth temperature control which improved the doping and thickness uniformity.

Hayafuji et al. (1986) reported on highly uniform growth of GaAs and $Al_xGa_{1-x}As$ in a multiwafer OMVPE reactor (Cambridge Instruments MR-200). They demonstrated a variation of growth rate ≤ 5% along the gas flow direction for two 5 cm diameter wafers under optimized carrier gas (H_2) flow conditions for GaAs and $Al_xGa_{1-x}As$ growth. The variation among twenty 5 cm diameter wafers was ≤ 8%. The variation of Al mole fraction in $Al_xGa_{1-x}As$ was ≤ 10% along the gas flow direction and among ten 5 cm diameter wafers.

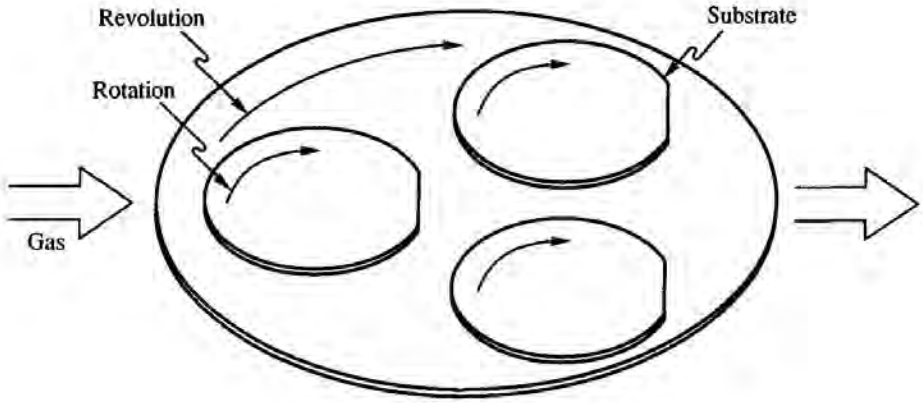


Fig. 6.4. Planetary motion including rotation and revolution of three substrates in a multiwafer OMVPE reactor. Planetary motion provides improved doping, thickness, and composition uniformity.

Excellent uniformity results were reported by Tanaka et al. (1987) using a multiwafer OMVPE system with 'planetary motion' of the wafers. The growth system was of the horizontal reactor type and can accommodate three 5 cm substrates. The planetary motion of the three substrates is illustrated in Fig. 6.4 as used by Tanaka et al. (1987). Each of the three substrates rotates around its own symmetry axis. Furthermore, a disk-shaped substrate holder rotates as well, subjecting the three wafers to a revolutionary motion. The disk-shaped substrate

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

holder in turn is mounted on the graphite susceptor. The growth of Si-doped $Al_xGa_{1-x}As$ was achieved by means of the precursors TMGa and TMAl for the cations, AsH_3 for the anions, and Si_2H_6 for Si doping. The $Al_{0.28}Ga_{0.72}As$ growth rate was 3.5 Å/s and the V/III ratio was 70.

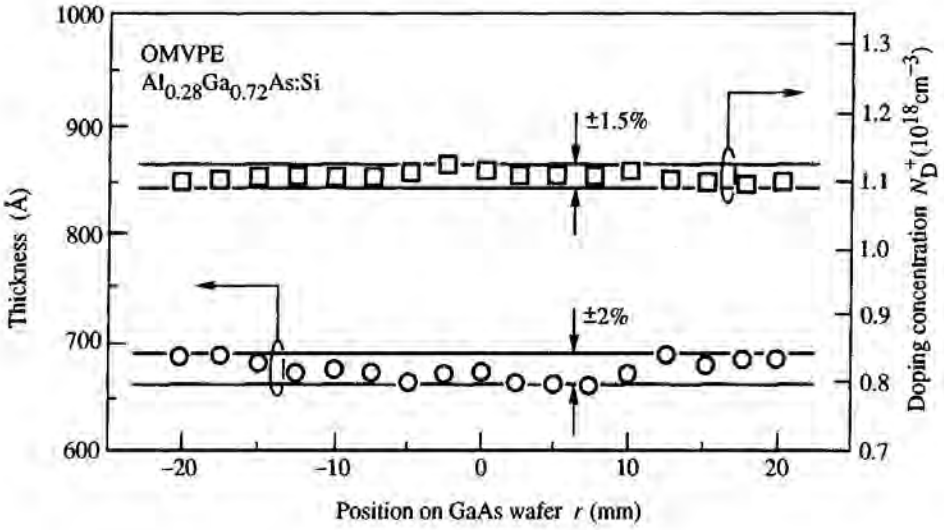


Fig. 6.5. Variation of epitaxial layer thickness and doping concentration of an $Al_{0.28}Ga_{0.72}As$ film grown by OMVPE. The 5 cm substrate was subjected to rotary and revolutionary motion during growth (after Tanaka et al., 1987).

The uniformity of thickness and doping concentration are shown in Fig. 6.5 as a function of position on one of the 5 cm substrates (Tanaka et al., 1987). Both the epitaxial layer thickness and the doping concentration were measured by capacitance–voltage measurements. The uniformity of the thickness and the carrier concentration of the n-type $Al_{0.28}Ga_{0.72}As$ film is better than ±2.0% and ±1.5%, respectively. The uniformity data is thus comparable to wafers grown by molecular-beam epitaxy. The variation of the thickness and the doping concentration among the three epitaxial layers grown simultaneously in the reactor is ±1.0%. The results demonstrate that impressive uniformities can be achieved in low-pressure OMVPE systems with planetary rotation and revolution of the substrates.

A theoretical model of growth rates as a function of the position on a wafer grown by OMVPE was developed by van de Ven et al. (1986) and Giling and van de Ven (1987). The authors considered a horizontal OMVPE reactor with a rectangular cross section and a laminar gas flow. They found that the growth rate

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

decreases along the gas flow axis. The rate of decrease was well described by an $x^{1/3}$ power law in the laminar gas flow domain where x is the position on the wafer along the gas flow direction. Depletion of the gas phase was identified as the reason for the reduced growth rate. Note that if the depletion of the gas phase were indeed the only reason for non-uniformities, planetary motion of the substrate could greatly reduce non-uniformities of doping, thickness, and composition.

6.3 Gaseous doping precursors

This section covers gaseous doping precursors used during epitaxial growth of III–V semiconductors. The doping precursors are classified according to their column in the periodic table, i.e. Sect. 6.3.1 Group-IIa dopants, Sect. 6.3.2 Group-IIb dopants, Sect. 6.3.3 Group-IV dopants, and Sect. 6.3.4 Group-VI dopants. These sub-sections emphasize the chemical characteristics of the doping precursors rather than the property of the resulting doping impurities in the semiconductor which are covered in Chap. 5.

Gaseous doping precursors are used during epitaxial growth with organometallic and hydride vapor-phase epitaxy (OMVPE and VPE) and chemical-beam epitaxy (CBE). In principle, both growth techniques can use the same doping precursors. However, the CBE growth temperature is typically 100–200 °C lower than the VPE growth temperatures. The pyrolysis of gaseous doping precursors depends exponentially on the substrate temperature. The dopant incorporation is therefore less efficient in CBE owing to the lower growth temperatures. In other words, thermally less stable doping precursors are desirable for CBE growth as compared to VPE growth. Finally, for a general discussion of impurity characteristics (saturation, autocompensation, etc.) the reader is referred to Chaps. 7 and 8.

6.3.1 Group-IIa precursors

Among the group-IIa elements, Be and Mg precursors are most popular.

Be precursors

Diethylberyllium, (DEBe, C_2H_5Be) has been used as a Be precursor for OMVPE (Bottka et al., 1984) as well as for CBE (Weyers et al., 1990). DEBe is a well-behaved precursor with desirable incorporation characteristics. The Be incorporation into GaAs using DEBe is insensitive to growth temperature and V/III flux ratio (Parsons et al., 1986). The characteristics indicate that DEBe decomposes at growth temperatures, does not reevaporate from the growing crystal surface, and fully incorporates into the growing layer. Doping concentrations ranging between 10^{15} cm^{-3} and 10^{20} cm^{-3} were achieved with

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

good surface morphology. The hole concentration in OMVPE-grown $Al_{0.12}Ga_{0.88}As$ is shown in Fig. 6.6 as a function of the DEBe mole fraction in the gas phase (Bottka et al., 1984). The V/III flux ratio during growth was six. DEBe was found to decompose readily and Be was concluded to occupy acceptor sites. The authors confirmed the efficient decomposition and incorporation of DEBe by a comparison with silane which required a ten times larger mole fraction in the gas phase to yield the same carrier density. For a DEBe mole concentration of 5×10^{-8}, the authors measured a net hole concentration in the mid 10^{18} cm^{-3} range, regardless of the V/III ratio used during the growth. The authors concluded that no competing chemical parasitic reactions occur simultaneously, which would hinder the incorporation of Be into the lattice.

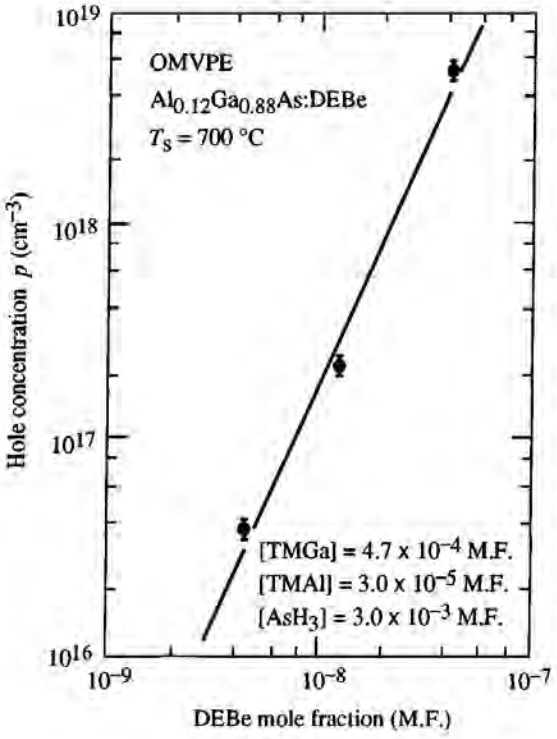


Fig. 6.6. Hole concentration in OMVPE-grown $Al_{0.12}Ga_{0.88}As$ as a function of DEBe mole fraction in the gas phase. A V/III flux ratio of six was used during growth (after Bottka et al., 1984).

The functional dependence of the hole concentration versus the DEBe mole fraction is shown in Fig. 6.6. The hole concentration depends superlinearly on the DEBe mole fraction. An increase of the DEBe mole fraction by a factor of 10 results in an increase of the free hole concentration by a factor of 100. Such a superlinear increase of the Be incorporation from DEBe was also found by Parsons et al. (1986). An increase of the hole concentration by a factor of 30 was found for a tenfold increase of the DEBe flow rate controlled by a H_2 bubbler.

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

The authors suggest that impurity compensation at low concentrations could explain the superlinear dependence. Dopant source gas manifolding was proposed as an alternative explanation.

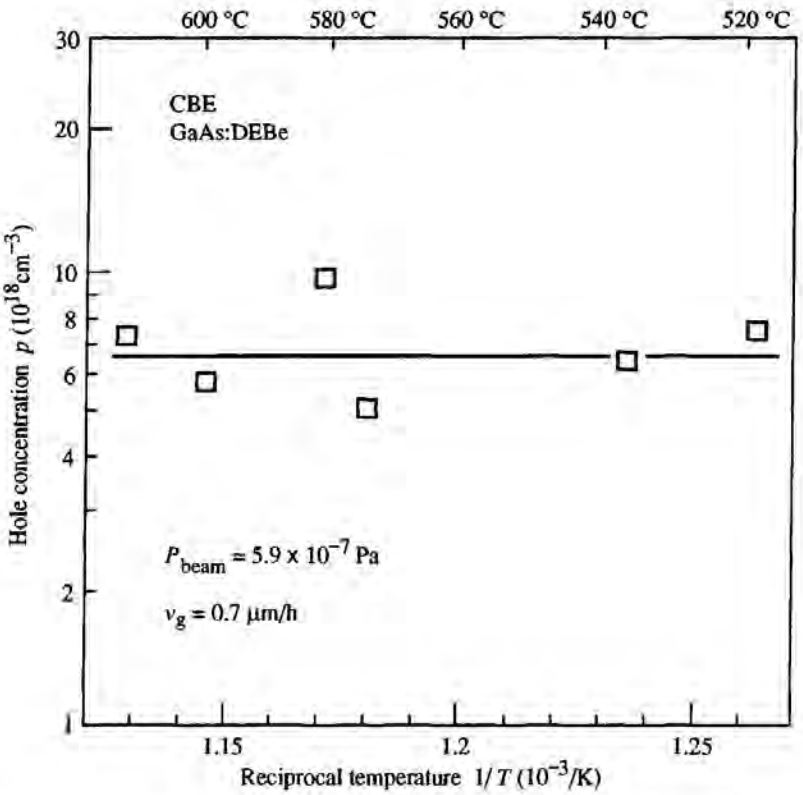


Fig. 6.7. Hole concentration (T = 300 K) of p-type GaAs grown by chemical-beam epitaxy using DEBe as a doping precursor as a function of reciprocal growth temperature (after Weyers et al., 1990).

The free hole concentration of Be-doped GaAs is shown in Fig. 6.7 as a function of the reciprocal growth temperature. The epitaxial GaAs was grown by chemical-beam epitaxy (Weyers et al., 1990) at a growth rate of 0.7 µm/hr using DEBe as the p-type dopant precursor. The hole concentration exhibits the desirable property of being independent of the growth temperature which makes DEBe a suitable precursor for CBE. The authors concluded that the doping efficiency of DEBe is close to unity, that DEBe is efficiently pyrolyzed on the surface, and that Be is subsequently incorporated into the lattice. The experiments were all performed at relatively low temperatures (520–620 °C) at which desorption of Be has not been observed. The authors concluded that these

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

characteristics make DEBe a near ideal doping precursor for GaAs and $Al_xGa_{1-x}As$.

Problems were encountered when DEBe was used as p-type precursor in InP during growth by CBE (Weyers et al., 1990). First, the doping level was not reproducible. Second, the measured hole mobilities were low. Third, the doping efficiency was approximately 10^{-3}. Here, the doping efficiency is defined as

$$\eta_{\text{Doping}} = \frac{\text{carrier concentration/beam pressure of dopant source}}{(4/a_0^3)\text{/beam pressure of group-III source}} \quad (6.1)$$

where $4/a_0^3$ is the cation concentration of the semiconductor. For GaAs, $a_0 = 5.6533$ Å, i.e. $4/a_0^3 \cong 2.2 \times 10^{22}$ cm^{-3}. However, the authors pointed out that other doping precursors such as DEZn or Cp_2Mg exhibited similar problems. The authors further suggested that oxygen contamination and possible parasitic gas-phase reactions are the cause for the unusual doping characteristics of InP.

Be, DEBe, BeO, and other compound containing Be are highly toxic (Parsons and Krajenbrink, 1983). DEBe inflames in air like most beryllium dialkyls and the resulting BeO is highly toxic.

Mg precursors

The most frequently used magnesium precursor is biscyclopentadienylmagnesium (Cp_2Mg, $Mg(C_2H_5)_2$) which is a white solid with a 281 °C melting point temperature. It lacks the strong toxicity of the Be precursors. Hole concentrations exceeding 10^{19} cm^{-3} using the Cp_2Mg doping precursor were achieved in GaAs and $Al_xGa_{1-x}As$ (Lewis et al., 1983; Kozen et al., 1986). Furthermore, Mg has not been reported to diffuse as strongly as e.g. Zn in the high concentration regime.

Another Mg precursor, bismethylcyclopentadienylmagnesium (MCp_2Mg, $Mg(C_2H_4-CH_3)_2$), was reported by Timmons et al. (1986). This compound is structurally similar to Cp_2Mg but has methyl groups connected to each carbon ring. The melting point of MCp_2Mg is 27 °C permitting a liquid source at moderate heating powers. Hole concentrations exceeding 10^{19} cm^{-3} were achieved in GaAs.

Doping experiments with Cp_2Mg revealed that the Mg concentration in GaAs increases monotonically up to substrate temperatures of approximately 580 °C (Weyers et al., 1990). At temperatures exceeding 600 °C an exponential decrease of the free hole concentration was observed. The decrease of the free carrier concentration was quantitatively modeled in terms of an Arrhenius characteristic with an activation energy of 5 eV. The reduction of the free hole

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

concentration was attributed to reevaporation of Mg from the crystal surface. The same activation energy of 5 eV was found for elemental Mg doping in GaAs grown by molecular-beam epitaxy (Wood et al., 1982) confirming the hypothesis of Mg reevaporation. The sensitivity of Mg incorporation on the growth temperature is an undesired property which makes accurate process control mandatory.

The incorporation of Mg from MCp_2Mg is superlinear. Timmons et al. (1986) measured the free carrier concentration as a function of the H_2 carrier gas flow through the liquid MCp_2Mg doping source. For free carrier concentrations in the $10^{17} - 10^{19}$ cm^{-3} range, the concentration increased superlinearly with the carrier gas flow. The increase of the carrier gas flow by a factor of 3 resulted in a hole concentration increase by a factor of 10.

MCp_2Mg and Cp_2Mg were found to adsorb on the reactor walls of OMVPE systems (Timmons et al., 1986; Roberts et al., 1984). Such an interaction of the precursor with the reactor walls results in memory effects and makes abrupt doping profiles difficult to achieve. Timmons et al. (1986) alleviated this problem by injecting MCp_2Mg near the quartz reactor and by using a run/vent approach. Using this approach, the feasibility of top-hat doping profiles with constant maximum concentration and abrupt turn on and off were demonstrated.

6.3.2 Group-IIb precursors

Among the group-IIb elements, Zn precursors are the most frequently used. These are well studied and the studies illustrate the interdependence of growth and doping parameters in OMVPE. Several organometallic Cd precursors are used in OMVPE. In general, the Cd doping properties were found to be inferior to those of Zn.

Zn precursors

Zinc is the most popular p-type doping impurity in OMVPE-grown III–V semiconductors. Zinc precursors are well studied and much is known about the interdependence of Zn doping precursors and crystal growth parameters. Dimethylzinc (DMZn, $(CH_3)_2Zn$) and Diethylzinc (DEZn, $(C_2H_5)_2Zn$) are used as precursors (see for example, Bass and Oliver, 1977; Manasevit and Thorsen, 1972).

Thermal decomposition of DMZn and DEZn is believed to be close to unity at substrate temperatures as low as 500 °C. Weyers et al. (1990) estimated the doping efficiency (see Eq. (6.1)) of DEZn to be approximately 0.3 at a temperature of 520 °C. Glew (1984) assumed near-complete pyrolysis of DMZn and DEZn at temperatures ≥ 500 °C. Thus, effective thermal decomposition and effective incorporation of Zn can be obtained at low growth temperatures.

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

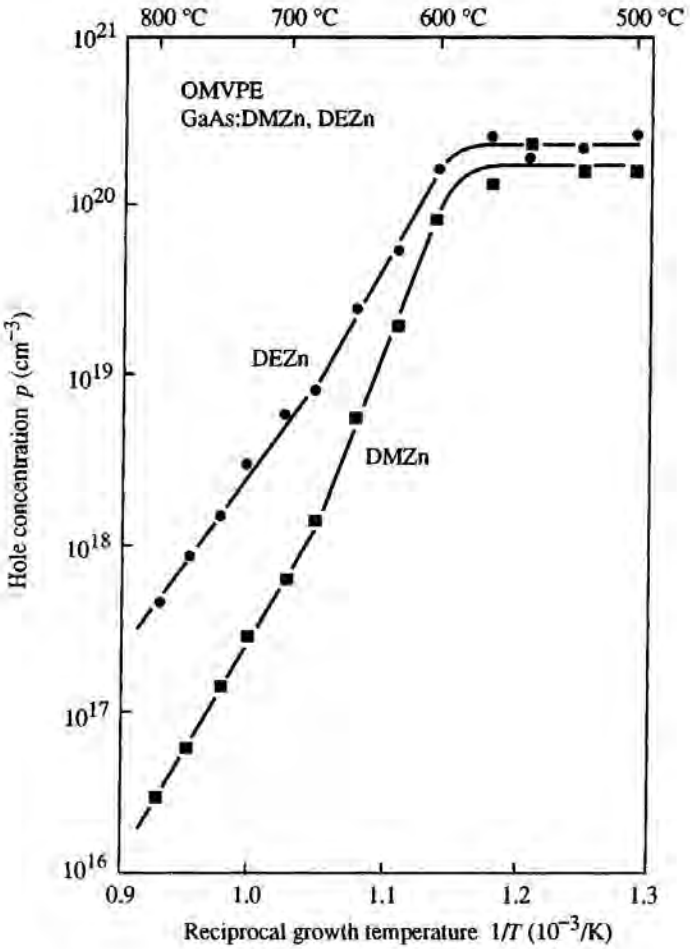


Fig. 6.8. Hole concentration of OMVPE-grown GaAs doped with DMZn and DEZn as a function of reciprocal growth temperature (after Glew, 1984).

For higher growth temperatures, the incorporation of Zn becomes strongly dependent on growth temperature. Figure 6.8 shows the hole concentration as a function of the growth temperature of OMVPE-grown GaAs (Glew, 1984). The hole concentration decreases rapidly with increasing substrate temperatures. Glew (1984) evaluated the activation energies associated with the slopes of the curves shown in Fig. 6.8. In the temperature range 575 °C to 675 °C, activation energies of 4.0 eV and 3.3 eV were obtained for DMZn and DEZn, respectively. In the temperature range 675 °C to 800 °C, activation energies of 2.7 eV and 2.1 eV were obtained for DMZn and DEZn, respectively. Similar values were reported by Bass and Oliver (1977).

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

The rapid decrease of the free hole concentration in Zn-doped GaAs grown at elevated temperatures can be attributed to Zn reevaporation from the crystal surface. Stringfellow (1989) pointed out that the hole concentration shown in Fig. 6.8 is inversely proportional to the vapor pressure of Zn, which increases by several orders of magnitude between 600 °C and 800 °C. Zinc doping is thus limited by desorption of Zn from the growing crystal surface.

Zinc incorporation also depends sensitively on the V/III flux ratio. Glew (1984) and Bass and Oliver (1977) reported the dependence of the free carrier concentration on the V/III flux ratio for DMZn and DEZn doping precursors. An increase of the free hole concentration was found for high V/III flux ratios. The enhanced hole concentration was attributed to an enhanced Zn incorporation on cation sites as concluded from the law of mass action. However, intermediate reaction steps involving As_4 and As_2 species cannot be excluded.

Conflicting reports were given on the dependence of Zn incorporation on the DMZn and DEZn mole fraction in the gas phase of the reactor. Glew (1984) found a sublinear dependence of DMZn incorporation in GaAs and a superlinear dependence of DEZn incorporation in GaAs. A previous report by Bass and Oliver (1977) found a linear dependence for the DMZn precursor in GaAs. For DEZn, a sublinear (Aebi et al., 1981), a linear (Keil et al. 1982), and a superlinear (André et al., 1977; Hallais, 1978) dependence of Zn incorporation on DEZn mole fraction were reported.

The complex interdependence of growth parameters and doping parameters in OMVPE growth was beautifully demonstrated in a series of experiments by Glew (1984). The author investigated the dependence of the growth rate on doping parameters such as doping species and doping precursor mole fraction in the gas phase. It is a desirable property of doping precursors that they not influence the growth rate, alloy composition (in ternary and quaternary semiconductors), etc. However, due to parasitic reactions, catalytic reactions, and complicated reaction pathways, interdependence of growth parameters and doping parameter results.

Figure 6.9 shows the dependence of the OMVPE GaAs growth rate on the reciprocal substrate temperature for the two doping precursors DMZn and DEZn (Glew, 1984). The dependence is characterized by three different regions. First, a low temperature regime in which the growth rate increases with temperature. Second, an intermediate growth temperature regime where the growth rate is constant. Third, a high temperature regime in which the growth rate decreases. The increase of the growth rate in the low temperature regime (500–575 °C) is attributed to the partial thermal cracking to TMGa (Nishizawa and Kurabayashi, 1983). With increasing temperature, the pyrolysis of TMGa is more complete and results in a higher growth rate. In the intermediate temperature region, the growth

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

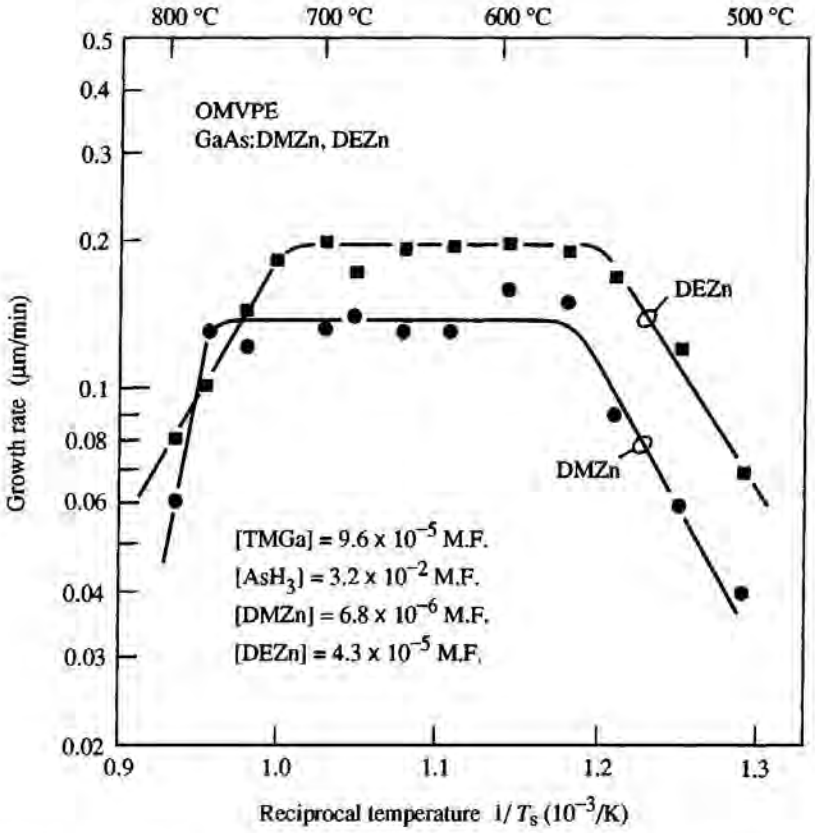


Fig. 6.9. Growth rate of OMVPE-grown GaAs as a function of reciprocal substrate temperature. The growth rate depends on the doping compound (DMZn or DEZn) used. The mole fraction (M.F.) of the gases are: [TMGa] = 9.6 x 10^{-5}M.F., [AsH_3] = 3.2 x 10^{-2}M.F., [DMZn] = 6.8 x 10^{-6}M.F., [DEZn] = 4.3 x 10^{-5}M.F., with a total flow of 3800 cm^3/min (after Glew, 1984).

rate is constant but clearly dependent on the doping species. A lower growth rate is found for the TMGa precursor as compared to the TEGa precursor. The transition to the high temperature regime depends critically on the doping species employed. Figure 6.9 reveals a transition temperature of 775 °C for DMZn and a lower temperature of 725 °C for DEZn. The origin of the qualitatively and quantitatively different dependence of DMZn and DEZn is not understood in detail. It is likely that parasitic cross reactions between products of the pyrolysis are the origin of this different growth rate behavior.

The dependence of the growth rate in OMVPE-grown GaAs on the doping precursor mole fraction in the gas phase is shown in Fig. 6.10 (Glew, 1984). The four growth temperatures 625, 650, 700, and 750 °C, and the two doping

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

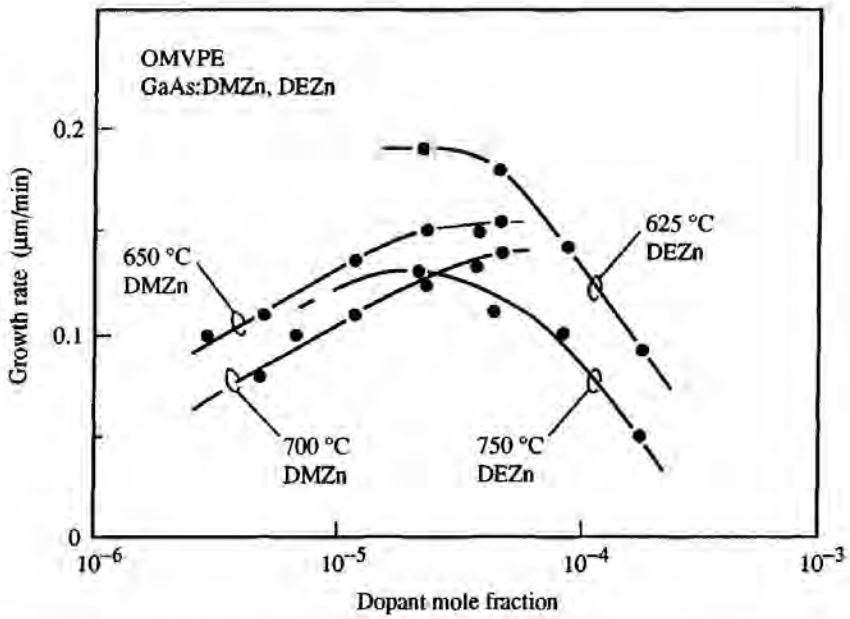


Fig. 6.10. Growth rate of GaAs grown by OMVPE as a function of the dopant concentration for the precursors DMZn and DEZn. The mole fractions in the gas phase are [TMGa] = 9.6 x 10^{-5}M.F., [AsH_3] = 3.2 x 10^{-2}M.F., with a total flow of 3800 cm^3/min (after Glew, 1984).

precursors DMZn and DEZn were used. Inspection of Fig. 6.10 reveals that the growth rate first increases with the Zn mole fraction and then, for higher Zn contents, decreases with Zn mole fraction. No conclusive explanation for the dependence of the growth rate on the dopant mole fraction was offered. A maximum of the growth rate is obtained at a dopant mole fraction of $2 \times 10^{-5} - 5 \times 10^{-5}$ which is similar to the maximum in growth rate for Sn-doped GaAs (Chang et al., 1983). Glew (1984) suggested that the decrease in growth rate at high Zn mole fractions is due to the interference of the high Zn concentration on the growing GaAs surface. In addition, parasitic, and catalytic gas phase reactions and reactions at the susceptor and substrate should be considered for the complicated dependence of the growth rate.

Cd precursors

Dimethylcadmium (DMCd, $(CH_3)_2Cd$) was employed by a number of research groups as a doping precursor for Cd (see, for example Blaauw et al., 1987; Glade et al., 1989; Matsumoto et al., 1989; Matsumoto et al., 1990; Nelson and Westbrook, 1984). The use of the DMCd doping precursor leads to incorporation of Cd on cation sites of the III–V lattice and to p-type conductivity of the

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

semiconductor. Cd was found to be a well-behaved dopant predominantly in InP compounds.

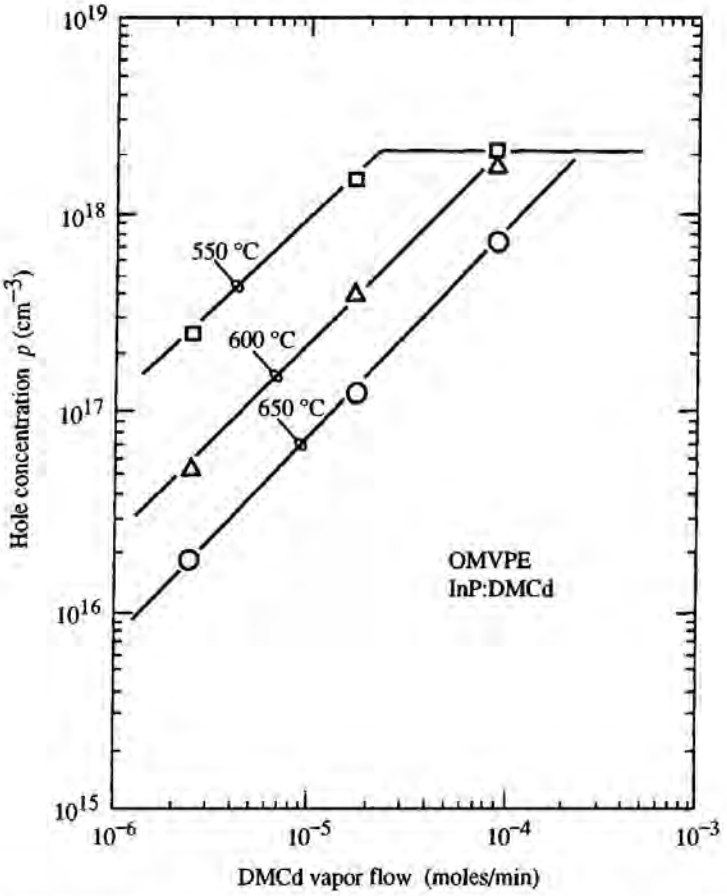


Fig. 6.11. Free hole concentration of DMCd-doped InP grown by OMVPE at 550 °C, 600 °C and 650 °C (after Nelson and Westbrook, 1984).

The incorporation of Cd from the DMCd precursor is limited by desorption of Cd from the growing surface for temperatures ≥ 550 °C. The free carrier concentration of InP grown by OMVPE is shown in Fig. 6.11 as a function of dopant mole fraction for three different growth temperatures (Nelson and Westbrook, 1984). The carrier concentration decreases by more than one order of magnitude as the growth temperature is increased from 550 °C to 650 °C. The decrease in hole concentration can be fully explained by Cd desorbing from the growth surface. The data shown in Fig. 6.11 does not reveal if all DMCd is decomposed during growth. Nevertheless, hole concentrations in the 10^{18} cm^{-3}

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

range can be achieved with Cd doping. Nelson and Westbrook (1984) ruled out that the change in doping concentration with growth temperature is due to a changing growth rate. Although the growth rate changes slightly with growth temperature, the decrease in doping concentration shown in Fig. 6.11 is too dramatic to be explained by a temperature-dependent growth rate.

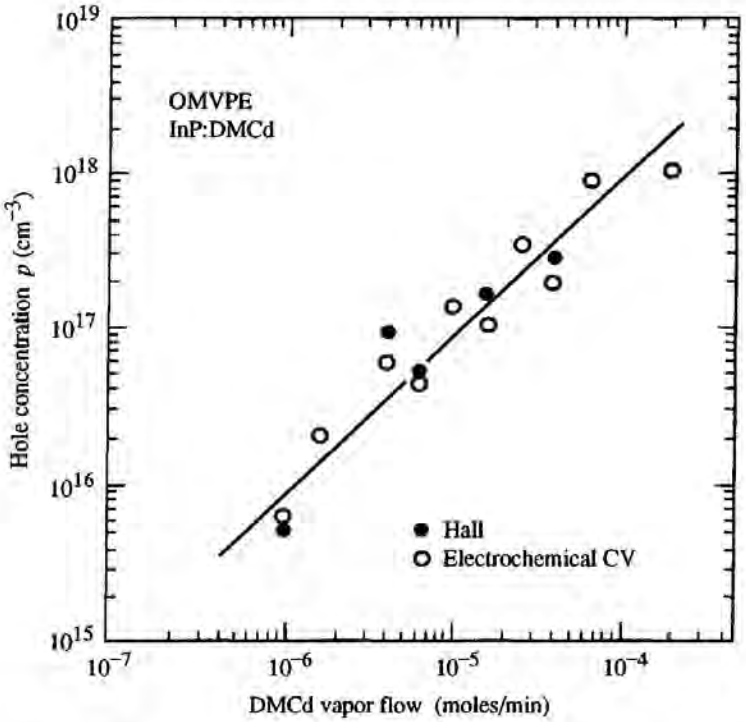


Fig. 6.12. Hall carrier concentration measured by the Hall effect and electrochemical capacitance–voltage profiling of OMVPE-grown InP as a function of DMCd mole fraction (after Nelson and Westbrook, 1984).

The hole concentration in OMVPE-grown InP depends linearly on the DMCd mole fraction in the gas phase. Figure 6.12 shows the free hole concentration measured by the Hall effect and electrochemical capacitance–voltage profiling as a function of the DMCd mole fraction in the gas phase for a growth temperature of 650 °C. A strictly linear dependence of the free hole concentration on the DMCd flow is obtained over more than two orders of magnitude in flow rate. Nelson and Westbrook (1984) confirmed the linear dependence by additional secondary ion mass spectrometry measurements. The authors concluded that the properties of DMCd discussed above make Cd a well-behaved and very useful p-type dopant for (adduct) OMVPE.

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

A linear dependence of the Cd acceptor concentration on the DMCd mole fraction was also found by Matsumoto et al. (1989) at temperatures of 680 °C during OMVPE growth. However, the linear dependence was not observed at higher temperatures. At a growth temperature of 720 °C the authors found an exponential dependence (rather than a linear one) of the acceptor concentration on the DMCd mole fraction in the gas phase. The unusual exponential dependence of the hole concentration was explained by Cd impurity induced two-dimensional nucleation.

A deterioration of the InP surface morphology was observed for low-pressure OMVPE-grown InP at high doping levels (Blaauw et al., 1987). DMCd was used as a p-type doping precursor. The degradation of the surface morphology was assessed by Normarsky microscopy. The degradation occurred at doping levels of 10^{17} cm^{-3} and above. Although the authors did not offer a conclusive explanation for the degradation in surface morphology, they concluded that Cd is not a suitable p-type dopant for low-pressure OMVPE-grown InP at high doping levels.

6.3.3 Group-IV precursors

Among the group-IV precursors, carbon, silicon, and tin precursors are of interest. Carbon precursors are used for p-type doping during chemical-beam epitaxial growth. Carbon is an exceptionally stable p-type impurity in GaAs. Silicon and tin are both used for n-type doping. The remaining group-IV elements Ge and Pb are of minor interest. Germanium is known to be strongly amphoteric in GaAs and cannot be used as efficient dopant (see Chap. 5). Lead, which is at the bottom of column IV, has a large atomic radius and is not expected to be a very soluble impurity in III–V semiconductors.

C precursors

The use of trimethylgallium ($(CH_3)_3Ga$, TMGa) in III–V semiconductors containing Ga during VPE and CBE results in a considerable incorporation of C. The efficient incorporation of C from TMGa can be exploited by using TMGa as a Ga- and simultaneously as a C-doping precursor. Very high C acceptor concentrations are possible with the TMGa precursor. Carbon is known to be a stable p-type impurity in GaAs which makes the TMGa precursor an attractive choice for highly p-type doped GaAs.

It was demonstrated in the 1970s that the choice of the organometallic Ga precursor affects the unintentional C background concentration of GaAs grown by OMVPE (Seki et al., 1975; see also Fraas et al., 1984). The first systematic comparative study of the TMGa and TEGa precursors with respect to C incorporation was made by Pütz et al. (1986) for GaAs grown by CBE. The

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

authors considered the Ga–C bond strength in TMGa and TEGa for a qualitative understanding of the C incorporation in the two cases.

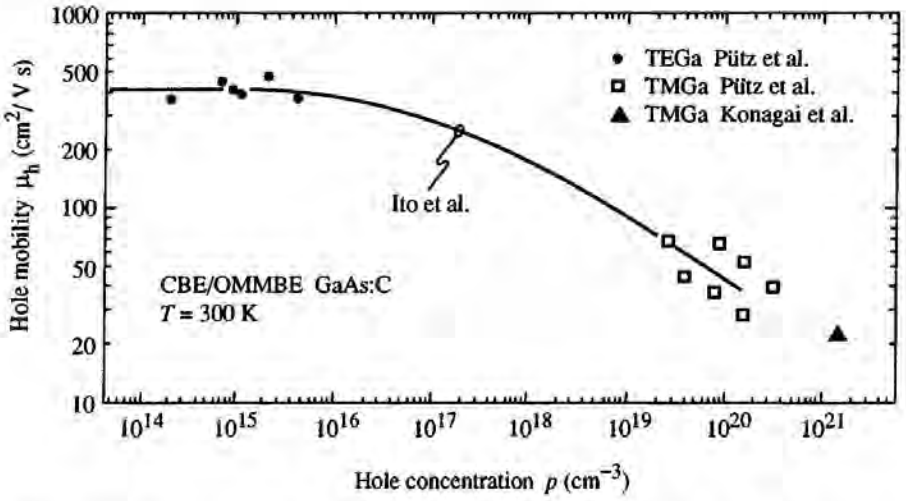


Fig. 6.13. Mobility as a function of hole concentration in C-doped GaAs grown by OMMBE using trimethylgallium (TMGa) and triethylgallium (TEGa) as Ga precursor (after Pütz et al., 1986; Konagai et al., 1990; Ito et al., 1984).

The sensitivity of C incorporation into GaAs on the choice of the Ga precursor is illustrated in Fig. 6.13. Pütz et al. (1986) used TMGa and TEGa as Ga precursors during CBE. Drastic differences were detected for the TMGa and the TEGa precursor with respect to the C doping concentration. The TEGa precursor results in moderately pure GaAs with residual background concentrations on the order of 10^{15} cm^{-3}. The background concentration was assumed to be C. The TMGa precursor, in contrast, results in very high C concentrations in the 10^{19} – 10^{20} cm^{-3} doping range. The high background doping concentration originates from the C contained in the alkyl radicals. The hole mobilities shown in Fig. 6.13 and their comparison with the best experimental mobility values for p-type GaAs (solid line, Wiley, 1975) indicate that the C-doped GaAs is not heavily compensated and is of high quality. These experimental results suggest that controlled C doping can be achieved by simultaneously using the two precursors TMGa and TEGa and by adjusting the TMGa/TEGa ratio to obtain the desired C-doping concentration. Doping concentrations ranging from 10^{14} cm^{-3} to 10^{21} cm^{-3} can be achieved using the two precursors.

Pütz et al. (1986) analyzed and discussed the chemistry of the TMGa and TEGa precursors which result in the dramatically different C incorporation. The

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

effective strength of the Ga–C bond is weaker in TEGa as compared to TMGa. The dissociation of an ethyl group in TEGa occurs according to the reaction

$$(C_2H_5)_3Ga \leftrightarrow (C_2H_5)_2GaH + CHCH_3 . \quad (6.2)$$

The dissociation of the ethyl group from TEGa and the transfer of a H atom to the Ga is known as the β-elimination process. The β-elimination process is not possible in TMGa. Therefore, the Ga–C bond in TEGa is effectively weaker as compared to TMGa (Pütz et al. 1986; Tuck, 1982; Yoshida et al., 1985). The stronger Ga–C bond in TMGa therefore leads to the incorporation of a fraction of C atoms simultaneously with the Ga atom.

The controlled C doping of GaAs using a balance of TMGa and TEGa has been further demonstrated by Weyers et al. (1986) for OMMBE and by Kuech et al. (1988) for OMVPE.

Very high C doping concentrations in GaAs were reported by Konagai et al. (1989, 1990) for OMMBE growth. The highest free hole concentration was 1.5×10^{21} cm^{-3} which is the highest free carrier concentration obtained in any III–V semiconductor. The corresponding Hall mobility and resistivity are 22 cm^2/V s and 1.9×10^{-4} Ω cm, respectively. This data point is included in Fig. 6.13. The high C concentration puts the GaAs lattice under tensile stress and results in a reduction of the lattice constant by −0.5% at a C concentration of 10^{21} cm^{-3}. Konagai et al. (1990) pointed out that the measured reduction in lattice constant is quantitatively in agreement with Vegard's law (see Sect. 12.3.2).

Low growth temperatures are required to achieve high C concentrations in GaAs. Figure 6.14 shows the free hole concentration of (001) and (111) oriented C-doped GaAs as a function of reciprocal temperature. The highest doping concentrations are achieved at low temperatures of 450 °C. In addition to the doping concentration, the growth rate changes from 2.5 μm/h to 0.5 μm/h for a temperature of 600 °C and 450 °C, respectively. Konagai et al. (1989) proposed that the TMGa molecule is less likely to thermally dissociate at low growth temperature which leads to an enhanced incorporation of C, i.e. to high C doping concentrations.

Other doping precursors proposed for C doping of III–V semiconductors during OMMBE are ethene (C_2H_2) and trimethylarsenic ($(CH_3)_3As$, TMAs). Weyers et al. (1990) showed that both ethene and TMAs are stable molecules that are not pyrolyzed effectively at typical growth temperatures and do not result in significant C incorporation. Thermal pyrolysis in a cracker cell upon entering the CBE or OMMBE growth system would result in a more efficient C incorporation from the ethene and TMAs precursors.

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

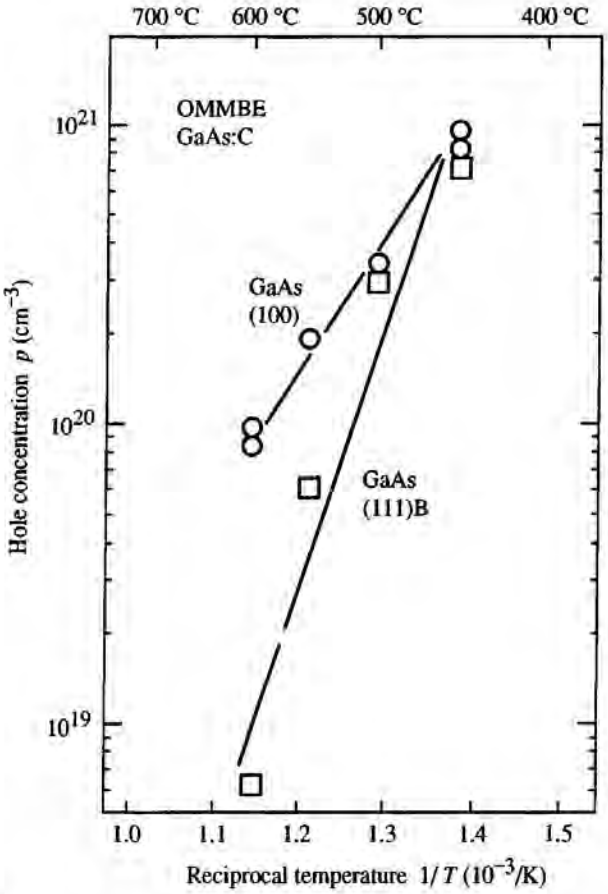


Fig. 6.14. Hole concentration of C-doped GaAs as a function of reciprocal growth temperature. TMGa is employed as both doping percursor and as group-III precursor (after Konagai et al., 1990).

Si precursors

A number of Si doping precursors are used for Si doping including silane (SiH_4), disilane (Si_2H_6), and triethylsilane ($(C_2H_5)_3SiH$, TESiH). The precursors are used for n-type doping of III–V semiconductors in which Si is a donor.

The most obvious Si doping precursor is silane (see, for example, Bass, 1979), which initially proved to be a well-behaved precursor. Ohno et al. (1984) demonstrated, using silane as a Si doping precursor, that abrupt Si doping profiles can be achieved in GaAs. However, subsequent reports showed that the incorporation of Si from silane depends strongly on the growth temperature. A higher growth temperature yields increased Si incorporation in the semiconductor. The temperature dependence of Si incorporation is shown in Fig. 6.15 for

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

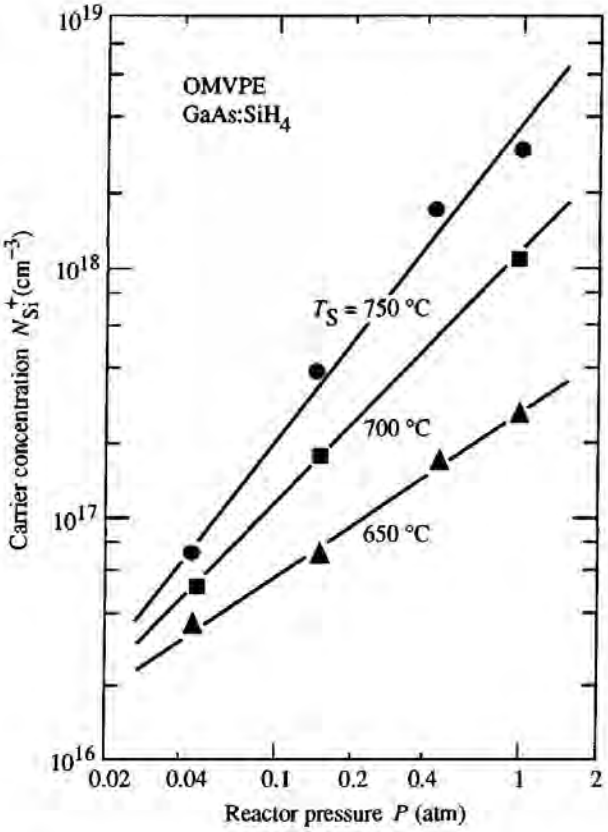


Fig. 6.15. Electron concentration of OMVPE-grown GaAs as a function of reactor pressure for different growth temperatures. Silane (SiH_4) is used as doping precursor (after Field and Ghandhi, 1986a).

OMVPE-grown GaAs doped using SiH_4 as a doping precursor (Field and Ghandhi, 1986a). The experimental data shows that the free carrier concentration (Si concentration) increases with growth temperature. The increase suggests that SiH_4 is not pyrolyzed in the entire temperature range. As the substrate temperature is increased, more Si is incorporated due to a more efficient thermal decomposition of silane. Field and Ghandhi (1986a) also developed a thermodynamical model which predicts the increasing Si concentration with reactor pressure depicted in Fig. 6.15.

The incomplete pyrolysis and the resulting dependence of the Si concentration on the growth temperature is an undesirable property of a doping precursor. Since the incomplete pyrolysis stems from the thermal stability of the SiH_4 molecule, a less stable molecule could alleviate this problem. Veuhoff et al. (1985) proposed that disilane (Si_2H_6), which is less stable than silane, should

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

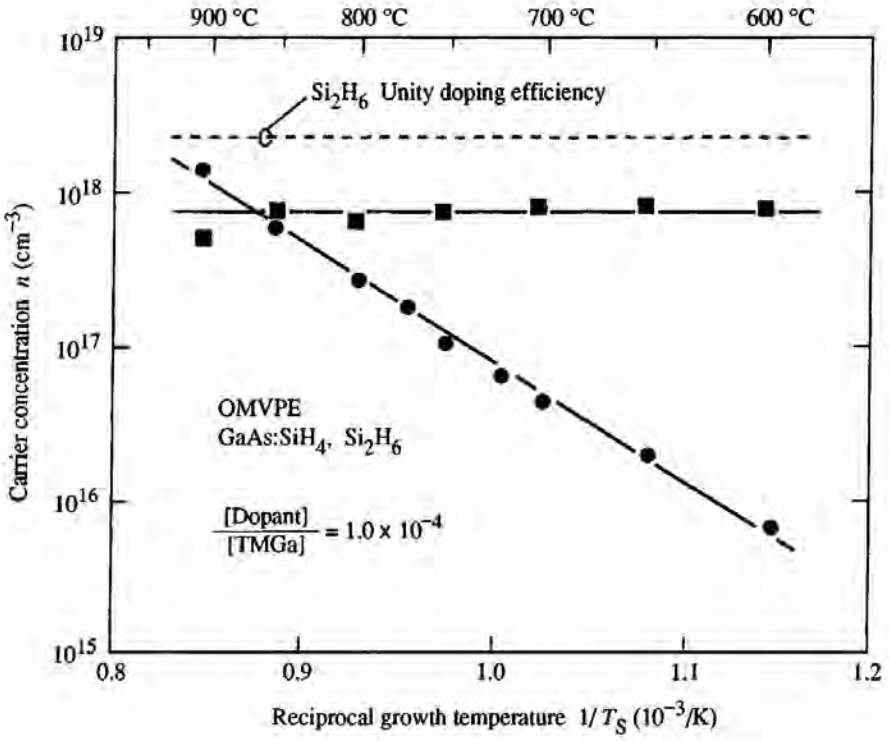


Fig. 6.16. Carrier concentration of OMVPE-grown GaAs as a function of reciprocal growth temperature for the two doping compounds silane (SiH_4) and disilane (Si_2H_6). The dashed line indicates unity doping efficiency for disilane. The ratio between the dopant and TMGa mole fraction is 1×10^{-4} (after Kuech et al., 1984).

therefore be used as a Si doping precursor. Previously, Kuech et al. (1984) demonstrated that the Si incorporation from the disilane precursor is independent of growth temperature in the temperature regime studied. The experimental results are shown in Fig. 6.16 which reveals that the incorporation of Si from the disilane precursor is independent of the growth temperature for the temperature interval 600 °C to 900 °C. The dopant mole fraction is kept constant for the entire temperature range studied. The experimental results further show that the incorporation efficiency (see Eq. (6.1)) of Si from the precursor disilane is approximately 0.5, i.e. much higher than the incorporation efficiency of Si from silane which can be $< 10^{-2}$.

A further advantageous property of the Si incorporation from disilane is the linearity of the incorporation. Figure 6.17 shows the free electron concentration as a function of the disilane mole fraction (Kuech et al., 1984). The free electron

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

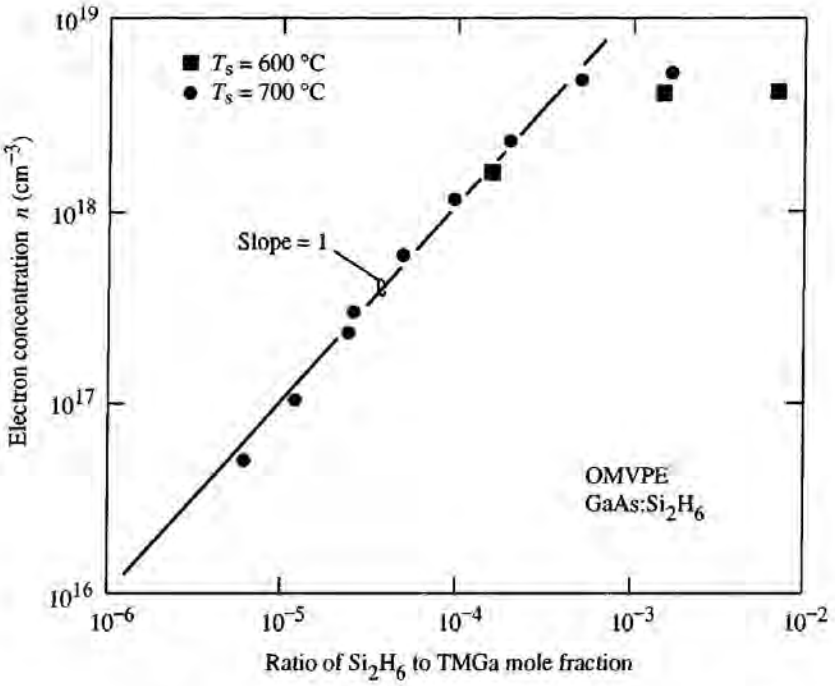


Fig. 6.17. Carrier concentration of OMVPE-grown GaAs as a function of Si_2H_6 concentration in the gas phase. A similar dependence was found for disilane (Si_2H_6) doping in $Al_xGa_{1-x}As$ (after Kuech et al., 1984).

concentration follows the dopant mole fraction in a linear fashion. The evaluation of the experimental data revealed a slope of unity. The incorporation of Si from the disilane precursor thus has several clear advantages namely (i) high pyrolytic efficiency, (ii) high incorporation efficiency, and (iii) an incorporation efficiency which is linearly dependent on the precursor mole fraction.

While advantageous properties were found for Si-doped GaAs and $Al_xGa_{1-x}As$, Si doping of InP using silane was reported to be problematic (Hsu et al., 1986). As for GaAs, the authors found a strong temperature dependence of the Si incorporation from silane. In the temperature range of 575 °C – 650 °C, an incorporation efficiency (see Eq. (6.1)) of $4.5 \times 10^{-3} - 9.5 \times 10^{-3}$ was found. Furthermore, the luminescence intensity decreased drastically with Si doping of InP. A reduction of more than one order of magnitude was found for Si doping levels in the 10^{17} cm^{-3} range. Concomitant with the increased Si concentration was the appearance of a broad 1.06 μm line which is indicative of midgap defect states.

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

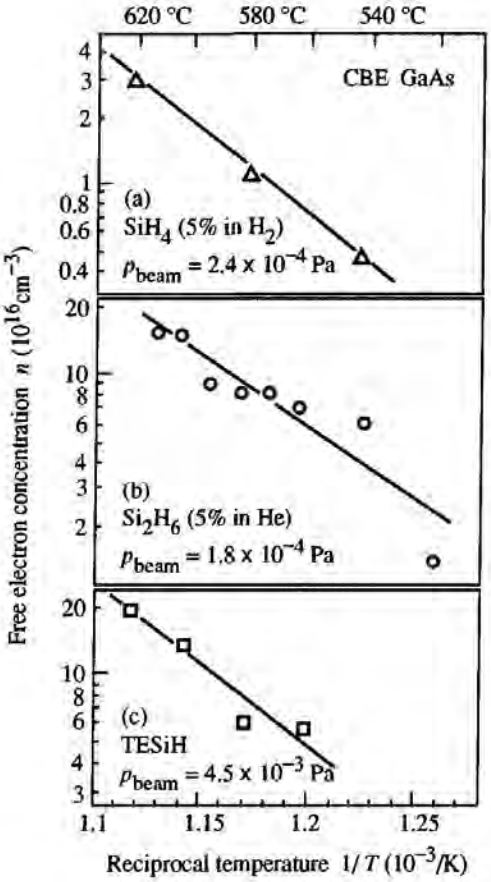


Fig. 6.18. Free electron concentration in n-type GaAs doped with (a) SiH_4 (5% in H_2), (b) Si_2H_6 (5% in He), and (c) TESiH grown by chemical-beam epitaxy as a function of reciprocal substrate temperature (after Weyers et al., 1990).

Gaseous Si precursors are also candidates for doping in CBE due to the doping anomalies observed with elemental doping effusion cells (Skevington et al., 1990). The authors reported the formation of SiC in the hot effusion cells which made controlled doping impossible. This problem can be overcome by gaseous doping precursors. Weyers et al. (1990) investigated the precursors silane, disilane, and triethylsilane (TESiH) for GaAs doping during CBE. The growth temperature dependence of the free carrier concentration is shown in Fig. 6.18 for the three doping precursors silane, disilane, and TESiH. The experimental results are summarized as follows. First, the free carrier concentration increases exponentially with reciprocal temperature for all of the three doping precursors. Second, silane results in a smaller free carrier concentration as compared to disilane. The experimental results indicate that the incorporation of Si from silane, disilane, and TESiH is limited by incomplete pyrolysis in the temperature range 540 – 620 °C. The Arrhenius behavior displayed in Fig. 6.18 is indicative of such a pyrolysis limited incorporation. The

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

experimental results further indicate that the incorporation of Si from silane is less effective than the incorporation from disilane. The weaker efficiency of silane can be attributed to the greater thermal stability of the SiH_4 molecule as compared to the Si_2H_6 molecule.

The three doping precursors silane, disilane, and trimethylsilane do not have all the desirable properties required for CBE applications. The lack of complete pyrolysis of the three precursors is due to the lower growth temperatures used during CBE of typically 400 – 600 °C. A solution to the incomplete pyrolysis could be thermal decomposition of the Si precursor in a cracker furnace upon entering the ultra-high vacuum reaction chamber. Upon cracking the precursor, elemental Si would reach the growing crystal surface and incorporate with high efficiency as known from elemental Si doping during MBE. Weyers et al. (1990) also mentioned the possibility of a significant difference between Si precursor decomposition during OMVPE and CBE. The authors suggested that disilane (during OMVPE) is decomposed in the gas phase before reaching the semiconductor surface. They suggested that disilane is decomposed at the semiconductor surface during CBE. However, experimental evidence was not presented to support this hypothesis.

Sn precursors

Tetramethyltin ($(CH_3)_4Sn$, TMSn) and tetraethyltin ($(C_2H_5)_4Sn$, TESn) are used as doping precursors for Sn doping in III–V semiconductors. Both TMSn and TESn have been used for OMVPE growth (see, for example, Roth et al., 1984 and Parsons and Krajenbrink, 1984). TESn has also been used for Sn doping during chemical beam epitaxy. Tin doping in GaAs is known to be problematic for elemental effusion cell doping during MBE (see Chap. 5). Strong diffusion and segregation of Sn are some of the commonly observed phenomena. However, in atmospheric pressure OMVPE the adverse properties are not as pronounced.

Parsons and Krajenbrink found that Sn-doped GaAs using TESn during atmospheric pressure OMVPE has well-behaved characteristics. These characteristics include (i) good control of the Sn concentration over four orders of magnitude (ii) no surface accumulation of Sn, and (iii) weak Sn diffusion which made possible the growth of abrupt doping profiles; (iv) Sn-doped GaAs did not exhibit any measurable memory effect; the study further revealed that (v) the Sn incorporation does not depend on the growth temperature in the range 650 – 700 °C and that (vi) the Sn incorporation depends linearly on the TESn mole fraction in the gas phase over four orders of magnitude of Sn concentration.

A degradation of the GaAs surface morphology was found for high Sn doping levels (Yakimova et al., 1984). Tin doping using a TMSn precursor at

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

concentrations $N_{Sn} \geq 10^{19}$ cm^{-3} resulted in a drastic increase of the etch pit density and a degradation of the surface morphology. The authors attributed the degradation to increased interstitial Sn incorporation at high doping concentrations. This interpretation was supported by a simultaneous saturation of the electrical activity in highly Sn-doped GaAs which occurred at $N_{Sn} \geq 10^{19}$ cm^{-3}.

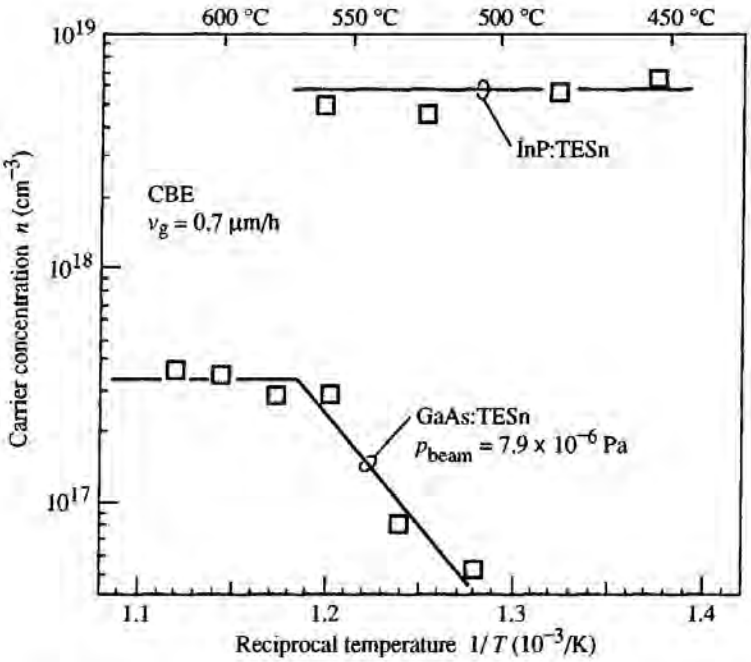


Fig. 6.19. Electron concentration in InP and GaAs grown by chemical-beam epitaxy as a function of reciprocal growth temperature. Tetraethyltin (TESn) is used as n-type doping percursor (after Weyers et al., 1990).

The incorporation of Sn from TESn in GaAs and InP during growth by CBE is more complicated. Figure 6.19 shows the free carrier concentration of Sn-doped GaAs and InP as a function of reciprocal growth temperature (Weyers et al., 1990). The free carrier concentration increases with the growth temperature for temperatures ≤ 570 °C. The increase indicates that the incorporation is limited by pyrolysis. At higher temperatures (T > 570 °C) the carrier concentration remains constant. The constant carrier concentration would suggest that TESn is completely decomposed at T > 570 °C. However, Weyers et al. (1990) concluded that the low incorporation efficiency (see Eq. (6.1)) of only 2×10^{-3} argues against such a complete decomposition. Instead the authors

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

suggest that reevaporation of Sn from the crystal surface balances out the more efficient pyrolysis at higher temperatures. The two processes have similar activation energies, i.e. 1.5 eV and 1.35 eV for pyrolysis and reevaporation, respectively. The similarity of the activation energies indicates that pyrolysis and reevaporation can indeed balance.

Figure 6.19 shows a constant Sn incorporation in InP grown by CBE as a function of temperature. Tetraethyltin is used as a Sn doping precursor. It is unlikely that the low temperatures used for growth of InP ($T < 560$ °C) result in a complete pyrolysis of TESn. The low incorporation efficiency of 10^{-4} (Weyers et al., 1990) and the relatively low growth temperature argue against such a complete pyrolysis. The constant carrier concentration of Sn-doped InP could stem also from a balance of reevaporation and pyrolysis. However, this explanation is not conclusive due to a lack of data for the reevaporation of Sn from the InP surface.

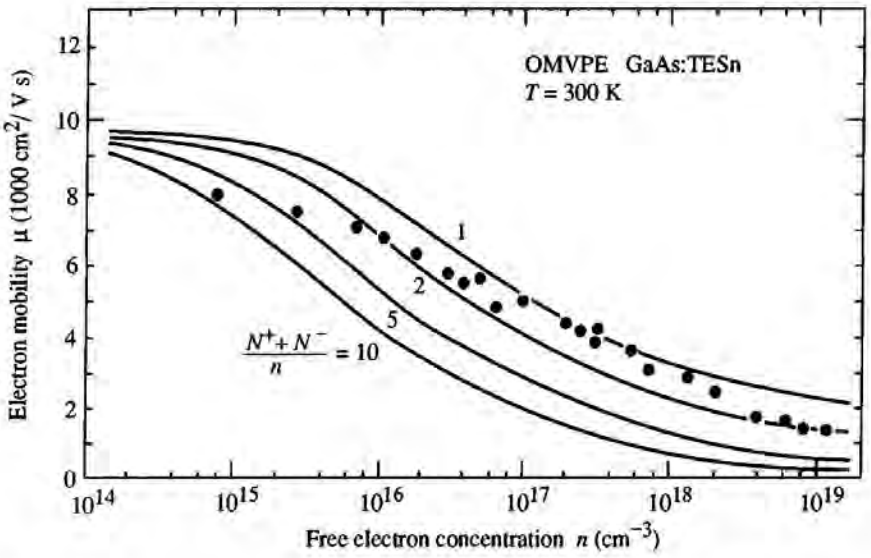


Fig. 6.20. Room-temperature Hall electron mobility as a function of free carrier concentration in Sn-doped GaAs grown by OMVPE (after Parsons and Krajenbrink, 1984).

Sn is known to autocompensate in GaAs only to a small extent. Electron mobilities are therefore expected to be high in Sn-doped GaAs. The electron mobilities of Sn-doped GaAs grown by OMVPE are shown in Fig. 6.20 (Parsons and Krajenbrink, 1984). The doping precursor used in this study is tetraethyltin. Also included in Fig. 6.20 are theoretical mobilities for four different

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

337-TA-1366 RX-0255

compensation ratios (solid lines). The uppermost of the four curves is for uncompensated material, i.e. $n = N^+$. The experimental curves closely follow the uncompensated line for doping concentrations in the 10^{17} cm^{-3} range. At low doping concentrations ($n \cong 10^{15}$ cm^{-3}), the mobilities are lower than the uncompensated curve, probably due to unintentional background impurities. At the high concentration end ($n \cong 10^{19}$ cm^{-3}) the mobilities are lower than the uncompensated curve probably due to autocompensation of Sn. The autocompensation of group-IV impurities at high doping concentration is a general phenomenon observed for all group-IV impurities and will be further discussed in Chap. 7.

6.3.4 Group-VI precursors

Among the group-VI elements, sulfur, selenium, and tellurium are of interest. The remaining elements of group VI, oxygen and polonium, are of no interest. Oxygen forms donor levels located around midgap in GaAs and other III–V semiconductors. Polonium has a large atomic radius and is not expected to be soluble in III–V semiconductors.

Group-VI impurities have high vapor pressures. Among the three elements S, Se, and Te, sulfur has the highest vapor pressure. Descending to elements further down the group, the vapor pressure decreases. Nevertheless, the vapor pressures of Se and Te are still higher than, for example, Si. The high vapor pressure of group-IV elements leads to reevaporation of impurities from the growing semiconductor surface. The donor concentration of GaAs indeed decreases with increasing growth temperature if doped with any of the group-VI impurities. The decrease can be attributed to reevaporation of impurities from the crystal surface.

S precursors

Hydrogen sulfide (H_2S) is the most common doping precursor used for S doping of III–V semiconductors. It is a toxic gas, which can be easily detected by its characteristic smell resembling that of rotten eggs. Sulfur occupies anion sites in III–V semiconductors and is a shallow donor.

Sulfur is characterized by the highest vapor pressure of S, Se, and Te. The possibility of thermally activated desorption of S from epitaxial surfaces is therefore of prime interest. The free carrier concentration of S-doped GaAs grown by OMVPE is shown in Fig. 6.21 as a function of reciprocal growth temperature (Field and Ghandhi, 1986b). Four different pressures of 0.046, 0.15, 0.46, and 1.0 atmospheres were used during the growth experiments. The experimental finding shows a definite decrease of the free electron concentration at high growth temperatures. At low pressures, the decrease in electron concentration exceeds one order of magnitude for a temperature interval of only

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

RX-0255.264

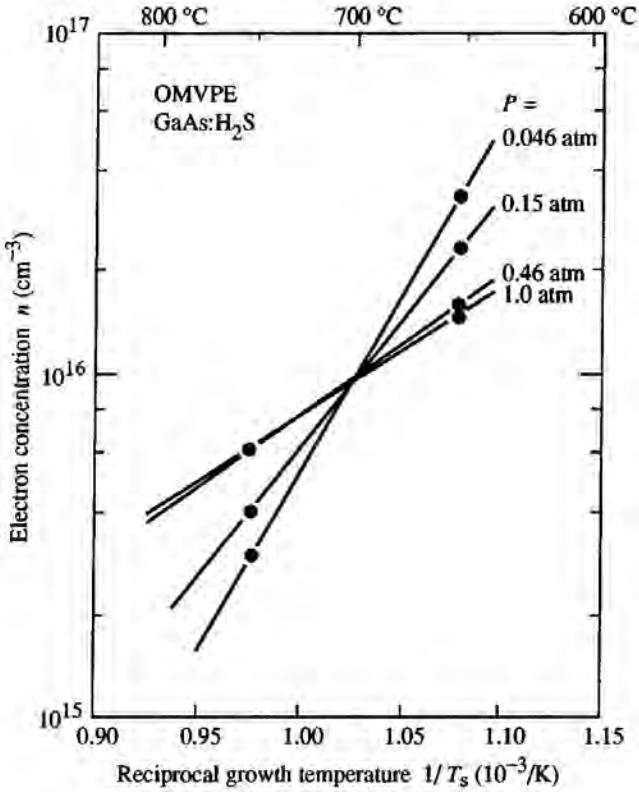


Fig. 6.21. Free electron concentration in OMVPE-grown GaAs as a function of reciprocal substrate temperature. Hydrogen sulfide (H_2S) is used as n-type doping precursor (after Field and Ghandhi, 1986b).

100 °C. The mole fractions of TMGa, AsH_3, and H_2S were kept constant during these experiments. Considering the high vapor pressure of S, it is very likely that the decrease of the free carrier concentration is due to a decrease in S incorporation which in turn is caused by increased reevaporation of S from the crystal surface.

Field and Ghandhi (1986b) also pointed out the relevance of the As vacancy concentration on the crystal surface. They assumed that the chemisorption and adsorption energies depend on the total coverages of H, S, AsH, and As_2. Since these coverages depend on temperature, a change in S incorporation is expected. The authors further commented on the pressure dependence of the S incorporation. They argued that the rate of desorption of S is limited by diffusion. The diffusion of S is faster and more efficient at low reactor pressures. Thus, a stronger decrease of the S concentration at low pressures can be understood on the basis of diffusion.

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

The doping behavior of S in GaAs grown by chloride vapor-phase epitaxy was investigated by Heyen et al. (1977). H_2S was used as the S doping precursor at a growth temperature of 750 °C. The study revealed that at a constant set of parameters the incorporation of S followed linearly the H_2S mole fraction in the gas phase over a wide range of concentrations. Further experiments showed, however, that the S incorporation also depends on the pressures of other reactants in the growth system including AsH_3, GaCl, and HCl.

In a subsequent study Veuhoff et al. (1981) confirmed the linear dependence of S incorporation and the H_2S mole fraction in the gas phase for chloride VPE grown, ⟨001⟩ oriented GaAs. The incorporation behavior was found to be different for GaAs grown on the (111) GaAs surface. For this orientation the authors found a sublinear dependence of the S incorporation versus H_2S mole fraction. The functional dependence of the S incorporation was found to follow a square-root dependence of the H_2S mole fraction.

The efficiency of S incorporation was found to depend on the V/III ratio during GaAs OMVPE growth (Bass and Oliver, 1977). The efficiency (see Eq. (6.1)) of S incorporation decreases with increasing V/III ratio. Since S occupies anion sites it competes with As. A high As/Ga flux ratio is therefore expected to reduce S incorporation. Hydrogen sulphide is gaseous at room temperature. Adsorption of H_2S at the walls of gas tubes, gas manifold, and reactor walls is not likely during doping with H_2S. Any memory effects are therefore not expected for H_2S doping and have not been reported.

Se precursors

Hydrogen selenide is the most commonly used doping precursor for Se doping in III–V semiconductors. Selenium occupies anion sites and is a shallow donor. The memory effect observed with H_2Se doping during OMVPE poses a constraint to the versatility of the H_2Se precursor.

The analysis of the doping characteristics of H_2Se in GaAs revealed that the precursor has a number of well-behaved characteristics. Glew (1982) showed that the incorporation of Se into GaAs follow a superlinear relationship with respect to the H_2Se mole fraction in the gas phase. The superlinear relationship was well described by a power law with an exponent of 3/2, i.e. $N_{Se} \propto [H_2Se]^{3/2}$ where the brackets represent the concentration (mole fraction) of H_2Se in the gas phase of the reactor. The 3/2 power law relation was observed at growth temperatures ranging from 650 °C to 800 °C and also for different V/III flux ratios during growth. The highest doping concentration achieved was 10^{19} cm^{-3} in GaAs.

The incorporation of Se strongly decreases with increasing V/III ratio. Glew (1982) found an approximately linear decrease of the Se concentration in the

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

GaAs crystal with increasing V/III flux ratio. A decrease of N_{Se} by a factor of 10 was found for an increase of the V/III flux ratio from 10 to 100. The sensitivity of Se incorporation to the V/III ratio can be understood on the basis of the law of mass action. Since Se occupies anion sites, it competes with As for lattice sites. Thus an increase of the V/III ratio leads to a reduction of N_{Se}.

One drawback of the H_2Se precursor for OMVPE is the occurrence of Ga_2Se_3 complexes during GaAs growth (Glew, 1982). It was assumed that a parasitic gas phase reaction above the susceptor forms Ga_2Se_3. The molecule Ga_2Se_3 was assumed to be incorporated as one entity into the crystal. Certainly the Ga_2Se_3 complexes do not act as shallow donors in GaAs.

Another major drawback of the H_2Se precursor for OMVPE is the pronounced occurrence of the ***memory effect***. The memory effect makes abrupt changes in the doping profile difficult. Furthermore, it results in an enhanced background Se impurity concentration in nominally undoped material. The memory effect of Se doping with the H_2Se precursor in OMVPE was investigated in detail by Lewis et al. (1984). They attributed the occurrence of the memory effect to the adsorption of H_2Se onto stainless steel surfaces of the gas manifold and onto the glass surfaces of the reactor. After termination of the doping process, H_2Se desorbs slowly from these surfaces and results in Se doping of nominally undoped material. The authors proposed several ways to reduce the memory effect. First, a low growth rate shortens the Se doping tails in the semiconductor after the intentional Se doping is terminated. Second, the reduction of the tubing length from the H_2Se gas cylinder to the reactor reduces the adsorption area and thus the memory effect. Third, the moderate heating of the H_2Se gas manifold and of the reactor walls prevents the adsorption of the precursor on the stainless steel and glass walls.

Te precursors

Diethyltellurium ($(C_2H_5)_2Te$, DETe) has been used as Te doping precursor for OMVPE growth of GaAs and InP (see for example: Houng and Low, 1986; Hsu et al., 1986). Tellurium incorporates on the anion site of III–V semiconductors and is a shallow donor. The DETe is a well-behaved precursor except for the memory effect which was found in DETe doping of GaAs (Houng and Low, 1986). DETe is also a candidate for n-type doping in CBE (Weyers et al., 1990).

Well-behaved doping characteristics of the DETe precursor in OMVPE-grown GaAs were found by Houng and Low (1986). The free electron concentration is shown in Fig. 6.22 as a function of the DETe mole fraction in the vapor. Free electron concentrations in the $10^{16} - 10^{18}$ cm^{-3} range were achieved. The relationship between the free carrier concentration (i.e. Te concentration) and the

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

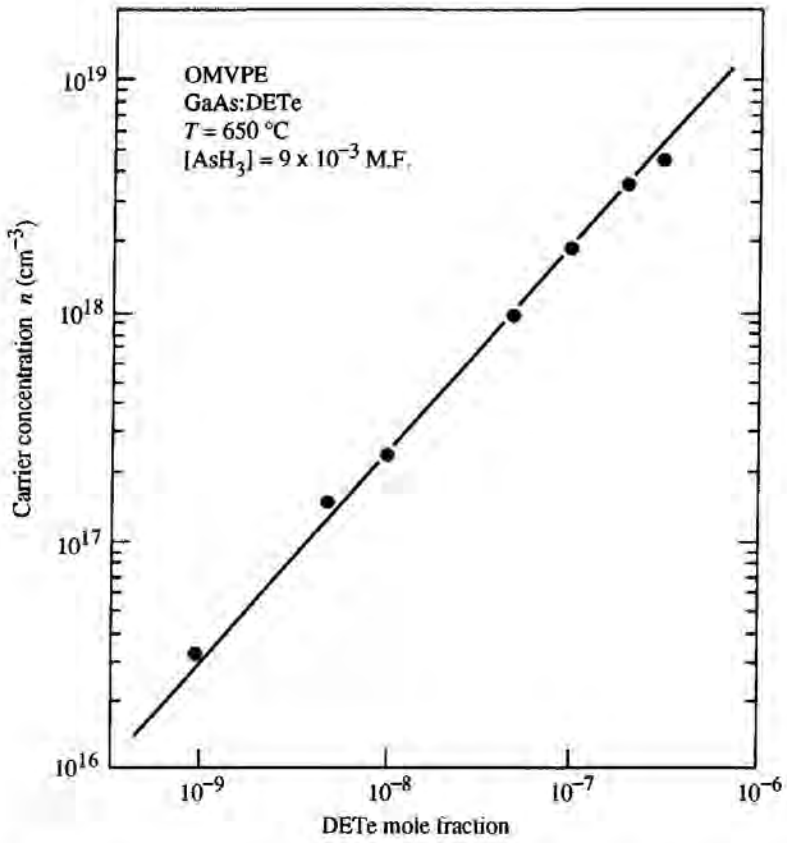


Fig. 6.22. Carrier concentration at 300 K versus DETe mole fraction in the gas phase of OMVPE-grown GaAs (after Houng and Low, 1986).

DETe mole fraction is linear which usually allows well-controlled doping of GaAs. The authors also evaluated the distribution coefficient for GaAs doping with DETe. The ***distribution coefficient*** is defined as $k = x_s/x_v$ where x_s is the fraction of lattice sites in the solid filled with Te impurity atoms and x_v is the ratio of DETe concentration and AsH_3 in the vapor phase of the reactor. Qualitatively the distribution coefficient is a measure of how efficiently Te is incorporated as compared to As. Note that the distribution coefficient is not identical but related to the doping efficiency defined in Eq. (6.1). However, if all incorporated impurities are electrically active, then the doping efficiency equals the distribution coefficient. If $k = 1$, for example, then Te has the same probability as As of being incorporated into the growing GaAs lattice. The authors inferred a distribution coefficient of $k = 8$ from the experimental data of Fig. 6.22.

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

The large distribution coefficient of $k = 8$ found for DETe doping of GaAs indicates that DETe is effectively decomposed at a growth temperature of $T = 650$ °C. Pyrolysis is not a limiting mechanism for the incorporation of Te. However, at lower growth temperatures of 520 – 580 °C a limitation of Te doping due to a lack of pyrolysis was found for GaAs and InP grown by CBE (Weyers et al., 1990).

Houng and Low (1986) also investigated the dependence of Te incorporation on growth parameters. A decreasing incorporation of Te was found with increasing arsine pressure in GaAs growth by OMVPE. Assuming that the free carrier concentration equals the Te concentration, a power law dependence of $N_{Te} \propto [AsH_3]^{-0.71}$ was found. The decrease of N_{Te} with increasing arsine concentration can be qualitatively explained on the basis of the law of mass action. Tellurium competes with As for the anion sites in the zincblende lattice. An increase of the AsH_3 mole fraction reduces the probability of Te incorporation. The authors also investigated the influence of the growth temperature on the doping concentration of the DETe precursor. The increase of the growth temperature from 600 °C to 750 °C resulted in a decrease of the free carrier concentration by approximately 20%. This weak decrease is probably due to reevaporation of Te from the growing surface.

The DETe precursor was used for n-type doping of InP (Hsu et al., 1986). The authors found that the carrier concentration increased superlinearly with the DETe mole fraction. The superlinear increase contrasts the (linear) DETe doping characteristics of GaAs shown in Fig. 6.22. Hsu et al. (1986) state that the physical origin of the superlinear dependence of Te incorporation with DETe mole fraction is unknown. A distribution coefficient of $k = 36$ was inferred from the doping characteristics of Te in InP. A comparison of the distribution coefficients of S, Se, and Te revealed that $k_{Te} > k_{Se} > k_S$. This qualitative dependence is consistent with the vapor pressure of Te, Se, and S, which increases for elements with smaller atomic weight within column VI of the periodic system.

High free electron concentrations were achieved for Te-doped InP. The electron mobilities of highly Te-doped InP grown by OMVPE are shown in Fig. 6.23 (Hsu et al., 1986). Concentrations of 10^{19} cm^{-3} with a mobility exceeding 1000 cm^2/Vs were achieved. The solid curve in Fig. 6.23 is a fit to the data points with the highest mobilities. The figure also shows that the electron mobilities obtained with Si doping are consistently lower than the mobilities obtained from Te-doped GaAs. The lower mobilities of Si-doped GaAs are probably due to autocompensation of Si which is known to be amphoteric.

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

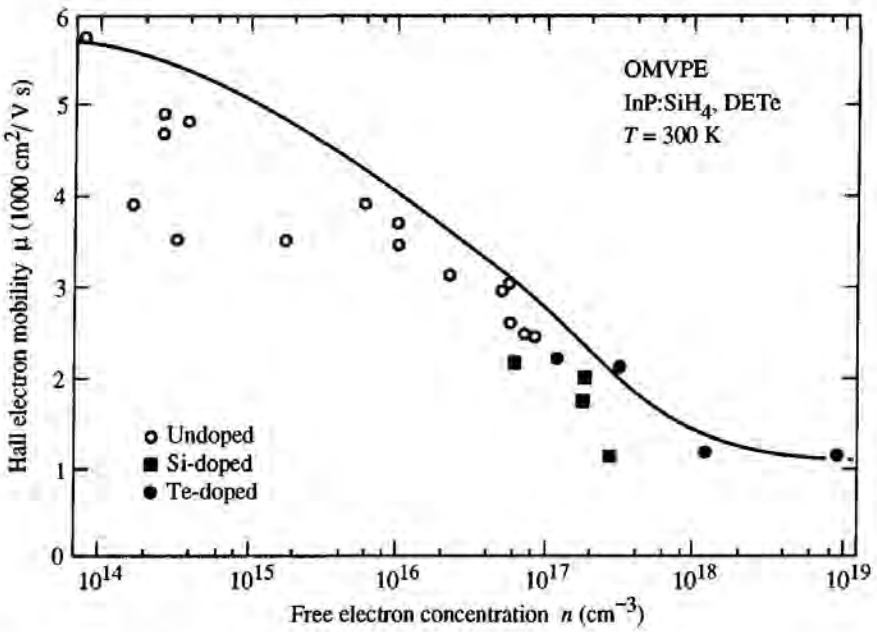


Fig. 6.23. Room-temperature electron mobility of OMVPE-grown unintentionally doped, Te-doped (diethyltellurium), and Si-doped (silane) InP as a function of carrier concentration (after Hsu et al., 1986).

The DETe precursor was also used for doping of GaAs and InP during CBE (Weyers et al., 1990). The authors found an increasing doping efficiency (see Eq. (6.1)) with growth temperature for GaAs and InP growth. The authors attributed this increasing doping efficiency to a more effective pyrolysis at higher growth temperature. Note that the growth temperatures during CBE are generally lower than during OMVPE. Weyers et al. (1990) employed GaAs growth temperatures of 500 – 620 °C and InP growth temperatures of 470 – 560 °C. At low temperatures, the thermal dissociation of DETe molecules is less likely. Furthermore, the very low pressures and the resulting molecular flow regime make gas-phase reactions less likely in CBE as compared to VPE.

https://doi.org/10.1017/CBO9780511599828.009 Published online by Cambridge University Press

7

Impurity characteristics

The incorporation and activation of impurities is governed by a number of physical and chemical laws. It is essential to know and apply them in order to predict experimental functional dependencies of impurity incorporation. For example, the incorporation of many impurities depends on the epitaxial growth temperature. How does the electrical activity, compensation, etc. depend on this growth temperature? How are the impurity characteristics influenced by the V/III flux ratio? These are some of the questions that are of interest. Furthermore, some incorporation characteristics are found only for certain impurity elements. Autocompensation is prominent among group-IV impurities. In order to minimize autocompensation, it is desirable to know the functional dependences of amphoteric impurity incorporation.

Another example of impurity characteristics is the *doping efficiency*. This is defined as the ratio of the free carrier concentration and the dopant concentration. The doping efficiency is unity for an ideal dopant element and an ideal doping procedure. However, it can be quite low. For example, the doping efficiency of impurities incorporated by implantation can be below 10% before activation. The understanding of the principles governing the doping efficiency is therefore desirable.

Finally, very high doping concentrations become increasingly important as the spatial dimensions of semiconductor structures shrink. What are the highest doping concentrations achievable in III–V semiconductors? What limits the maximum impurity concentration? These are some of the topics and questions that will be addressed in this chapter.

Dopants can be incorporated into semiconductors by using several doping techniques including (i) doping during epitaxial growth, (ii) doping by

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

implantation, and (iii) doping by diffusion. The nature of the incorporation can strongly affect the impurity characteristics. For example, the electrical activity of dopants is very different for doping during epitaxial growth or doping by implantation. The different processes of impurity incorporation are therefore considered first.

7.1 Doping incorporation during growth

The incorporation of doping impurities during epitaxial crystal growth depends on the nature of the epitaxial growth method. Dopant incorporation is quite different for the two prominent epitaxial growth techniques molecular-beam epitaxy (MBE) and organometallic vapor-phase epitaxy (OMVPE). Doping incorporation in OMVPE is dominated by the chemistry and thermodynamics of the doping and growth process. In MBE, doping incorporation is mainly controlled by kinetics such as impurity arrival rate, reevaporation rate etc. Therefore, the incorporation of impurities will be discussed separately for the two growth techniques. The law of mass action is very useful for the evaluation and prediction of trends in impurity incorporation and will be discussed first.

The ***law of mass action*** is frequently invoked to derive characteristics of impurity incorporation. The dependence of impurity incorporation on growth temperature and on V/III flux ratio are frequently analyzed in terms of the law of mass action. Due to the importance of the law of mass action, the basic properties of this law will be reviewed. A more exhaustive description of the law can be found in textbooks (see, for example, Kittel and Kroemer, 1980; Kroger, 1973; see also Sect. 3.4). Consider the chemical reaction of the starting reagents A and B and of the reaction products C and D.

$$A + B \leftrightarrow C + D \tag{7.1}$$

where A, B, C, and D can be elements or compounds, i.e. general products of a chemical reaction. The chemical reaction at constant temperature is associated with a change in energy of the system which is given by

$$\Delta G = \Delta H - T\Delta S \tag{7.2}$$

where ΔH represents the changes of different contributions to the energy (i.e. chemical, mechanical, electrostatic, etc.), T is the temperature, and ΔS is the change in entropy occurring in the reaction of Eq. (7.1). The term ΔG is ***Gibbs free energy*** of the system. The term ΔH in Eq. (7.2) is called ***enthalpy***, formation energy, heat of formation, or heat content. It describes the total energy set free by the chemical reaction.

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

337-TA-1366 RX-0255

The quantity ΔH is the total change in enthalpy incurred during the chemical reaction of Eq. (7.1). ΔH can be written as the sum of different contributions to the enthalpy, that is

$$\Delta H = \sum_j \Delta E_j + P\Delta V \tag{7.3}$$

where P and V are the pressure and the volume of the system and $P\Delta V$ is the mechanical work done by the system. Examples for contributions to the change in internal energy ΔE_j are as follows.

1. Chemical binding energies. The formation energy of chemical bonds of the components A and B versus C and D (see Eq. (7.1)) can be different.
2. Electrostatic and magnetostatic energies. Electrostatic and magnetostatic energy densities are given by $\varepsilon\mathcal{E}^2$ and μH^2 where $\mathcal{E}$ and H are the electric and magnetic field, respectively.
3. Electronic energies. If electronic transitions are involved in Eq. (7.1), they also enter the change in energy.
4. Elastic energies. For impurity incorporation in a semiconductor, elastic energies can be important, especially if the impurity has a substantially different size as compared to the lattice atom it substitutes for. Furthermore, steric hindrance of impurities having a different valence structure than atoms of the host crystal contributes to the elastic energy.

The second term of Eq. (7.2) includes the change in entropy ΔS. The entropy is given by

$$S = k \ln w \tag{7.4}$$

where k is Boltzmann's constant. The parameter w is called the *disorder parameter*; it is the probability that a system will exist in the state it is in, relative to all the states it could be in. Equation (7.4) connects the macroscopic (or thermodynamic) quantity of entropy with a statistical or microscopic quantity, the probability. Any chemical reaction of a system that starts in one equilibrium state and ends in another equilibrium state *increases* the entropy of the system (Kittel and Kroemer, 1980).

Considering the chemical reaction of Eq. (7.1) and the change in total free energy given in Eq. (7.2), the ***law of mass action*** is given by

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

RX-0255.273

$$\frac{[C][D]}{[A][B]} = K = ce^{-\Delta G/kT} \tag{7.5}$$

where the brackets [...] represent the concentrations of the compounds and c is a constant. Thus the ratio of the products of the concentrations of initial and final reaction products is a constant at any given temperature. The law of mass action applies to *diluted* chemical solutions and to electronic 'reactions' in *non-degenerate* semiconductors. In the following sections, the law of mass action will be applied to study impurity incorporation during MBE and OMVPE.

7.1.1 Impurity incorporation in MBE

It is frequently argued that the growth of III–V semiconductors by MBE is mainly controlled by kinetics and that thermodynamics is not applicable to MBE growth. At low growth temperatures, incorporation of atoms would be limited by kinetic barriers that make it impossible for atoms to reach their thermal equilibrium positions. The lack of thermal equilibrium does not, however, apply to semiconductor *surfaces*. Heckingbottom (1985) showed that it is reasonable to assume that thermal equilibrium is indeed reached at growing semiconductor surfaces. At a growth temperature of 900 K, the vibrational frequency of an atom can be obtained from $\hbar\omega \cong kT$, which yields an attempt rate of $f = kT/2\pi\hbar \cong 10^{13}\ s^{-1}$. It is further assumed that a hop-attempt for surface diffusion is successful with the probability $p \cong \exp(-E_a/kT)$ where E_a is the activation barrier for the diffusion hop. Using an approximate activation energy of 1 eV for a surface hop (Heckingbottom, 1985; Neave and Joyce, 1983), one obtains the number of successful diffusion hops per second as $fp = 5 \times 10^7\,s^{-1}$. At a growth rate of one monolayer per second atoms on the surface can complete 5×10^7 diffusion hops before they are buried by atoms impinging on the growing surface. The large number of diffusion hops makes it likely that lattice atoms as well as impurity atoms reach low-energy sites, i.e. their equilibrium positions. It is thus very reasonable to assume that the semiconductor surface is in equilibrium at growth temperatures of 900 K (627 °C).

Once the atoms are incorporated into the growing crystal the atoms become practically immobile. The activation energy for Ga self diffusion in GaAs was estimated to be 6 eV (Tan and Gösele, 1988; Goldstein, 1961). Calculating the frequency of successful diffusion hops according to $f_s = kT/2\pi\hbar \exp(-6\ \text{eV}/kT)$ one obtains at $T = 900$ K a frequency of $f_s = 5 \times 10^{-21}\,s^{-1}$ which means that the atoms are practically immobile upon being incorporated into the growing semiconductor.

The clearly different regimes of surface and bulk have led to the concept of ***local equilibrium*** (Heckingbottom, 1985). The concept of local equilibrium

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

suggests that *equilibrium is reached at the surface but not in the bulk.* Thermodynamics can thus be applied to the surface during crystal growth. On the other hand, structures already incorporated in the semiconductor have high kinetic barriers which results in (meta) stability of the structure. Such meta-stability is evidenced by atomically abrupt doping profiles and interfaces which can be grown by MBE.

The concept of local equilibrium cannot be applied to low growth temperatures. Using again $f_s = kT/2\pi\hbar \exp(-E_a/kT)$ for the rate of successful diffusion hops, the rate is obtained as $f_s \cong 2/s$ for a growth temperature of 400 K (127 °C) and an activation energy of $E_a = 1$ eV. Obviously thermal equilibrium positions cannot be reached at this growth temperature unless the growth rate is exceedingly small. The above discussion of surface equilibrium, surface diffusion, etc. has been kept brief. A detailed discussion of surface processes in epitaxy was given by Kreuzer (1991).

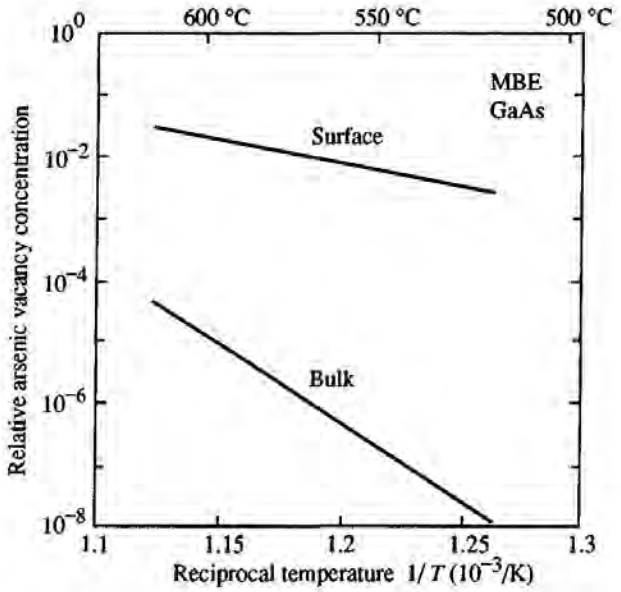


Fig. 7.1. Relative arsenic vacancy concentration on a GaAs surface and in GaAs bulk as a function of substrate temperature for a constant arsenic surface concentration (after Heckingbottom, 1985).

Surface chemistry of epitaxially grown III–V semiconductors plays a crucial role in impurity incorporation. The surface chemistry is determined by a number of factors, most importantly the abundance of adsorption (physisorption and chemisorption) products, the species of adsorption products, and the

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

concentration of group-III and group-V vacancies. In addition, the concentration of vacancies depends (i) on the V/III flux ratio and (ii) on the growth temperature. For GaAs, for example, the As coverage and the concentration of As vacancies depend strongly on the growth temperature. The vapor pressures of Ga, As_2, and As_4 on a GaAs surface were measured by Arthur (1967) and reviewed by Cho and Arthur (1975). The As_2 and As_4 vapor pressures are strongly temperature dependent. The vapor pressure of Ga is smaller than that of As at growth temperatures of 500-600°C. Both elements have the same pressure, when evaporating from a GaAs surface at about 630 °C, which is called the *congruent evaporation temperature*. As a result of the strong temperature dependence of the As vapor pressure, the concentration of As vacancies also depends strongly on temperature. The concentration of As vacancies is shown in Fig. 7.1 as a function of temperature (Heckingbottom, 1985). Figure 7.1 reveals that the As vacancy concentration increases by more than one order of magnitude for a temperature interval of 100 °C. Analogously, at a constant V/III flux ratio the As coverage decreases rapidly with increasing temperature. The temperature dependence of the As vacancy concentration will be used in the further analysis of impurity incorporation.

Impurities ideally incorporate only on substitutional sites. Any incorporation as defect, complex, or on an interstitial site is undesirable. We now consider the conditions that result in incorporation on substitutional sites. As an example, we consider an impurity I that incorporates in GaAs predominantly on a substitutional As site (I_{As}). We also consider the second possibility that the impurity incorporates on an interstitial site (I_{Int}) and neglect all other possibilities. The atomic reaction is then written as

$$\mathrm{Int} + \mathrm{I_{As}} \leftrightarrow \mathrm{I_{Int}} + \mathrm{V_{As}} \tag{7.6}$$

where Int and V_{As} represents an interstitial site and an As vacancy, respectively. Application of the law of mass action to Eq. (7.6) yields

$$\frac{[\mathrm{I_{Int}}][\mathrm{V_{As}}]}{[\mathrm{I_{As}}][\mathrm{Int}]} = c\mathrm{e}^{-\Delta E/kT}. \tag{7.7}$$

The concentration of interstitial sites in a crystal, [Int], is very large and is not significantly altered if some of the impurities occupy interstitial sites. We can therefore rewrite Eq. (7.7) as

$$[\mathrm{I_{Int}}]/[\mathrm{I_{As}}] = [\mathrm{V_{As}}]^{-1} c^{*} \mathrm{e}^{-\Delta E/kT} \tag{7.8}$$

where $c^{*} = c[\mathrm{Int}]$ is a constant.

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

337-TA-1366 RX-0255

Two important consequences can be inferred from Eq. (7.8). First, in order to incorporate impurities predominantly on As sites, a large As vacancy concentration is advantageous. This can be achieved in terms of a low V/III flux ratio during growth. This result can be generalized as follows: *The incorporation of impurities on group-V (group-III) sites can be enhanced by a low V/III (high V/III) flux ratio.* Low V/III flux ratios can be advantageous for the incorporation of e.g. S, Se, Te, and C which occupy anion sites. On the other hand, a high V/III flux ratio can be advantageous for e.g. Be-acceptors, Zn and Si which occupy cation sites. It is worthwhile noting that interstitials were considered in Eq. (7.6) as electrically inactive impurities. Similar equations can be written for other electrically inactive configurations such as complexes, impurity pairs, or other defects. Considering other non-active impurity configurations, one arrives at the same conclusion with respect to the V/III flux ratio.

The second consequence of Eq. (7.8) concerns the temperature dependence of the incorporation. Eq. (7.8) predicts that $[I_{Int}]/[I_{As}] \propto \exp(-\Delta E/kT)$, where ΔE is the change in free energy in the reaction of Eq. (7.6). In order to maximize the concentration of impurities on the desired As site, $[I_{As}]$, growth must occur at a low temperature. At low temperature, the exponential term $\exp(-\Delta E/kT)$ is small, i.e. the concentration of impurities on the desired As sites, $[I_{As}]$, is large (see Eq. (7.8)). Consider now the other extreme of very high growth temperatures. For such temperatures (which may be impractical for other reasons) the exponential term $\exp(-\Delta E/kT)$ approaches unity, i.e. the concentration of interstitial impurity atoms will be high. That is, the thermal energy kT is so large that a small energy gain ΔE does not represent a strong driving force for impurities to occupy the 'right' substitutional As site.

The law of mass action thus indicates that low growth temperatures are advantageous to drive impurities on the lowest energy site which is, in the above example, a substitutional site. The lower limit of growth temperatures is not arbitrary but given by kinetic limitations. That is, the growth temperature must be sufficiently high to enable impurities to diffuse on the crystal surface in order to find low energy sites. Growth temperatures that are too low would impose a kinetic limitation on the impurity motion and prevent the impurities from equilibrating at the lowest energy sites.

It has been a general experimental observation that the epitaxial growth of III–V semiconductors at very high temperatures favors a multitude of defects (see, e.g. Razeghi, 1989) including impurity-related defects. This tendency can be explained on the basis of the law of mass action. Most impurity-related defects have lower formation energies than the formation energy of a substitutionally incorporated impurity (less energy is gained by forming an

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

RX-0255.277

impurity-related defect compared to forming a substitutional impurity). Since the probability of defect creation is proportional to the exponential term of the law of mass action, $\exp(-\Delta E/kT)$, high temperatures will generally increase the formation of defects. Let us consider an example. Assume the difference in formation energy of a specific defect and a substitutional impurity, ΔE, equals 0.5 eV. Consider further the two growth temperatures $T_1 = 500$ K and $T_2 = 1000$ K. The probability of forming the defect at T_2 compared to T_1 increases by a factor of $\exp(\Delta E/kT_1 - \Delta E/kT_2) \cong 330$. That is, for the specific example illustrated above, the concentration of defects increases by about a factor of 330 at the higher of the two growth temperatures.

A further effect of high growth temperatures is the preferred reevaporation of the group-V constituent. The elements P, As, and Sb have a higher vapor pressure than their group-III counterparts. That is, the surface coverage of compound semiconductors with group-V constituents decreases rapidly at high growth temperatures. It can be shown by rate equations that the surface coverage of e.g. GaAs with As is inversely proportional to the As vapor pressure. Consequently, impurity surface reactions that are affected by the group-V element concentration on the surface are strongly influenced by the growth temperature via the group-V surface coverage.

As an example we consider the amphoteric impurity Si in GaAs and consider the exchange reaction on the surface

$$Si_{Ga} + V_{As} \leftrightarrow Si_{As} + V_{Ga} \tag{7.9}$$

where Si_{Ga} and Si_{As} represent Si on a Ga site (donor) and As site (acceptor), respectively; V_{As} and V_{Ga} represent an As and Ga vacancy, respectively. The Si acceptors (Si_{As}) compensate the Si donors (Si_{Ga}). This effect is called autocompensation. The ***autocompensation ratio*** is defined as the ratio of Si-acceptor concentration to Si-donor concentration, i.e. $[Si_{As}]/[Si_{Ga}]$. Application of the law of mass action yields

$$\frac{[Si_{As}][V_{Ga}]}{[Si_{Ga}][V_{As}]} = c \exp(-\Delta E/kT) . \tag{7.10}$$

This equation has two terms which strongly depend on temperature, which are $[V_{As}]$ and the exponential term. Both terms increase exponentially at higher temperatures. Since the Ga reevaporation rate is smaller, we assume, within a limited temperature interval, that $[V_{Ga}] \simeq$ const. Thus the temperature dependence of $[V_{As}]$ and of the exponential term result in increased autocompensation at high growth temperatures. Such an increased autocompensation has indeed been observed experimentally by Chai et al. (1981)

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

who found increased autocompensation of Si in GaAs at low V/III flux ratio and high growth temperatures. Qualitatively similar findings were made for Ge- and C-doped GaAs (Künzel et al., 1982; Hoke et al., 1991a and 1991b). A further discussion of autocompensating impurities is given in Sect. 7.3.

The requirement of low growth temperatures is especially important at high doping concentrations. To illustrate the importance of low growth temperatures we consider again the law of mass action and attribute the energy difference, ΔE, to the difference in formation energy between the desired *substitutional* configuration and an undesired, electrically *non-active* configuration, e.g. an impurity complex. (Entropy is neglected since it is usually a small part of the free energy). Next, it will be shown that the difference in formation energy between the two configurations decreases at very high doping concentrations. As mentioned previously, the formation energy is composed of several parts including chemical bond energies, mechanical energies, and electronic energies. Due to band-filling, the electronic energy contribution to the formation energy increases with the free carrier concentration. Electrons entering the conduction band must be excited to the (degenerate) Fermi level rather than to the bottom of the conduction band. That is, electrons (in an n-type semiconductor) require an energy of (see Chap. 3)

$$E_F - E_D \cong E_F - E_C \cong \frac{\hbar^2}{2m^*}(3\pi^2 n)^{2/3} \tag{7.11}$$

where E_F, E_D, and E_C are the Fermi energy, donor level energy, and conduction band edge energy, respectively. The expression on the right-hand side of Eq. 7.11 is valid for extreme degeneracy in an isotropic parabolic conduction band. Equation 7.11 shows that the electronic energy increases with the free carrier concentration with $n^{2/3}$. At high electron concentrations, the Fermi energy can be several hundred meV above the bottom of the conduction band. Thus, the electronic energy required for incorporating a donor increases at high doping concentrations due to the increase in Fermi energy. See Sect. 7.4 for a more detailed discussion of the electronic energies. A second factor which may increase the energy required to incorporate a substitutional impurity at high doping concentrations is elastic strain energy. High doping concentrations tend to increase the elastic strain which can result in an increase in energy to incorporate substitutional impurities (see Chap. 7.5).

The decrease in the difference of formation energies, ΔE, makes electrically non-active impurity defects more likely. In order to reduce the probability of defect formation, low growth temperatures are required, especially at high doping concentrations. This is because the impurity defect is formed with a probability

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

337-TA-1366 RX-0255

$\propto \exp(-\Delta E/kT)$ which is smallest at low temperatures. It is, however, important to keep in mind that the growth temperature cannot be arbitrarily low. The growth temperature must be sufficiently high to enable the growing semiconductor surface to reach thermal equilibrium.

Another aspect of low growth temperatures is the higher group-V coverage occurring at low growth temperature due to the strong temperature dependence of the vapor pressure of group-V elements. Group-VI donor impurities (S, Se, Te) which occupy group-V sites, compete with anions. Thus, low growth temperatures may affect the efficient group-VI impurity incorporation due to the high group-V atom coverage at low growth temperatures.

It has been found experimentally for many impurities that the highest electrically active doping concentrations are indeed achieved at low growth temperatures. Chai et al. (1981) found that the highest free carrier concentrations in Si-doped GaAs are obtained at low growth temperatures. Under optimized low temperature conditions, free carrier concentrations as high as 1.6×10^{19} cm^{-3} were reached in Si-doped GaAs (Malik et al., 1992). Hamm et al. (1989) reported very high Be doping levels in $Ga_{0.47}In_{0.53}As$ at growth temperatures as low as 366 °C. Similar results were reported for Sn-doped $Ga_{0.47}In_{0.53}As$ (Kawaguchi and Nakashima, 1989; Panish et al., 1990a).

The incorporation of impurities has been modeled in terms of simple kinetic rate equations. Wood et al. (1982), and Poole et al. (1988) took into account an impurity arrival rate on the epilayer (per cm^2 per s), J_I, an impurity surface density, C_I, an impurity desorption rate from the surface, DC_I, and an incorporation rate into the epitaxial layer, KC_I. The change in surface concentration is then given by

$$\frac{dC_I}{dt} = J_I - DC_I - KC_I . \tag{7.12}$$

The authors further assumed that the impurity desorption rate is thermally activated, i.e. $D = D_0 \exp(-E_a/kT)$ where E_a is the activation energy for desorption. The incorporation constant was assumed to be proportional to the growth rate, ν_g, according to $K = K'\nu_g$, where K′ is a constant.

Under steady-state equilibrium conditions, the concentration of impurities on the surface does not change, i.e. $dC_I/dt = 0$. Under such steady state conditions one obtains (Poole et al., 1988)

$$J_I = C_I D_0 \, e^{-E_a/kT} + K' C_I \nu_g \tag{7.13}$$

which allows one to calculate the incorporated impurity concentration in the growing semiconductor according to

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

RX-0255.280

$$N_I = (J_I/v_g)/[1 + (D_0 e^{-E_a/kT}/K' v_g)] \quad (7.14)$$

where J_I/v_g is the maximum impurity concentration achievable at the growth rate v_g and the impurity flux J_I. In order to achieve the highest doping concentration J_I/v_g, the desorption rate must be zero.

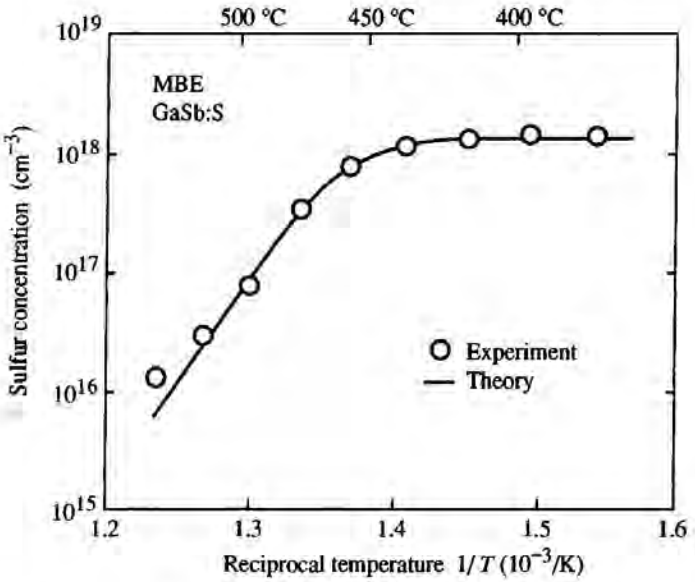


Fig. 7.2. Experimental and theoretical S concentration in GaAs grown by MBE as a function of reciprocal growth temperature. The theoretical curve is obtained from a set of rate equations (after Poole et al., 1988).

Poole et al. (1988) calculated the incorporation of sulfur into GaSb using the rate equation derived above. The result of their experimental and calculated S concentration is shown in Fig. 7.2 as a function of the reciprocal growth temperature. Experimental and theoretical points are in agreement. The parameters used for the calculation were as follows. Maximum impurity concentration (without desorption): $J_I v_g = 1.3 \times 10^{18}$ cm^{-3}, $(D_0/K' v_g) = 1.24 \times 10^{24}$, and $E_a = 3.5$ eV. The authors pointed out that the choice of the activation energy for S reevaporation of $E_a = 3.5$ eV is similar to the value of $E_a = 3.1$ eV found for S reevaporation from the GaAs surface reported by Andrews et al. (1983).

7.1.2 Impurity incorporation in OMVPE

The incorporation of doping impurities in OMVPE is more complicated as compared to MBE. It was shown in the previous section that the rate equation for

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

dopant incorporation consists of different terms, which include the arrival rate, the reevaporation rate, and the incorporation rate. Although the basic rate equation is also valid for OMVPE, the process of doping incorporation is more complicated. In OMVPE, doping precursors are chemical *compounds* rather than doping *elements* as in MBE. The chemical doping compounds in OMVPE may, for example, not react efficiently and may not decompose if they are very stable molecules. Furthermore, the incorporation of doping impurities may depend on the nature of the majority constituents, i.e. the group-III and group-V compounds. Parasitic reactions between the majority constituents and the minority constituents (i.e. doping precursors) can influence the doping incorporation as well.

The ***thermodynamic distribution coefficient*** K_t is used to quantitatively describe the efficiency of dopant incorporation. The distribution coefficient is defined as the ratio of the mole fraction of the doping impurity in the semiconductor and the partial pressure of the doping precursor in the gas phase of the reactor at the gas–semiconductor interface, i.e. (Stringfellow, 1985, 1986, 1989)

$$K_t = x_D^s / P_D^i \tag{7.15}$$

where the superscripts s and i refer to 'solid' and 'interface', respectively. The doping composition x_D^s is dimensionless and is the fraction of sublattice sites that are occupied by a doping impurity atom. The thermodynamic distribution coefficient has thus the dimension of reciprocal pressure. In addition to the thermodynamic distribution coefficient given by Eq. 7.15, the compositional distribution coefficient is frequently used (see, for example, Kisker and Zawadzki, 1988). The ***compositional distribution coefficient*** K_c is simply the ratio of the doping mole fraction in the solid semiconductor and of the doping precursor in the gas phase of the reactor, i.e.

$$K_c = x_D^s / x_D^g = x_D^s / (P_D^g / P_{III}^g) \ . \tag{7.16}$$

Let us be more specific about the quantities x_D^s and x_D^g. For dopants occupying group-III sites, x_D^s is the fraction of group-III sites in the solid occupied with doping impurities. An analogous statement is valid for impurities occupying group-V sites. For dopants occupying group-III or group-V sites, x_D^g is the ratio of doping precursor and group-III precursor molecules. The x_D^g is thus the ratio of partial pressures of the doping and the group-III precursor. Note that K_c is dimensionless. Doping precursors with $K_c << 1$ do not readily incorporate. For example, C does not readily incorporate from CH_4 due to the thermal stability of the methane molecule. Doping impurities with $K_c >> 1$ readily incorporate into

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

the semiconductor. A distribution coefficient of $K_c \cong 1$ is desirable for the sake of process controllability. The use of the thermodynamic versus the compositional distribution coefficient has advantages and disadvantages. The *compositional* distribution coefficient can be easily determined from the initial source gas doping composition and the doping concentration in the semiconductor. However, the compositional distribution coefficient depends on many other parameters such as pressure, fluid dynamics, properties of the stagnant layer, etc. The thermodynamic distribution coefficient is more appropriate to rigorously model the doping incorporation during OMVPE. The *thermodynamic* distribution coefficient depends on the chemical processes at the interface only. However, it is more complicated to determine the true thermodynamic distribution coefficient experimentally.

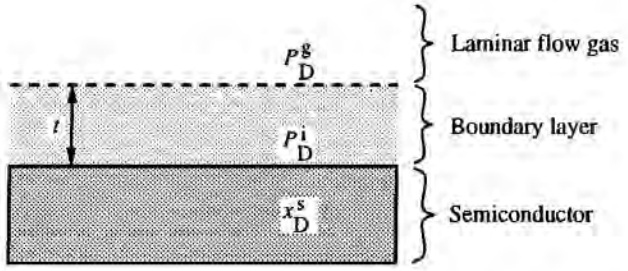


Fig. 7.3. Schematic illustration of the laminar flow region, the stagnant boundary layer of thickness t, and the epitaxial layer grown by OMVPE. The partial pressures of the doping precursor are P_D^g and P_D^i in the gas phase and at the interface, respectively. The doping mole fraction in the solid is x_D^s.

A further complication of the OMVPE doping is given by the *fluid dynamical processes* occurring in OMVPE reactors. These fluid dynamical properties result in a stagnant gas layer in the vicinity of the growing semiconductor surface. The stagnant boundary layer was discussed and illustrated in Chap. 6. The transport of doping precursor molecules and of majority precursor constituents is of clear interest for the modeling of doping incorporation. Stringfellow (1985) considered purely diffusive transport of precursor molecules through the stagnant boundary layer. It was further assumed that no turbulences occur in the source gases, the boundary layer or at their interface. The diffusion transport flux through the boundary layer of thickness t, schematically illustrated in Fig. 7.3, can be expressed as

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

$$j = \frac{D_D(P_D^g - P_D^i)}{RTt} \tag{7.17}$$

where D_D is the diffusion coefficient of the doping precursor, R is the gas constant, and T is the temperature. The partial pressures of the doping precursor are P_D^g in the gas phase above the stagnant boundary layer and P_D^i at the interface between the boundary layer and the semiconductor.

Pyrolysis of doping precursors or of majority constituents can either occur in or near the boundary layer or on the growing semiconductor substrate during OMVPE. Tsang (1989) pointed out that this is in contrast to CBE where the pyrolysis occurs only at the epitaxial crystal surface. The lack of knowledge of the exact location of pyrolysis further complicates the modeling of the doping incorporation process in OMVPE.

Stringfellow (1989) developed a model for the incorporation of a single impurity species in a binary III–V semiconductor and assumed that (i) the group-III element determines the growth rate and that the group-III element is depleted at the semiconductor–gas interface, (ii) that the V/III flux ratio is larger than unity, and (iii) that the diffusion coefficients of all gaseous precursor molecules through the boundary layer are equal. These assumptions allow one to obtain on the basis of Eq. (7.17) a simple expression for the doping concentration in the semiconductor, i.e.

$$x_D^s = \frac{j_D}{j_{III}} = \frac{P_D^g - P_D^i}{P_{III}^g} \tag{7.18}$$

where j_D and j_{III} are the fluxes of the doping precursor and the group-III precursor through the stagnant boundary layer, respectively. Note that the validity of Eq. (7.18) requires complete pyrolysis of all constituents. Inserting the thermodynamic distribution coefficient (see Eq. (7.15)) into Eq. (7.18) yields

$$x_D^s = \frac{P_D^g - (x_D^s/K_t)}{P_{III}^g} \tag{7.19}$$

where thermodynamic equilibrium conditions are assumed. Stringfellow (1989) differentiated between two cases.

In the first case, the vapor pressure of the impurity is low and all dopants reaching the crystal surface are incorporated, i.e. the dopant is depleted at the interface $P_D^i << P_D^g$. Equation (7.18) then yields the simple expression

$$x_D^s = P_D^g/P_{III}^g \,. \tag{7.20}$$

In this case the dopant distribution coefficient is inversely proportional to P_{III}^g; i.e.

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

inversely proportional to the growth rate. This case is called the *mass-transport-limited* case, since the incorporation of dopants is limited by the transport of dopants through the boundary layer.

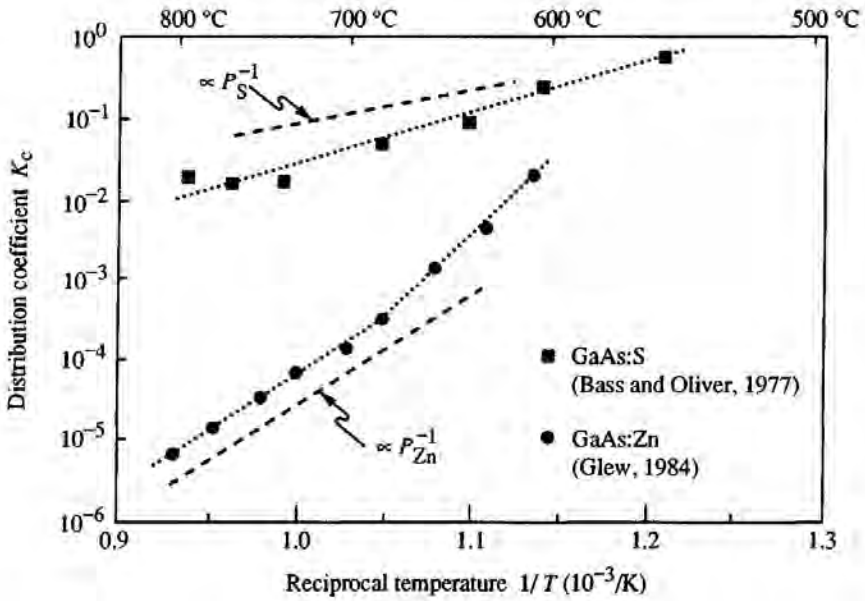


Fig. 7.4. Compositional distribution coefficient versus reciprocal growth temperature for S- and Zn-doped GaAs grown by OMPVE. The dotted lines represent a fit to the experimental data. The dashed lines are the reciprocal vapor pressures of S and Zn in arbitrary units (after Stringfellow, 1986).

In the second case, the dopant vapor pressure is high and most of the dopant is reevaporated from the epitaxial crystal surface. For this condition, $P_D^i \equiv P_D^g$ and one obtains from Eq. (7.15)

$$x_D^s = K_t P_D^g \,. \tag{7.21}$$

It is of practical value to compare Eq. (7.21) with the compositional distribution coefficient of Eq. (7.16). Elimination of x_D^s from the two equations yields

$$K_c = K_t P_{III}^g \tag{7.22}$$

which is frequently used to describe the incorporation of doping impurities. In the second case, the thermodynamics at the epitaxial crystal-to-gas interface determines the incorporation of doping impurities. Furthermore, the distribution coefficient is independent of P_{III}^i and the growth rate. The dependence of the dopant incorporation on the group-V partial pressure was also considered

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

337-TA-1366 RX-0255

(Stringfellow, 1986, 1989).

Doping impurities with a high vapor pressure tend to reevaporate before they are incorporated into the semiconductor. Since the vapor pressure depends on temperature, the distribution coefficient of such dopants decreases with increasing temperature. It is obvious that the distribution coefficient of such impurities is inversely proportional to the vapor pressure of the impurity element. Experimental data of Glew (1984) and Bass and Oliver (1977) on the Zn and S dopant incorporation in GaAs were used by Stringfellow (1986) for a comparison between the distribution coefficients of these dopants and their vapor pressure. The comparison in shown in Fig. 7.4. The dotted lines are a fit to the experimental points. The dashed lines reflect the temperature dependence of the inverse vapor pressure. Comparison of the dotted (experimental) and dashed (theoretical) lines in Fig. 7.4 reveal that the two functional dependences are in good qualitative agreement. The agreement suggests that the incorporation of high vapor pressure doping elements such as Zn and S are indeed determined by reevaporation of the impurities from the epitaxial crystal surface. However, the kink in the experimental data of Zn indicates that the vapor pressure is not the only quantity controlling the doping incorporation.

7.2 Doping incorporation by implantation

Ion implantation is a powerful technique for incorporating doping impurities into III–V semiconductors. In this technique, charged impurity atoms are accelerated in an electric field and penetrate into the semiconductor target. Typical acceleration energies range from 10 keV to 500 keV. Typical doses of impurity atoms implanted in III–V semiconductors range from 10^{11} cm^{-2} to 10^{15} cm^{-2}. The scattering processes of impurity ions in the solid cause considerable structural damage which results in a small electrical activity of implanted impurities. Subsequent high-temperature anneals of the semiconductors reduce the damage and can increase the electrical efficiency of the impurity ions to near 100%. Typical annealing cycles are 700 to 1100 °C for 5 to 100 s. Ion implantation is also used to form highly resistive regions in semiconductors using hydrogen ions (protons) and oxygen ions. Hydrogen implantation leaves defect traces which, if not annealed, act as deep traps. Oxygen forms deep levels in III–V semiconductors and renders the material highly resistive. Several reviews have become available on ion implantation in III–V semiconductors and post-implantation annealing (see, for example, Anholt et al., 1988; Gill and Sealy, 1986; Sharma, 1989).

Ion implantation makes possible controlled doping with lateral selectivity. The lateral selectivity can be achieved by either scanning and blanking the ion

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

Innoscience1366-00138959

RX-0255.286

beam or by implantation masks. The use of implantation masks allows one to obtain lateral doping distributions that are self-aligned with respect to the edge of the mask. Furthermore, the implantation dose and implantation depth can be precisely controlled. Excellent uniformities can be obtained over large areas of the semiconductor. A drawback of doping by ion implantation is the inherent width of the doping profiles in longitudinal and lateral direction. The finite width of the doping distribution is caused by the statistical nature of scattering processes of doping ions with lattice atoms (see, for example, Feldman and Mayer, 1986). Abrupt doping profiles are not feasible at high acceleration energies, i.e. large implantation depths.

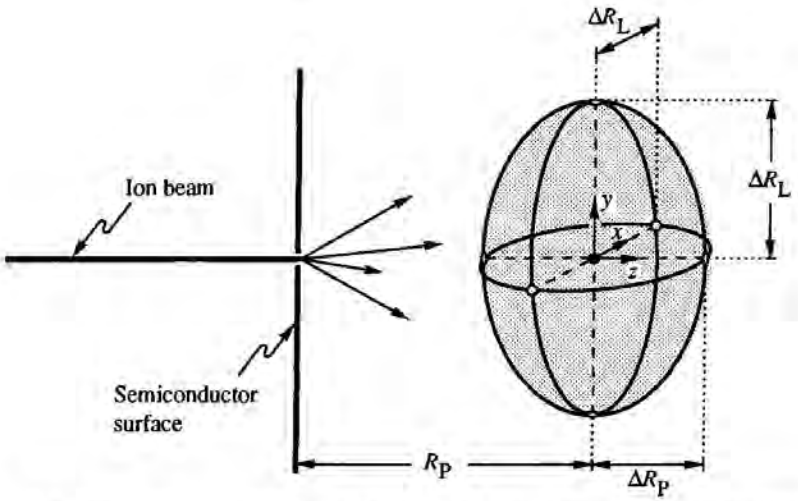


Fig. 7.5. Schematic illustration of the impurity distribution resulting from a focused ion beam propagating along the z-axis. The projected range of the implantation is R_P and the projected straggle is ΔR_P. The lateral straggle (along the x- and y-directions) is ΔR_L.

The spatial distributions of impurities implanted into the semiconductor can be described, in the simplest approximation, by gaussian distributions (see, for example Pickar, 1975). The basic implantation geometry of an infinitesimal small beam diameter is shown in Fig. 7.5. Ions propagate along the z-direction and enter the semiconductor at the origin of the cartesian coordinate system. The mean depth of the impurities is called the ***projected range***, R_P. The position standard deviation of the doping distribution in the projected and the lateral direction are ΔR_P and ΔR_L, respectively; ΔR_P and ΔR_L are called ***projected straggle*** and ***longitudinal straggle***, respectively. The quantities are illustrated in Fig. 7.5. The distribution of ions implanted into a semiconductor is thus given by

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

337-TA-1366 RX-0255

(Pickar, 1975)

$$N(x,y,z) = \frac{N_t}{(2\pi)^{3/2}\Delta R_P \Delta R_L^2} \exp\left[-\left(\frac{z - R_P}{\sqrt{2}\,\Delta R_P}\right)^2\right] \exp\left[-\frac{x^2 + y^2}{(\sqrt{2}\,\Delta R_L)^2}\right] \tag{7.23}$$

where $N(x,y,z)$ is the ion concentration (per cm^3) and N_t is the total number of implanted ions. Thus $\int\int\int N(x,y,z)\,dxdydz = N_t$. Note that the distribution of Eq. (7.23) can be expressed in the cylindrical coordinates r and z with $r^2 = x^2 + y^2$.

If the ion beam is scanned uniformly over a wafer, the resulting impurity distribution depends only on the z-direction not on the x- and y-directions. The doping concentration is then given by

$$\boxed{N(z) = \frac{N^{2D}}{\sqrt{2\pi}\,\Delta R_P} \exp\left[-\left(\frac{z - R_P}{\sqrt{2}\,\Delta R_P}\right)^2\right]} \tag{7.24}$$

where N^{2D} is the implantation dose, i.e. the density of implanted ions (per cm^2).

The gaussian function of Eq. (7.24) is a simple approximation for true doping distribution. The two parameters R_P and ΔR_P, which are the median and the standard deviation of the gaussian distribution, can be determined from measured impurity depth profiles of ion implanted semiconductors. Secondary ion mass spectrometry is frequently used for such an impurity depth profiling. The depth profiling yields a measured value for the impurity concentration, $N(z)$, which in turn allows one to determine the projected range and projected straggle. The median and the standard deviation of a given sample are calculated according to (see, for example, Kreyszig, 1972; Ryssel and Ruge, 1986; Ryssel and Biersack, 1986; Selberherr, 1984)

$$R_P = \langle z \rangle = (N^{2D})^{-1} \int_0^\infty z\,N(z)\,dz\,, \tag{7.25}$$

and

$$\Delta R_p = \sqrt{\langle z^2 \rangle - \langle z \rangle^2} \tag{7.26a}$$

where

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

RX-0255.288

$$\langle z^2 \rangle = (N^{2D})^{-1} \int_0^\infty z^2 \, N(z) \, dz \,, \tag{7.26b}$$

and $\langle z \rangle^2$ is obtained by taking Eq. (7.25) to the second power.

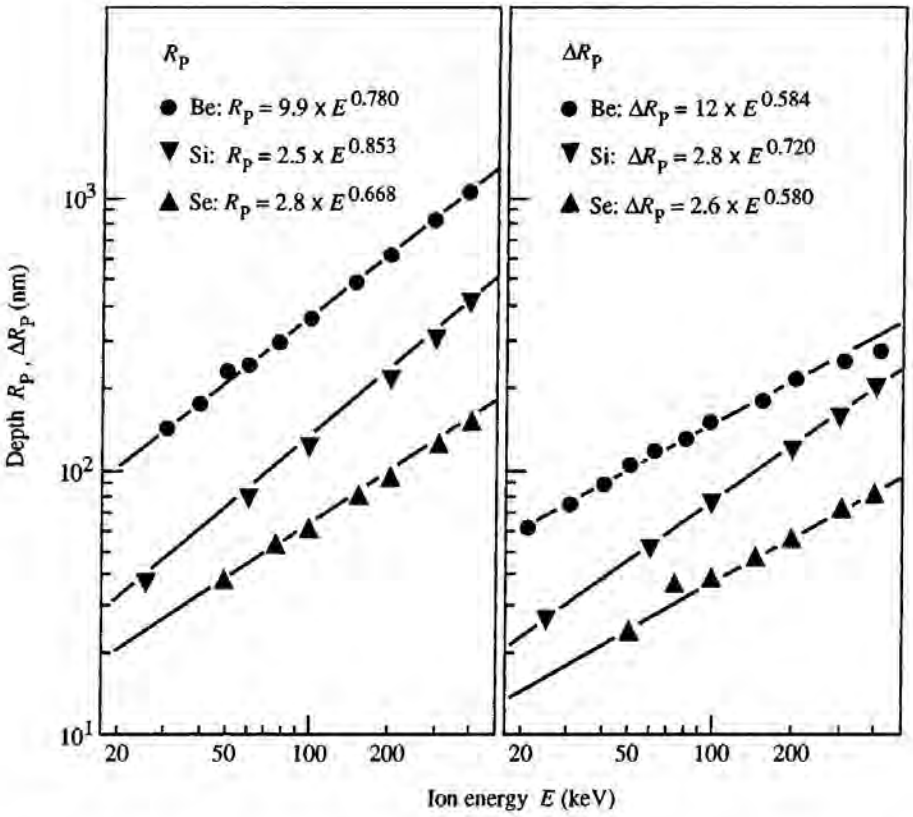


Fig. 7.6. Measured projected range (R_p) and projected straggle (ΔR_p) for Be, Si and Se implanted into GaAs at ion energies ranging from 20 keV to 400 keV. The six equations for R_p and ΔR_p are power-law fits in which R_p, ΔR_p and E are measured in nm and keV, respectively. The impurities with a dose of 1×10^{13}cm^{-2} were implanted into bare wafers at a tilt and rotation angle of 7° and 9°, respectively (after Anholt et al., 1988).

Experimental values for the projected range and the projected straggle of Be, Si, and Se ions implanted into GaAs are shown in Fig. 7.6 as a function of the ion energy (Anholt et al., 1988). The projected range and the projected straggle are found to follow a power-law dependence. The equations shown in Fig. 7.6 are valid if the units of the energy E are in keV and the units of R_P and ΔR_P are in nm. The projected straggle is typically 0.2 to 0.5 times the projected range.

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

Impurity profiles of ion-implanted III–V semiconductors are frequently asymmetric (i.e. non-gaussian) and skewed towards or away from the semiconductor surface. The skewedness of the implanted profiles is characterized by the parameter γ (Anholt et al., 1988) given by

$$\gamma = \langle (z - R_p)^3 \rangle / \Delta R_p^3 . \tag{7.27}$$

Positive values of γ of the order of unity and larger describe distributions decreasing approximately exponentially at large depths. Negative values of γ describe distributions skewed towards shallower depths.

Monte Carlo computer simulations were employed to calculate the distribution of ions implanted in semiconductors (see, for example, Biersack and Haggmark, 1980; Biersack and Ziegler, 1982; Cristel et al., 1980; Gibbons et al., 1975; Ziegler et al., 1985). The numerical Monte Carlo calculations are performed due to the lack of reliable analytic formulas for the ion distribution in implanted semiconductors. The traces and final positions of 25 H^+ ions implanted into GaAs at an ion energy of 50 keV are shown in Fig. 7.7. The plot was obtained using the TRIM Monte Carlo Code (Ziegler et al., 1991).

If the Monte Carlo calculation shown in Fig. 7.7 is performed with several thousand ions, a reliable ion distribution is obtained. The distributions of Si and Be in GaAs and InP are shown in Fig. 7.8 for implantation energies ranging from 10 keV to 500 keV. 50 000 ions were used for each distribution calculated by TRIM (Ziegler et al., 1991). The implantation depth increases with decreasing atomic number for the same implantation energy. Si and Be are popular doping impurities used for n-type and p-type doping, respectively. Hydrogen implantation (see Chap. 9) is used to render the semiconductor highly resistive. Such proton bombardment is frequently used, for example, in semiconductor lasers to confine the current flow to a well-defined area.

Donor and acceptors implanted into III–V semiconductors are not fully activated after the implantation. The lack of activation is due to (i) damage introduced during the implantation process and (ii) non-substitutional sites occupied by the donors and acceptors. The degree of activation is characterized by the ***activation efficiency*** η_a. The activation efficiency is defined as the ratio of free carrier density (n, p) and implanted ion density (N_D, N_A), that is

$$\eta_a = \frac{n^{2D}}{N_D^{2D}}; \quad \eta_a = \frac{p^{2D}}{N_A^{2D}} \tag{7.28}$$

for n-type and p-type implantation, respectively. The superscript 2D (two-dimensional) refers to carrier and ion densities per unit area. The activation efficiency after implantation but before annealing can be as low as 10%.

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

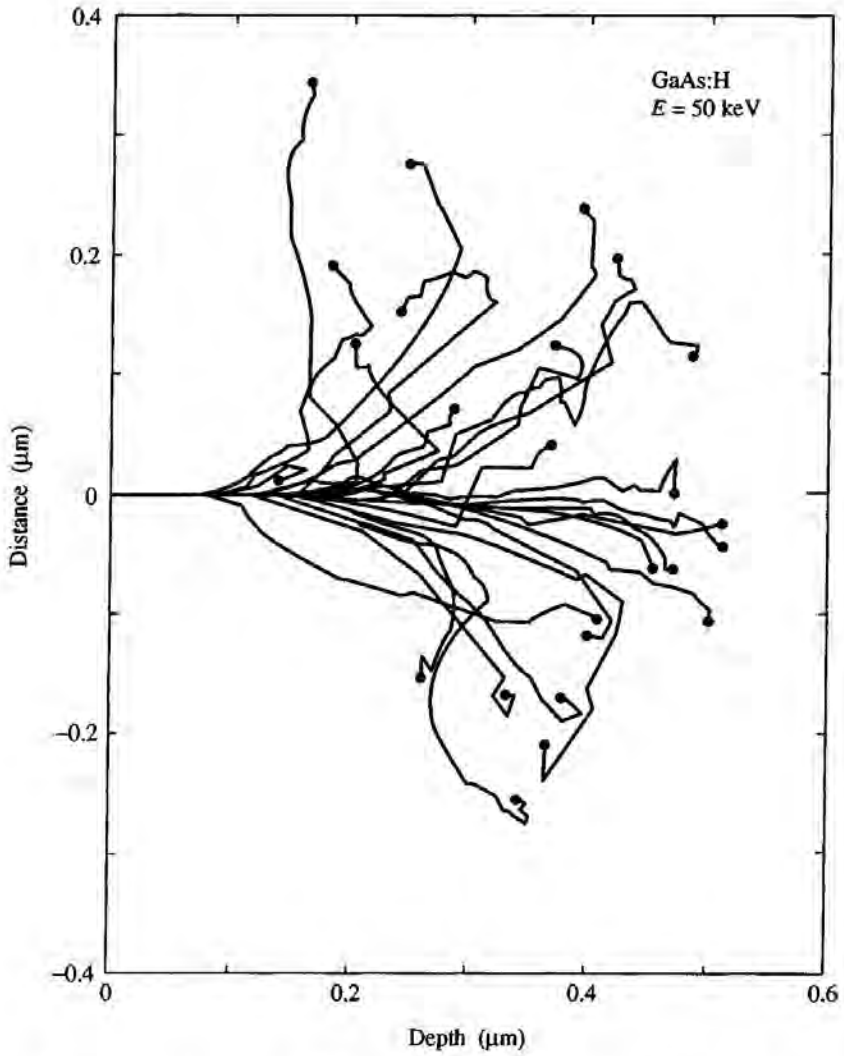


Fig. 7.7. Collision cascades of 25 hydrogen ions in GaAs involving many elastic and inelastic scattering events with the GaAs lattice. All hydrogen ions enter the lattice at the same point.

Ion implantation is usually performed with the implant target at room temperature. However, implantation at higher temperatures may be advantageous for a high donor and acceptor activation. Woodhouse et al. (1984) reported improved activation of Se in InP for an implantation target heated to 180 °C as compared to room temperature implantation.

Post-implantation annealing is used to reduce the concentration of defects and to increase the dopant activation. Rapid thermal annealing, i.e. a short, high-

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

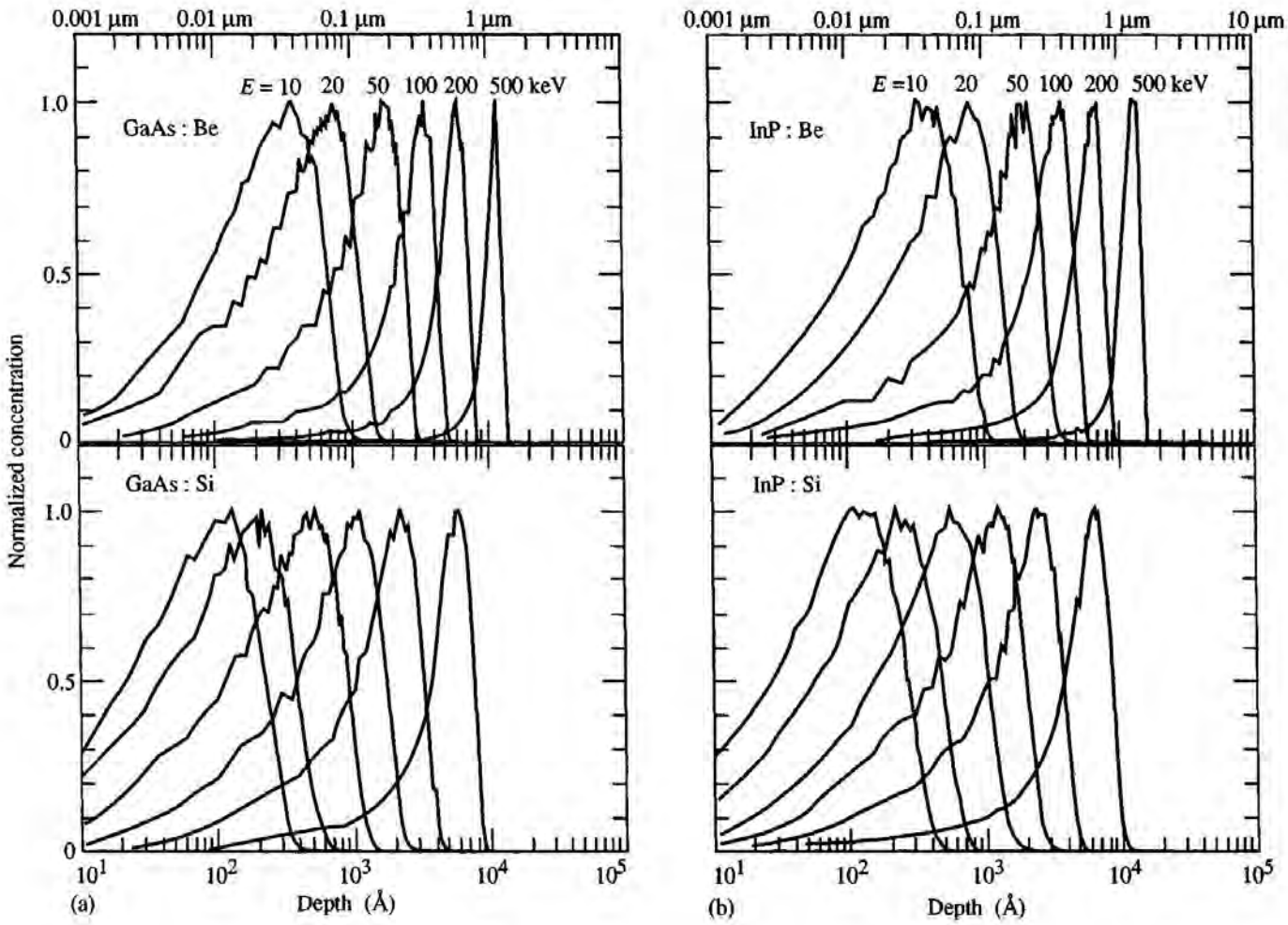


Fig. 7.8 Calculated distributions of Be and Si ions implanted into (a) GaAs and (b) InP at different implantation energies (TRIM '91; Ziegler et al., 1991).

Innoscience1366-00138965

temperature annealing cycle is the preferred method. Rapid thermal annealing results in less dopant redistribution as compared to furnace annealing (see, for example, Vescan et al., 1982; Maier and Selders, 1986). The annealing atmosphere is usually a gas with reducing characteristics such as forming gas (85% N_2/15% H_2). The annealing cycle can also be performed in an atmosphere with excess As or P pressure, in order to avoid the preferential evaporation of anions (As, P, or Sb) from the III–V semiconductor surface. In addition to rapid thermal annealing, furnace annealing and laser annealing (see, for example, Ferris et al., 1979) have been employed for implanted semiconductors. Furnace annealing yields lower activation efficiencies as compared to rapid thermal annealing. Laser annealing is mostly used for annealing of implants in Si at very high temperatures. In III-V semiconductors, laser annealing has not been used extensively.

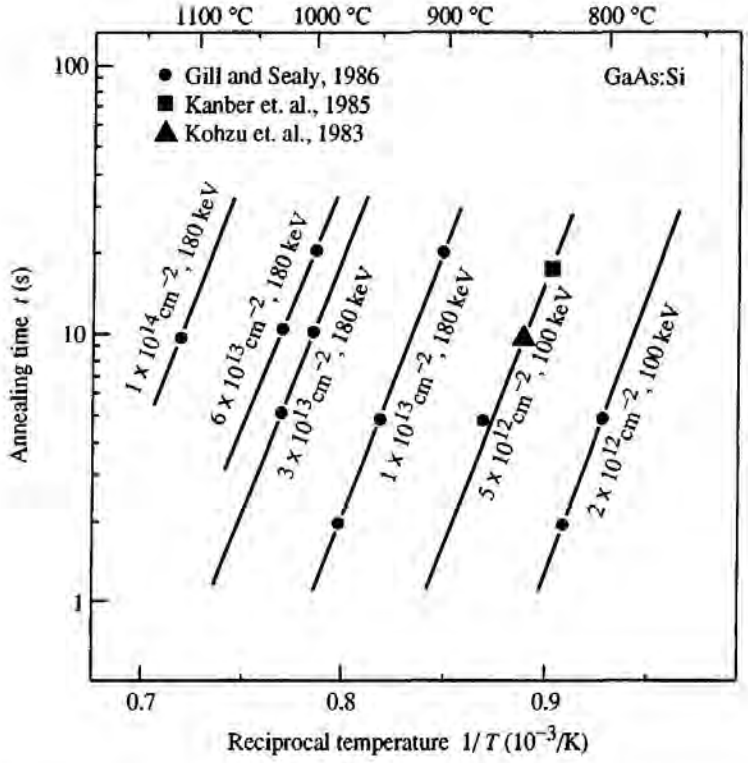


Fig. 7.9. Optimum annealing times and temperatures (for high activation and low diffusion) for Si-implanted GaAs with doses of 2 x 10^{12}cm^{-2} to 1 x 10^{14}cm^{-2} and implantation energies of 100 keV to 180 keV (after Gill and Sealy, 1986, reprinted by permission of the publisher, The Electrochemical Society, Inc.).

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

337-TA-1366 RX-0255

The annealing parameters, i.e. temperature and time, must be optimized in order to achieve maximum activation efficiency of implanted impurities. It is desirable to select a long annealing time and a high temperature, in order to anneal out the damage created by implantation and to maximize the activity of the dopant. On the other hand, short annealing times and low temperatures avoid excessive diffusion of the doping impurity. The optimum annealing times for GaAs implanted with Si donors at different implantation energies and doses are shown in Fig. 7.9 (Gill and Sealy, 1986). The plot also contains data points of Kanber et al. (1985) and Kohzu et al. (1983). Inspection of Fig. 7.9 shows that the optimum annealing temperature increases with both the implantation energy and the implantation dose. The requirement of higher temperature can be qualitatively explained by the larger damage caused by higher implantation energies and doses. Figure 7.9 further illustrates that optimum annealing times range from 2 to 20 s for GaAs:Si.

Dopants used for intentional p- and n-type doping have different characteristics which depend on the chemical nature of the implanted impurities. Sharma (1989) summarized the characteristics of Si, Se, Be, Mg, and Hg in InP. In the following, the characteristics of these dopants are summarized. The reader is also referred to Chap. 5 for further discussion.

Silicon is used as an n-type dopant in many III–V semiconductors. Si, as are all group-IV impurities, is amphoteric. The Si activation efficiency prior to annealing is low. A significant increase of the activity occurs during annealing (Bahir et al., 1989). Rapid thermal annealing is the preferred activation process in order to reduce anion out-diffusion. Furthermore, rapid thermal processing was reported to reduce the Si diffusion tail (Krautle, 1988). Relatively little redistribution is generally observed for Si-implanted GaAs and InP during annealing (Bahir and Merz, 1988).

Selenium (Se) is not amphoteric (in contrast to Si), and has a high activation efficiency. Peak electron concentrations $>10^{19}$ cm^{-3} were obtained in InP (Donnelly and Hurwitz, 1980; Gill and Sealy, 1984). On the other hand, Se is a heavier atom than Si and causes more damage during the implantation process. Hot implants were used to minimize the generation of damage during implantation (Woodhouse et al., 1984). Furthermore, Se is a faster diffuser than Si which results in broader doping profiles after annealing.

Beryllium is a popular p-type dopant used for implantation. Be is a stable impurity at low and intermediate doses ($\leq 10^{13}$ cm^{-2}). At high doses (e.g. 10^{14} cm^{-2}), strong redistribution occurs during annealing in GaAs and InP (see, for example, Wilson et al., 1988). Furthermore, Be is a small atom and causes less damage as compared to other p-type dopants (Sharma, 1989). It is interesting

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

RX-0255.294

to note that As and P co-implantation results in a higher electrical activation of Be (Wang, 1987). A higher anion concentration reduces the probability of Be occupying anion sites, i.e. enhances the probability of substitutional, active Be on cation sites.

Magnesium has been used as a p-type dopant in III–V semiconductors. It is heavier than Be and therefore causes more damage. High temperatures during implantation help to reduce the damage incurred during Mg implantation (Inada et al., 1981).

Ion implantation is used for a wide range of purposes in electronic III–V semiconductor structures and devices including (i) implantation for ohmic contact formation (ii) self-aligned implantation in field-effect transistors to reduce source and drain parasitic resistances, (iii) implantation of the electron-channel in field-effect transistors, (iv) formation of pn-junctions in junction field-effect transistors, (v) base regions in hetero and homo bipolar transistors, and (vi) highly conductive sub-collector regions in bipolar transistors. In photonic III–V semiconductor devices, ion implantation is frequently used to create highly resistive regions for current confinement. Oxygen and hydrogen ions (protons) are used for such isolation implantations.

7.3 Doping incorporation by diffusion

Doping impurities can be incorporated into III–V semiconductors by diffusion from a surface reservoir into the bulk. This method had great importance in the infancy of III–V device processing (see, for example, Weimer, 1962; Heime, 1967). However, more powerful doping techniques such as doping during epitaxial growth or by ion implantation rendered doping incorporation by solid-state diffusion less important. The major drawback of doping incorporation by diffusion are the limited possibilities for determining the shape of the doping profile; the concentrations of doping profiles created by diffusion are highest at the surface and continuously decrease into the bulk. Doping profiles with near arbitrary shape oppose the very nature of the diffusion process. Thus, doping incorporation by diffusion will be discussed in brevity due to the limited freedom offered by this doping technique and due to its limited use in present III–V processing technology.

The basic experimental configuration for impurity diffusion is shown in the inset of Fig. 7.10. The dopant reservoir can be considered as infinite, since it is assumed that only a minor fraction of doping impurities diffuses into the semiconductor. The fickian diffusion equation (see Chap. 8) must be solved for

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

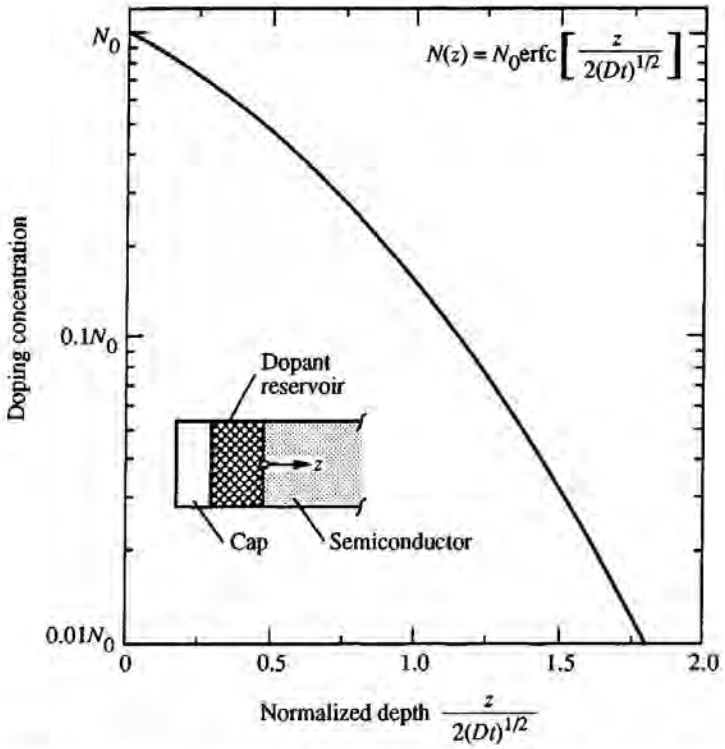


Fig 7.10. Complementary error function (erfc) describing the diffusion of impurities with surface concentration N_0 from an 'infinite' reservoir into a semiconductor. The inset shows an experimental layer sequence consisting of protective cap, dopant reservoir, and semiconductor.

$$N(z) = N_0 \quad \text{for } z<0 \text{ and } t \gtrless 0\,, \tag{7.29a}$$

$$N(z) = 0 \quad \text{for } z>0 \text{ and } t<0\,. \tag{7.29b}$$

Note that the concentration of impurities on the surface equals N_0 independent of the time. The solution of the fickian diffusion equation for these boundary conditions is

$$\boxed{N(z) = N_0 \text{ erfc}\left(\frac{z}{2\sqrt{Dt}}\right)} \tag{7.30}$$

where erfc is the complementary error function and D is the diffusion coefficient of the doping impurity in the semiconductor. The complementary error function can be expressed in terms of an integral

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

$$\text{erfc}(\alpha) = 1 - \text{erf}(\alpha) = \frac{2}{\sqrt{\pi}} \int_{\alpha}^{\infty} e^{-t^2}\, dt \qquad (7.31)$$

where $\alpha = \frac{1}{2}z/\sqrt{Dt}$ and erf is the error function. The integral of Eq. (7.31) can be easily solved numerically. The doping impurity profile of the erfc function is shown in Fig. 7.10 on a normalized depth scale. Note that the complementary error function decreases with depth more rapidly than the exponential function. Frequently, the square root $\sqrt{Dt}$ is called the ***diffusion length***, L_D. At the depth of $z = L_D = \sqrt{Dt}$, the abscissa of Fig. 7.10 has a value of $\frac{1}{2}z/\sqrt{Dt} = \frac{1}{2}$ at which the concentration has dropped to approximately half its interface value N_0.

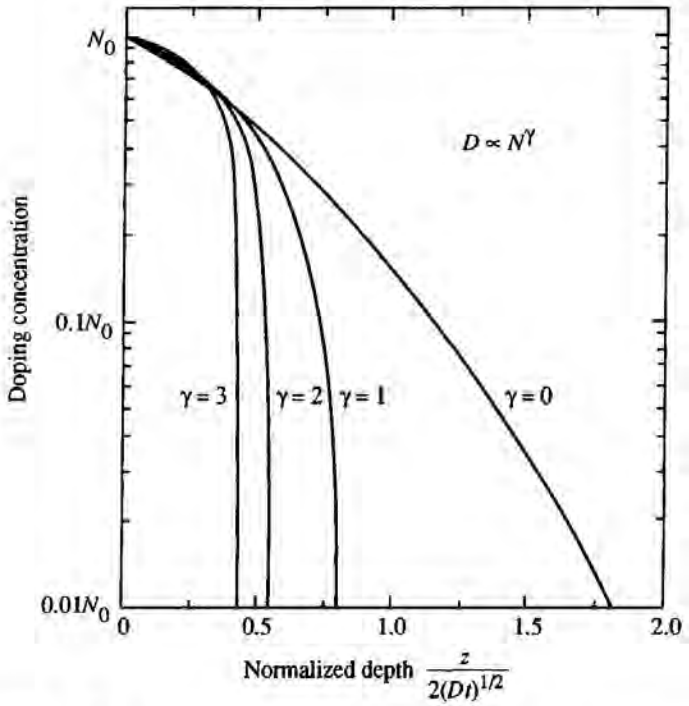


Fig. 7.11. Profile of impurities diffused into a semiconductor from an "infinite" reservoir with surface concentration N_0. The diffusion coefficient is assumed to depend on the concentration according to the power law $D \propto N^\gamma$ with $\gamma = 0, 1, 2$ and 3.

Equation (7.30) is valid for impurities with a diffusion coefficient that is independent of concentration. For many experimental cases, however, the diffusion coefficient depends on the concentration (see Chap. 8). The diffusion from an infinite reservoir with a concentration dependent diffusion coefficient was numerically solved by Weisberg and Blanc (1963). The result of the diffusion

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

calculation is displayed in Fig. 7.11 for a diffusion coefficient that depends on concentration according to a power law, i.e. $D \propto N^{\gamma}$ with $\gamma = 0, 1, 2$, and 3. The profiles become increasingly abrupt for higher values of γ.

The simple theoretical analysis provided above illustrates that doping in terms of impurity diffusion from an infinite reservoir has only two degrees of freedom which are N_0 and $L_D = \sqrt{Dt}$. If N_0 is a fixed value such as the solubility limit, only one degree of freedom, namely the diffusion length remains. Thus, diffusion is much less versatile as compared to the more flexible techniques of doping during epitaxial growth and by implantation.

The high temperatures required for doping incorporation by diffusion can deteriorate the surface and near-surface crystalline quality of the semiconductor. If III–V semiconductors are heated to high temperatures, preferential evaporation of group-V elements occurs. The loss of P, As, or Sb has several consequences such as non-stoichiometry, increase in point defect (vacancy) concentration, formation of group-III precipitates, and degradation of the surface morphology. The loss of group-V elements can be avoided by (i) appropriate counter pressures in the diffusion furnace or (i) coating of the sample with a capping layer (see Fig. 7.10) that prevents the out-diffusion of group-V elements.

Various capping layers were employed to reduce the out-diffusion of group-V elements from the semiconductor. Singh et al. (1988) employed SiO_2, borosilicate glass, phosphosilicate glass, and Al_2O_3 to prevent preferential As loss from a heated GaAs surface. The same layers were also employed by Camlibel et al. (1984) for capping InP and $Ga_{0.47}In_{0.53}As$ during Zn diffusion. Al_2O_3 has also been used as a capping layer by Schmitt et al. (1984) for Zn diffusion in InP. Silicon nitride (Si_3N_4) films are frequently used as a capping layer on GaAs and InP based compounds, particularly for diffusion and activation processes following ion implantation (see, for example, De Souza et al., 1990; Valco et al., 1989). An alternative to separate capping layers are single layers that act as both dopant reservoir and capping layers. Emulsions have been employed for this purpose (Arnold et al., 1980, 1984). Such emulsions can be spun on by a conventional resist spinner (Heime, 1989). Upon heating, the emulsions solidify forming a solid mixture of SiO, SiO_2 and the diffusing element. The SiO and SiO_2 fulfill two purposes, namely passivation of the surface and preventing group-V out-diffusion.

Doping incorporation by diffusion can be done in an ***open*** or ***closed system***. The use of capping layers discussed above makes possible the use of 'open' systems, i.e. conventional furnaces with an inert or reducing atmosphere surrounding the wafer. In contrast, a 'closed' system can contain the wafer in a sealed ampoule which is either evacuated or contains an inert gas. In addition, the

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

closed system can contain a solid group-V element (P, As, or Sb). Upon heating the ampoule, the group-V element sublimes and generates a group-V-rich atmosphere which reduces out-diffusion of group-V elements from the wafer.

Diffusing impurity elements can be provided from either the gas phase or the solid phase. In the case of gas phase provision, the doping precursor flows over the sample, pyrolyzes, and incorporates into the semiconductor. In the case of solid phase provision, the diffusing element is contained in a solid film on the wafer. Inorganic materials have been employed for the diffusion source. Examples for the solid sources are fluorides (ZnF_2, see Camlibel et al., 1984), nitrides (Zn_2N_3, see Arnold et al., 1984), and phosphides (Zn_3P_2, see Schmitt et al., 1984). In addition to unmasked diffusion of an entire wafer, masked diffusion was also reported (Arnold et al., 1984). The authors used Al_2O_3 as a mask. However, problems arose due to crystallization of the diffusion source at the edges of the Al_2O_3 mask which made subsequent photolithographic processes difficult.

The technology of ***ohmic contacts*** in III–V semiconductors is partially based on the diffusion of doping impurities from the surface into the bulk. It is quite clear from several reviews of ohmic contact technology in III–V semiconductors that processes other than diffusion play a major role (see, for example, Braslau, 1981; Kuphal, 1981). Detailed studies show that new phases with varying domain size appear after the alloying process. The metallization used for ohmic contact formation in III–V semiconductors is summarized in Table 7.1.

Table 7.1. *Metallization for ohmic contacts*

Metallization	Typical composition	Conductivity
AuGe	88%/12%	n-type
AuSn	95%/5%	n-type
AuBe	99.1%/0.9%	p-type
AuZn	95%/5%	p-type

Frequently multilayer metallizations are used to provide a better surface morphology and lower contact resistances after annealing. As an example, we consider the AuGe/Ni/Au metallization for n-type semiconductors. Typical thicknesses are 1000 Å, 500 Å, and 2000 Å for the eutectic AuGe, Ni, and Au, respectively. Ni has a sufficiently low surface tension preventing the underlying

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

AuGe film from 'balling up'. Without the Ni, AuGe contacts have the tendency to form islands during the alloying process (balling up) which degrades the electrical characteristics of the contact. The final Au top layer provides a low resistance contact pad suitable for bonding and probing.

7.4 Amphotericity and autocompensation

Group-IV impurities such as C, Si, Ge, and Sn can occupy substitutional cation as well as anion sites in III–V semiconductors. These impurities are shallow donors and shallow acceptors on group-III and group-V sites, respectively. This characteristic of group-IV impurities in III–V semiconductors is called ***amphotericity***. The origin of this word lies in the Greek word *amphora* which represents an ancient jar with two identical handles opposing each other; the word ***amphoteric*** means 'having two characters'. The two-sided characteristic of amphoteric impurities is undesirable and results in a partial or full compensation of donors by acceptors or vice versa. ***Autocompensation*** occurs if the compensation is due to the same chemical species of impurities, i.e. for amphoteric group-IV impurities in III–V semiconductors. It is obvious that the characteristics of amphotericity and autocompensation are uniquely related to C, Si, Ge, and Sn in III–V semiconductors and cannot occur in elemental semiconductors such as Si. In the following, the incorporation of an amphoteric impurity into a compound semiconductor is considered. The discussion is based on the law of mass action. A discussion of the thermodynamics of amphoteric impurities has also been given by Teramoto (1972) and by Munoz-Yague and Baceinedo (1982).

The degree of compensation of amphoteric impurities in III–V semiconductors depends on the doping concentration. It is generally found that the degree of compensation of such amphoteric impurities increases at high doping concentrations. Next, the dependence of the compensation on the doping concentration is analyzed quantitatively. As an illustrative example, Si-doped GaAs is considered. However, the results of the analysis have general validity and thus apply to all amphoteric impurities (C, Si, Ge, Sn) in III–V semiconductors. The incorporation of Si into the host GaAs is assumed to occur during epitaxial growth. Three different types of reactions can be distinguished for such an incorporation, namely the (i) host chemical reactions, the (ii) impurity incorporation reactions, and the (iii) electronic reactions. The host chemical reactions for incorporation of Ga and As from the gas phase are given by

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

$$Ga_{gas} + V_{Ga} \leftrightarrow Ga_{Ga} \tag{7.32a}$$

$$\tfrac{1}{2}As_{2,gas} + V_{As} \leftrightarrow As_{As} \tag{7.32b}$$

$$Ga_{gas} + \tfrac{1}{2}As_{2,gas} \leftrightarrow GaAs \tag{7.32c}$$

where V_{Ga} and V_{As} are a Ga vacancy and an As vacancy, respectively. The formation of point defects (e.g. Ga_{As}) is neglected. Amphoteric impurities can be incorporated either on anion or on cation sites. The impurity incorporation reactions are then given by

$$Si_{gas} + V_{Ga} \leftrightarrow Si_{Ga} \leftrightarrow Si_{Ga}^{+} + e \tag{7.33a}$$

$$Si_{gas} + V_{As} \leftrightarrow Si_{As} \leftrightarrow Si_{As}^{-} + h \tag{7.33b}$$

where e and h represent a free electron and hole, respectively. Note that any incorporation of Si into interstitial sites is neglected. We finally have the electronic reaction

$$e + h \leftrightarrow 0\,. \tag{7.34}$$

In an n-type semiconductor this generation/recombination reaction is associated with an energy of magnitude $E_F - E_V$.

We assume that the Si donor and acceptor impurities are ionized, i.e. we restrict our considerations to room temperature or typical growth temperatures. The sum of the impurity incorporation reactions (Eqs. (7.33a and b)), taking into account the electronic reaction (Eq. (7.34)), yields

$$Si_{Ga}^{+} + 2e + V_{As} \leftrightarrow Si_{As}^{-} + V_{Ga}\,. \tag{7.35}$$

Finally, elimination of V_{Ga} and V_{As} by means of the host chemical reactions (Eqs. (7.32)) yields

$$Si_{Ga}^{+} + 2e + As_{As} + Ga_{gas} \leftrightarrow Si_{As}^{-} + Ga_{Ga} + \tfrac{1}{2}As_{2,gas}\,. \tag{7.36}$$

We apply the law of mass action to this reaction and obtain

$$\frac{N_A^{-}\,[Ga_{Ga}]\,[\tfrac{1}{2}As_{2,gas}]}{N_D^{+}\,n^2\,[As_{As}][Ga_{gas}]} = K \tag{7.37}$$

where the brackets [...] represent the *concentrations* of the reagents and $N_A^{-} = [Si_{As}^{-}]$ and $N_D^{+} = [Si_{Ga}^{+}]$. The equilibrium constant K depends only on temperature. Equation (7.37) can be simplified by using $[Ga_{Ga}]/[As_{As}] = 1$ which is valid for stoichiometric GaAs. Further simplification results by

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

assuming that the V/III flux ratio is constant and independent of the doping level. Then Eq. (7.37) can be written as

$$\boxed{\frac{N_A^-}{N_D^+} = K^* n^2} \tag{7.38}$$

where K^* is a constant. Thus, the compensation ratio N_A^-/N_D^+ increases with the square of the free carrier concentration. The equation illustrates that the compensation ratio becomes *vanishingly small* for sufficiently small free carrier concentrations, i.e. $N_A^- \ll N_D^+$ for small N_D^+. Next we consider the high concentration limit and substitute $n = N_D^+ - N_A^-$ into Eq. (7.38) which yields

$$\boxed{\frac{N_A^-}{N_D^+} = K^*(N_D^+ - N_A^-)^2} \tag{7.39}$$

Analysis of this equation reveals that $N_D^+ > N_A^-$ for all concentrations. Simultaneously the equation can only be satisfied at very high concentrations if $N_D^+ \equiv N_A^-$. As a result, the compensation ratio approaches $N_A^-/N_D^+ \lesssim 1$ at high doping concentrations (see also Sect. 7.5.4). Thus, strong compensation is expected to occur at high doping concentrations for all amphoteric impurities in III–V semiconductors.

Note that the above considerations are valid only if equilibrium conditions are indeed established on the surface as well as in the bulk of the semiconductor. Furthermore, the reaction paths of the impurities must be complete, i.e. all relevant reaction paths must be included. If, for example, a major fraction of impurities were to incorporate on interstitial sites, then this reaction path would have to be included as well.

The above considerations are valid if the entire system is in equilibrium. During actual epitaxial growth, the crystal *surface* is in equilibrium, while the *bulk* may be far from equilibrium (see Sect. 7.1). Before an impurity is incorporated into the semiconductor, the impurity atom is chemisorbed to a (reconstructed) lattice site on the semiconductor surface. If the surface site is either an anion or cation site, the following exchange reaction can be written for Si on a GaAs *surface*.

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

$$Si_{As} + V_{Ga} \overset{\Delta G_s}{\leftrightarrow} Si_{Ga} + V_{As} \tag{7.40}$$

where ΔG_s is the Gibbs free energy, i.e. the sum of formation energy and the change in entropy ($T\Delta S$) that the system experiences during the surface reaction. The equilibrium concentration of the reaction products is determined by the law of mass action

$$\frac{[Si_{As}][V_{Ga}]}{[Si_{Ga}][V_{As}]} = K_s = c \exp\left[-\frac{\Delta G_s}{kT}\right] \tag{7.41}$$

where the brackets [...] refer to the concentration of the respective quantity and c is a constant. Since Si in the above example is a doping impurity, its concentration is much smaller than the concentration of both As and Ga vacancies on the surface. Because of the small impurity concentration, the vacancy concentration, $[V_{As}]$ and $[V_{Ga}]$ is not changed significantly during the reaction of Eq. (7.40). Thus the vacancy concentration can be considered as constant and the law of mass action can be rewritten as

$$\frac{[Si_{As}]}{[Si_{Ga}]} = \frac{[V_{As}]}{[V_{Ga}]}\, c \exp\left[-\frac{\Delta G_s}{kT}\right] \equiv \text{const.} \tag{7.42}$$

The equation illustrates the relevant factors for impurity incorporation. First, the ***energy factor*** $\exp(-\Delta G_s/kT)$ determines the degree of autocompensation of amphoteric impurities. That is, if $\Delta G_s/kT$ is a large quantity, autocompensation will be negligible for a wide range of growth conditions. An example for an impurity dominated by energies is C in GaAs; the autocompensation of this impurity depends little on the growth conditions (see Chap. 5 and Hoke et al., 1991a and 199b). Second, the ***kinetic factor*** $[V_{As}]/[V_{Ga}]$ which depends on the growth technique and the growth conditions, determines the incorporation and compensation of amphoteric impurities, provided the energy factor does not dominate. An example for an impurity dominated by kinetics is Ge in GaAs; the autocompensation of Ge depends strongly on the growth conditions, such as the V/III ratio and growth temperature (see Chap. 5 and Künzel et al., 1982). It is useful to keep in mind that the incorporation of impurities is a very complex process. Molecular dynamics calculations (Gilmer, 1992) show that the concentrations of Si_{As} and Si_{Ga} on the surface are not necessarily transferred into the bulk since Si in the near-surface region (1-3 monolayers) has a higher diffusivity than in the bulk. In the near-surface region, Si can change sites more easily than in the bulk.

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

The incorporation of amphoteric impurities on cation sites occurs preferably *at high V/III-flux ratios*. High V/III-flux ratios reduce $[V_{As}]$ in Eq. (7.42) which reduces the ratio $[Si_{As}]/[Si_{Ga}]$, i.e. increases $[Si_{Ga}]$. This tendency has been confirmed experimentally for the impurities C, Si, and Ge, and probably applies also to Sn (see Chap. 5). A second method to decrease $[V_{As}]$ is the use of an As-dimer source rather than an As-tetramer source. As_2 molecules are known to be more reactive than As_4 molecules which results in the reduction of As vacancies on the surface. The increased incorporation of impurities on group-III sites for the use of As_2 sources has been demonstrated for Si and Ge (see Chap. 5). The influence of the growth temperature is twofold. First, *low growth temperatures decrease the exponential factor* $\exp(-\Delta G_s/kT)$. As a result, the law of mass action can be dominated by the energy factor which increases the Si_{Ga} concentration (less autocompensation). Experimental observations agree with this tendency for GaAs:Si, since the highest free electron concentrations (i.e. minimal autocompensation) are obtained at low growth temperatures (see Chap. 5). Second, *low growth temperatures reduce the concentration of group-V vacancies*. The group-V concentration at the surface increases at low growth temperatures due to a reduction of reevaporation of group-V atoms from the surface. In Si-doped GaAs, decreasing autocompensation is expected at low growth temperatures, i.e. decreased $[V_{As}]$. This trend is indeed observed experimentally (see Chap. 5). Note that the two effects of the growth temperature result in reduced autocompensation for GaAs:Si. In contrast, for GaAs:C the two effects counteract. C predominantly occupies group-V sites in GaAs. A reduction of the growth temperature increases C_{As} due to the energy term $\exp(-\Delta G_s/kT)$. However, the concomitant decrease of V_{As} occurring at low growth temperatures makes C_{Ga}, i.e. autocompensation, more favorable. At present, sufficient experimental data are not available to assess the dominance of either of the two effects.

The incorporation of amphoteric impurities can be very different, if equilibrium conditions are achieved in the *bulk* rather than only on the *surface*. Several exchange reactions can occur in the bulk including

$$Si_{As} + V_{Ga} \overset{G_{b1}}{\leftrightarrow} Si_{Ga} + V_{As} \tag{7.43a}$$

$$Si_{As} \overset{G_{b2}}{\leftrightarrow} V_{As} + Si_{Ga} + I_{Ga} \tag{7.43b}$$

$$Si_{Ga} \overset{G_{b3}}{\leftrightarrow} V_{Ga} + Si_{As} + I_{As} \tag{7.43c}$$

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

where I_{Ga} and I_{As} represent a Ga and As interstitial, respectively. Note that Eq. (7.43a) is formally identical to Eq. (7.40). However, the change in free energy in the bulk, ΔG_b, can be different from the change in free energy for the surface reaction (Eq. (7.40)). The two other reactions (Eqs. (7.43b and c)) are so-called '*kick-out*' reactions. In these reactions a lattice atom is 'kicked out' of the lattice site into a interstitial position by the impurity atom. Other reactions, e.g. in which the impurity is kicked out by an interstitial lattice atom, can occur as well but will not be further considered. Note that the change in energy, is different for the three Eqs. (7.43a, b and c). Application of the law of mass action to Eqs. (7.43) yields

$$\frac{[Si_{As}][V_{Ga}]}{[Si_{Ga}][V_{As}]} = K_{b1} = c_1 \exp\left[-\frac{\Delta G_{b1}}{kT}\right], \tag{7.44a}$$

$$\frac{[Si_{As}]}{[V_{As}][Si_{Ga}][I_{Ga}]} = K_{b2} = c_2 \exp\left[-\frac{\Delta G_{b2}}{kT}\right], \tag{7.44b}$$

$$\frac{[Si_{Ga}]}{[V_{Ga}][Si_{As}][I_{As}]} = K_{b3} = c_3 \exp\left[-\frac{\Delta G_{b3}}{kT}\right]. \tag{7.44c}$$

The equations allow one to estimate trends for dopant incorporation under different growth conditions. For example, autocompensation is minimized for the reaction of Eq. (7.44a) at low concentrations of V_{As}. Low concentrations of V_{As} decrease $[Si_{As}]$ in Eq. (7.44b) but increase $[Si_{As}]$ in Eq. (7.44c). Note that $[Si] = [Si_{Ga}] + [Si_{As}]$ is assumed to be valid in Eqs. (7.43) and (7.44). Note also that the reactions of Eqs. (7.43a) and (7.44a) were estimated to be the most prominent exchange reactions (Munoz-Yague and Baceinedo, 1982). The above considerations illustrate that the incorporation of impurities depends on the location at which equilibrium is achieved, i.e. if equilibrium is achieved on the surface, in the bulk, or both. This can explain why different growth methods result in drastically different incorporation of amphoteric impurities. For example, Si is predominantly a donor in MBE and OMVPE grown GaAs, while Si is an acceptor in liquid-phase epitaxy (LPE) grown GaAs. Since the surface chemistry (especially the V/III flux ratio) is very different for these growth techniques, differences in impurity incorporation are expected as well.

Note that the charge states of the reaction products in Eqs. (7.40) and (7.43) have been neglected. The charge states of impurities on the semiconductor surface are unknown. In addition, the charge states of vacancies in the bulk have been calculated (Baraff and Schlüter, 1985) but have not been determined

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

experimentally. Assuming the charge states V_{As}^{0}, V_{Ga}^{0}, Si_{Ga}^{+}, and Si_{As}^{-} in Eq. (7.43a) and taking into account the electronic reaction $e+h \leftrightarrow 0$, results in the same dependence of the compensation ratio on the free carrier concentration as given in Eq. (7.38), i.e. $N_A^- / N_D^+ \propto n^2$.

The incorporation of doping impurities is not expected to depend on the epitaxial growth ***orientation***. Under equilibrium conditions, the compensation can be inferred from Eq. (7.44a) for impurities already incorporated in the bulk. Obviously, Eq. (7.44a) does not depend on the orientation of the growing crystal face. Nevertheless, experimental studies on the incorporation of Si impurities into GaAs grown along different crystallographic orientations revealed a clear dependence of the Si doping characteristics on the direction of the growth axis. Ballingall and Wood (1982) reported very different Si doping incorporation characteristics for epitaxial growth on ⟨001⟩, ⟨011⟩, and ⟨111⟩B oriented GaAs. The authors found that Si incorporates as a donor in ⟨001⟩ GaAs and as acceptor and donor for crystals with a (011) growth face. Low growth temperatures resulted in n-type conductivity of ⟨011⟩ oriented GaAs, while higher growth temperatures (> 550 °C) resulted in p-type conductivity. Growth on the (111) B face resulted in highly resistive material. Subsequent studies on Si doping incorporation were published by Bose et al. (1988) and Lee et al. (1989a, 1989b, 1990). Kadoya et al. (1990) reported Si doping on the (111)A face of GaAs during epitaxial growth. The conductivity was found to be n-type for a growth temperature of 480 °C and/or high V/III flux ratios. P-type conductivity was found for high growth temperatures and low V/III flux ratios. P-type conductivity for Si doping of the GaAs (111)A surface was also found by Miller and Asbeck (1987). All groups reporting growth of ⟨111⟩ oriented GaAs by MBE obtained textured or rough surface morphologies indicating that epitaxial growth on these surfaces is non-trivial.

It is useful to visualize the microscopic nature of the (001), (011), (111)A, and (111)B surfaces of III–V semiconductors. These different surfaces are shown in Fig. 7.12 where we have neglected any surface reconstruction. The (001) GaAs growth surface consists of either a layer of As atoms or a layer of Ga atoms; the (001) surface is a polar surface. In contrast, the (011) GaAs surface is non-polar. The (111)A and (111)B surfaces are again polar surfaces. While the (001) surface strongly favors Si donor incorporation, the two other surfaces favor an amphoteric incorporation of Si. Inspection of the surfaces shown in Fig. 7.12 indicates the reason for this peculiar Si incorporation characteristic. A Si impurity atom chemisorbed at the different surfaces has two bonds to host surface atoms for the (001) surface but only one bond to a host surface atom for the (011), (111)A, and (111B) surface. The weaker bonds of the impurity atom to the semiconductor

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

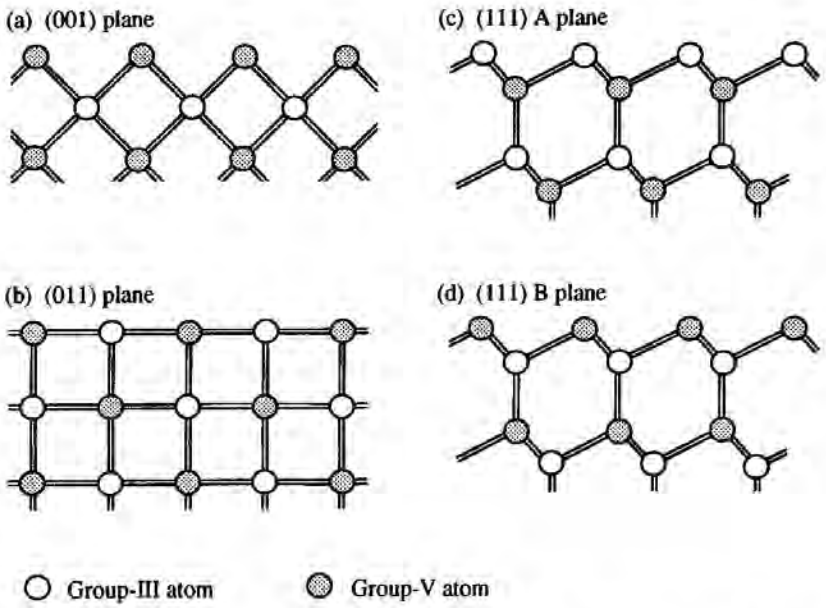


Fig. 7.12. Schematic illustration of the atomic structure of different surfaces of III–V semiconductors (sideview; surface reconstruction neglected).

surface make site switching more likely. The weak surface binding with only one bond makes the difference of Si_{Ga} and Si_{As} formation energy less relevant on the (011) and (111) surface as compared to the (001) surface. It has been indeed observed experimentally that the incorporation and amphotericity of Si in GaAs depends strongly on the growth condition for the (011) and (111) face (Ballingall and Wood, 1982; Miller and Asbeck, 1987; Kadoya et al., 1991). All reports indicate that high V/III flux ratios and low growth temperature favor Si donor incorporation while low V/III flux ratios and high growth temperatures favor Si acceptor incorporation. This finding can be fully explained on the basis of the law of mass action. High V/III flux ratios and low growth temperatures decrease the concentration of As vacancies which results in a decreased $[Si_{As}]$, i.e. less autocompensation of the Si donors as inferred from Eq. (7.42).

Assume that thermal equilibrium is reached in the bulk and that the concentration of vacancies and defects in the bulk does not depend on the epitaxial growth face. Under these assumptions, the incorporation of an amphoteric impurity (e.g. Si) is not expected to depend on the growth orientation. To explain the experimental facts which are in contrast to this expectation, two possibilities should be considered. First, the concentration of vacancies may depend on the orientation of the growth surface. Such a difference in vacancy and

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

337-TA-1366 RX-0255

defect concentration may occur because of the different surface chemistry and the different morphology of the epitaxial layer grown on different faces. Differences in the vacancy concentration result in different compensation ratios as evident from Eq. (7.42a). Second, while thermal equilibrium is achieved on the surface, it may *not* be achieved in the bulk at sufficiently low growth temperatures. While the surface equilibrium is determined by Eq. (7.42), the bulk equilibrium is given by Eq. (7.44a). Note that the free energies, ΔG_s and ΔG_b, are different for the surface and the bulk. If equilibrium is achieved only on the surface but not in the bulk (in agreement with Sect. 7.1), then impurities remain on their original site when incorporated in the bulk, i.e. remain metastable over long periods of time.

If amphoteric impurities are implanted into III–V semiconductors, it is desirable to keep autocompensation minimal. After implantation and subsequent activation (annealing) some fraction of the implanted impurity atoms may not be electrically active. Consider again the example of Si in GaAs; the Si impurities are either donors, acceptors, or electrically inactive (e.g. Si–Si pairs, Si clusters, etc.). The free electron concentration is then given by

$$n = [\mathrm{Si}] - 2[\mathrm{Si}_{\mathrm{acceptor}}] - [\mathrm{Si}_{\mathrm{inactive}}] \tag{7.45}$$

where the brackets refer to the concentrations, and [Si] is the total Si concentration in the semiconductor. Assuming that the Si donor concentration is given by

$$[\mathrm{Si}_{\mathrm{donor}}] = [\mathrm{Si}] - [\mathrm{Si}_{\mathrm{acceptor}}] - [\mathrm{Si}_{\mathrm{inactive}}] \tag{7.46}$$

the free electron concentration from Eqs. (7.45) and (7.46) is

$$n = N_D - N_A \tag{7.47}$$

where $N_D = [\mathrm{Si}_{\mathrm{donor}}]$ and $N_A = [\mathrm{Si}_{\mathrm{acceptor}}]$. The law of mass action (Eq. (7.44a)) indicates that a reduction in As vacancies decreases the autocompensation of Si in GaAs. ***Co-implantation*** of As in GaAs can reduce As vacancies in order to decrease the Si concentration on Ga sites. The amphotericity of other impurities can also be reduced by co-implantation. For example, the co-implantation of Ga in GaAs decreases the concentration of C donors in C-implanted GaAs. There are several reports on co-implantation of III–V semiconductors with As or P to improve the activation of impurities which preferentially occupy group-III sites (Kim et al., 1987; Wang, 1987, 1989). The authors report an improvement in the activation of Be-implanted InP. Note that Be is not an amphoteric impurity. Nevertheless, the co-implantation increases $[\mathrm{Be}_{\mathrm{In}}]$, which results in improved activation. The activation of Si implanted into GaAs is improved upon co-implantation with As (Tell, 1991). The improvement

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

Innoscience1366-00138981

RX-0255.308

in activation is usually relatively small and is therefore not used in the production of III–V semiconductor devices. Another method to improve Si activation in implanted GaAs is the employment of As overpressures during the activation anneal (see, for example, Pearton, 1988). The overpressure prevents As out-diffusion from the near surface region of GaAs and thus reduces the formation of As vacancies. As a result, a higher Si activation is achieved.

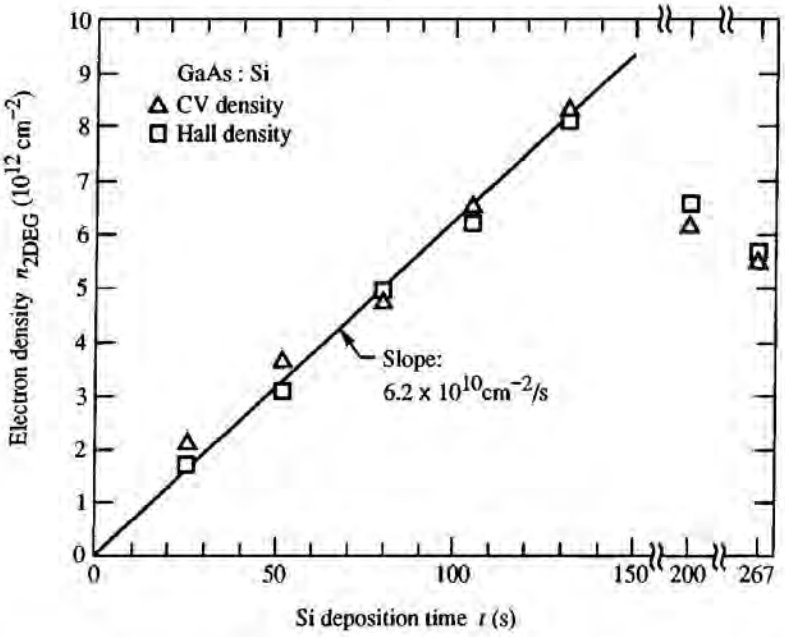


Fig. 7.13. Ionized impurity density (measured by capacitance–voltage technique) and free electron density (measured by the Hall effect) of GaAs doped with Si during growth interruption versus Si deposition time. The free carrier and the ionized Si density decrease simultaneously at high doping densities.

Autocompensation in III–V semiconductors doped with amphoteric impurities *increases at high doping concentration.* The increase can be quite drastic as in the case of GaAs doped with Si in the 10^{19} cm^{-3} range (see, for example, Maguire et al., 1987). For other impurities, e.g. C in GaAs the increase with doping concentration is less pronounced. An illustrative example of compensating and inactive Si in δ-doped GaAs is shown in Fig. 7.13 (Schubert et al., 1990a), which shows the free electron density as a function of Si deposition time. The Si impurities were evaporated onto the GaAs surface during growth interruption with a deposition rate of 6.2×10^{10} cm^{-2}/s. The free carrier density was measured by the Hall effect and by capacitance–voltage (CV)

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

measurements. The experimental results of Fig. 7.13 show that the free electron density does not follow the Si doping density at densities $N_{Si} > 8 \times 10^{12}$ cm^{-2}. If compensation effects were absent, the experimental densities would follow the solid line in Fig. 7.13 which represents unity activation. However, both, the Hall density and the CV density decrease at high doping density. At a Si density of $N_{Si} = 1.66 \times 10^{13}$ cm^{-2} a free electron density of only 5.5×10^{12} cm^{-2} is measured indicating a compensation ratio of $N_A^- / N_D^+ = N_A^- / (N_{Si} - N_A^-) \equiv 50\%$.

The maximum free carrier concentration shown in Fig. 7.13 is $n_{max}^{2D} = 8 \times 10^{12}$ cm^{-2}. This two-dimensional density corresponds to a three-dimensional concentration of $n = (n^{2D})^{3/2} \equiv 2.2 \times 10^{19}$ cm^{-3}. Note that this three-dimensional concentration is higher than the highest free carrier concentration reported for homogeneously Si-doped GaAs of 1.8×10^{19} cm^{-3} (Ogawa and Baba, 1985).

Beall et al. (1988) arrived at similar conclusions about the saturation and decrease of the free carrier density in highly Si-doped GaAs, doped during growth interruption. The authors concluded that Si incorporates at high densities as electrically inactive Si clusters, Si–Si pairs and Si acceptors. Local-mode spectroscopy was used to identify the nature of Si impurities in δ-doped GaAs at high concentrations. The local-mode spectrum is shown in Fig. 7.14 for as-grown and annealed (600 °C, 30 min) GaAs. The local-mode spectrum allows one to identify the presence of several Si configurations including substitutional Si_{Ga}, compensating Si_{As}, and Si–Si pairs. Figure 7.14 reveals that the concentration of substitutional Si on Ga sites, Si_{Ga}, increases drastically upon annealing. Beall et al. (1989) explained the experimental results as follows. On the as-grown sample the doping profile is narrow with very high concentrations ($\equiv 2.5 \times 10^{20}$ cm^{-3}). Most of the Si impurities are electrically inactive or compensating at high concentrations. Upon annealing, the concentration decreases due to diffusion of Si impurities out of the original impurity plane. As a result of the lower concentration, the largest fraction of Si occupies substitutional lattice sites ($[Si_{Ga}] = 2 \times 10^{13}$ cm^{-2}).

The amphoteric and compensating behaviour of group-IV impurities is a general observation made by many research groups (see also Chap. 5). The compensation is most pronounced at the highest Si concentrations. The most notable compensation ratios were reported by Greiner and Gibbons (1984). At a Si concentration of 2×10^{20} cm^{-3} the authors found an electrically active Si donor concentration of 2×10^{18} cm^{-3}, which corresponds to a compensation ratio of approximately 99%.

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

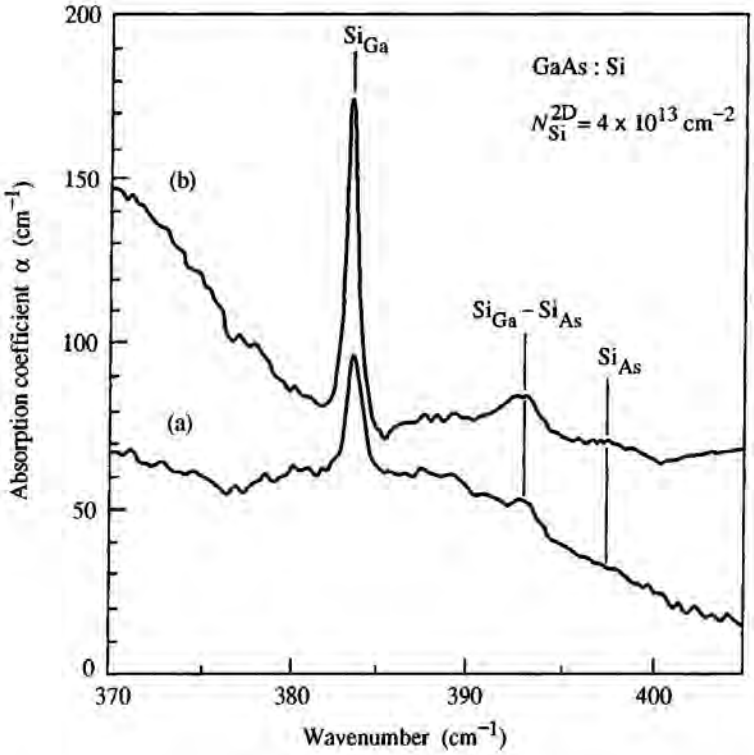


Fig. 7.14. Local vibrational mode (LVM) spectra for a [001] oriented GaAs epitaxial layer doped with Si of density 4×10^{13} cm^{-2} deposited during growth interruption. Spectrum (a) and (b) are obtained on the as-grown sample and after annealing (600 °C, 30 min), respectively (after Beall et al., 1989).

The increase in autocompensation of amphoteric impurities at high doping concentrations, which has been observed experimentally, agrees with the trend predicted by Eqs. (7.38) and (7.39). A quantitative analysis of experimental data on autocompensation has, however, not been performed.

It is useful to consider the total formation energy of an amphoteric impurity to be composed of a *chemical* contribution and an *electronic* contribution. The chemical contribution to the formation energy is due to the formation of bonds of the impurity in the host as well as strain and steric effects. The electronic contribution is due to the electronic processes associated with the donor and acceptor formation. For the creation of an ionized donor in an n-type semiconductor, the electron must occupy an empty conduction band state, i.e. the electron must be lifted from the donor level to the Fermi level. Thus, the electronic formation energy for a donor is given by

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

$$E_{\text{electronic}}^{\text{Donor}} = -(E_F - E_D) \,. \tag{7.48a}$$

If an acceptor level is formed in an n-type semiconductor, an electron from the conduction band can recombine onto the acceptor level. The electronic energy *gained* by an acceptor formation is thus

$$E_{\text{electronic}}^{\text{Acceptor}} = E_F - E_A \,. \tag{7.48b}$$

The total formation energy is the sum of the chemical and electronic part. The *difference* in formation energies between a donor and an acceptor is then given by

$$\begin{aligned} \Delta E &= E_{\text{chem}}^{\text{Donor}} - E_{\text{chem}}^{\text{Acceptor}} + E_{\text{electronic}}^{\text{Donor}} - E_{\text{electronic}}^{\text{Acceptor}} \\ &= \Delta E_{\text{chem}} - 2E_F + E_D + E_A \\ &\cong \Delta E_{\text{chem}} - 2(E_F - E_C) - E_g \end{aligned} \tag{7.48c}$$

where we assumed that $E_C - E_D \cong 0$ and $E_C - E_A \cong E_g$. The relative occupancy of cation and anion sites by amphoteric impurities depends on the difference in formation energies between the two sites. Assuming a Boltzmann distribution for the occupancy yields

$$\frac{N_A^-}{N_D^+} = K_B \, e^{-\Delta E/kT} \tag{7.49}$$

where ΔE is given by Eq. (7.48c) and K_B is a constant. Insertion of Eq. (7.48c) into Eq. (7.49) and using non-degenerate statistics for the conduction band, i.e. $E_F - E_C = kT \ln (n/N_c)$, yields

$$\frac{N_A^-}{N_D^+} = K_B^* \, n^2 \, e^{-\Delta E_{\text{chem}}/kT} \,. \tag{7.50}$$

Note that the equation has the identical dependence on the free carrier concentration as Eq. (7.38) which is, of course, not accidental.

The chemical formation energies of an amphoteric impurity are different for cation and anion site occupation. One can estimate the difference in formation energy from the different bond strengths of the impurity with the constituent anions and cations of the semiconductor. The bond strengths are known for some impurities, e.g. the C–Ga and the C–As bonds (Hartley and Patai, 1982) which are 61 kcal/mole ≅ 2.65 eV/atom and 57 kcal/mole ≅ 2.47 eV/atom, respectively. For tetrahedrally bonded C in GaAs, the difference in chemical binding energy is thus $\Delta E \cong 4 \times (2.65 \text{ eV} - 2.47 \text{ eV}) = 690$ meV. However, this energy must be

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

used with some caution because isolated binding energies are not necessarily representative for the tetrahedrally bonded C in GaAs since steric and other effects are not taken into account.

Another estimate of the difference in formation energy of amphoteric impurities between cation and anion site occupancy can be obtained from concentration measurements of the two species. Photoluminescence measurements can be used to obtain an estimate of the impurity concentrations on the different sites; that is, acceptors and donors of the same amphoteric impurities have distinctly different luminescence signals. The quantitative evaluation of luminescence spectra from Si-doped GaAs at a low doping level of $N_{Si} = 10^{16}$ cm^{-3} reveals the presence of $10^{12} - 10^{14}$ cm^{-3} Si acceptors (Harris and Schubert, 1990). Since $N_A/N_D \propto \exp(-\Delta E/kT)$, the difference in formation energy is estimated to be on the order of 300–500 meV.

7.5 Maximum doping concentration

The scaling of the spatial dimensions of semiconductor devices necessitates that the doping concentrations are increased as well. Certainly, the maximum attainable doping concentration has fundamental limits which depend on the doping method, the host semiconductor and the doping element. Equally important is the electrical activity of impurities at high doping concentrations. That is, high impurity atom concentrations are useful only if complete or near complete electrical activity is achieved. We therefore focus on the maximum *electrically active* impurity concentration rather than on the maximum impurity concentration. Table 7.2 summarizes the highest free carrier concentrations measured for various impurity elements and III–V semiconductors.

Table 7.2 shows that the maximum doping concentration depends strongly on the semiconductor hosting the impurities and on the impurity element. The highest doping concentrations are in the 10^{18} to 10^{20} cm^{-3} range. The highest value has been reported for GaAs:C with a hole concentration of 1.5×10^{21} cm^{-3}. Since GaAs has an atom concentration of $8/a_0^3 = 4.4 \times 10^{22}$ cm^{-3}, the electrically active C concentration corresponds to approximately 3% of the atom concentration of the host.

In the following sections, we consider several mechanisms that limit the maximum electrically active impurity concentration in semiconductors. The physical origins of the limitations discussed are quite different. The models are based on a concentration dependent diffusion coefficient (Sect. 7.5.1), coulombic repulsion (Sect. 7.5.2), defect models (Sects. 7.5.3 and 7.5.4), autocompensation (Sect. 7.5.5), and the solubility limit (Sect. 7.5.6).

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

7.5.1 Diffusion limited maximum concentration

The concentration of impurities in semiconductors can be limited to some maximum value, if the diffusion coefficient of impurities increases *drastically* at high concentrations. To illustrate this mechanism, assume that the diffusion coefficient of an impurity is very high for concentrations exceeding some critical concentration, N_{crit}, while it is much lower for concentrations below the critical concentration. Regions in the semiconductor with $N > N_{crit}$ are then characterized by strong diffusive motion of impurities. Redistribution occurs until impurities reach regions where $N < N_{crit}$, i.e. where the diffusion is slow.

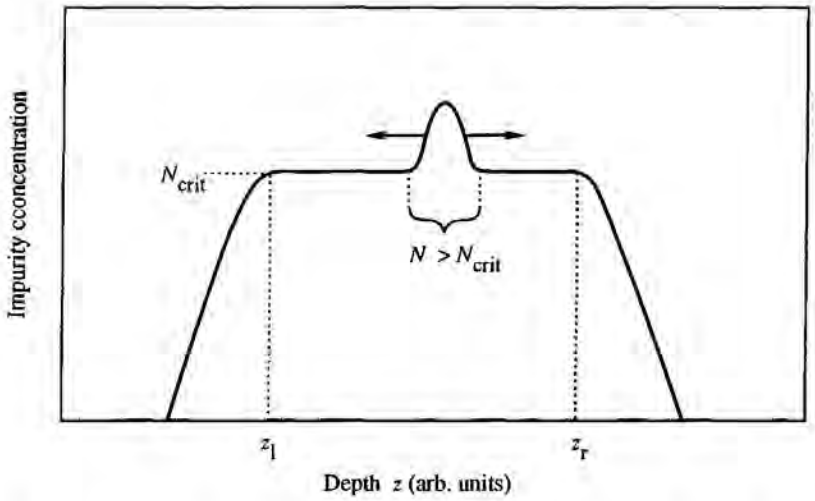


Fig. 7.15. Schematic illustration of a one-dimensional doping profile which locally exceeds a critical concentration N_{crit}, above which rapid diffusion occurs. Impurity atoms strongly diffuse until $N \leq N_{crit}$ for all values of z.

The concept of diffusion limited maximum concentration is illustrated in Fig. 7.15 which shows a one-dimensional impurity profile. The impurity profile has three distinct regions, that is, regions where $N < N_{crit}$, $N = N_{crit}$, and $N > N_{crit}$. The regions where $N \leq N_{crit}$ are assumed to be stable against rapid diffusion. In contrast, the region where $N > N_{crit}$ (i.e. $z_l \leq z \leq z_r$) is assumed to be highly diffusive. Impurities in this region strongly redistribute until they reach the regions $z < z_l$ and $z > z_r$. That is, strong diffusion occurs until $N \leq N_{crit}$ in the entire semiconductor.

A diffusion coefficient that increases with concentration is obtained from several models. The substitutional-interstitial diffusion model (see Chap. 8) predicts a power-law dependence of the diffusion coefficient on concentration,

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

337-TA-1366 RX-0255

Table 7.2. *Maximum free carrier concentrations in III–V semiconductors*

Semiconductor	Impurity	Concentration	Reference
$Al_{0.25}Ga_{0.75}As$	Si (n-type)	4×10^{18} cm^{-3}	Drummond et al. (1982)
$Al_{0.30}Ga_{0.70}As$	Be (p-type)	2×10^{19} cm^{-3}	Fujita et al. (1980)
$Al_{0.80}Ga_{0.20}As$	Be (p-type)	9×10^{18} cm^{-3}	Fujita et al. (1980)
$Al_{0.48}Ga_{0.52}As$	Sn (n-type)	2×10^{19} cm^{-3}	Cheng et al. (1981)
$Al_{0.50}In_{0.50}P$	Be (p-type)	3×10^{18} cm^{-3}	Nakajima et al. (1989)
GaP	Si (n-type)	8×10^{18} cm^{-3}	Baillargeon et al. (1989)
	Zn (p-type)	5×10^{19} cm^{-3}	Mottram et al. (1971)
GaAs	Te (n-type)	2.6×10^{19} cm^{-3}	Wu et al. (1988)
	Si (n-type)	1.8×10^{19} cm^{-3}	Ogawa and Baba (1985)
	C (p-type)	1.5×10^{21} cm^{-3}	Yamada et al. (1989)
	Be (p-type)	2×10^{20} cm^{-3}	Lievin et al. (1988)
$Ga_{0.47}In_{0.53}As$	Sn (n-type)	1×10^{20} cm^{-3}	Panish et al. (1990a)
	Be (p-type)	5×10^{20} cm^{-3}	Hamm et al. (1989)
$Ga_{0.5}In_{0.5}P$	Se (n-type)	8.3×10^{18} cm^{-3}	Iwamoto et al. (1984)
	Be (p-type)	4×10^{19} cm^{-3}	Nakajima et al. (1989)
GaSb	Te (n-type)	6×10^{18} cm^{-3}	Chin and Bonner (1982)
	Zn (p-type)	1×10^{20} cm^{-3}	Aliev et al. (1967)
InP	Sn (n-type)	8×10^{19} cm^{-3}	Astles et al. (1973)
	Sn (n-type)	1×10^{20} cm^{-3}	Panish et al. (1990b)
	Be (p-type)	7×10^{18} cm^{-3}	Benchimol et al. (1989)
	Zn (p-type)	2×10^{18} cm^{-3}	Tsang et al. (1991)
InAs	Te (n-type)	3×10^{19} cm^{-3}	Harrison and Houston

e.g. $D \propto N^2$ or $D \propto N^3$. However, such a dependence is too weak to explain the rigid upper limit for the impurity concentration observed in many III–V semiconductors.

A model that predicts a much stronger dependence of the diffusion coefficient on the impurity concentration was proposed by Mathiot and Pfister (1982, 1984) for impurities in Si. The model is based on the combination of the conventional vacancy/interstitial diffusion and diffusion in a percolation cluster. Recently, Larsen and Larsen (1990) tested the percolation/vacancy/interstitial model for

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

Innoscience1366-00138988

RX-0255.315

337-TA-1366 RX-0255

several impurities in Si and found very good agreement with experimental results. The percolation/vacancy/interstitial model assumes that an infinite percolation cluster is formed when the impurity concentration exceeds a critical concentration N_{crit}. The impurity/vacancy potential is reduced in the semiconductor by the quantity ΔE_A. The enhancement of the impurity diffusion for $N > N_{crit}$ is due to this reduced binding energy; the enhancement factor F is equal to $\exp(\Delta E_A/kT)$. Mathiot and Pfister (1982, 1984, 1989) expressed the impurity diffusion coefficient as

$$D_{eff} = p_\infty D_{percolation} = (1 - p_\infty) D \tag{7.51}$$

where D is the regular diffusion coefficient of the impurity (which may be concentration dependent) and $D_{percolation}$ is the diffusion coefficient of the percolation cluster. The probability of an impurity belonging to a percolation cluster is given by p_∞ which approaches unity at high concentrations. The percolation diffusion coefficient is proportional to the enhancement factor, i.e. $D_{percolation} \propto F = \exp(\Delta E_A/kT)$. The authors extracted the enhancement factor F and the critical concentration N_{crit} from experimental impurity profiles. Critical concentrations on the order of $N_{crit} \cong 10^{20}$ cm^{-3} and enhancement factors F of several thousands were found.

Although Mathiot and Pfister (1982) developed the percolation cluster model for Si, it is likely that it can also be applied to semiconductors other than Si. The original percolation cluster model was not intended to explain the maximum concentration in semiconductors. It is, however, quite evident that for sufficiently large enhancement factors F the critical concentration N_{crit} limits the maximum concentration of impurities in semiconductors.

7.5.2 Coulombic repulsion

The repulsive coulombic interaction of impurities within the depletion region of a growing semiconductor can limit the incorporation of impurities during growth. As a result, a certain maximum impurity concentration cannot be exceeded during epitaxial growth (Schubert et al., 1992). The model is of general nature and applies to semiconductors whose Fermi level is pinned at the surface.

The region near to the surface of a semiconductor is usually depleted of free carriers. The free carrier depletion is a result of the pinning of the Fermi level at a semiconductor surface. The pinning of the Fermi level is the result of electronic states (so-called Bardeen states) which are energetically located close to the middle of the fundamental gap for most of the III–V semiconductors. For GaAs, the Fermi-level pinning is found for different surface orientations, e.g. the ⟨001⟩, ⟨011⟩, and ⟨111⟩ surface orientation. Chiang et al. (1983) showed that the Fermi

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

RX-0255.316

level at a GaAs surface is pinned at 0.55 ± 0.1 eV above the valence band maximum for all reconstructed surfaces studied by the group. Fermi-level pinning was found to be insensitive to temperature in the temperature range 20–500 °C, consistent with the fact that the thermal energy kT is much smaller than other relevant energies (Nanichi and Pearson, 1969).

The coulombic energy in an n-type semiconductor is the sum of donor–donor, donor–electron, and electron–electron interaction energies. Since the concentration of free carriers can be neglected in the depletion region, the coulombic interaction energy is given by all donor–donor interactions, i.e.

$$E = \frac{1}{2} \sum_{i \neq j} \frac{e^2}{4\pi\varepsilon r_{ij}} \tag{7.52}$$

where ε is the permittivity of the semiconductor and r_{ij} is the separation of two donor ions. The sum is carried out for all impurities in the depletion region. Note that the coulombic interaction energy is reminiscent of the Madelung energy of ionic crystals. Unscreened Coulomb potentials can be used in Eq. (7.52) due to the low free carrier concentration in the depletion region. Screened Coulomb potentials must be used to calculate the interaction in the neutral part of the semiconductor. The screened interaction is, however, weaker. In any case, the coulombic interaction favors large distances between impurities.

The driving force can be calculated using the depletion approximation. The band diagram of a depleted n-type semiconductor surface is shown in Fig. 7.16. The Fermi level is pinned at an energy $e\phi_B$ below the conduction band edge. Impurities within the depletion region experience an electrostatic driving force towards the surface which results in a drift motion of the impurities with a velocity

$$v_d = \mu \mathcal{E} \tag{7.53}$$

where $\mathcal{E}$ is the field inside the depletion region and μ is the mobility of ionized impurities. The mobility can be expressed in terms of the diffusion coefficient using the Einstein relation $\mu = eD/kT$. The effect of the surface field was shown to redistribute impurities towards the surface (Schubert et al., 1990b). The maximum electric field occurs at the surface and is given by

$$\mathcal{E}_{max} = (2eN_D\phi_B/\varepsilon)^{1/2} . \tag{7.54}$$

The equation shows that the driving force is proportional to the square root of the doping concentration.

At some doping concentration the surface driving force becomes so strong that impurities are no longer incorporated into the crystal. Impurities then drift

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

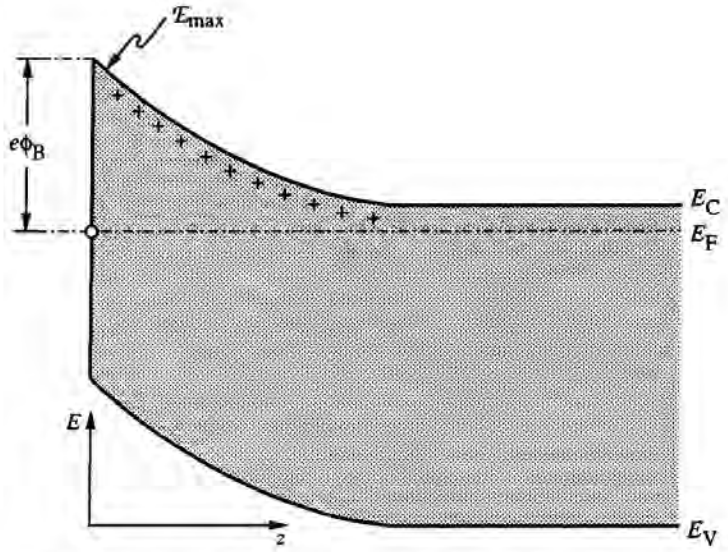


Fig. 7.16. Illustration of the band diagram of an n-type semiconductor whose Fermi level is pinned at the semiconductor surface. Ionized impurities in the space charge region interact repulsively due to their Coulomb potentials.

along the growth direction with the growing semiconductor surface. The condition can be quantified by assuming that the growth rate (i.e. growth velocity) equals the drift velocity

$$v_d = v_g \,. \tag{7.55}$$

The doping concentration at which this condition is fulfilled is obtained from Eqs. (7.53), (7.54), and the Einstein relation ($\mu = eD/kT$) and is obtained as

$$\boxed{N_{max} = \left(v_g \frac{kT}{eD}\right)^2 (\varepsilon/2e\phi_B)} \tag{7.56}$$

This concentration represents the maximum achievable doping concentration in semiconductors. Incorporation of impurities at a concentration higher than N_{max} is not possible, since impurities would drift to the surface at a faster rate than the growth rate.

We now consider the example of Be-doped GaAs. The diffusion coefficient of Be in GaAs has been determined to be $\cong 3 \times 10^{-16}$ cm^2/s at typical growth temperatures of 500–600 °C (Schubert et al. 1990c). With a growth rate of

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

0.8 μm/hr = 2.2Å/s and a barrier height of $e\phi_B = 0.7$ eV, the maximum doping determined from Eq. (7.56) is $N_{max} \cong 1.5 \times 10^{20}$ cm^{-3}. This value is in good quantitative agreement with experimentally achieved maximum concentrations of Be in GaAs (see Table 7.2).

The equation for the maximum doping concentration (Eq. (7.56)) also predicts how the doping concentration depends on (i) growth rate and (ii) the diffusion coefficient. Impurities with lower diffusion coefficients are expected to yield higher doping concentrations. This trend is indeed found for acceptors in GaAs. Carbon doping in GaAs is possible up to $\cong 10^{21}$ cm^{-3}; C is also the impurity with the lowest diffusion coefficient. Beryllium doping is possible up to $\cong 10^{20}$ cm^{-3} and Be is a moderately fast diffuser. Zinc doping is possible up to $\cong 10^{19}$ cm^{-3}; at the same time, Zn is known to be a fast diffuser. The dependence of the maximum doping concentration on the growth rate has not yet been thoroughly studied in III–V semiconductors. Initial experiments in Si grown by molecular-beam epitaxy at different growth rates indicate that N_{max} increases with enhanced growth rate in agreement with the theoretical expectation (Gilmer and Gossmann, 1993).

The coulombic interaction of impurities in the depletion region of a semiconductor has been calculated using Monte Carlo techniques (Schubert et al., 1992). In the calculation, the interaction of impurity ions and charged carriers trapped in surface states is considered. A calculated impurity profile is shown in Fig. 7.17. The calculation simulates a growth process in which impurities are deposited on the non-growing surface of a semiconductor. Subsequently, epitaxial growth is resumed. The calculation demonstrates that the rate of impurity incorporation is limited to some maximum value. If this maximum concentration is locally exceeded, repulsive interaction drives more impurities back to the semiconductor surface. In addition, the repulsive interaction prevents impurities from being incorporated from the surface into the bulk. Both mechanisms stabilize the impurity concentration in the bulk.

A further result of the repulsive Coulomb interaction is that impurities are not distributed randomly in the semiconductor crystal. It is usually assumed that impurities are distributed according to the ***Poisson distribution*** (see, for example, Shockley, 1961). However, closely spaced impurity configurations are unfavorable at high impurity concentrations due to repulsive Coulomb interaction. Inspection of the impurity distribution calculated by the Monte Carlo technique revealed that the distribution is non-poissonian. Specifically, configurations with two or more closely spaced impurities occur less frequently than expected on the basis of the Poisson distribution.

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

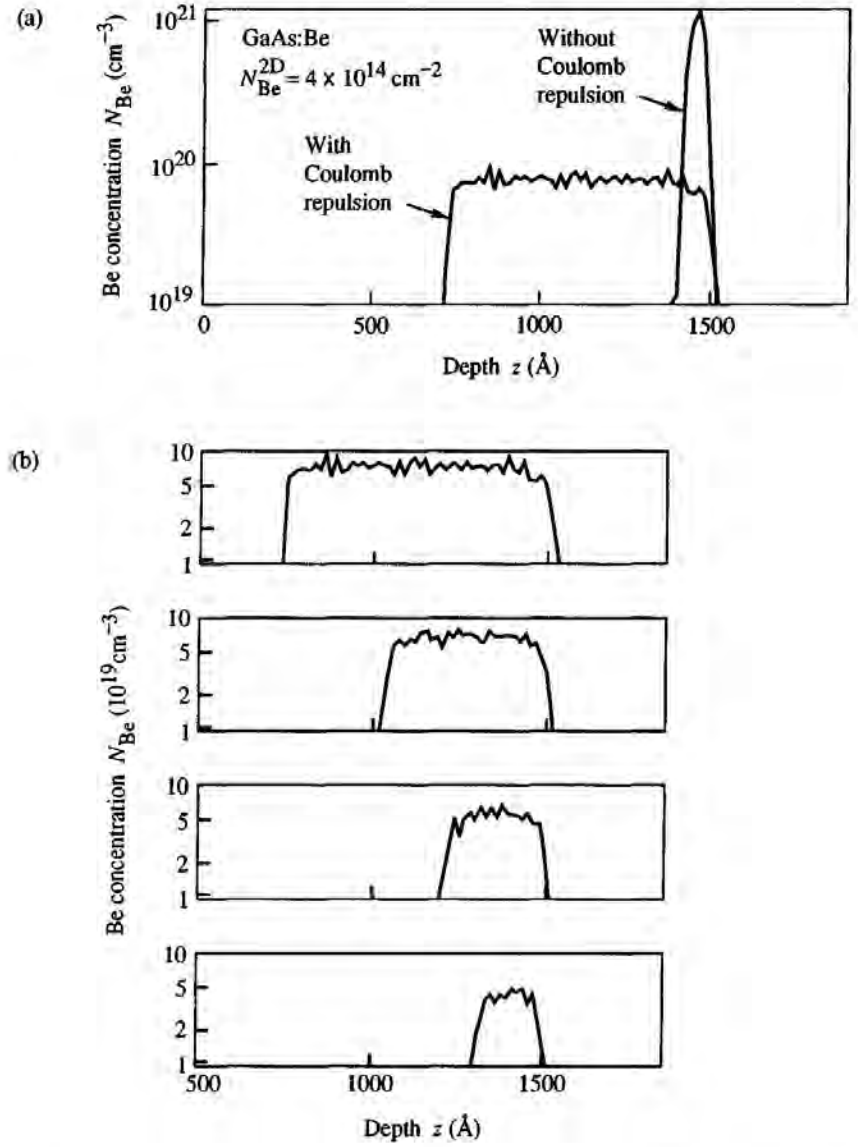


Fig.7.17. Calculated impurity distribution in a semiconductor doped during growth interruption (a) with and without taking into account Coulomb correlation effects. A growth rate of 1Å/s and a diffusion coefficient of 2×10^{16}cm^2/s was used in the calculation. (b) Calculated profiles for different Be doses.

7.5.3 Compensating native defects

Native defects in III–V semiconductors such as vacancies, interstitials, and antisite defects can have compensating characteristics. That is, many of the native defects are positively charged in p-type semiconductors and negatively charged in n-type semiconductors. The change in the charge state of these defects

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

reduces the concentration of free carriers. Furthermore, the abundance of such compensating native defects can increase with doping concentrations. Thus, for sufficiently high concentrations of the defects, the free carrier concentration in semiconductors can be limited to a maximum value.

The formation energy and the charge states of several native defects was calculated by Baraff and Schlüter (1985). As an example, we consider the formation of a Ga vacancy and a Ga interstitial by moving a substitutional Ga atom to an interstitial site in a GaAs lattice. The reaction equation is given by

$$0 \leftrightarrow V_{Ga} + I_{Ga}^{Td} \tag{7.57}$$

where V_{Ga} is a Ga vacancy; I_{Ga}^{Td} is a Ga interstitial which occupies an interstitial site of tetrahedral symmetry. The equilibrium concentration of the native defects is obtained from the law of mass action

$$[V_{Ga}][I_{Ga}^{Td}] = c \exp(-\Delta G/kT) \tag{7.58}$$

where the brackets [...] refer to the concentration of the respective defect and ΔG is the change in Gibbs free energy associated with the reaction. The formation energy ΔE is typically much larger than the entropy term $T\Delta S$ which can therefore be neglected.

It proved useful to represent the total formation energy as a sum of a *chemical* energy term and an *electronic* energy term. Assume that the defect has the charge state Z with $Z = -2, -1, 0, +1, +2$ and so on. The electronic energy term is then given by $-(E_F - E_D)Z$ where E_F and E_D are the Fermi-level and the defect-level energy and Z is the charge state of the defect. If the defect can accept two electrons ($Z = -2$), the reduction in electronic defect energy amounts to $2(E_F - E_D)$ since two electrons can drop from the Fermi level to the defect level. The defect reaction energy can be written as

$$\Delta E(Z, E_F) = \Delta E(Z) - E_F \sum_n \sigma(E_F - E_{D,n}) \tag{7.59}$$

where $\Delta E(Z)$ is the *chemical* part of the formation energy and the second term of the right-hand side describes the *electronic* part of the energy. The sum of the second part is carried out over all energy levels, $E_{D,n}$, of the defect below the Fermi level E_F; σ is the step function which is zero for $E_F < E_{D,n}$ and unity for $E_F \geq E_{D,n}$. If the defect contains Z occupied levels and the occupancy remains constant in the range of Fermi levels considered, then Eq. (7.59) simplifies to (Baraff and Schlüter, 1985)

$$\Delta E(Z, E_F) = \Delta E(Z) - E_F Z \tag{7.60}$$

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

where Z is the charge state of the defect. The energy $\Delta E(Z)$ is the total energy of the defect in the charge state Z and is independent of E_F.

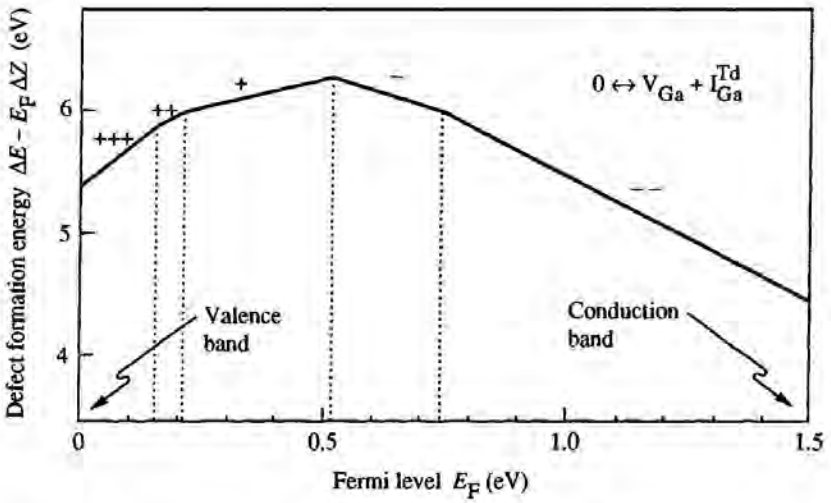


Fig. 7.18. Total reaction energies, $(\Delta E - E_F \Delta Z)$, for the defect reaction $0 \leftrightarrow V_{Ga}+I_{Ga}^{Td}$, where I_{Ga}^{Td} represents an interstitial Ga atom with tetrahedral symmetry surrounded by four As atoms (after Baraff and Schlüter, 1985).

The total defect formation energy for the Ga vacancy/interstitial defect in GaAs is shown in Fig. 7.18 as a function of the Fermi level. The reference level $E_F = 0$ is at the top of the valence band. Figure 7.18 illustrates two important features of the native defect. First, the charge state becomes more negative as the Fermi level increases. For $E_F \equiv E_V$ (p-type GaAs), the defect is positively charged, i.e. traps holes; for $E_F \equiv E_C$ (n-type), the defect traps electrons. The defect thus has a ***compensating*** nature regardless of the conductivity type of the semiconductor. Second, the total defect formation energy continuously decreases for either highly n-type or p-type conductivity. According to the law of mass action (Eq. (7.58)), the abundance of defects increases for high doping concentrations where $E_F \cong E_C$ or $E_F \cong E_V$, i.e. a lower formation energy prevails. Both of these characteristics clearly contribute to the compensation of intentional shallow impurities in semiconductors. The compensating nature was also found for other native defects in GaAs such as antisite defects and As vacancies (Baraff and Schlüter, 1985). Furthermore, the formation energy of compensating defects was found to decrease for Fermi levels close to the band edges (see Fig. 7.18). This characteristic was found for defects that are likely to occur in As-rich GaAs (e.g.

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

As_{Ga}) as well as in Ga-rich (e.g. Ga_{As}) GaAs.

Walukiewicz (1989) proposed that the nature of compensation native defects explains the free carrier saturation occurring at high doping concentrations. The author considered the two defect reactions in GaAs

$$V_{Ga} + As_{As} \leftrightarrow As_{Ga} + V_{As} \tag{7.61a}$$

$$Ga_{As} + V_{Ga} \leftrightarrow V_{As} + Ga_{Ga} \tag{7.61b}$$

and showed that the defects have a donor-like (acceptor-like) characteristic in p-type (n-type) material, i.e. compensate the intentional doping. Thus, the defects have the tendency of shifting the Fermi level away from the band edges into the forbidden gap. For a sufficiently high defect concentration, the Fermi level may be stabilized at some energy which is close to the middle of the fundamental gap in GaAs. This energy is called the stabilized Fermi level, E_{FS}. If the Fermi level moves away from E_{FS}, the formation energy of the compensating native defects decreases. As a result, more compensating defects are formed which opposes the movement of the Fermi level. At some Fermi level the rate of defect generation equals the increase in free carrier concentration which results in a saturation of the free carrier concentration.

The position of the stabilized Fermi level is of critical importance to the maximum achievable concentration. If E_{FS} is already close to the conduction band, the Fermi level can move into the conduction band without an abundance of compensating defects being formed. As a result, high values of n/N_c can be achieved, where N_c is the effective density of states at the bottom of the conduction band. If, however, E_{FS} is far below the conduction band edge, an abundance of compensating defects are formed, if the Fermi level moves into the conduction band. As a result, only low values of n/N_c can be achieved. Similar conclusions can be drawn for the position of E_{FS} relative to the valence band edge in p-type semiconductors.

A systematic study of the maximum carrier concentrations in semiconductors was published by Tokumitsu (1990). Figure 7.19 shows the stabilization position of the Fermi level in several elemental and compound semiconductors. The stabilization position is within the forbidden gap for most of the semiconductors; E_{FS} is within the conduction band for InAs. The author found that high values of n/N_c were achieved in semiconductors in which $E_C - E_{FS}$ is small; these semiconductors include InAs, $Ga_{0.47}In_{0.53}As$, and InP. On the other hand, high values of p/N_v were achieved in semiconductors in which $E_{FS} - E_V$ is small such as GaSb, Ge, Si, and GaAs. The experimental data was fitted to an exponential function. The maximum carrier concentration was expressed by

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

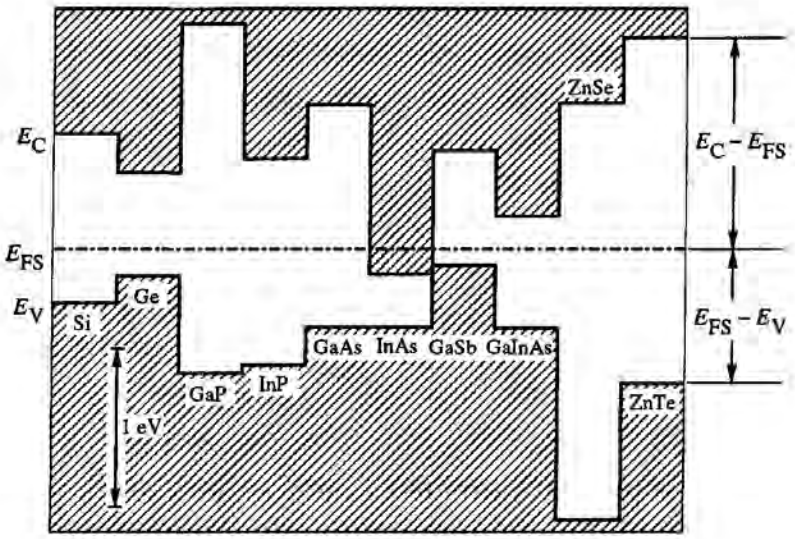


Fig. 7.19. Schematic illustration of the Fermi-level stabilization energies within the band gap of different semiconductors (after Tokumitsu, 1990).

$$n/N_c = 2.7 \times 10^3 \, e^{-5.5(E_C - E_{FS})} \text{ (n-type)}, \tag{7.62a}$$

$$p/N_v = 4.0 \times 10^2 \, e^{-6.1(E_{FS} - E_V)} \text{ (p-type)}. \tag{7.62b}$$

These empirical relations yielded good agreement with many experimental values for the maximum free carrier concentrations.

The model of compensating native defects is a general model in which the species of the impurity does not enter the model. Differences found between different impurity elements in the same semiconductor cannot be explained by the model. For example, the maximum hole concentration achievable in p-type GaAs varies between 10^{19} cm^{-3} (GaAs:Zn) and 10^{21} cm^{-3} (GaAs:C). Such differences are not expected on the basis of the model.

7.5.4 Limitations for amphoteric dopants

Amphoteric impurities are known to autocompensate at high impurity concentrations. The amphoteric impurities C, Si, Ge, and Sn can occupy either cation or anion sites in the zincblende lattice on which they act as donors or acceptors, respectively. The conditions of impurity incorporation can usually be selected in such a way that the incorporation of impurities occurs predominantly on one preferred site. For example, C occupies predominantly acceptor sites in GaAs when incorporated during epitaxial growth.

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

We next consider the incorporation at high doping concentrations and use the result of the law of mass action for the impurity incorporation (see Eq. (7.38))

$$\frac{N_A^-}{N_D^+} = K^* n^2 . \tag{7.63}$$

We assume that the amphoteric Si impurity can only be a donor or an acceptor, i.e. $N_{Si} = N_D^+ + N_A^-$, and that both types of impurities are always ionized. Using $n = N_D^+ - N_A^-$ and $x = N_A^- / N_{Si}$, Eq. (7.63) can be written as

$$x = K^* N_{Si}^2 (1 - 2x)^2 (1 - x) . \tag{7.64}$$

Solving this equation for very large Si concentrations yields the solution $x = N_A^- / N_{Si} = ½$. Thus, autocompensation increases with doping concentration until the number of impurities incorporated as donors equals those incorporated as acceptors, i.e. $N_A^- \cong N_D^+$. The free carrier concentration then ***saturates***, i.e. a maximum of the free carrier concentration is reached.

For a given saturation concentration the constant K^* in Eq. (7.63) can be determined. Using a saturation concentration of 10^{18} cm^{-3} for Si in GaAs and $N_A^- \cong N_D^+$, one obtains $K^* \cong 10^{-36}$ cm^6. Using this value for the constant in Eq. (7.63) allows us to quantitatively estimate the compensation ratio at lower concentrations. As an example, we choose $n = 10^{16}$ cm^{-3} and obtain $N_A^- / N_D^+ = 10^{-4}$. This ratio is lower than the compensation ratio determined by optical techniques (Lu et al., 1990; Kamiya and Wagner, 1977) but in reasonable agreement with a recent study by Harris and Schubert (1990).

Another characteristic of amphoteric impurities is the tendency of pair formation. The pair formation can be driven by several effects including chemical, strain or coulombic effects. Coulombic interaction effects can be quantified in a straightforward way. Assume a semiconductor with randomly distributed positive donor ions and randomly distributed electrons. The coulombic interaction energy is zero for such a neutral semiconductor.* If, however, two impurities, one being a donor and the other one being an acceptor, form a closely spaced pair, a reduction in coulombic energy of $e^2/4\pi\varepsilon r$ results. For a nearest neighbor pair $r \cong 2.4$ Å, which yields an energy of $\cong 460$ meV. Considering the magnitude of this energy, pairing of amphoteric impurities is a quite attractive mechanism. (The chemical bond energies and strain energies may also change drastically for pairs but this is neglected here). The probability of impurity-pair formation increases *quadratically* with the impurity concentration

* Note that electrons are not distributed randomly due to correlation effects and screening. As a result the net coulombic interaction energy is *cohesive* for any doped semiconductor.

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

337-TA-1366 RX-0255

since *two* impurities are required to form a pair. Thus, the pair formation is expected to increase with impurity concentration. At some concentration all impurities incorporate as paired species. As a result, the free carrier concentration saturates or may even decrease for increasing doping concentration. Such a decrease has been observed in Si-doped GaAs (See Chap. 5).

It is necessary to compare the changes in electronic and coulombic energies with the difference in formation energies. As shown in Sect. 7.4, the difference in formation energies between the two sites of the GaAs lattice is typically several hundred meV. For C in GaAs the difference was estimated to be $\Delta E \cong 690$ meV. Other amphoteric impurities such as Si, Ge, and Sn have probably lower values of ΔE due to their higher atomic weight, i.e. lower reactivity. An energy difference of several 100 meV for ΔE is indeed comparable in magnitude to the increase in electronic energy and to the coulombic pair formation energy. The roughly comparable magnitude in energy suggests that the two mechanisms can impose a limit on the maximum free carrier concentration in III–V semiconductors doped with amphoteric impurities.

7.5.5 Concentration limited by DX-centers

A model for the maximum free electron concentration based on the DX-center was proposed by Theis et al. (1988). The model was proposed for GaAs in which DX-centers are known to form localized states which are in resonance with the conduction band. Substitutional donors can undergo a transition in which the initially substitutional donor moves to a threefold-coordinated interstitial site (Chadi and Chang, 1988). The displacement of the shallow donor from its substitutional site is accompanied by the breakage of one donor atom bond and the transfer of the electrons to the DX-center. The energy levels of shallow donors and DX-centers are schematically illustrated in Fig. 7.20. Shallow donor levels are indicated as well as the DX-levels which are within the conduction band. For increasing free carrier concentration the Fermi level rises until it coincides with the DX-level. At this point a substitutional donor undergoes lattice relaxation and captures two electrons according to (Chadi and Chang, 1988)

$$2D^0 \leftrightarrow 2D^+ + 2e \leftrightarrow D^+ + DX^- \qquad (7.65)$$

where D and DX represent the shallow donor and the DX-center configuration. The properties of the DX-center will be discussed in more detail in Chap. 9.

The Fermi level cannot exceed the DX-center level since electrons of additional donors are captured by donors undergoing the transition to DX-centers.

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

Innoscience1366-00138999

RX-0255.326

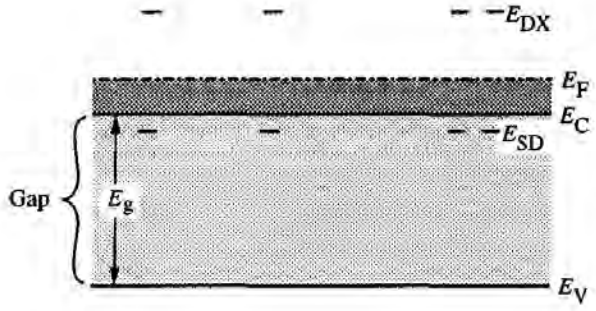


Fig. 7.20. Schematic illustration of shallow donor levels and DX levels of substitutional Si impurities in GaAs.

Additional electrons are captured according to Eq. (7.65). Using degenerate statistics, the maximum free electron concentration is given by (see Chap. 3)

$$n_{max} = \frac{1}{3\pi^2}\left[\frac{2m^*(E_{DX} - E_C)}{\hbar^2}\right]^{3/2} \tag{7.66}$$

where E_{DX} is the energy of the donor in the DX-configuration. As an example, we assume $E_{DX} - E_C = 200$ meV, $m^* = 0.067\ m_0$, and a single, isotropic, parabolic conduction band minimum; the maximum free carrier concentration is then obtained from Eq. (7.66) as $n_{max} \cong 7 \times 10^{18}$ cm^{-3}. Theis et al. (1988) extrapolated a maximum free electron density in GaAs limited by the DX-configuration to be $1.5 \times 10^{19} - 2 \times 10^{19}$ cm^{-3}.

The energy level of the DX-configuration depends on the doping concentration. Etienne and Thierry-Mieg (1988) found that the DX-level moves away from the conduction band minimum at high doping concentrations. The authors further found in $Al_xGa_{1-x}As$ that the DX-center and the persistent photoconductivity effects (see Chap. 9) *decrease* rather than increase with doping density. These findings together with the fact that strong persistent photoconductivity has not been found in highly doped GaAs indicate that the DX-center is not the dominant mechanism limiting the free carrier concentration in n-type GaAs.

7.5.6 Solubility and strain

The ***solid solubility limit*** is not well known for impurities in III–V semiconductors. In contrast, the solid solubility limit of impurities in the group-IV semiconductor Si have been well documented (Beadle et al., 1985) for shallow donors and acceptors. The solid solubility limits the concentration of impurities in Si and applies to high-temperature growth from the melt. However, epitaxial III–V semiconductors are rarely grown from the melt (except in liquid-phase

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

Innoscience1366-00139000

epitaxy) Most epitaxial III–V semiconductors are grown from the gas phase at lower temperatures such as 400–700 °C. It is questionable if the solubility limit can be applied at such low temperatures. The solid solubility is a property of equilibrium which is not necessarily reached at low temperatures.

The solid solubility limit of one compound A in another compound B is defined as the maximum concentration of A in B at which A is homogeneously distributed in B without any precipitates. The presence of precipitates of a doping element in a semiconductor can thus be used as an indicator that the solid solubility limit has been exceeded. Precipitates can be detected in III–V semiconductors, for example, by transmission electron microscopy (TEM) or by secondary ion mass spectrometry (SIMS). There have been a number of reports on the presence of precipitates or impurity clusters in III–V semiconductors, for example Si precipitates in GaAs (Chai et al., 1981), or Be precipitates in $Ga_{0.47}In_{0.53}As$ (Hamm et al., 1989). However, the occurrence of precipitations depends strongly on the growth conditions such as substrate temperature and V/III flux ratio. It is not useful to establish a certain concentration as a solubility limit if this 'limit' depends strongly on growth conditions. Furthermore, impurity concentrations were found to be limited to some concentration without the observation of precipitates (Schubert et al., 1990c; Panish, 1990). The limitation of impurity concentrations by the 'solubility limit' has to be considered with caution.

Elastic ***strain*** is another physical mechanism that can limit the impurity concentration in semiconductors. An impurity atom with a different atomic size than the host lattice atom either exerts *compressive* or *tensile* strain on the semiconductor lattice, if the impurity is larger or smaller than the lattice atom, respectively. To accommodate the impurity on a lattice site, the strain energy must be provided during the incorporation process. If the strain energy (per impurity atom) strongly increases with impurity concentration, then impurities may not incorporate at all above some critical concentration. In order to evaluate the concentration dependence, the interaction energy must be calculated. Morgan (1972) investigated the elastic interaction between two defects separated by a distance r. The author found that the strain interaction energy is proportional to r^{-3}. The absolute magnitude of this energy is not known in covalent materials. Lannoo and Bourgoin (1981) assumed that it is similar in covalent materials and metals. An interaction energy of <10 meV was found for a defect separation of 10 Å. This relatively small interaction energy indicates that elastic strain can limit the impurity concentration only in the extreme concentration regime, i.e. for concentrations $\geq 10^{21}$ cm^{-3} ($r \leq 10$ Å) where the interaction energy assumes higher values.

https://doi.org/10.1017/CBO9780511599828.010 Published online by Cambridge University Press

8

Redistribution of impurities

Spatial redistribution of impurities in semiconductors occurs at sufficiently high temperatures. That is, impurities may not remain on the site of their initial incorporation. As discussed in the previous chapter, the diffusive redistribution can even be used to incorporate impurities from the surface into the bulk of the semiconductor. However, impurity redistribution is most frequently an undesired effect; it leads to the smearing of doping profiles or, for example, can result in the displacement of the location of a pn-junction.

Redistribution of impurities can be ***isotropic*** or ***anisotropic***. Conventional diffusion is an example for isotropic redistribution. On the other hand, anisotropic redistribution of impurities occurs in the presence of energy gradients. Such gradients can be the result of electric potential gradients, strain gradients, surfaces, etc. Preferential redistribution of impurities occurs along the energy gradient, i.e. towards regions of lower energy.

The vicinity of a semiconductor surface can lead to preferential redistribution of impurities. Some impurities tend to remain on the surface during epitaxial growth and do not readily incorporate into the bulk. The impurities ***segregate*** to the ***surface***. Surface segregation is an undesired effect since it complicates the growth of well-defined impurity distributions. The effect of preferential redistribution of impurities (in the vicinity of a surface) towards the surface is called ***surface migration***. This can be the result of potential gradients or defect gradients in the vicinity of a surface.

The mechanisms of impurity redistribution can be quite different in nature. ***Substitutional diffusion*** occurs if an impurity hops from one lattice site to a second, vacant lattice site. During ***substitutional–interstitial diffusion***, an impurity leaves its substitutional site and hops to an interstitial lattice site, and

https://doi.org/10.1017/CBO9780511599828.011 Published online by Cambridge University Press

337-TA-1366 RX-0255

then again to a vacant lattice site. The impurity on the interstitial site may also ***kick-out*** a regular lattice atom and take its place. The exact mechanism of diffusion plays, as we shall see later, an important role in the dependence of the diffusion mechanism on impurity concentration.

8.1 Diffusion and brownian motion

Diffusion of impurities and defects in semiconductors is the result of random hops of the impurity atoms and defects. The random hops are thermally activated. The thermally stimulated random motion of particles is known as ***brownian motion*** and was first observed for small carbon particles floating on water. The similarity of brownian motion and solid-state diffusion was pointed out by Einstein (1905, 1906) who developed the *theory of brownian motion*. An exhaustive review on the theory of brownian motion was given by Risken (1984).

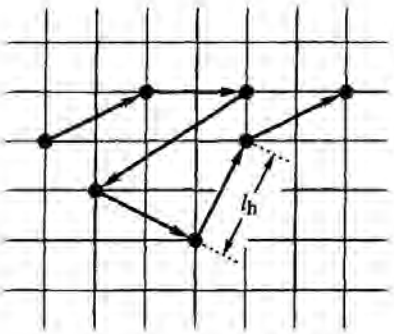


Fig. 8.1. Illustration of diffusion hops of an impurity atom between substitutional sites of a lattice. The average hop distance in the illustration is approximately two lattice constants.

The random movement of an impurity in the lattice matrix of a semiconductor is schematically illustrated in Fig. 8.1. The impurity hops from lattice site to lattice site. The ***average hop length*** is denoted as l_h and is approximately two lattice constants for the example shown in Fig. 8.1. We now consider a single, successful hop of an impurity from one lattice site to a second, *vacant* lattice site. The diffusion hop can be characterized by an ***attempt rate*** and a ***success probability***. The attempt rate is the number of attempts (per unit time) of the impurity to leave its lattice site. The attempt rate is given by the vibrational frequency, i.e. each vibration of the impurity atom represents one attempt to leave the lattice site. The attempt frequency is given by

$$\hbar\omega = \hbar 2\pi f = (3/2)\ kT\ . \tag{8.1}$$

At a temperature of 600 °C, the attempt rate obtained is $f = 2.7 \times 10^{13}\ \mathrm{s}^{-1}$.

The probability that the attempt is successful is modeled by an energy barrier

https://doi.org/10.1017/CBO9780511599828.011 Published online by Cambridge University Press

RX-0255.330

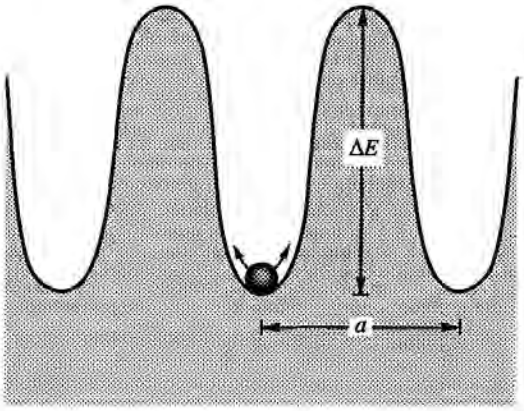


Fig. 8.2. Diffusion of impurity atoms in a periodic crystal potential with period a occurs via random jumps to adjacent energy minima. The energy barrier between the minima is ΔE.

that the impurity atom must overcome in order to leave the original lattice site. The success probability is given by

$$p = \mathrm{e}^{-\Delta E/kT} \tag{8.2}$$

where ΔE is the magnitude of the energy barrier. A particle in a periodic potential is schematically shown in Fig. 8.2. In order to reach an adjacent lattice site, an energy barrier of magnitude ΔE must be overcome. For an energy barrier of vanishing magnitude, $\Delta E = 0$, the success probability equals unity, $p = 1$, i.e. each attempt would be a successful attempt. The rate of successful diffusion hops is given by the product of attempt rate and success probability. The ***diffusion constant D*** (also called diffusion coefficient) of an impurity is defined as (see, for example, Risken, 1984)

$$D = l_{\mathrm{h}}^2 \, fp \tag{8.3}$$

where fp is the rate of successful diffusion hops and l_{h} is the average hop length. Using Eq. (8.2) the diffusion constant can be written as

$$\boxed{D = D_0 \, \mathrm{e}^{-\Delta E/kT}} \tag{8.4}$$

where $D_0 = l_{\mathrm{h}}^2 f$. The temperature dependence of D_0 is much weaker as compared to the dependence of the exponential term $\exp(-\Delta E/kT)$. It is therefore frequently assumed that D_0 is a constant, i.e. independent of temperature. Comparison of $D_0 = l_{\mathrm{h}}^2 f$ with Eq. (8.3) reveals that D_0 is the diffusion constant if all hop attempts were successful, i.e. $p = 1$.

https://doi.org/10.1017/CBO9780511599828.011 Published online by Cambridge University Press

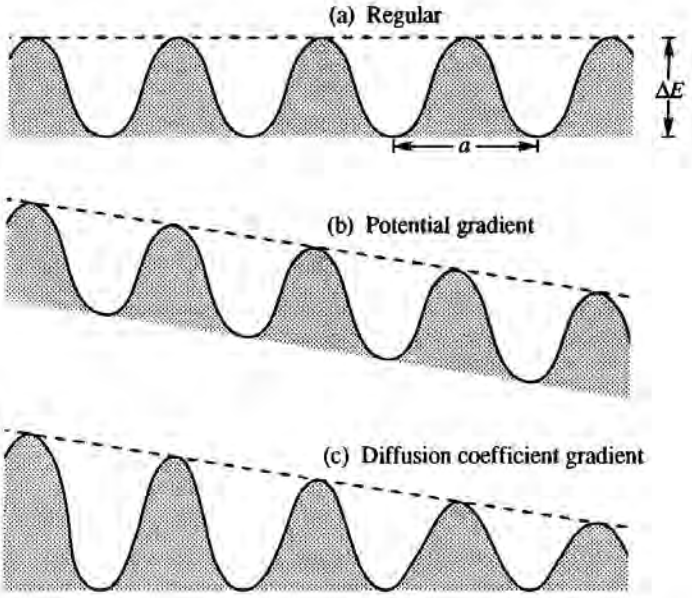


Fig. 8.3. Schematic energy diagram of a periodic crystal potential for (a) regular diffusion, (b) diffusion in a potential gradient (electric field), and (c) diffusion in a potential with a diffusion coefficient gradient, i.e. a spatially changing diffusion coefficient.

Pure diffusion is isotropic due to the random nature of the diffusion process. Anisotropic diffusion results in the presence of additional energy gradients. Consider a positively charged impurity in an electric field. A diffusion hop along the field distribution decreases the electrostatic energy by $e\mathcal{E}\,l_h$, while the diffusion against the field direction increases the energy by $e\mathcal{E}\,l_h$, where $\mathcal{E}$ is the electric field. Thus, a diffusion hop along the field direction is more likely than against the field direction. Diffusion events along the field direction are not impeded by the field. However, the probability for a successful attempt against the field direction is reduced by $\exp(-e\mathcal{E}\,l_h/kT)$ in analogy to Eq. (8.2). The preferential direction of the diffusion hops results in a net motion of the positively charged impurities along the field direction. The ***drift velocity*** of the impurities is characterized by a mobility μ and is given by

$$v = \mu\mathcal{E}. \tag{8.5}$$

The mobility μ is related to the diffusion constant according to

https://doi.org/10.1017/CBO9780511599828.011 Published online by Cambridge University Press

$$\boxed{\mu = \frac{eD}{kT}} \tag{8.6}$$

which is called the **Einstein relation** (Einstein, 1905, 1906). It relates the random, isotropic diffusive motion with the directed motion in the presence of a potential energy gradient. The periodic crystal potentials associated with regular diffusion and with diffusion in a potential gradient are schematically shown in Fig. 8.3(a) and (b), respectively. It is evident that diffusion in a potential gradient occurs preferably to the low-energy side, i.e. right-hand side of Fig. 8.3(b). A potential resulting in a spatially varying diffusion constant is shown in Fig. 8.3(c). The height of the energy barriers changes, resulting in a spatially changing diffusion coefficient. Preferential diffusion towards the side with the lower barrier results. That is, diffusion occurs preferentially along the ***diffusion constant gradient***. Diffusion in a material with a diffusion constant gradient will be analyzed quantitatively in the next section.

Up to now we have considered the diffusion process from an initial lattice site to a final site. It was implicitly assumed that the final site is vacant in order to be able to accommodate the diffusing impurity. The diffusion constant D has been used as a purely phenomenological constant, i.e. without considering the microscopic nature of the diffusion process. The microscopic mechanisms of diffusion processes are, however, of great importance. The dependence of the diffusion constant D on crystal parameters, for example the vacancy concentration, can be derived from the nature of the microscopic diffusion process. As an example, we consider the diffusion of an impurity on substitutional lattice sites. First, a vacant lattice site must be available for a successful diffusion hop. That is, the diffusion constant is proportional to the *concentration of vacancies*. Second, if there is no vacancy in the vicinity of the impurity, it cannot make a diffusion jump. However, a vacancy can diffuse into the vicinity of the impurity, thus enabling the impurity to make a successful hop. Therefore, the impurity diffusion constant is also proportional to the *vacancy diffusion constant* D_V (Manning, 1968; see also Sects. 8.3.1 and 8.6), i.e.

$$D \propto [V] D_V \tag{8.7}$$

where [V] is the concentration of vacancies. Depending on the detailed diffusion mechanism, other factors may also enter. Different diffusion processes are discussed further in Sect. 8.3.

https://doi.org/10.1017/CBO9780511599828.011 Published online by Cambridge University Press

8.2 Fickian diffusion

In the simplest form of diffusion, impurities make random atomic jumps and the movement of impurities is *uncorrelated*. For such uncorrelated motion the movement of impurities does not depend on other impurities, that is the diffusion is *independent of the concentration* of the impurities in the semiconductor. If the atomic jumps are random and uncorrelated then the net motion of impurities is proportional to the concentration gradient. The impurity flux (per unit time and unit area) is then given by

$$j(x,y,z) = -D\ \nabla N(x,y,z) \tag{8.8a}$$

where $N(x,y,z)$ is the impurity concentration and D is the diffusion constant. Even though D depends on the temperature, D is a *constant* with respect to the spatial variables x, y, and z of the differential equation. For an impurity distribution which depends only on the z-direction of a cartesian coordinate system, the equation simplifies to

$$\boxed{j(z) = -D\ \frac{\mathrm{d}}{\mathrm{d}z}\ N(z)} \tag{8.8b}$$

which is ***Fick's first law***. Note that $j(z)$ is the flux of impurities in the z-direction and the flux depends only on z. Fick's first law states that the impurity flux is proportional to the impurity gradient. If the gradient is zero (i.e. constant impurity concentration for all x,y,z), then there is no net movement of impurities ($j=0$) despite the many random atomic jumps which occur.

In the simplest case D is a constant and does not depend on the spatial position or on the impurity concentration. In this case, Eq. (8.8) is a linear differential equation of first order, and the diffusion process is referred to as ***linear fickian diffusion***. However, the diffusion constant may depend on the position and on the impurity concentration, i.e. $D = D(N,z)$. This case is referred to as ***non-linear fickian diffusion***. In both cases, the impurity flow is along the concentration gradient. The flow along the gradient is the result of random, uncorrelated hops. As an example, consider two regions 1 and 2 with impurity concentrations N_1 and N_2. If $N_2 > N_1$, there will be a net flow from 2 to 1 because there are more impurities in 2 than in 1 and consequently the hop rate $2 \rightarrow 1$ will be larger than $1 \rightarrow 2$.

In addition to Fick's first law, the ***continuity equation*** must be taken into account. For one-dimensional impurity distributions that depend only on the z-direction, the continuity equation is given by

https://doi.org/10.1017/CBO9780511599828.011 Published online by Cambridge University Press

337-TA-1366 RX-0255

$$\boxed{\frac{\partial}{\partial z} j(z) = -\frac{\partial}{\partial t} N(z,t)} \tag{8.9}$$

The equation states the fact that the increase (decrease) of impurities in any volume element equals the net flow of impurities into (out of) the volume element per unit time.

Insertion of Eq. (8.8) into Eq. (8.9) yields

$$\frac{\partial}{\partial t} N(z,t) = \frac{\partial}{\partial z} D \frac{\partial}{\partial z} N(z,t) . \tag{8.10a}$$

Under the assumptions that (i) D is a constant which does not depend on z, or that (ii) the dependence of D on z is much weaker than the dependence of N on z, Eq. (8.10a) simplifies to the linear, second-order partial differential equation

$$\boxed{\frac{\partial N(z,t)}{\partial t} = D \frac{\partial^2 N(z,t)}{\partial z^2}} \tag{8.10b}$$

which is known as ***Fick's second law***. In the following we consider the solution of Fick's second law for some specific impurity distributions. Specifically, the cases of constant impurity density and constant impurity surface concentration will be analyzed.

We first consider the case of a ***finite reservoir***, that is the total number of diffusing atoms does not change with time. Assume a semiconductor which has a sheet of impurities at $z = 0$ in the xy-phase. The two-dimensional density of the impurities is denoted as N^{2D} (in units of cm^{-2}). Such a doping distribution can be expressed in terms of the Dirac δ-function

$$N(z) = N^{2D} \delta(z) . \tag{8.11}$$

Upon diffusion, i.e. the random, uncorrelated motion of impurities, the impurity distribution broadens.

The solution of Fick's second equation with an initial δ-function distribution is given by the ***gaussian distribution***. After impurities with a diffusion constant D diffuse for a time t in a semiconductor, the impurity distribution is given by

$$\boxed{N(z) = \frac{N^{2D}}{\sqrt{2\pi}\sqrt{2Dt}} \exp\left[-\frac{z^2}{4Dt}\right]} \tag{8.12}$$

https://doi.org/10.1017/CBO9780511599828.011 Published online by Cambridge University Press

RX-0255.335

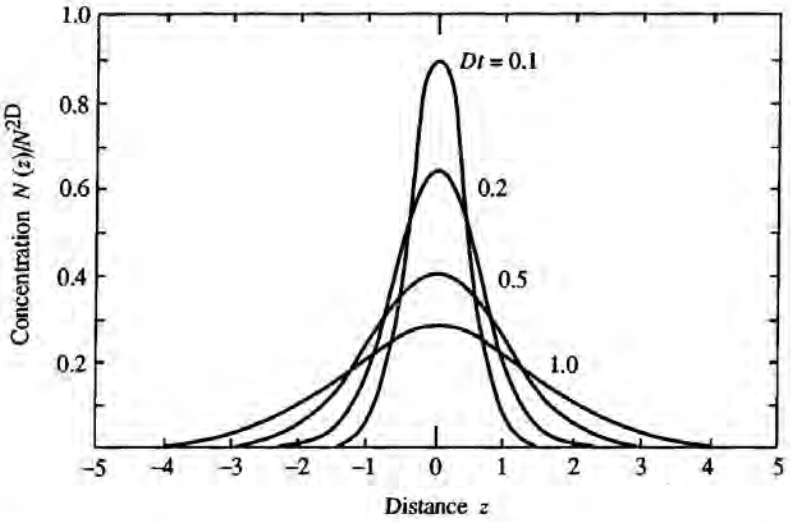


Fig. 8.4. Gaussian distribution of diffusing impurities for different values of Dt. For $t \to 0$ the impurity distribution becomes δ-function-like.

which is a solution of Fick's second law. The quantity $L_D = \sqrt{Dt}$ is called the ***diffusion length***. The gaussian impurity distribution is shown in Fig. 8.4 for different values of $L_D = \sqrt{Dt}$. The standard deviation of the gaussian distribution, σ, is related to the diffusion length by $\sigma = \sqrt{2}\, L_D = \sqrt{2Dt}$. The full-width half-maximum (FWHM) of the gaussian distribution is given by $\text{FWHM} \cong 2.355\sigma \cong 2.355\sqrt{2Dt}$. Note that any doping distribution can be expressed as a superposition of δ-functions or gaussian functions. Since the diffusion processes are uncorrelated, the doping distribution after diffusion can be obtained by superposition of the diffused individual gaussian distributions.

Next we consider the diffusion of impurities in a semiconductor with ***constant surface concentration***, i.e. the concentration of impurities at some location $z = 0$ (surface) is maintained at all times (infinite reservoir). Such conditions can be achieved experimentally with a thin film of impurities on a semiconductor (e.g. a Si film on GaAs). If the total amount of impurities diffusing into the semiconductor is small as compared to the amount of impurities in the surface film, then the surface concentration remains approximately constant. The initial conditions of the diffusion with a constant surface concentration are given by

https://doi.org/10.1017/CBO9780511599828.011 Published online by Cambridge University Press

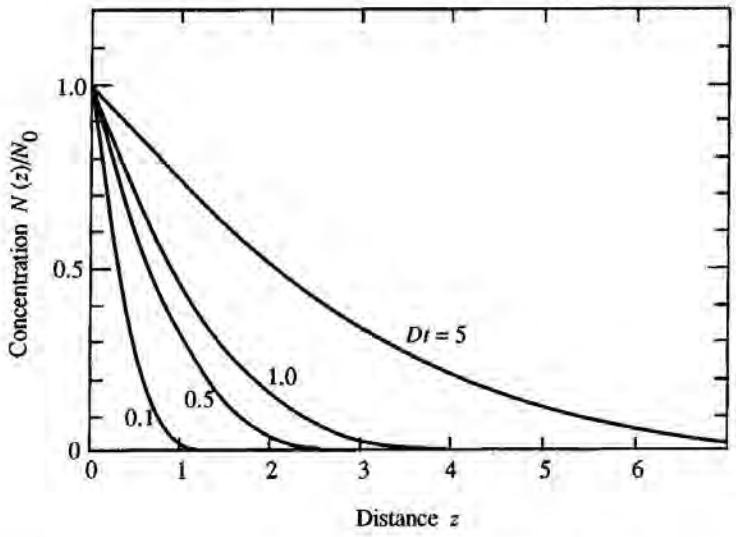


Fig. 8.5. Impurity concentration versus depth for different values of Dt. The impurity concentration at the surface $N(z = 0) = N_0$ is a constant. The impurity distribution is given by the error function complement (erfc).

$$N(z,t) = N_0 \quad \text{at } z = 0 \text{ for all } t\,, \tag{8.13a}$$

$$N(z,t) = 0 \quad \text{at } z > 0 \text{ and } t = 0\,. \tag{8.13b}$$

The solution of Fick's second law with the constant surface concentration boundary condition is given by the error-function-complement (erfc). The concentration of diffusing impurities with diffusion coefficient D at the location z and the time t is given by

$$\boxed{N(z,t) = N_0 \operatorname{erfc} \frac{z}{2\sqrt{Dt}}} \tag{8.14}$$

Examples of such diffusion profiles are shown in Fig. 8.5 for different values of Dt. The two analytical solutions of Fick's second law given above (Eqs. (8.12) and (8.14)) are the most relevant for practical diffusion problems. The distribution of impurities with more complex initial profiles can be synthesized by superposition of several simple (e.g. gaussian) distributions.

Fick's first and second laws were obtained for a diffusion constant that does not depend on the spatial position. This assumption may not be valid for all experimental conditions. In some cases, the diffusion constant may actually

https://doi.org/10.1017/CBO9780511599828.011 Published online by Cambridge University Press

337-TA-1366 RX-0255

depend on the spatial location. Consider a semiconductor whose vacancy concentration increases in the near-surface region. Such an increased vacancy concentration in the vicinity of the surface can be due to out-diffusion of anions or cations, which can occur during annealing without a protective cap. Since the substitutional diffusion coefficient is proportional to the vacancy concentration, the diffusion coefficient is expected to increase in the near-surface region and thus depends on the spatial coordinate. For the one-dimensional, z-dependent case

$$D = D(z) . \tag{8.15}$$

Fick's laws need to be modified in the case of a diffusion coefficient which depends on the spatial coordinate. ***Generalization*** of ***Fick's first law*** yields

$$j(z) = -D(z) \frac{\mathrm{d}N(z)}{\mathrm{d}z} - N(z) \frac{\mathrm{d}D(z)}{\mathrm{d}z} . \tag{8.16}$$

The new second term of the right-hand side of the equation represents a net flux of impurities to regions with a smaller diffusion coefficient. Impurities in a highly diffusive region diffuse rapidly until they reach a slowly diffusive region. As a result, the equilibrium impurity concentration will be *higher* in the region with a *smaller* diffusion coefficient.

Application of the continuity equation to Eq. (8.16) yields the ***second generalized fickian law*** as

$$\begin{aligned} \frac{\partial N(z,t)}{\partial t} &= \frac{\partial^2}{\partial z^2} [D(z)\, N(z,t)] \\ &= D(z) \frac{\partial^2 N(z,t)}{\partial z^2} + 2 \frac{\partial D(z)}{\partial z} \frac{\partial N(z,t)}{\partial z} + N(z,t) \frac{\partial^2 D(z)}{\partial z^2} . \end{aligned} \tag{8.17}$$

The third term of the sum is usually very small and can therefore be neglected. As an example, we consider the equilibrium impurity distribution in a semiconductor with a spatially varying diffusion constant. In equilibrium, $\partial N/\partial t = 0$, and consequently

$$\frac{\partial D(z)}{\partial z} N(z) = -D(z) \frac{\partial N(z)}{\partial z} . \tag{8.18}$$

The equation implies that a diffusion constant that decreases along some spatial direction results in an increased impurity concentration along that direction.

The linear fickian diffusion theory assumes uncorrelated, random diffusion hops of impurities. The assumption of uncorrelated motion is unlikely to be fulfilled at very high concentrations where pair formation may occur. Thus, in a

https://doi.org/10.1017/CBO9780511599828.011 Published online by Cambridge University Press

RX-0255.338

337-TA-1366 RX-0255

general treatment, the diffusion coefficient must be considered to be concentration dependent. Some diffusion mechanisms indeed yield such concentration dependence (see substitutional–interstitial diffusion in the following section). In this case the first generalized fickian equation (Eq. (8.16)) applies. It can be written as

$$j(z) = -D(z)\,\frac{\mathrm{d}N(z)}{\mathrm{d}z} - N(z)\,\frac{\mathrm{d}D[N(z)]}{\mathrm{d}N(z)}\,\frac{\mathrm{d}N(z)}{\mathrm{d}z}\,. \qquad (8.19)$$

If D has a non-linear functional dependence on N or z, then Eq. (8.19) becomes non-linear. Diffusion processes following Eq. (8.19) are called ***non-linear fickian*** diffusion processes. There are no general analytic solutions to the non-linear diffusion equation. Usually, numerical and iterative techniques are employed to solve it. Weisberg and Blanc (1963) solved the equation for a linear, quadratic, and cubic dependence of D on the concentration N (i.e. $D \propto N$, $D \propto N^2$, and $D \propto N^3$). A frequently used numerical method to solve the non-linear diffusion equation is the ***method of finite differences*** (see, for example, Tuck, 1988). The analysis of impurity profiles originating from non-linear diffusion is not straight forward. A method frequently invoked to analyze such diffusion profiles is the ***Boltzmann–Matano method***, which allows one to extract the concentration-dependence of D (see, for example, Tuck, 1988).

8.3 Diffusion mechanisms

The atomic mechanism for impurity diffusion in III–V semiconductors depends on the physical and chemical properties of the respective impurity and its host semiconductor. The magnitude of the diffusion process can depend on a large variety of parameters such as vacancy concentration, concentration of other point defects, atomic size of the impurity, free carrier concentration, position of the Fermi level, stoichiometry, etc. In the following, several diffusion mechanisms are discussed and their functional dependences are analyzed.

To illustrate the relevance of different characteristics, we consider two examples of diffusion processes, namely substitutional and interstitial diffusion of impurities, as schematically depicted in Fig. 8.6. In the substitutional diffusion (Fig. 8.6a) process, the impurity hops from one substitutional site to a second, vacant substitutional site. Since one vacancy is required for each diffusion hop, it is expected that the diffusion coefficient for this process is proportional to the vacancy concentration. In the interstitial process (Fig. 8.6b), an impurity hops from a first to a second interstitial site. It is reasonable to assume that atoms with a small atomic diameter can readily propagate via interstitial sites. The concentration of vacancies plays a minor role in interstitial diffusion. The two

https://doi.org/10.1017/CBO9780511599828.011 Published online by Cambridge University Press

RX-0255.339