High-Speed Electronics and Optoelectronics

Sheila Prasad Hermann Schumacher and Anand Gopinath

CAMERIDGE

MBRIDGE www.cambridge.org/9780521862837

This page intentionally left blank

High-Speed Electronics and Optoelectronics

This authoritative account of electronic and optoelectronic devices operating at frequencies greater than 1 GHz covers the concepts and fundamental principles of operation, and, uniquely, their circuit applications too.

Key features include:

- a comprehensive coverage of electron devices, such as MESFET, HEMT, RF MOS-FET, BJT and HBT, and their models;
- discussions of semiconductor devices fabricated in a variety of material systems, such as Si, III–V compound semiconductors and SiGe;
- a description of light-emitting diodes, semiconductor lasers and photodetectors;
- an executive summary at the beginning of each chapter;
- plentiful real-world examples; and
- end-of-chapter problems to test understanding of the material covered.

From crystal structure to atomic bonding, recombination and radiation in semiconductors to p–n junctions and heterojunctions, a wide range of critical topics is covered. Moreover, a chapter on analogue circuit applications provides an introduction to scattering parameter theory, followed by descriptions of different types of amplifier and oscillator utilising HBTs and HEMTs. Optimisation algorithms, such as simulated annealing and neural network applications, and parameter extraction of electronic device equivalent circuit models are also discussed. Graduate students in electrical engineering, industry professionals and researchers will all find this a valuable resource.

Sheila Prasad is Professor Emeritus in the Electrical and Computer Engineering Department at Northeastern University. Her current research interests include microwave and high-speed semiconductor devices and circuits, and optoelectronic circuits. She has co-authored the book *Fundamental Electromagnetic Theory and Applications* with Ronold W. P. King and has authored over 130 journal and conference publications.

Hermann Schumacher is Professor and Director of the Competence Center on Integrated Circuits in Communications, Institute of Electron Devices and Circuits, University of Ulm. He is also the Director of the International Master Program on Communications Technology at the University of Ulm, and has authored over 150 journal and conference publications.

Anand Gopinath is Professor in the Department of Electrical and Computer Engineering at the University of Minnesota. He is Life Fellow of the IEEE, Fellow of the OSA and Fellow of IET (UK). His research is in the areas of RF/microwave and optical semiconductor devices, integrated optics and metamaterials.

High-Speed Electronics and Optoelectronics: Devices and Circuits

SHEILA PRASAD

Northeastern University, Boston

HERMANN SCHUMACHER University of Ulm, Germany

ANAND GOPINATH University of Minnesota, Minneapolis



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

Cambridge University Press The Edinburgh Building, Cambridge CB2 8RU, UK

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

www.cambridge.org Information on this title: www.cambridge.org/9780521862837

© Cambridge University Press 2009

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

First published in print format 2009

ISBN-13 978-0-511-57724-6 eBook (Adobe Reader) ISBN-13 978-0-521-86283-7 Hardback

Cambridge University Press has no responsibility for the persistence or accuracy of urls for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Contents

2

	Preface		Page vii
	Ackno	owledgements	ix
Part Or	ne Devices	5	1
1	Revie	Review of semiconductor materials and physics	
	1.1	Executive summary	3
	1.2	Semiconductor materials	4
	1.3	Types of solids	5
	1.4	Crystal structure	5
	1.5	Crystal directions and planes	6
	1.6	Atomic bonding	8
	1.7	Atomic physics	9
	1.8	The de Broglie relation	11
	1.9	Quantum mechanics	12
	1.10	Statistical mechanics	16
	1.11	Electrons in a semiconductor	16
	1.12	The Kronig–Penney model	16
	1.13	Semiconductors in equilibrium	18
	1.14	Direct and indirect semiconductors	20
	1.15	Recombination and radiation in semiconductors	25
	1.16	Carrier transport in semiconductors	29
	1.17	p–n junction	30
	1.18	Schottky diode	34
	1.19	Heterostructures	35
	1.20	Silicon-germanium heterostructures	40
	1.21	Problems	43
	Refei	rences	45
2	Electronic devices		46
	2.1	Executive summary	46
	2.2	MESFET	46
	2.3	High electron mobility transistor	67

	2.4	Radio Frequency MOSFETs	95
	2.5	Bipolar and hetero-bipolar transistors	115
	2.6	Problems	156
	Refe	rences	158
3	Optimisation and parameter extraction of circuit models		163
	3.1	Executive summary	163
	3.2	Optimisation of device models	163
	3.3	Simulated annealing	164
	3.4	Neural networks applied to modelling	168
	3.5	Optimisation of neural networks by the genetic algorithm	180
	3.6	Structured genetic algorithm	181
	3.7	Semi-analytical device parameter extraction	184
	3.8	Basic expressions for small-signal parameter extraction	203
	3.9	Small-signal model of the collector-up (inverted) HBT	212
	3.10	Problems	214
	Refe	rences	218
4	Optoelectronics		221
	4.1	Executive summary	221
	4.2	Optical sources	221
	4.3	Photodetectors	261
	4.4	Problems	284
	Refe	rences	285
Part Tv	vo Circuits	3	289
5	Build	ling blocks for high-speed analogue circuits	291
	5.1	Executive summary	291
	5.2	Basic relations for two-port networks	291
	5.3	Noise in two-ports	306
	5.4	Transistor amplifiers	313
	5.5	Oscillators	383
	5.6	Mixers	396
	5.7	Baluns, unbals and hybrids	412
	5.8	Problems	419
	Refe	rences	421
	Index		423

Preface

Starting from the development of transistor technology to laser technology, the field of solid state devices and their circuit applications has advanced rapidly. The silicon bipolar junction transistor was first applied to low frequency circuits. The subsequent advances in materials science made it possible to fabricate compound semiconductor transistors capable of operating at microwave frequencies and high speeds. This presented the capability of applications in both analogue and digital circuits. At the same time, the wide choice of high performance semiconductor materials also enabled the development of optoelectronic devices such as lasers and light-emitting diodes. The communications industry continues to grow and diversify, thus necessitating the design of circuits which will satisfy the requirements of mobile telephones which are becoming more and more sophisticated in their performance. Circuit design has applications in other areas such as optical communications.

This book focusses on high-speed electronics and optoelectronics where the devices operate at frequencies $\geq 1 \text{ GHz}$. It is presented in two parts with devices being discussed in the first part and the circuit applications in the second part. In Part One, semiconductor devices fabricated in a variety of material systems – Si, III–V compound semiconductors and SiGe – are presented. We discuss the concepts and the fundamental principles of operation. We do not attempt to present the latest results as they will already be obsolete by the time the book is published. It is assumed that the reader has had a course in fundamental solid state physics.

Chapter 1 reviews semiconductor materials and physics. For the reader who is familiar with the topics, this chapter will be a brief review. If not, the reader can go to the references section to get a detailed coverage of the topics. Semiconductor materials are described followed by brief discussions of crystal structure and bonding. The section on quantum mechanics is intended to present only the important concepts and is not a comprehensive treatment of the subject. Semiconductor properties are described followed by types of semiconductors. Semiconductor junctions are treated in detail as they are the basis of the devices to be treated in subsequent chapters.

Chapter 2 presents high-frequency/high-speed electronic devices starting with the MESFET, which was the first transistor to operate at microwave frequencies. The development of the high electron mobility transistor (HEMT) represented a major advance in technology and is presented here in detail. The recent application of MOSFETs to radio frequency has been successful and the properties are covered in detail. Finally, bipolar and heterojunction bipolar transistors (HBTs) are described. Models for the transistors are presented and their method of implementation is described.

Chapter 3 presents the optimisation and parameter extraction of the circuit models of the electronic devices. The simulated annealing algorithm is discussed followed by the application of neural networks to circuit modelling. The genetic algorithm is defined and its application to optimisation is shown. Parameter extraction methods are given for circuit models using semi-analytical methods and basic expressions are derived.

Chapter 4 deals with various optical sources such as light-emitting diodes and lasers, giving details of their physical properties and their modes of operation. The discussion of emitters is followed by an extensive coverage of a variety of photodetectors.

In Part Two of the book, we discuss analogue circuits at the gate level. We will assume that the reader has a background (at the undergraduate level) in fundamental analogue circuit theory. Chapter 5 (Part Two of the book) deals with the components of high-speed analogue circuits. After a review of scattering parameter theory, the power and noise relations for two-port networks are discussed. Transistor amplifiers are covered in detail, showing the application of the devices described in Chapter 2. This is followed by a discussion of oscillators and mixers for high-speed circuits. Important passive components of high-speed circuits complete this chapter.

We have a layered approach to each chapter in the book. There is an executive summary at the beginning of each chapter. This will make the book valuable also for technical managers who may not want to go through the chapter content in detail. We have extensive problems at the end of each chapter, which will give the student applications of the theory. This book should be useful to research engineers and graduate students. Results from various research papers are presented, many of which are only available in journals which are referenced extensively. However, the reader need not go to the original papers as the results are given in sufficient detail to give a good understanding of the material.

Acknowledgements

Sheila Prasad

I would like to express my gratitude to Professor Clifton G. Fonstad, Jr, at the Massachusetts Institute of Technology. My long collaboration with him started with the first sabbatical leave at MIT when I worked in his group. It initiated my work on HBTs at microwave frequencies and the continued support he provided to me and my students in his laboratory resulted in this successful research. I acknowledge my colleague at Northeastern University, Dr Michael Vai (now at MIT Lincoln Laboratory), with whom I performed research on optimisation and modelling techniques. Many students worked with me on various aspects of the research reported in this book. I would particularly like to acknowledge the work of Dr Bin Li whose research results continue to be cited in publications. I would also like to acknowledge my student Kofi Deh for his help with the figures and manuscript editing. Dr Henry Choy and Dr Wojtek Giziewicz, both of whom were students at MIT, gave me invaluable suggestions for the book material and also helped me with MATLAB, graphics programmes and Latex when needed. I acknowledge my colleague at Northeastern University, Professor Jeff Hopwood (now at Tufts University), with whom I had many useful discussions about the content of the book. It has been a great experience to work with both of my co-authors. Last but not least, I would like to thank my husband, Fred Hinchey, for his great patience and support in the course of this book project.

Hermann Schumacher

I gratefully acknowledge the valuable assistance of Dr Andreas Trasser, Dr Wolfgang Haag and Ms Ursula Winter in proofreading the original manuscript. Their helpful suggestions had a significant impact. Most importantly, I am eternally indebted to my wife Christiane. Without her patience and loving care, this book would never have materialised.

Anand Gopinath

I acknowledge the valuable discussions on lasers and photodiodes with my past and present graduate students including Ross Schermer, Prakash Koonath, William Berglund, Jaesang Oh, Kang-Hyun Baek, Klein Johnson and others. I am grateful to my wife Marian for her patience and support while this book was being written.

Joint acknowledgements

Finally, we would like to thank our publisher Dr Julie Lancashire for her patience and her guidance throughout this book project. We also thank Dr Phil Meyler who encouraged us to initiate this project by submitting a book proposal which was accepted by Cambridge University Press. Thanks are also due to Ms Sarah Matthews who has been very helpful in the last and most difficult stages of the project.

Part One

Devices

1 Review of semiconductor materials and physics

1.1 Executive summary

Semiconductor devices are fabricated using specific materials that offer the desired physical properties. There are three classes of solid state materials: insulators, semiconductors and conductors. This distinction is based on the electrical conductivity of these materials with insulators having the lowest and conductors having the highest conductivity. Semiconductors fall in between and their conductivity is affected by several factors such as temperature, the incidence of light, the application of a magnetic field and impurities. This versatility makes semiconductors very important in electronics and optoelectronics applications.

Semiconductors themselves are divided into two classes: elemental and compound. Each type has distinctive physical properties which are exploited in device design. Typical elemental semiconductor device materials are silicon and germanium; examples of compound semiconductors are GaAs, InP, AlGaAs and SiGe. The single crystal structure of these materials is that of a periodic lattice and this determines the properties of the semiconductors. Silicon has the diamond crystal structure and the compound semiconductors have the zincblende lattice structure. The bonding between atoms in a crystal of the semiconductors is termed *covalent bonding*, where electrons are shared between atoms. Fundamental principles of quantum mechanics are applied to determine the energy band structure of the semiconductor.

The basic device physics involves the description of the energy band structure, the density of states, the carrier concentration and the definition of donors and acceptors. Semiconductors are categorised as direct or indirect depending on the bandgap. The absorption mechanism is described and radiation and recombination processes important to device performance are detailed. The two carrier transport processes are drift and diffusion. The currents due to these transport processes are expressed in terms of the applied electric field, the carrier mobility and the carrier concentration. The junction formed by p-type semiconductor (excess holes) and n-type semiconductor (excess electrons) is described and the characteristics of such a junction are given. The important Schottky diode, a junction formed by a metal and a semiconductor layer (n-doped in this case) is characterised.

Heterostructures formed by dissimilar semiconductors are important in device design. The properties of heterojunctions of semiconductor materials are presented. Silicon– germanium heterojunctions are of particular interest as high performance electronic

Period	Group III	Group IV	Group V
2	В	С	N
	Boron	Carbon	Nitrogen
3	Al	Si	Р
	Aluminium	Silicon	Phosphorus
4	Ga	Ge	As
	Gallium	Germanium	Arsenic
5	In	Sn	Sb
	Indium	Tin	Antimony

 Table 1.1
 Portion of the periodic table showing semiconductor material elements

Table 1.2 Elemental and binary compound semiconductors

Elements	IV-IV Binary compounds	III-V Binary compounds
Si Silicon Ge Germanium	SiC Silicon carbide SiGe Silicon germanium	AlAs Aluminium arsenide AlP Aluminium phosphide AlSb Aluminium antimonide BN Boron nitride GaAs Gallium arsenide GaSb Gallium antimonide InAs Indium arsenide InP Indium phosphide InSb Indium antimonide

Table 1.3	Ternary and	quaternary	semiconductors
-----------	-------------	------------	----------------

Ternary compounds	Quaternary compounds
$Al_xGa_{1-x}As$	$Al_xGa_{1-x}As_ySb_{1-y}$
Aluminium gallium arsenide $GaAs_{1-x}P_x$	Aluminium gallium arsenic antimonide $Ga_xIn_{1-x}As_{1-y}P_y$
Gallium arsenic phosphide	Gallium indium arsenic phosphide

devices have been designed using this material alloy. This chapter gives a detailed discussion of these heterojunctions.

1.2 Semiconductor materials

Materials used for semiconductors fall into two categories: elemental semiconductors and compound semiconductors. Table 1.1 shows the section in the periodic table which has the semiconductor elements and Table 1.2 lists examples for elemental and binary compound semiconductors. Some ternary and quaternary semiconductors are listed in Table 1.3.

1.3 Types of solids

There are three types of solids: crystalline, polycrystalline and amorphous. The arrangement of atoms is periodic in three dimensions in a crystalline solid with forces binding the atoms together. This periodicity exists over the entire crystal and it will appear the same regardless of the region where the crystal is viewed. If the periodicity of the atoms occurs over a small region of the solid and changes in different regions of the solid, the solid is termed to be *polycrystalline*. Atoms in amorphous solids exhibit no periodicity. Figure 1.1 shows the three different types of solids.

1.4 Crystal structure

Semiconductor materials such as Si, Ge and GaAs that are to be used for devices are crystalline, that is, a single crystal. This periodic arrangement of atoms in a crystal is termed a *lattice* and the distance between the atoms is the *lattice constant*. The unit cell is a fundamental unit in the crystal and a repetition of the unit cell generates the entire lattice. The unit cell is not unique and can be chosen in various ways as shown in Figure 1.2(a). This is a two-dimensional representation of the crystal lattice. The entire lattice can be constructed by translations of any of the three unit cells in two coordinate directions. The primitive unit cell is the smallest unit cell. A generalised primitive three-dimensional unit cell is shown in Figure 1.2(b). The coordinate directions are **a,b,c**. In cubic structures, these would be the rectangular coordinates. The basic cubic crystal structures are (a) the simple cubic, (b) the body-centred cubic and



(a) Crystalline

(b) Polycrystalline



(c) Amorphous

Fig. 1.1 Schematic arrangement of atoms in solids.



(a) Two-dimensional lattice – shaded areas show possible unit cells



(b) Generalised primitive unit cell





(a) Simple cubic

(b) Body-centred cubic

(c) Face-centred cubic

Fig. 1.3 Types of cubic lattices.



Fig. 1.4 (a) Diamond lattice and (b) Zincblende lattice. (S. M. Sze, *Semiconductor Devices: Physics and Technology*, John Wiley & Sons, 1985). Reprinted with permission of John Wiley & Sons, Inc.

(c) the face-centred cubic shown in Figure 1.3. The simple cubic lattice has an atom at each corner of the cube, where the length of a side of the cube is *a*, the lattice constant. The body-centred cubic lattice (BCC) has an additional atom in the centre of the cube and the face-centred cubic lattice (FCC) has an additional atom in the centre of each face of the cube. The two most important semiconductor crystal structures are the diamond lattice structure and the zincblende structure. Silicon and germanium have the diamond lattice structure and most of the binary compound semiconductors such as GaAs have the zincblende lattice structure. The only difference between the diamond and the zincblende structures is that the latter has two different types of atoms as seen in Figure 1.4. The diamond structure consists of two inter-penetrating FCC sublattices of atoms. The second FCC cube is shifted by one-fourth of the body diagonal, which is the longest diagonal. In the zincblende structure of GaAs, one sublattice has gallium atoms and the other has arsenic atoms.

1.5 Crystal directions and planes

Crystals are of finite size and hence have surfaces. It is necessary to define the planes at the crystal surfaces and the crystallographic directions, both of which determine the



Fig. 1.5 Representation of plane with Miller indices [6, 5, 8].



Fig. 1.6 Representation of direction with Miller indices [6, 5, 8].

properties of semiconductor devices. The rectangular coordinate system defines the cubic crystal and the plane surfaces and directions are described by a set of indices called the *Miller indices*. Planes are described by the indices (h,k,l) and the directions perpendicular to these planes are described by the same indices [hkl].

Example: Find the Miller indices of the plane which makes intercepts 3a,4a,2a along the coordinate axes in a cubic crystal, where *a* is the lattice constant. Draw the direction vector with the same Miller indices.

Solution: The intercepts are 3, 4 and 2. The reciprocals are 1/3, 1/4 and 1/2. Multiplication by the lowest common denominator, which is 12, yields (4,3,6). These are the Miller indices which define the plane shown in Figure 1.5. It can be shown that parallel planes are described by the same Miller indices.

The Miller indices of the direction are given as [436]. The intercepts on the three coordinate axes are 3, 4 and 2. The direction vector is drawn and seen to be perpendicular to the planes shown in Figure 1.6.

The basic planes in cubic crystals are shown in Figure 1.7. It is also important to describe specific directions in a crystal in addition to the planes. As in the case of the crystal plane, a crystal direction is also described by three integers which are the components of a vector drawn in the particular crystal direction. The crystal planes and directions of most interest are shown in Figure 1.8. The [hkl] direction is perpendicular to the (hkl) plane.



Fig. 1.7 Basic crystal planes.



Fig. 1.8 Important crystal planes and directions.

1.6 Atomic bonding

Atoms are held together by bonding forces to form solids. When the attractive and repulsive forces are equivalent, the atoms are in equilibrium and maintain the spacing characterised by the lattice constant, *a*. There are different bonding classifications which are described by the dominant force of attraction. When one of the atoms gives up an electron in the outer shell to another atom, positive and negative ions are produced. There is a Coulomb interaction force of attraction between them. This is termed *ionic bonding*. At equilibrium, the forces of attraction and repulsion are equivalent. Sodium chloride (NaCl) and Potassium chloride (KCl) are examples of ionic bonding after the formation of the Na⁺ and Cl⁻ ions.

1.6.1 Covalent bonding

This type of bonding results when electrons are shared by neighbouring atoms. The hydrogen atom is the simplest example of covalent bonding. Each of the two electrons bonds with the other to complete the lowest energy shell as shown in Figure 1.9.

Each atom in a diamond or zincblende lattice has four nearest neighbours. Each atom has four electrons in the outer orbit. These are the valence electrons and each atom shares these valence electrons with its four neighbours. The interaction between the shared electrons results in bonding forces which are quantum mechanical in nature. In other words, each electron pair constitutes a covalent bond. Elements in group IV such as Si and Ge have four valence electrons as shown in the references [8, 14, 15]. These are available for bonding as seen in Figure 1.10. Compound semiconductors such as





Fig. 1.10 Covalent bonding in silicon.

Fig. 1.9

GaAs exhibit both covalent as well as ionic bonding. This is due to the fact that Ga and As occur in two different groups in the periodic table and hence there is a transfer of charge resulting in some ionic bonding.

1.7 Atomic physics

The theories of atomic physics were based on experimental observations. These theories subsequently explained the experiments and led to the understanding of atoms in matter.

1.7.1 The photoelectric effect

The measurements of Planck on a heated sample of material indicated that energy is radiated in discrete units called *quanta* as shown in Equation (1.1).

$$E = h\nu, \tag{1.1}$$

where *h* (Planck's constant) = 6.63×10^{-34} J · s and *v* is the frequency of the radiation. Heinrich Hertz discovered the photoelectric effect in 1887. The experiments performed by Philipp Lenard, a former student of Hertz, showed that if light shines on a metal surface in vacuum, some of the electrons receive enough energy so that they are emitted from the surface into the vacuum. They were interpreted by Albert Einstein, who received the Nobel Prize for his work in 1921. This is termed the *photoelectric effect* and the maximum energy is a function of the frequency of the incident light. The quantised units of light energy are called *photons*.

Based on further experimental observations of Davisson and Germer (USA) and Thompson (UK) on the diffraction of electrons by the atoms in a crystal, de Broglie related the wavelength of a particle of momentum p = mv, where *m* is the mass of the particle as seen in Equation (1.2):

$$\lambda = \frac{h}{p} = \frac{h}{mv}.$$
(1.2)

1.7.2 The Bohr model of the atom

A model of the atom was first proposed by Bohr. In his model, the electrons move in stable circular orbits about the nucleus and the electron may move to an orbit of higher or lower energy. The electron would either gain energy or lose energy by the absorption or emission of a photon of energy hv. Bohr further proposed that the angular momentum of the electron moving in a circular orbit was an integral multiple of Planck's constant as seen in Equation (1.3).

$$p_{\theta} = \frac{nh}{2\pi} = n\hbar, \ n = 1, 2, 3, \dots$$
 (1.3)

The hydrogen atom with one electron and the nucleus illustrates this concept in a simple manner as seen in Figure 1.11.

Assuming that the electron of mass m rotates in a stable orbit of radius r with velocity v, the angular momentum is written in Equation (1.4):

$$p_{\theta} = mvr = n\hbar. \tag{1.4}$$

The electrostatic force between the charge on the nucleus and the charge on the electron must be equal to the centripetal force for the electron to remain in stable orbits. This yields the expression in Equation (1.5) for the energy of the electron [15]:

$$E_{\rm n} = -\frac{mq^4}{2(4\pi\epsilon_0)^2 n^2 \hbar^2}.$$
 (1.5)



Fig. 1.11 Bohr model of the hydrogen atom.



Fig. 1.12 Electron orbits in Bohr model (not to scale).

The electron orbits in the Bohr model are shown in Figure 1.12.

1.8 The de Broglie relation

The initial theoretical and experimental results of Planck, Einstein and Bohr laid the foundation for the development of quantum mechanics. It was de Broglie, however, who first postulated that if waves were seen to behave as particles then it could be that particles might behave like waves.

In the Bohr formulation, the electron which travels in a circular orbit of radius r is assumed to behave like a wave with a wavelength λ . It travels in a circular path equal in length to the circumference $2\pi r$, which will be an integral number of wavelengths so that

$$n\lambda = 2\pi r. \tag{1.6}$$

The Bohr formulation yielded the linear velocity of the electron to be

$$v = \frac{q^2}{4\pi\epsilon_0 n\hbar}.\tag{1.7}$$

Using this velocity relation, the wavelength can be written as

$$\lambda = \frac{h}{mv} = \frac{h}{p},\tag{1.8}$$

where p is the linear momentum of the electron. Thus, de Broglie postulated that the relationship between the wavelength and the linear momentum p of a particle is given by Equation (1.2).

$$p = \frac{h}{\lambda} = \frac{h}{2\pi} \frac{2\pi}{\lambda} = \hbar k.$$
(1.9)

This is the *de Broglie relationship*. For free electrons, the energy–momentum relationship is as follows:

$$E = \frac{mv^2}{2} = \frac{p^2}{2m}; \ p = \sqrt{2mE}.$$
 (1.10)

Hence, the experiments of Davisson and Germer and of Thompson were verified by the de Broglie relationship.

1.9 Quantum mechanics

Newtonian mechanics can be used to describe physical behaviour that is macroscopic. Typical examples of this are planetary motion, the classical electromagnetic fields and fluid motion. The motion of electrons and the interaction of electrons in atoms in semiconductor materials cannot, however, be described thus since we are dealing with microscopic behaviour. This physical behaviour on the atomic scale can only be described by quantum mechanics rather than Newtonian mechanics. Quantum or wave mechanics had as its basis the physical understanding developed by Planck and de Broglie. The classical laws of the conservation of energy, momentum and angular momentum are also assumed to be valid in quantum mechanics. Hence, the physics involved in the interaction between atoms can be described mathematically by quantum mechanics.

1.9.1 Probability and the uncertainty principle

When the motion of the particle is microscopic, the parameters cannot be described exactly but rather in terms of average (expectation) values. Hence we have, for example, the expectation values of position, momentum and energy of an electron. So, we have a probabilistic rather than an exact description of the particle behaviour. There is, thus, an inherent uncertainty in the position and momentum of the particle. This was formulated by Heisenberg and is termed the *Heisenberg uncertainty principle*. The uncertainty in the measurement of the position and momentum of particle motion is given as

$$(\Delta x)(\Delta p_{\mathbf{x}}) \ge \hbar. \tag{1.11}$$

The uncertainty in energy is related to the time at which the energy was measured and is given by

$$(\Delta E)(\Delta t) \ge \hbar. \tag{1.12}$$

These equations show that the simultaneous measurements of position and momentum on the one hand and energy and time on the other hand cannot be performed with arbitrary accuracy.

It follows that we can only determine the probability of finding an electron in a certain position or having a certain momentum. This leads to the definition of a probability density function. The probability of finding a particle in a range, say, from x to x + dx is given by

Classical variable	Quantum operator
x	x
f(x)	f(x)
Momentum $p(x)$	$\frac{\hbar}{i}\frac{\partial}{\partial x}$
Kinetic energy $\frac{p^2}{2m}$	$\frac{-\hbar^2}{2m}\frac{\partial^2}{\partial x^2}$
Potential energy V	V
Total energy E	$\frac{-\hbar}{j} \frac{\partial}{\partial t}$

 Table 1.4
 Classical variables and quantum operators

$$\int_{-\infty}^{\infty} P(x)dx = 1, \qquad (1.13)$$

where P(x) is a normalised function. The average value of a function x is defined as

$$\langle f(x)\rangle = \int_{-\infty}^{\infty} f(x)P(x)dx = 1.$$
(1.14)

The correspondence between classical and quantum mechanical quantities is shown in Table 1.4.

The basic principles of quantum mechanics will now be reviewed. Each particle in a physical system is described by a wave function $\Psi(x, y, z, t)$. The function and its space derivatives are continuous, finite and single-valued.

The probability of finding a particle with wave function Ψ in the volume dxdydz is $\Psi^*\Psi dxdydz$. Then we have the following definition for three-dimensional space:

$$\int_{-\infty}^{\infty} \Psi^* \Psi dx dy dz = 1.$$
 (1.15)

The expectation value of any physical quantity X can be written as

$$\langle X \rangle = \int_{-\infty}^{\infty} \Psi^* X_{\text{oper}} \Psi dx dy dz,$$
 (1.16)

where X_{oper} is the operator corresponding to the variable X.

The classical equation for energy conservation is Kinetic energy + Potential energy = Total energy:

$$\frac{p^2}{2m} + V = E.$$
 (1.17)

1.9.2 The wave equation

We obtain the quantum mechanical energy equation by substituting the corresponding operators which operate on the one-dimensional wave function $\Psi(x, t)$:

$$\frac{-\hbar^2}{2m}\frac{\partial^2\Psi(x,t)}{\partial x^2} + V(x)\Psi(x,t) = E\Psi(x,t) = \frac{-\hbar}{j}\frac{\partial\Psi(x,t)}{\partial t}.$$
 (1.18)



Fig. 1.13 Infinite potential well, width = 2a.

This is the one-dimensional Schrödinger wave equation. The three-dimensional wave equation is

$$\frac{-\hbar^2}{2m}\nabla^2\Psi + V(x)\Psi = E\Psi = \frac{-\hbar}{j}\frac{\partial\Psi}{\partial t}.$$
(1.19)

The wave equation is applied to the solution of various physical problems. The problem of the infinite potential well provides an understanding of the method of solution and an insight into the discrete energies of a single electron [14, 15].

This basic physical concept is important since quantum wells can be fabricated using semiconductor structures for devices. A general solution of the one-dimensional wave equation can be written as follows:

$$\Psi(x,t) = \psi(x) \exp\left(\frac{-jEt}{\hbar}\right).$$
(1.20)

We consider the infinite quantum well of width 2a with zero potential outside the well as shown in Figure 1.13.

On solving the one-dimensional wave function, we obtain n solutions and the discrete energy levels are given by [14, 15],

$$E_{\rm n} = \frac{\pi^2 \hbar^2 n^2}{8m_0 a^2},\tag{1.21}$$

where m_0 is the rest mass of the electron and a is the lattice constant of the crystal. The one-dimensional problem of a particle in a finite potential well can also be solved and the allowed energies of the particle determined [10]. The phenomenon of tunnelling wherein an electron with energy E tunnels through a potential barrier with barrier height V_0 greater than E is also explained by quantum mechanics. Classically, the electron would not be able to show this behaviour. If we have a potential barrier of width a, the one-dimensional Schrödinger equation can be solved in the three regions I, II and III as



Fig. 1.14 Potential barrier.

shown in Figure 1.14. There are three regions for the problem. Regions I and III have zero potential. Say region II has a potential V_0 , then the solutions in the three regions are given by:

Region I:
$$\psi(x) = A \exp(jkx) + B \exp(-jkx); k^2 = \frac{2mE}{\hbar^2}$$
 (1.22)

Region II:
$$\psi(x) = C \exp(-\alpha x) + D \exp(+\alpha x); \ \alpha^2 = \frac{2m(V_0 - E)}{\hbar^2}$$
 (1.23)

Region III:
$$\psi(x) = F \exp(jkx); k^2 = \frac{2mE}{\hbar^2}.$$
 (1.24)

Using the conditions that the wave function and its derivatives are continuous at the boundaries, x = 0 and x = a, the tunnelling probability is of the form:

$$T = \left|\frac{F}{A}\right|^2 = \frac{4}{4\cosh^2(\alpha d) + \left(\frac{\alpha}{k} - \frac{k}{\alpha}\right)^2 \sinh^2(\alpha d)}.$$
 (1.25)

Boundary conditions are matched at the two boundaries and T, the tunnelling probability is determined.

The method of solution is the same regardless of the shape of the barrier. Triangular and trapezoidal barriers have a simple geometry and hence give us exact solutions. When the barriers are of arbitrary shape, the tunnelling probability is solved using the Wentzel–Kramers–Brillouin (WKB) approximation:

$$T \cong \exp\left[-2\int_{d_1}^{d_2} |f(x)| dx\right]$$
(1.26)

with

$$f(x) = \frac{2m_0}{\hbar^2} [V(x) - E], \qquad (1.27)$$

where V(x) is the arbitrary potential. The limits of the integral d_1 to d_2 represent the classically forbidden region, where the potential energy is larger than the total particle energy.

1.10 Statistical mechanics

1.10.1 The free electron

When the three-dimensional Schrödinger equation is solved, the general solution gives the wave function for the electron in motion in a region of zero potential. The behaviour of electrons in semiconductor crystals can be assumed to be like that of socalled *free electrons* under certain conditions, hence the importance of this result. The time-independent wave function solution is given by

$$\psi(\mathbf{r}) = \mathbf{A} \exp(\mathbf{k} \cdot \mathbf{r}), \qquad (1.28)$$

where **A** is a complex quantity and is the amplitude, **k** is the wave vector and **r** is the three-dimensional space vector. This results in energies of the same form as Equation (1.21).

1.10.2 Fermi–Dirac distribution

The Fermi–Dirac distribution function f(E) gives the probability that states with energy E are occupied by particles [10]:

$$f(E) = \frac{1}{1 + \exp\left(\frac{E - E_{\rm F}}{kT}\right)},\tag{1.29}$$

 $E_{\rm F}$ represents the Fermi energy where f(E) becomes equal to 1/2.

1.11 Electrons in a semiconductor

Since semiconductors have periodic lattice structures, the electrons are subjected to a periodic potential. Hence the Schrödinger equation must be solved for a periodic potential [10]. The *Bloch theorem* states that the one-dimensional wave function for an electron in a periodic potential is given by

$$\psi(x) = V_{\mathbf{k}}(x) \exp(jkx), \qquad (1.30)$$

where $V_k(x)$ is a periodic potential with the same periodicity as the semiconductor crystal with lattice constant *a* such that

$$V_{\rm k}(x) = V_{\rm k}(x + na),$$
 (1.31)

where *n* is an integer.

1.12 The Kronig–Penney model

An important model for the band structure is the Kronig–Penney model (Figure 1.15). The one-dimensional periodic potential is given by

١

$$V(x) = 0, \ 0 \le x \le a \tag{1.32}$$

$$V(x) = V_0, \ -b \le x \le 0 \tag{1.33}$$



Fig. 1.15 Periodic potential for Kronig–Penney model.

The periodicity distance is d = a + b. The wave equation is solved in the three regions and the continuity conditions for the wave function and its derivatives are applied. The non-trivial solutions are obtained when the electron energy is less than and greater than the potential V_0 [10, 14]. The transcendental equation to be solved is

$$\cos k_{\rm x} d = \cos a\alpha \cosh b\delta - \frac{\alpha^2 - \delta^2}{2\alpha\delta} \sin a\alpha \sinh b\delta, \ 0 < E < V_0$$
(1.34)

$$\cos k_{\rm x} d = \cos a\alpha \cos b\delta - \frac{\alpha^2 + \delta^2}{2\alpha\delta} \sin a\alpha \sin b\delta, \ E > V_0 \tag{1.35}$$

with

$$\alpha = \sqrt{\frac{2m_0 E}{\hbar^2}}, \ \beta = \sqrt{\frac{2m_0 (E - V_0)}{\hbar^2}}, \ \delta = \sqrt{\frac{2m_0 (V_0 - E)}{\hbar^2}}.$$
 (1.36)

The solution of the equation gives the energy E. The allowed energy bands are separated by band gaps with no allowed energies. It follows that there are forbidden energy regions for an electron which is subjected to a periodic potential in a semiconductor crystal.

1.12.1 Effective mass

When the centre of mass of a classical particle moves with a velocity v, we define a phase velocity. If we have a packet of travelling waves with a centre frequency ω and a wavenumber k, we have the classical dispersion relation for the group velocity:

$$v_{\rm g} = \frac{d\omega}{dk}.\tag{1.37}$$

In the quantum mechanical formulation, the wavepacket is the analogue of the classical particle in a given region of space. This wavepacket consists of constant-energy wave

function solutions and a centre energy is defined. Hence the wavepacket group velocity in the quantum-mechanical formulation can be written as in Equation (1.20):

$$v_{\rm g} = \frac{1}{\hbar} \frac{dE}{dk}.$$
 (1.38)

Using the force-momentum relations, we define the effective mass of an electron in a crystal as

$$m^* = \left(\frac{1}{\hbar^2} \frac{d^2 E}{dk^2}\right)^{-1}.$$
 (1.39)

Section 1.14.1 defines heavy and light holes corresponding to wide and narrow bands respectively.

1.12.2 Carriers in semiconductors

The two types of carriers in semiconductors are the conduction band electrons and the valence band holes. The electrons occupy the conduction band when the temperature is raised above 0 K. The unoccupied states in the valence band are holes and are defined to have a positive charge with the same magnitude as the electronic charge. Hence, we consider electrons in determining the conduction band properties and holes in determining the valence band properties. The band structures of several semiconductors are given by Pierret, and Streetman and Banerjee [10, 15] and others.

1.13 Semiconductors in equilibrium

1.13.1 Intrinsic semiconductors

A semiconductor is described as being intrinsic when there are no impurities and no defects in the crystal. The concentration of electrons in the conduction band is equal to the concentration of holes in the valence band. At 0 K, the electrons occupy all the available energy states in the valence band and all the states in the conduction band are empty. This follows from the fact that at 0 K, each electron is in the lowest possible energy state. As the temperature is increased the electrons are excited due to the acquired thermal energy and move into the conduction band leaving behind holes in the valence band. Therefore, the equilibrium concentration of electrons in the conduction band p_0 in intrinsic semiconductors [2, 15]:

$$n_0 = p_0 = n_{\rm i},\tag{1.40}$$

where n_i is simply referred to as the intrinsic concentration of holes and electrons.

1.13.2 Extrinsic semiconductors

When impurity atoms are added to the intrinsic semiconductor such that the electron concentration is no longer equal to the hole concentration, it becomes an extrinsic



Fig. 1.16 Band diagram.

semiconductor and $n_0 \neq p_0$. Thus the doping of a semiconductor with impurities can produce excess electrons or holes. These atoms can be either donors or acceptors. If the dopant produces an excess of electrons, the dopant is referred to as a *donor*, the semiconductor becomes n-type material with n > p and the current is predominantly due to the negatively charged electrons. If, on the other hand, the dopant generates holes, the dopant is referred to as an *acceptor*, the result is a p-type semiconductor with p > nand the current is predominantly due to the positively charged holes. Note that the hole charge has the same magnitude as the electronic charge [2, 8, 15].

1.13.3 Semiconductor band diagrams

The band diagrams for p- and n-type semiconductors at thermal equilibrium are given in Figure 1.16. The bottom of the conduction band is E_c , the top of the valence band is E_v , the intrinsic energy level is at mid-band and is denoted by E_i and the Fermi level is E_F .

1.13.4 Electron and hole distribution

The distribution of electrons in the conduction band and holes in the valence band is obtained using the Fermi–Dirac probability function. The electron distribution in the conduction band is written as

$$n(E) = g_{c}(E)f(E),$$
 (1.41)

where $g_c(E)$ is the density of quantum states in the conduction band and f(E) is the Fermi–Dirac probability function given in Equation (1.29). The hole distribution in the valence band can be written in a similar way:

$$p(E) = g_{v}(E)[1 - f(E)].$$
(1.42)

The density of states functions are written as

$$g_{\rm c}(E) = \frac{m_{\rm n}^* \sqrt{2m_{\rm n}^*(E - E_{\rm c})}}{\pi^2 \hbar^3}, \ E \ge E_{\rm c}$$
(1.43)

$$g_{\rm v}(E) = \frac{m_{\rm p}^* \sqrt{2m_{\rm p}^*(E_{\rm v} - E)}}{\pi^2 \hbar^3}, \ E \le E_{\rm v}.$$
 (1.44)

The equilibrium concentration of electrons can now be written as

$$n_0 = \int_{E_c}^{\infty} n(E) \mathrm{d}E, \qquad (1.45)$$

where n(E) is given by Equation (1.41). Similarly, the equilibrium hole concentration is written as

$$p_0 = \int_{-\infty}^{E_v} p(E) \mathrm{d}E, \qquad (1.46)$$

where p(E) is given by Equation (1.42). The equilibrium electron and hole concentrations in the conduction and valence bands respectively are written as

$$n_0 = N_{\rm c} \exp\left(\frac{-(E_{\rm c} - E_{\rm F})}{kT}\right) \tag{1.47}$$

$$p_0 = N_{\rm v} \exp\left(\frac{-(E_{\rm F} - E_{\rm v})}{kT}\right),\tag{1.48}$$

where N_c and N_v are the effective density of states functions in the conduction and valence bands respectively.

$$N_{\rm c} = 2 \left(\frac{2\pi m_{\rm n}^* kT}{h^2}\right)^{3/2}$$
(1.49)

$$N_{\rm v} = 2 \left(\frac{2\pi m_{\rm p}^* kT}{h^2}\right)^{3/2}.$$
 (1.50)

The intrinsic carrier concentration n_i is given by

$$n_{\rm i}^2 = n_0 p_0. \tag{1.51}$$

By substitution of Equations (1.47) and (1.48), we can write the intrinsic concentration as

$$n_{\rm i}^2 = N_{\rm c} N_{\rm v} \exp\left(\frac{-(E_{\rm c} - E_{\rm v})}{kT}\right) \tag{1.52}$$

$$= N_{\rm c} N_{\rm v} \exp \frac{-E_{\rm g}}{kT},\tag{1.53}$$

where E_{g} is the bandgap energy.

1.14 Direct and indirect semiconductors

When light illuminates a semiconductor, and the photon energy is equal to or larger than the band gap, the light is absorbed, and creates hole–electron pairs. These holes and electrons are equal in number to maintain charge neutrality, and since they are not in equilibrium, in due course they recombine; this recombination may be radiative or nonradiative. Radiative recombination, when a photon is emitted usually at the bandgap energy, only occurs in direct bandgap material, whereas non-radiative recombination may occur in both direct and indirect bandgap semiconductors. In indirect semiconductors, this non-radiative recombination requires a phonon to mediate the process. Non-radiative processes in direct bandgap material are usually through traps or due to surface recombination. The direct or indirect band gap defines whether the lowest position of the conduction band aligns with the maximum of the valence band along momentum space, where the effective momentum value k is equal to zero.

Direct bandgap semiconductors are capable of photon emission, by radiative recombination, but indirect semiconductors have a low probability of radiative recombination. However, indirect bandgap semiconductors may have isoelectronic impurity states which are direct, and therefore the recombination from these states may also be radiative. GaP, which is an indirect gap semiconductor, may be doped with zinc oxide or nitrogen to produce these states, and the widely used green or red light–emitting diodes are examples of this emission.

1.14.1 Absorption processes

Considering Figure 1.17, we note that three different valence bands are shown. Equation (1.39) had linked the carrier mobility the second derivative of energy with respect to k. It is then easy to understand why the band with less curvature at k = 0 (V_1) is called the "heavy hole band", while the one which is more strongly bent (V_2) is called "light hole band". In most bulk semiconductors, the light and heavy hole bands concide at k = 0 – they are degenerate. The band V_3 is called the "split-off band". Note



Fig. 1.17 Band diagram of bulk Ge (left) and Si (right). Note that the minimum of the conduction band is not aligned with maximum of the valence band at k = 0, indicating that these are indirect semiconductors. S. Wang, *Fundamentals of Semiconductor Theory and Device Physics*, 1st Edition, pp. 233–234, ©1989. Reprinted by permission of Pearson Education, Inc., Upper Saddle River, NJ.



Fig. 1.18 Band diagram of bulk GaAs. Note that the minimum of the conduction band is aligned with maximum of the valence band at k = 0, indicating that this is a direct gap semiconductor. S. Wang, *Fundamentals of Semiconductor Theory and Device Physics*, 1st Edition, pp. 233–234, ©1989. Reprinted by permission of Pearson Education Inc., Upper Saddle River, NJ.

that in each of these cases, the minimum of the conduction band is not aligned with the maximum of the valence band where k = 0, which implies that these are indirect semiconductors. In Figure 1.18, the band diagram of the direct semiconductor GaAs is shown; here the conduction band minimum is along the k = 0 axis and aligned with the maximum of the valence band. Note the degenerate valence band of heavy holes and light holes and the split-off band.

Photo-excitation of semiconductors with photons energies equal to or greater than the bandgap energy of the material results in absorption, which in turn causes the creation of hole–electron pairs for each photon. The major source of absorption in semiconductors is the valence band to conduction band transition. In the case of direct semiconductors, the transition occurs when the photon energy is at the bandgap value or larger and results in the transition of an electron in the valence band to the conduction band. In indirect semiconductors, the absorption has to be mediated by phonons. In addition to the band to band absorption, transitions take place from acceptor to donor levels, from acceptor to conduction band, valence to donor level, all of which result in absorption below the bandgap energy. Figure 1.19 shows schematically the band to band transition. The conduction band to valence band and impurity band to impurity band transitions are shown schematically in Figure 1.21. Not shown in this figure are the impurity band to conduction and valence band transitions, all of which lead to absorption and emission of the appropriate photon energies.



Fig. 1.19 Schematic diagram of the conduction and valence bands of a direct semiconductor and the transitions.







Fig. 1.21 Transitions possible with a semiconductor with impurity donor and acceptor bands: conduction band to valence band and impurity band to impurity band are illustrated. Others, conduction band to impurity band and valence band to impurity band have not been shown.

The absorption rate of the band to band transitions, for both the direct and indirect transitions may be calculated using quantum theory, but is not included here.

Free carrier absorption also occurs in most semiconductors as the carrier density is always non-zero. The absorption of a photon by a carrier within a band results in the carrier having a larger energy. The absorption coefficient is proportional to the carrier density [3]. This effect is important in the design of waveguide devices, where typically this may result in absorption of the order of 1 dB cm^{-1} when the carrier densities are high in the 10^{18} cm^{-3} region.

1.14.2 Exciton absorption

In pure semiconductors, the absorbed photon with bandgap energy or larger may create excitons, which are electron-hole pairs that are bound, and in the binding process give up the binding energy. The binding energy of these excitons is of the order of about 4.5 meV, and at low temperatures, an excitonic absorption peak is seen a little below the band to band absorption energy. At room temperature, this peak is not seen in bulk material, because the thermal broadening due to optical phonons is comparable, and the excitons that are created dissociate very rapidly. In quantum wells, however, the excitons remain extant at room temperature due to enhanced binding energies, which are typically two or three times that of the thermal broadening energy. Thus, the absorption characteristics of the material with quantum wells also show the excitonic absorption in addition to the usual band to band absorption. When a transverse electric field is applied to the quantum well, the absorption edge shifts to a longer wavelength. A simple explanation of this phenomenon is shown in Figure 1.22, where the schematic wave functions of the electron and hole in the quantum well are shown. When the transverse field is applied, then the quantum well bands tilt, and the resulting gap between the electron-hole wave functions decreases, which results in the absorption edge moving to a smaller energy and thus a longer wavelength.



Fig. 1.22 A schematic diagram of the wave functions in a quantum well, and the effect of applying a field across the well, resulting in tilting of the wells. This so-called Quantum Confined Stark Effect reduces the effective band gap of the material.



Fig. 1.23 Absorption coefficient for various semiconductors (M. Shur, *Physics of Semiconductor Devices*, Prentice Hall, 1990 ©Prentice Hall).

Other absorption mechanisms are due to valence to impurity band, impurity band to other impurity band or impurity band to conduction band transition, intraband absorption between different levels in the same band, and free carrier absorption.

The absorption spectra of different semiconductors is summarised in Figure 1.23.

1.15 Recombination and radiation in semiconductors

The absorption of photons by the semiconductor results in the generation of electrons and holes, which disturbs the equilibrium status of the semiconductor. Electrical injection also results in this non-equilibrium of an excess of electrons in the conduction band and an equal number of holes in the valence band. These recombine, both non-radiatively and radiatively, the latter in direct gap semiconductors. In general, the radiative transitions are dominated by the conduction band to valence band emission and therefore define the energy of the emitted photons. Other recombination processes include exciton recombination, donor to acceptor and other impurity recombinations. The radiation spectrum from recombination is generally shifted to lower energy from the absorption spectrum, and this is termed the *Stokes* or the *Franck–Condon shift* due to imperfections in materials or interfaces. In general the excess electrons and holes decay at some rate, resulting in the density varying as $\exp(-t/\tau)$, where τ is defined as the lifetime of the carriers. The decay of these carriers results in transfer of energy to the lattice in the form of phonons for the non-radiative decay and transfer of energy to photons for radiative decay.

The corresponding lifetimes are labelled as τ_{nr} and τ_r for the non-radiative and radiative decay, respectively, and the corresponding non-radiative and radiative rates are R_{nr} and R_r , respectively. Thus, the total lifetime constant τ is given as:

$$\frac{1}{\tau} = \frac{1}{\tau_{\rm nr}} + \frac{1}{\tau_{\rm r}}.$$
 (1.54)

The corresponding total spontaneous rate of recombination is given by

$$R_{\rm spon} = R_{\rm nr} + R_{\rm r}.$$
 (1.55)

Devices such as the light-emitting diode (LED) largely depend on spontaneous emission, and in this case the internal quantum efficiency is given by

$$\eta_{\text{internal}} = \frac{R_{\text{r}}}{R_{\text{nr}} + R_{\text{r}}}.$$
(1.56)

The exponential decay rate of the excess carriers is inversely proportional to recombination rate, and if the excess of electron is Δn , then the recombination rates R_r and R_{nr} are given by the expressions: $R_r = \Delta n/\tau_r$ and $R_{nr} = \Delta n/\tau_{nr}$. Then internal quantum efficiency may also be written as

$$\eta_{\text{internal}} = \frac{\frac{1}{\tau_{\text{r}}}}{\frac{1}{\tau_{\text{r}}} + \frac{1}{\tau_{\text{nr}}}} = \frac{1}{1 + \frac{\tau_{\text{r}}}{\tau_{\text{nr}}}} = \frac{\tau_{\text{nr}}}{\tau_{\text{r}} + \tau_{\text{nr}}}.$$
(1.57)

The total spontaneous recombination rate is given by the equation:

$$R_{\text{total}} = A\Delta n + B\Delta n^2 + C\Delta n^3.$$
(1.58)

The first term is the Shockley–Read–Hall recombination due to defects and traps, the second is the spontaneous emission due to radiative transition, and the third is the Auger recombination term. Auger recombination is non-radiative, occurs at high injection levels, and is a three-particle process. It becomes important in ternary and quaternary compounds of InP-based materials, and is evident in the long wavelength laser structures.

1.15.1 Spontaneous and stimulated emission

The radiative recombination process discussed above occurs spontaneously, and this is used in traditional LED structures. In lasers, stimulated emission is the source of light, and in this section the relationship between absorption, spontaneous emission and stimulated emission, first outlined by Einstein in 1917, is discussed. The derivations given here follow the approach outlined by Casey and Panish [4] and Agrawal [1].

It can be shown that the blackbody radiation law is given by

$$P(E) = \frac{8\pi \bar{n}^{3} E^{2}}{h^{3} c^{3} (e^{E/kT} - 1)},$$
(1.59)



Fig. 1.24 Transitions from level E_1 to E_2 for absorption and from E_2 to E_1 for emission, F_1 and F_2 are the electron and hole quasi-Fermi levels respectively.

where \overline{n} is the index of the material under consideration, *h* is Planck's constant, *c* is the speed of light in vacuum, *E* is the energy given by hv, v is the frequency and *k* is Boltzmann's constant. This is the expression for the energy density blackbody radiation P(E), and is in thermal equilibrium, when the input radiation is equal to the outgoing radiation.

For a semiconductor, consider the transitions from the conduction band to the valence band and also the reverse. The energy levels in each of these bands have to obey the Pauli exclusion principle, which implies only two carriers at each level. Thus, the band is a series of levels, as shown in Figure 1.24, and the transition energy for an electron from a level E_1 in the valence band to a level E_2 in the conduction band requires that an incident photon has energy given by $hv = E_2 - E_1$. Let the probability of this transition taking place be given by B_{12} , and let f_1 be the probability that an electron exists at level E_1 and $(1 - f_2)$ be the probability that a vacancy occurs at level E_2 . Also assume that the radiation density of photon energy incident on the semiconductor is given by $P(E_{21})$. Then the upward transition rate is given by

$$r_{12} = B_{12}f_1(1 - f_2)P(E_{21}). (1.60)$$

Note that f_1 and f_2 take the form of the Fermi–Dirac distribution

$$f_{\rm i} = \frac{1}{e^{(E_{\rm i} - F_{\rm i})/kT} + 1},\tag{1.61}$$

where F_i is the corresponding quasi-Fermi level, k is Boltzmann's constant and T the temperature in Kelvin.

Similarly, the downward transition rate, now called the *stimulated transition*, is given as

$$r_{21}(\text{stim}) = B_{21}f_2(1 - f_1)P(E_{21}), \qquad (1.62)$$

where B_{21} is the transition probability, f_2 is the probability that an electron is present at E_2 and $(1 - f_1)$ is the probability that there is vacancy at E_1 .

Finally, there is the spontaneous transition from E_2 to E_1 , without any incident radiation involved, given by

$$r_{21}(\text{spon}) = A_{21}f_2(1 - f_1). \tag{1.63}$$

In thermal equilibrium, the input radiation is equal to the output, the Fermi levels $F_1 = F_2$, and hence

$$r_{12} = r_{21}(\text{spon}) + r_{21}(\text{stim}). \tag{1.64}$$

Equating, simplifying and noting that $P(E_{21})$ is the blackbody radiation term,

$$P(E_{21}) = \frac{8\pi \bar{n}^3 E^2}{h^3 c^3 (e^{E_{21}/kT} - 1)}$$
(1.65)

$$=\frac{A_{21}f_2(1-f_1)}{B_{12}f_1(1-f_2)-B_{21}f_2(1-f_1)}$$
(1.66)

$$=\frac{A_{21}}{B_{12}e^{E_{21}/kT}-B_{21}}.$$
(1.67)

Equating Equations (1.65) and (1.67) and separating them into temperature-dependent and temperature-independent terms give the following results:

$$A_{21} = \frac{8\pi\bar{n}^3 E^2}{h^3 c^3} B_{21} \tag{1.68}$$

and

$$B_{21} = B_{12}. (1.69)$$

These are Einstein's coefficients and their relationships with each other.

The condition under which stimulated emission dominates is an interesting one. This requires a non-equilibrium condition in which the presence of incident radiation is required. This results in the population densities in the conduction and valence bands to be different from the equilibrium condition. For stimulated emission to dominate, the stimulated emission rate, r_{21} (stim), needs to exceed the absorption rate r_{12} . Substituting from Equations (1.60) and (1.62),

$$B_{21}f_2(1-f_1)P(E_{21}) > B_{12}f_1(1-f_2)P(E_{21}).$$
(1.70)

Since $B_{21} = B_{12}$, this equation becomes

$$f_2(1-f_1) > f_1(1-f_2).$$
 (1.71)

Substituting for f_1 and f_2 from Equation 1.61, this equation becomes

$$e^{(F_2 - F_1)/kT} > e^{(E_2 - E_1)/kT}$$
(1.72)

or

$$F_2 - F_1 > E_2 - E_1. \tag{1.73}$$

This implies that the difference in the quasi-Fermi levels is greater than the emission energy of the photon. If the emission is at bandgap energy, then the difference between quasi-Fermi levels needs to be greater than the bandgap energy E_{g} .

1.16 Carrier transport in semiconductors

Drift and diffusion are the two mechanisms whereby carriers are transported in semiconductors such that there is current flow. It will be assumed that thermal equilibrium will not be disturbed during these processes [2, 8, 10, 15].

1.16.1 Drift current

When an external electric field is applied to a semiconductor, it produces a force that will accelerate the electrons and holes in opposite directions as long as there are available energy states in the conduction and valence bands. The net drift of charge will produce a current which is the drift current. If the electric field is denoted as \mathcal{E} , the drift current densities for electrons and holes are written as

$$J_{\rm n(drift)} = q n v_{\rm ndr} = q \mu_{\rm n} n \mathcal{E}$$
(1.74)

$$J_{\rm p(drift)} = q p v_{\rm pdr} = q \mu_{\rm p} n \mathcal{E}, \qquad (1.75)$$

where q is the charge on a particle (electron or hole), J is the surface density of current, v_{ndr} and v_{pdr} are the drift velocities of electrons and holes, respectively, and μ is the mobility.

1.16.2 Diffusion current

Electrons flow from a region of higher concentration to a region of lower concentration, producing a flux of electrons and an electron diffusion current which is in the opposite direction to the flux. The hole flow is such that the hole flux and the hole diffusion current are in the same direction since the holes are positively charged. The diffusion current densities for electrons and holes are given by

$$J_{\rm ndiff} = q D_{\rm n} \frac{dn}{dx} \tag{1.76}$$

$$J_{\rm pdiff} = -q D_{\rm p} \frac{dp}{dx},\tag{1.77}$$

where D_n and D_p are the electron and hole diffusion coefficients respectively. The diffusion coefficient is related to the mobility μ by the Einstein relation:

$$D = \frac{\mu kT}{q}.$$
(1.78)

Hence,

$$\frac{D_{\rm n}}{\mu_{\rm n}} = \frac{D_{\rm p}}{\mu_{\rm p}} = \frac{kT}{q}.$$
(1.79)

Adding Equations (1.74)–(1.77),

$$J = (q\mu_{\rm n}n + q\mu_{\rm p}p)\mathcal{E} + qD_{\rm n}\frac{dn}{dx} - qD_{\rm p}\frac{dp}{dx}.$$
(1.80)

1.17 p–n junction

When a junction is formed by a p-type and an n-type semiconductor, holes move from the p to the n side across the metallurgical junction and electrons move in the opposite direction. There are concentration gradients of electrons and holes giving rise to diffusion. Furthermore, when the electrons leave the n region, positively ionised donor atoms remain behind and, similarly, negatively ionised acceptor atoms remain in the p region. These ionised donors and acceptors reside on both sides of the metallurgical junction and are not mobile. The length of this region increases as the diffusion continues. The resultant electric field is directed from the positive charge to the negative charge. This field builds up in such a way as to oppose the diffusion of the carriers in both directions across the junction. An equilibrium condition is reached and there is no net flow of current across the junction. The ionised region on both sides of the metallurgical junction is called the *depletion* or *space charge region*. The p–n junction is seen in Figure 1.25. The band diagram and space charge distribution of a p–n homojunction are shown in Figure 1.26. The general form of Poisson's equation for an abrupt junction where there is an abrupt change in the doping concentration is

$$\frac{d\mathcal{E}}{dx} = \frac{q}{\epsilon_{\rm s}}(p - n + N_{\rm D} - N_{\rm A}),\tag{1.81}$$

where *p* is the hole concentration, *n* is the electron concentration and N_D and N_A are the ionised donor and acceptor concentrations respectively. If the metallurgical junction is the origin, the depletion region on the p side extends to $-x_p$ and on the n side to x_n . Poisson's equation is written for the depletion regions on either side of the metallurgical junction as

$$\frac{d\mathcal{E}}{dx} = -\frac{q}{\epsilon_{\rm s}} N_{\rm A}, \quad -x_{\rm p} < x \le 0 \tag{1.82}$$

$$\frac{d\mathcal{E}}{dx} = \frac{q}{\epsilon_{\rm s}} N_{\rm D}, \quad 0 < x \le x_{\rm n}, \tag{1.83}$$

where ϵ_s is the permittivity of the semiconductor material. The regions outside the depletion region are neutral regions and hence the electric field is zero. Equations (1.82) and (1.83) are solved using the boundary condition on the electric field to get



Fig. 1.25 p–n junction at equilibrium.



Fig. 1.26 p–n junction without externally applied voltage – band diagram (centre) and space charge distribution (bottom).

$$\mathcal{E} = -\frac{qN_{\rm A}}{\epsilon_{\rm s}}(x+x_{\rm p}), \quad -x_{\rm p} < x \le 0 \tag{1.84}$$

$$\mathcal{E} = \frac{qN_{\rm D}}{\epsilon_{\rm s}}(x - x_{\rm n}), \quad 0 < x \le x_{\rm n}.$$
(1.85)

The potential is related to the electric field by the equation

$$\mathcal{E} = -\frac{dV}{dx}.\tag{1.86}$$

Integrating Equations (1.84) and (1.85)

$$V(x) = \frac{qN_{\rm A}}{2\epsilon_{\rm s}}(x+x_{\rm p})^2, \quad -x_{\rm p} < x \le 0$$
(1.87)

$$V(x) = -\frac{qN_{\rm D}}{2\epsilon_{\rm s}}(x - x_{\rm n})^2, \quad 0 < x \le x_{\rm n}.$$
(1.88)

1.17.1 The built-in potential

The built-in potential on the p side of the junction is the potential difference across the depletion region. It is determined similarly on the n side.

$$V_{\rm bip} = \frac{qN_{\rm A}}{2\epsilon_{\rm s}} x_{\rm p}^2 \tag{1.89}$$

$$V_{\rm bin} = \frac{qN_{\rm D}}{2\epsilon_{\rm s}} x_{\rm n}^2. \tag{1.90}$$

The total built-in potential across the junction is

$$V_{\rm bi} = (V_{\rm bip} + V_{\rm bin}) \tag{1.91}$$

$$= \frac{q}{2\epsilon_{\rm s}} [N_{\rm A} x_{\rm p}^2 + N_{\rm D} x_{\rm n}^2].$$

$$(1.92)$$

The continuity of the electric field across the junction at x = 0 requires that

$$N_{\rm A}x_{\rm p} = N_{\rm D}x_{\rm n}.\tag{1.93}$$

It is assumed that the dopants are fully ionised and the total ionised positive charge per unit area on the n side is equal to the total ionised negative charge per unit area on the p side. At thermal equilibrium, there is no net current flow and hence the drift and diffusion currents are equal. The electron current is

$$J_{\rm n} = 0 \tag{1.94}$$

$$= J_{\rm ndrift} + J_{\rm ndiff} \tag{1.95}$$

$$=q\mu_{\rm n}n\mathcal{E}+qD_{\rm n}\frac{dn}{dx}.$$
(1.96)

The hole current is written as

$$J_{\rm p} = 0$$
 (1.97)

$$= J_{\rm pdrift} + J_{\rm pdiff} \tag{1.98}$$

$$= q\mu_{\rm p}p\mathcal{E} - qD_{\rm p}\frac{dp}{dx}.$$
(1.99)

When the net hole current is zero, and with the electric field equal to the gradient of the potential, it may be shown that

$$V_{\rm bi} = \frac{kT}{q} \ln \frac{N_{\rm A} N_{\rm D}}{n_{\rm i}^2}.$$
 (1.100)

It has been assumed that there is full ionisation of the dopant impurity levels such that the majority carrier concentrations are the doping concentrations and the equilibrium concentrations are related by

$$n_0 p_0 = n_i^2. (1.101)$$

1.17.2 The depletion layer width

The widths of the depletion layer in the p- and n-type semiconductors may be calculated. The maximum electric field occurs at the metallurgical junction, x = 0. This is given by

$$\epsilon_{\rm s} \mathcal{E}_{\rm max} = q N_{\rm D} x_{\rm n} = q N_{\rm A} x_{\rm p}. \tag{1.102}$$

Using the Equations (1.92) and (1.93) with

$$|V_{bi}| = \frac{\mathcal{E}_{max}}{2}[x_n + x_p]$$
 (1.103)

$$W = x_{\rm n} + x_{\rm p} \tag{1.104}$$

$$=\sqrt{\frac{2\epsilon_{\rm s}}{q}}\left(\frac{1}{N_{\rm D}}+\frac{1}{N_{\rm A}}\right)\mid V_{\rm bi}\mid.$$
(1.105)

1.17.3 The depletion capacitance

The depletion capacitance is the capacitance at the p–n junction. The depletion layer is modelled as a parallel plate capacitor. The capacitance is written as

$$C_{\rm j} = \frac{\epsilon_{\rm s} A}{W},\tag{1.106}$$

where A is the area of the p–n junction and W is the depletion layer width given by Equation (1.105). The junction capacitance is given by

$$C = A \sqrt{\frac{q\epsilon_{\rm s} N_{\rm A} N_{\rm D}}{2V_{\rm bi}(N_{\rm A} + N_{\rm D})}}.$$
(1.107)

1.17.4 p–n junction under bias

At thermal equilibrium, the total electrostatic potential across the p-n junction is the built-in potential, V_{bi} , and the potential difference between the p and n regions is $q V_{bi}$. If now a voltage V_A is applied with the positive terminal connected to the p side and the negative to the n side, the junction is forward-biased and the total electrostatic potential across the junction is $V_{bi} - V_A$, resulting in a reduction of the depletion layer width. A potential barrier was formed at thermal equilibrium restricting the motion of the majority carriers. The application of the forward bias reduces the height of the barrier. If, on the other hand, a voltage is applied with the positive terminal connected to the n side and the negative terminal to the p side, the electrostatic potential across the junction is $V_{bi} - (-V_A)$ and the height of the barrier is increased with the reverse bias. The depletion widths and the energy band diagrams are shown in the figure. The width of the depletion layer is given by

$$W = \sqrt{\frac{2\epsilon_{\rm s}}{q} \left(\frac{N_{\rm D} + N_{\rm A}}{N_{\rm A}N_{\rm D}}\right) (V_{\rm bi} \mp V_{\rm A})}.$$
(1.108)

1.17.5 Current–voltage characteristics

The total current density in a p-n junction is given as:

$$J = q \left[\frac{D_{\rm n} n_{\rm p0}}{L_{\rm n}} + \frac{D_{\rm p} p_{\rm n0}}{L_{\rm p}} \right] \left[\exp\left(\frac{q V_{\rm A}}{kT}\right) - 1 \right],\tag{1.109}$$