

Ying Fu

---

# Physical Models of Semiconductor Quantum Devices

*Second Edition*

# Physical Models of Semiconductor Quantum Devices

Ying Fu

# Physical Models of Semiconductor Quantum Devices

Second Edition

 Springer

Ying Fu  
Royal Institute of Technology  
Stockholm, Sweden

ISBN 978-94-007-7173-4

ISBN 978-94-007-7174-1 (eBook)

DOI 10.1007/978-94-007-7174-1

Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2013946430

© Springer Science+Business Media Dordrecht 1999, 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

# Preface

The history of technology development is epitomized in Moore's law. Industrial deep-submicron and laboratorial nanometer process technologies have already been fabricating electronic and optical components containing only a few active electrons, and the geometrical sizes of these components are comparable with the characteristic wavelength of the electrons. However, the advanced multimedia infrastructure and service in the future demand further developments in the chip's capability.

Photonic integrated circuits (PICs) are currently orders of magnitude larger in physical dimensions than their microelectronic counterparts. Field-effect-type transistors have reached lengths on the order of 50 nm, while in contrast, passive optical devices, also those based on photonic crystals, have sizes on the order of one photon wavelength. The sizes of active devices are even larger, essentially depending on the matrix element of the interaction. In order to pursue the steady increase in integration density in photonics such that it rivals the microelectronic footprint size, nanostructure-based high index of refraction and metallic behavior (negative epsilon) are two mostly studied fundamental issues to shrink optical component sizes and to tackle the sub-wavelength limit.

Nanotechnology has been named as one of the most important areas of forthcoming technology because they promise to form the basis of future generations of electronic and optoelectronic devices. From the point view of technical physics, all these developments greatly reduce the geometric sizes of devices, and thus the number of active electrons in the system. Quantum mechanical considerations about electronic states, electron transports and various scattering processes including light-matter interaction, are thus crucial. However, the theoretical study is extremely difficult. My first numerical simulation work about a three-dimensional energy band structure calculation in 1995 took more than 6 months to complete for one bias-configuration of a nanoscale metal-oxide-semiconductor field-effect transistor (MOSFET). With today's computation workstations the CPU time is reduced to be less than 24 hours.

In general, today's experimental and theoretical works are very much separated. The laboratory works are still largely based on try-and-error, while the theoretical models are over simplified as compared with the complexity of real devices. Ideally to be cost effective, experimental and theoretical works are to be coordinated in

such a complementary way that we try to analyze and understand the experimental results, then use the understanding to guide further experimental works, which in their turn serve as the feedback to modify and improve the theoretical model. By this, we expect an optimized device and a valid as well as effective theoretical device model.

The main purpose of the book is to discuss electrons and photons in and through nanostructures by the first-principles quantum mechanical theories and fundamental concepts (a unified coverage of nanostructured electronic and optical components) behind nano-electronics and optoelectronics, the material basis, physical phenomena, device physics, as well as designs and applications. The combination of viewpoints presented within the book can help to foster further research and cross-disciplinary interaction needed to surmount the barriers facing future generations of technology design.

Many specific technologies are presented, including quantum electronic devices, resonant tunneling devices, single electron devices, heterostructure bipolar transistors (HBTs) and high electron mobility transistors (HEMTs), detectors, and infrared sensors, lasers, optical modulators. It contains essential and detailed information about the state-of-the-art theories, methodologies, the way of working and real case studies, helping students and researchers to appreciate the current status and future potential of nanotechnology as applied to the electronics and optoelectronics industry.

In nanophotonics we will concentrate on local electromagnetic interactions between nanometric objects and optical fields (non-linear optics in nano- and microstructured photonic crystals) at the level of systems of nanostructures, into larger density on interfaces, which in turn leads to intriguing collective effects, such as plasmonics or multiple reflection and refraction phenomena.

The major task here is that the system at working condition is no longer static. Rather, it can only properly be described by including dynamic Maxwell and time-dependent Schrödinger equations. Furthermore, because the numbers of atoms and electrons in the real devices are huge, while the quantum mechanical Monte Carlo simulation requires too much computer memory and computer time, we will introduce top-down and bottom-up numerical ways that fundamentally we emphasize the quantum mechanical Monte Carlo simulation, while at the same time, we apply the large-system (cluster) tight-binding numerical method to study the device performance property (where the input parameters in the tight-binding method come from the study of bridging nano to micro scales).

Finally we will examine the processing—structure relationship. The state of nanostructures during the period that one monolayer exists—before being buried in the next layer—determines the ultimate structure of the nanostructure, and thus its properties. This part of the book takes into consideration the following potential influencing factors in solid-state growth techniques such as metalorganic vapour phase epitaxy (MOVPE): crystal defects, void structure, grain structure, interface structure in epitaxial films, reaction-induced structure, strain-induced self-formed quantum dot structures, through the use of MOVPE to produce quantum structured semiconductors.

This book provides a solid foundation for the understanding, design, and simulation of nano-electronic and optoelectronics devices. It will be of interest to researchers and specialists in the field of solid state technology, electronics and optoelectronics. It can also serve as a textbook for graduate students and new entrants in the exciting field. This book takes the reader from the introductory stage to the advanced level of the construction, principles of operation, and application of these devices, and puts readers immediately in a position to take their first steps in the field of computational nano-engineering and design. Results and conclusions of detailed nano-engineering studies are presented in an instructive style. Numerous references, illustrations, basic computation subroutines provide further support in this fast-emerging field. This book is designed as a self-contained introduction to both the understanding and solution of theoretical and practical design problems in nano devices.

Stockholm, Sweden  
May 2013

Ying Fu

# Contents

<b>1</b>	<b>Semiconductor Materials</b>	1
1.1	Atoms and Solids	3
1.2	Bulk and Epitaxial Crystal Growth	14
1.3	Bloch Theorem of Electrons in Solids	19
1.4	$sp^3s^*$ Tight-Binding Model	21
1.5	Bandedge States	27
1.6	Eight-Band $k \cdot p$ Model	33
1.7	Strain Field in Nanostructures	40
1.8	Heterostructure Material and Envelope Function	49
1.9	Dimensionality and Density of States	56
	References	62
<b>2</b>	<b>Electron Transport</b>	67
2.1	Quantum Mechanical Wave Transport	67
2.2	Scattering Theory	73
2.3	Time-Dependent Perturbation	82
2.4	Acceleration Theorems	84
2.5	Impurities and Fermi Level of Doped Semiconductor	89
2.6	Boltzmann Equation	97
2.7	Drift, Diffusion and Ballistic Transport	100
2.8	Carrier Scatterings	103
2.8.1	Phonon Scattering	103
2.8.2	Carrier-Carrier Interaction	109
2.8.3	Impurity Scattering	110
	References	110
<b>3</b>	<b>Optical Properties of Semiconductors</b>	111
3.1	Electromagnetic Field	111
3.2	Electron in Electromagnetic Field	118
3.3	Optical Spectrum of Nanostructure	126
3.4	Exciton and Its Optical Properties	134

3.5	Exciton in Quantum Well . . . . .	142
3.6	Colloidal Quantum Dot . . . . .	146
3.7	Exciton Polariton . . . . .	157
3.8	Multiphoton Process . . . . .	168
3.9	Auger Recombination and Impact Ionization . . . . .	174
	References . . . . .	181
<b>4</b>	<b>Electronic Quantum Devices . . . . .</b>	<b>185</b>
4.1	$p - n$ Junction and Field-Effect Transistor . . . . .	185
4.2	Semiclassical vs Quantum Considerations . . . . .	191
4.3	Resonant Tunneling Diode . . . . .	193
4.3.1	$I - V$ Characteristics at Steady State . . . . .	196
4.3.2	Response to a Time-Dependent Perturbation . . . . .	203
4.3.3	Phonon-Assisted Tunneling . . . . .	207
4.4	Heterostructure Barrier Varactor . . . . .	213
4.4.1	Conduction Current . . . . .	216
4.4.2	$C - V$ Characteristics . . . . .	218
4.5	High-Electron-Mobility Transistor . . . . .	224
4.5.1	Remote Impurity Scattering . . . . .	226
4.5.2	$\delta$ -Doped Field-Effect Transistor . . . . .	230
4.6	Nano-scale Field-Effect Transistor . . . . .	233
4.6.1	Wave Characteristics and Threshold Voltage . . . . .	234
4.6.2	Steady-State Wave Transport . . . . .	238
4.6.3	Interface Roughness . . . . .	246
4.6.4	Time-Dependent Wave Packet Transmission . . . . .	249
4.6.5	Nanometer MOSFET Architectures . . . . .	253
4.7	Coulombic Blockade and Single-Electron Transistor . . . . .	257
4.8	Future Perspectives . . . . .	261
4.8.1	Carbon Nanotubes . . . . .	261
4.8.2	Molecular Devices . . . . .	262
4.8.3	Metallic Devices . . . . .	262
4.8.4	Ferromagnetic Devices . . . . .	263
	References . . . . .	263
<b>5</b>	<b>Nanostructured Optoelectronics . . . . .</b>	<b>271</b>
5.1	Optical Transition and Quantum Selection Rule . . . . .	271
5.2	Intraband Optical Transition . . . . .	275
5.3	Optical Grating and Crosstalk . . . . .	289
5.4	Nanostructure Infrared Photodetector . . . . .	302
5.4.1	Quantum Well Infrared Photodetector . . . . .	302
5.4.2	Quantum Wire Infrared Photodetector . . . . .	307
5.4.3	Quantum Dot Infrared Photodetector . . . . .	308
5.5	SiGe Heterostructure Internal Emission Infrared Photodetector . . . . .	313
5.6	Quantum Dot Solar Cell . . . . .	316
5.7	Exciton-Polariton Photonic Crystal . . . . .	323
5.8	Resonant Tunneling Light Emitting Diode . . . . .	335

- 5.9 Nanostructure Laser . . . . . 342
- 5.10 Light Emission from Highly Strained  $\text{In}_{0.3}\text{Ga}_{0.7}\text{As}/\text{GaAs}$ 
  - Quantum Wells by Dipole  $\delta$  Doping . . . . . 345
- 5.11 Quantum Dot Biomarker . . . . . 348
- References . . . . . 350
- 6 Numerical Recipes . . . . . 353**
  - 6.1 Fermi-Dirac Integral . . . . . 353
  - 6.2 Amplitude of Transmitted Wave . . . . . 356
  - 6.3 Localized State . . . . . 366
  - 6.4 Local Density of States: Recursion Method . . . . . 368
  - 6.5 Time-Dependent Wave Packet Transmission . . . . . 377
  - References . . . . . 380
- Appendix A Fundamental Constants . . . . . 381**
- Appendix B Quantum Physics . . . . . 383**
  - B.1 Black Body Radiation . . . . . 383
  - B.2 The Compton Effect . . . . . 383
  - B.3 Electron Diffraction . . . . . 384
  - B.4 Operators in Quantum Physics . . . . . 384
  - B.5 The Schrödinger Equation . . . . . 384
  - B.6 The Uncertainty Principle . . . . . 386
  - B.7 Parity . . . . . 386
  - B.8 The Tunneling Effect . . . . . 387
  - B.9 Harmonic Oscillator . . . . . 388
  - B.10 Angular Momentum and Spin . . . . . 388
  - B.11 Hydrogen Atom . . . . . 390
  - B.12 Interaction with Electromagnetic Fields . . . . . 391
  - B.13 Time-Independent Perturbation Theory . . . . . 391
  - B.14 Time-Dependent Perturbation Theory . . . . . 392
  - B.15  $N$ -Particle System . . . . . 392
  - B.16 Quantum Statistics . . . . . 393
- Appendix C Electricity & Magnetism . . . . . 395**
  - C.1 The Maxwell Equations . . . . . 395
  - C.2 Force and Potential . . . . . 396
  - C.3 Electromagnetic Waves . . . . . 396
- Appendix D Solid State Physics . . . . . 399**
  - D.1 Crystal Structure . . . . . 399
  - D.2 Crystal Binding . . . . . 400
  - D.3 Crystal Vibrations . . . . . 400
  - D.4 Free Electron Fermi Gas . . . . . 401
    - D.4.1 Thermal Heat Capacity . . . . . 401
    - D.4.2 Electric Conductance . . . . . 402
  - D.5 Energy Bands . . . . . 402
- Index . . . . . 405**

# Chapter 1

## Semiconductor Materials

**Abstract** In this chapter we present a brief introduction and description of electrons in semiconductor materials of which devices are made. The chapter starts with the basic electronic energy band structure of a single atom, the modifications of the energy band structure when more atoms are brought together to form bulk materials, i.e., solid states, which is the basis for understanding the electronic and optical properties of semiconductor materials. We then focus on the theoretical descriptions of electron states at the conduction and valence bandedges of both bulk and heterostructure materials. Key contents: Bloch theorem and Schrödinger equation for the envelope function of electrons in solids.

Semiconductors are materials that have moderately good conductivity, which is higher than that of insulators and lower than that of metals. The conductivity of sufficiently pure semiconductors decays by orders of magnitude when they are cooled down from room temperature to liquid helium temperature (at absolute zero temperature, the conductivity almost vanishes). A semiconductor in a very pure state resembles an insulator, whereas in a highly polluted state it acts like a metal. Furthermore, irradiation with light can transform the semiconductor from insulator-like behavior to metal-like behavior. The optical absorption spectra of semiconductors normally exhibit a threshold. Below the threshold frequency, light can pass through with practically no losses, whereas above it the light is strongly absorbed.

All these macroscopic properties of a semiconductor can be traced back to a common microscopic origin: its energy band structure and the electron distribution in the energy bands. The energy band structure of a semiconductor consists of energy bands separated by bandgaps. At absolute zero temperature, a pure semiconductor is characterized by having only completely occupied and completely empty energy bands. It is this common microscopic feature that underlies the totality of macroscopic material properties that uniquely define a semiconductor.

The first reference to a characteristic semiconductor property dates back to Faraday, who in 1833 observed an increase of the electric conductivity of silver sulfide with increasing temperature. The term “semiconductor” was introduced in 1911 by Königsberg and Weiss subsequent to a similar term used in a similar context employed by Ebert in 1789 and Bromme in 1851.

In 1874, Braun discovered that contacts between certain metal sulfides and metal tips exhibited different electrical resistance upon reversal of the polarity of the applied voltage. Such point contact structures were used as rectifiers in radio receivers at the beginning of the 20th century. Similar rectifying action was also found for selenium and copper oxide. Copper oxide was used in 1926 by Grondahle as a rectifier, followed by rectifiers using selenium. The first practical application of copper oxide in photocells was accomplished in 1932 by Lang.

The decisive events for the entire future development of semiconductor physics were the invention of the germanium-based bipolar transistor in 1949 and realization of the field-effect transistor, with the help of silicon at the end of the 1950s. With the introduction of silicon, the development of semiconductor microelectronics began. Later, a similar role was played by compounds involving elements of III–V groups in the periodic table, such as GaAs for the development of semiconductor optoelectronics.

Today's advanced information technology is mainly attributed to the electronic representation and processing of information in a low-cost, high-speed, very compact, and highly reliable fashion. The quest for and accomplishment of continual miniaturization and integration of solid-state electronics have been the key to the success of the computer industry and computer applications.

As the number of transistors integrated in a circuit continues to increase, discrete device dimensions have begun to reach the nanometer regime. Such a down-scaling progress of individual device components has been tremendous over the last 30 years: a 1.0- $\mu\text{m}$  gate length metal-oxide-semiconductor field-effect transistor (MOSFET) was reported in 1974 by Dennard et al. [1]; 0.1- $\mu\text{m}$  gate length by Sai-Halasz et al. [2] in 1987, 70-nm gate length by Hashimoto et al. [3] in 1992, 40-nm gate length by Ono et al. [4] in 1995, 30-nm gate length in 1998 by Kawaura et al. [5]. Vertical *p*-MOS transistors also have been extensively investigated, and transistors with channel lengths of 130 and 100 nm based on Si as well as GaAs have been fabricated [6–11].

However, the advanced multimedia infrastructure and service in the future demand further reduction in the chip size. Chip density, represented by memory technology, has followed Moore's law and roughly doubled every other year. The trend remains strong and definite. For example, a 0.15- $\mu\text{m}$  process technology was implemented in the first 4-Gb dynamic random access memory (DRAM) in 1997 and the feature size of DRAM transistors is projected to be 0.18  $\mu\text{m}$  (1 Gb) in 2001, 0.13  $\mu\text{m}$  (4 Gb) in 2004, 0.10  $\mu\text{m}$  (16 Gb) in 2007, and 0.07  $\mu\text{m}$  (64 Gb) in 2010 [12, 13].

When the size of a system becomes comparable to the electron wavelength, quantum effects become dominant [14]. This occurs when transistors are down-scaled and their characteristic dimensions reach the nanometer regime, leading to various new phenomena, for example, electron interference [15], additive parallel conductance in the absence of magnetic field [16–18], conductance oscillation [19, 20] and abrupt period changes of conductance oscillation with applied magnetic field [21, 22], as well as novel electronic and optoelectronic devices such as resonant tunneling diodes [23, 24] and quantum well infrared photodetectors [25, 26], based on quantum mechanisms.

For nanoelectronics to become a reality, it is essential that new devices and circuits be fabricated with nanometer precision, and that devices and circuits can be designed accurately. Nanofabrication technology [27, 28] has made impressive advances by producing artificial semiconductor structures using molecular-beam epitaxy (MBE), metal-organic chemical vapor deposition (MOCVD), and chemical-beam epitaxy. Accurately controlled feature sizes as small as monolayers of atoms in the growth direction for dissimilar semiconductor materials, or heterostructure systems, have been achieved. Nanoscale lithography and patterning by electron-beam lithography have also been highly developed in the direction perpendicular to the growth direction. Soft lithography [29] can be used to make devices smaller than 100 nm and can pattern curved surfaces and functional materials other than photoresists. It can also fabricate three-dimensional structures and chemically modify surfaces.

The advances in nanofabrication technology have brought quantum effect device concepts to reality and have presented a great challenge for device physicists in the theoretical analysis of nanoelectronic devices [30–32]. In this chapter, we present quantum mechanical descriptions about electron states in both bulk and heterostructure semiconductor materials.

## 1.1 Atoms and Solids

In 1913 Niels Bohr presented a model of the hydrogen atom, which has one electron. Bohr stated two postulates.

1. *The electron moves only in certain circular orbits, called stationary states.*

Figure 1.1 shows an electron of mass  $m_0$  and charge  $-e$ , moving at speed  $v$  in a stable circular orbit of radius  $r$ , around a nucleus of charge  $+e$ . The centripetal force is provided by the Coulombic attraction

$$-\frac{e^2}{4\pi\epsilon_0 r^2}$$

between the electron and the nucleus, where  $\epsilon_0$  is the permittivity of free space. From Newton's second law we have

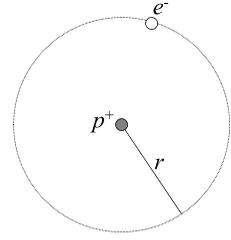
$$\frac{m_0 v^2}{r} = \frac{e^2}{4\pi\epsilon_0 r^2} \quad (1.1)$$

and then the total energy of the electron is

$$E = \frac{1}{2} m_0 v^2 - \frac{e^2}{4\pi\epsilon_0 r} = -\frac{e^2}{8\pi\epsilon_0 r} \quad (1.2)$$

2. *Radiation occurs only when the electron goes from one allowed orbit to another of lower energy. The energy of the radiation is  $\hbar\omega = E_m - E_n$ , where  $E_m$  and  $E_n$  are the energies of two allowed electron orbits.*

**Fig. 1.1** Bohr model of the hydrogen atom. The negatively charged electron  $e^-$  is in a circular orbit of radius  $r$  around the positively-charged proton  $p^+$



To restrict the allowed values of the orbital radius, we need the “third” postulate:  
3. *The angular momentum of the electron is restricted to integer  $n$  multiples of  $\hbar$ :*

$$m_0 v r = n \hbar \quad (1.3)$$

When  $v = n\hbar/m_0 r$  from the above equation is equated to

$$v = \sqrt{\frac{e^2}{4\pi\epsilon_0 m_0 r}}$$

of Eq. (1.1), we find the radius of the  $n$ th orbit is

$$r_n = \frac{4\pi\epsilon_0 \hbar^2 n^2}{m_0 e^2} = n^2 a_0 \quad (1.4)$$

where

$$a_0 = \frac{4\pi\epsilon_0 \hbar^2}{m_0 e^2} = 0.529 \text{ \AA} \quad (1.5)$$

is the Bohr radius. The energy of the  $n$ th orbit is

$$E_n = -\frac{m_0 e^4}{32\pi^2 \epsilon_0^2 \hbar^2 n^2} = -\frac{R_y}{n^2} \quad (1.6)$$

where

$$R_y = \frac{m_0 e^4}{32\pi^2 \epsilon_0^2 \hbar^2} = \frac{\hbar^2}{2m_0 a_0^2} = 13.6 \text{ eV}$$

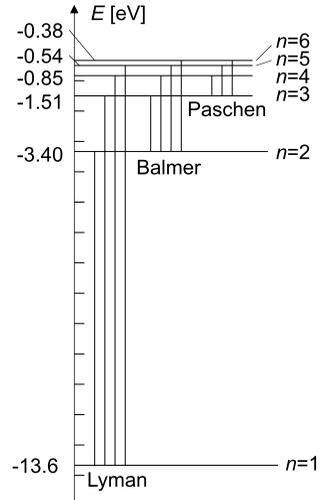
is the Rydberg constant.

Bohr’s theory may be applied to other single electron systems such as  $\text{He}^+$  or  $\text{Li}^{++}$ , provided the nuclear charge is replaced by  $Ze$ , where  $Z$  is the atomic number. And the energy of the  $n$ th state is

$$E_n = -\frac{R_y Z^2}{n^2} \quad (1.7)$$

The energy state diagram for hydrogen ( $Z = 1$ ) is shown in Fig. 1.2. Each state is characterized by the integer  $n$ , which is called the principal quantum number.

**Fig. 1.2** The energy state diagram of the hydrogen atom. Light is emitted or absorbed when an electron makes a transition between two states



When the atom is unexcited, the electron is in the ground state with  $n = 1$ . The electron may be raised to a higher level, normally referred to as an excited state, by a collision with another electron or by absorbing a photon. Note that at steady state, the photon energy must correspond exactly to the energy difference between the two states involved in the transition. The electron may return from an excited state to the ground state in one step or via intermediate levels. A photon with a single frequency is emitted in the first case, there are two or more frequencies in the second case. Lyman's series corresponds to transitions from higher levels to  $n = 1$ , transitions to level  $n = 2$  form the Balmer series; those to  $n = 3$  form the Paschen series.

The hydrogen atom is described rigorously by the Schrödinger equation in its time-independent form

$$\left( -\frac{\hbar^2 \nabla^2}{2m_0} - \frac{Ze^2}{4\pi\epsilon_0 r} \right) \Psi(\mathbf{r}) = E\Psi(\mathbf{r}) \quad (1.8)$$

Because of the spherical symmetry of the potential energy, the wave function can be expressed as

$$\Psi_{n\ell m}(r, \theta, \phi) = R_{n\ell}(r)Y_{\ell m}(\theta, \phi) \quad (1.9)$$

in spherical polar coordinate.  $Y_{\ell m}(\theta, \phi)$  are the angular momentum eigen functions. The first few normalized spherical harmonics  $Y_{\ell m}(\theta, \phi)$  ( $m = -\ell, -\ell + 1, \dots, \ell - 1, \ell$ ) are

$$Y_{00} = \frac{1}{\sqrt{4\pi}}$$

$$Y_{11} = -\sqrt{\frac{3}{8\pi}} \sin\theta e^{i\phi}$$

$$\begin{aligned}
Y_{10} &= \sqrt{\frac{3}{4\pi}} \cos \theta \\
Y_{22} &= \sqrt{\frac{15}{32\pi}} \sin^2 \theta e^{i2\phi} \\
Y_{21} &= -\sqrt{\frac{15}{8\pi}} \sin \theta \cos \theta e^{i\phi} \\
Y_{20} &= \sqrt{\frac{5}{16\pi}} (3 \cos^2 \theta - 1) \\
Y_{33} &= -\sqrt{\frac{35}{64\pi}} \sin^3 \theta e^{3i\phi} \\
Y_{32} &= \sqrt{\frac{105}{32\pi}} \sin^2 \theta \cos \theta e^{2i\phi} \\
Y_{31} &= -\sqrt{\frac{21}{64\pi}} \sin \theta (5 \cos^2 \theta - 1) e^{i\phi} \\
Y_{30} &= \sqrt{\frac{7}{16\pi}} (5 \cos^3 \theta - 3 \cos \theta)
\end{aligned} \tag{1.10}$$

The radial Schrödinger equation is

$$\frac{d^2 R_{n\ell}(r)}{dr^2} + \frac{2}{r} \frac{dR_{n\ell}(r)}{dr} - \frac{\ell(\ell+1)}{r^2} R_{n\ell}(r) + \frac{2m_0}{\hbar^2} \left( E_n - \frac{Ze^2}{4\pi\epsilon_0 r} \right) R_{n\ell}(r) = 0 \tag{1.11}$$

By introducing the Bohr radius  $a_0$ , see Eq. (1.5), the first few normalized radial wave functions are

$$\begin{aligned}
R_{10}(r) &= 2 \left( \frac{Z}{a_0} \right)^{3/2} e^{-Zr/a_0} \\
R_{20}(r) &= \frac{1}{\sqrt{2}} \left( \frac{Z}{a_0} \right)^{3/2} \left( 1 - \frac{Zr}{2a_0} \right) e^{-Zr/2a_0} \\
R_{21}(r) &= \frac{1}{2\sqrt{6}} \left( \frac{Z}{a_0} \right)^{3/2} \frac{Zr}{a_0} e^{-Zr/2a_0} \\
R_{30}(r) &= \frac{2}{3\sqrt{3}} \left( \frac{Z}{a_0} \right)^{3/2} \left( 1 - \frac{2Zr}{3a_0} + \frac{2Z^2 r^2}{27a_0^2} \right) e^{-Zr/3a_0} \\
R_{31}(r) &= \frac{8}{27\sqrt{6}} \left( \frac{Z}{a_0} \right)^{3/2} \left( 1 - \frac{Zr}{6a_0} \right) \frac{Zr}{a_0} e^{-Zr/3a_0} \\
R_{32}(r) &= \frac{4}{81\sqrt{30}} \left( \frac{Z}{a_0} \right)^{3/2} \frac{Z^2 r^2}{a_0^2} e^{-Zr/3a_0}
\end{aligned} \tag{1.12}$$

**Table 1.1** Shell structure of atomic states of electrons

$n$	Shell	$\ell$	Subshell
1	$K$	0	$s$ (sharp)
2	$L$	1	$p$ (principal)
3	$M$	2	$d$ (diffuse)
4	$N$	3	$f$ (fundamental)
5	$O$	4	$g$
6	$P$	5	$h$
.	.	.	.

The energy of each state depends only on the principal quantum number  $n$ , which varies from 1 to  $\infty$ , as shown by Eq. (1.7). The magnitude of the orbital angular momentum,  $L$ , of a state is determined by the orbit quantum number  $\ell$ ,

$$L = \sqrt{\ell(\ell + 1)}\hbar \quad (1.13)$$

where the maximum value of  $\ell$  is restricted by the value of  $n$ :  $\ell = 0, 1, 2, \dots, (n - 1)$ . In order to specify the direction of the angular momentum vector, we need to set up a preferred axis, say, the  $z$  axis. The component of the orbital angular momentum along this axis is also quantized

$$L_z = m\hbar \quad (1.14)$$

where the values of the orbital magnetic quantum number  $m$ , are restricted to  $m = 0, \pm 1, \pm 2, \dots, \pm \ell$ .

All states with a given value of  $n$  are said to form a shell. All these states are referred as degenerate since they have the same energy value. And it is easy to see that the degeneracy of these states is  $n^2$ . States with a given value of  $\ell$  form a subshell. The designations are listed in Table 1.1. The first four letters for the subshells are historical (sharp, principal, diffuse, and fundamental).

In addition, the electron has an intrinsic property called spin that manifests itself according to the following rules. The magnitude of the spin angular momentum,  $S$ , of the electron is determined by its spin quantum number,  $s = 1/2$ :

$$S = \sqrt{s(s + 1)}\hbar = \frac{\sqrt{3}}{2}\hbar \quad (1.15)$$

In a magnetic field, the  $z$  component of the spin can assume only two values

$$S_z = m_s\hbar \quad (1.16)$$

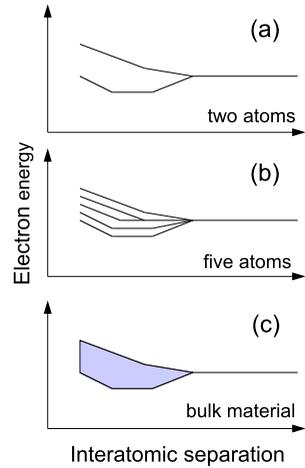
where the spin magnetic quantum number,  $m_s = \pm 1/2$ . The introduction of spin doubles the number of states allowed for each value of  $n$ .

Now, four quantum numbers,  $n$ ,  $\ell$ ,  $m_\ell$ , and  $m_s$  may be used to classify the states of electrons in all atoms, although the energy associated with a given set of values depends on the atom. The question naturally arises as to why all electrons in an

**Fig. 1.3** A simple, but approximate, mnemonic for the filling of sublevels

▲ 1s  
 ▲ 2s 2p  
 ▲ 3s 3p 3d  
 ▲ 4s 4p 4d 4f  
 ▲ 5s 5p 5d 5f 5g  
 ▲ 6s 6p 6d  
 ▲ 7s

**Fig. 1.4** (a) As two atoms are brought closer together, a single atomic level splits into two states with different energies. (b) A single atomic level splits into five when five atoms are in close proximity. (c) In a crystal each atomic level splits into an essentially continuous band of energies



atom do not fall to the ground state. A study of the classification of spectral lines led W. Pauli in 1925 to make an important statement, now called the Pauli exclusion principle:

*No two electrons in an atom can have the same four quantum numbers  $n$ ,  $\ell$ ,  $m_\ell$ , and  $m_s$ .*

With the aid of the exclusion principle one can see how electrons fill shells ( $n$ ) and subshells ( $\ell$ ). For each value of  $\ell$  there are  $(2\ell + 1)$  values of  $m_\ell$  and each subshell can accommodate  $2(2\ell + 1)$ . A simple useful, but approximate, mnemonic that tells us the order in which the subshells are first filled is shown in Fig. 1.3.

The ground-state electron configurations are indicated in the periodic table by the number of electrons in a subshell as a superscript. For example,  $2p^3$  means that there are three electrons in subshell  $\ell = 1$ .

In isolated atoms the energy levels are sharply defined. Now suppose that two atoms are brought close to each other so that their electron wave functions overlap. As a result of the interaction between the electrons, it turns out that each single state of the isolated atom splits into two states with different energies. As Fig. 1.4 shows, the degree of splitting increases as the interatomic separation decreases. Similarly, if five atoms are placed in close proximity, each original energy level splits into five new levels. The same process occurs in a solid, where there are roughly  $10^{28}$

**Table 1.2** Electron configurations of typical elements making up common semiconductors

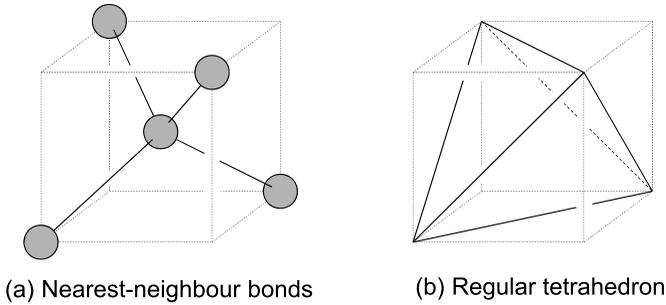
		Group IV			
		Core electrons		Valence electrons	
C		$1s^2$		$2s^2 2p^2$	
Si		$1s^2 2s^2 2p^6$		$3s^2 3p^2$	
Ge		$1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10}$		$4s^2 4p^2$	
		Group III		Group V	
		Core electrons		Core electrons	
		Valence electrons		Valence electrons	
Al	$1s^2 2s^2 2p^6$	$3s^2 3p^1$		N	$1s^2$
Ga	$1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10}$	$4s^2 4p^1$		P	$1s^2 2s^2 2p^6$
In	[Kr] $4d^{10}$	$5s^2 5p^1$		As	$1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10}$
				Sb	[Kr] $4d^{10}$
					$5s^2 5p^3$
		Group II		Group VI	
		Core electrons		Core electrons	
		Valence electrons		Valence electrons	
Zn	$1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10}$	$4s^2$		O	$1s^2$
Cd	[Kr] $4d^{10}$	$5s^2$		S	$1s^2 2s^2 2p^6$
Hg	[Xe] $4f^{14} 5d^{10}$	$6s^2$		Se	$1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10}$
				Te	[Kr] $4d^{10}$
					$5s^2 5p^4$

atoms/m<sup>3</sup>: The energy levels associated with each state of the isolated atom spread into essentially continuous energy bands separated from each other by energy gaps.

Before further examining the various properties of semiconductors it is extremely useful to examine the electron configurations of some of the elements which make up the various semiconductors as listed in Table 1.2.

A very important conclusion can be drawn about the elements making up the semiconductors: The outmost valence electrons are made up of electrons in either the *s*- or *p*-type orbitals. While this conclusion is strictly true for elements in the atomic form, it turns out that even in the crystalline semiconductors the electrons in the valence and conduction bands retain this *s*- or *p*-type character. The core electrons are usually not of interest, except of some special characterization-type experiments.

Here we have assumed that solids are composed of ion cores, i.e., nuclei, and those core electrons so strongly bound as to be negligibly perturbed from their atomic configuration by their environment in the solid, and valence electrons, i.e., those electrons whose configuration in the solid may differ significantly from that in the isolated atom. However, it is to be remembered that the distinction between core and valence electrons in Table 1.2 is one of convenience. For example, the en-



**Fig. 1.5** (a) Si atom interacts with four neighboring Si atoms positioned at the vertices of (b) a regular tetrahedron by sharing its four valence electrons ( $3s^23p^2$ ) with the four neighboring Si atoms

ergy of  $3d$  states is lower than  $4p$  according to Fig. 1.3 so that, instead of being core electrons in Table 1.2, electrons at  $3d^{10}$  in Ge should be categorized as valence electrons. This is one of major reasons that the energy band structure of Ge is much more complicated than Si.

In general it is found that when atoms exchange or share valence electrons so that the complement of quantum states is completed, they have a lower electrostatic energy for their combined electron patterns than when they are separate. For example, Si has four valence electrons grouped in two closely spaced energy levels ( $3s$  and  $3p$ , see Table 1.2), they can combine with themselves by sharing four valence electrons with four surrounding Si atoms in an endless array. The four nearest neighboring Si atoms around any one Si atom are positioned at the vertices of a regular tetrahedron, forming four tetrahedral bonds with the central atom, see Fig. 1.5. This creates the diamond crystal structure.

The intrinsic property of a crystal is that the environment around a given atom or a group of atoms is exactly the same as the environment around another atom or a similar group of atoms. To understand and to define the crystal structure, two important concepts are introduced, i.e., the Bravais lattice and the basis.

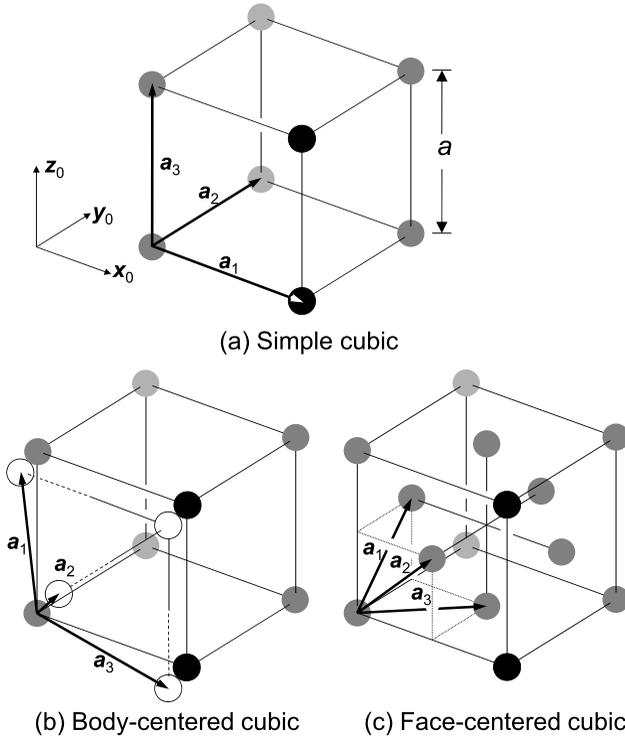
The *Bravais lattice* represents a set of points in the space which form a periodic structure. Each point sees exactly the same environment. A building block of atoms, called the *basis*, is then attached to each lattice point, yielding a crystal structure.

An important property of a Bravais lattice is the ability to define three vectors,  $\mathbf{a}_1$ ,  $\mathbf{a}_2$ , and  $\mathbf{a}_3$ , such that any lattice point  $\mathbf{R}'$  can be obtained from any other lattice point  $\mathbf{R}$  by a translation

$$\mathbf{R}' = \mathbf{R} + m_1\mathbf{a}_1 + m_2\mathbf{a}_2 + m_3\mathbf{a}_3 \quad (1.17)$$

where  $m_1$ ,  $m_2$ , and  $m_3$  range through all integral values (negative, zero, as well as positive). The translation vectors,  $\mathbf{a}_1$ ,  $\mathbf{a}_2$ , and  $\mathbf{a}_3$  are called primitive vectors that generate the Bravais lattice (which will be simply referred to as the lattice, or crystal lattice).

There are 14 types of lattices in the three dimensional space. We shall focus on the cubic lattice which is the structure taken by commonly used semiconductors.



**Fig. 1.6** (a) Simple cubic, (b) body-centered cubic, and (c) face-centered cubic lattices.  $a$  is the lattice constant

There are three kinds of cubic lattices: simple cubic, body-centered cubic and face-centered cubic, see Fig. 1.6. The simple cubic lattice, see Fig. 1.6(a), is generated by the primitive vectors of

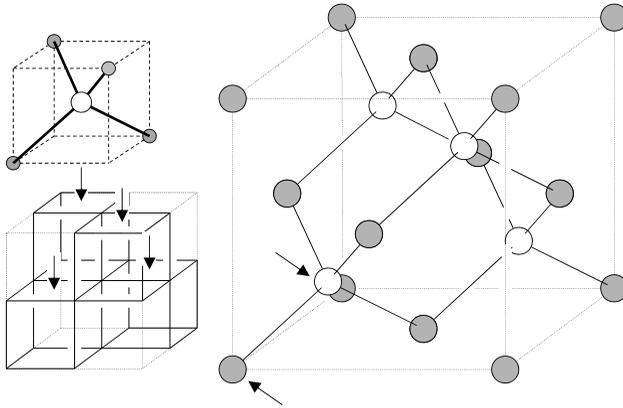
$$\mathbf{a}_1 = a\mathbf{x}_0, \quad \mathbf{a}_2 = a\mathbf{y}_0, \quad \mathbf{a}_3 = a\mathbf{z}_0 \quad (1.18)$$

where  $\mathbf{x}_0$ ,  $\mathbf{y}_0$ , and  $\mathbf{z}_0$  are the three unit vectors of a normal rectangular Cartesian coordinate,  $a$  is the lattice constant.

The body-centered cubic (bcc) lattice is formed by adding to the simple cubic lattice an additional lattice point at the center of the simple cube, see Fig. 1.6(b). A symmetric set of primitive vectors for the bcc lattice is

$$\mathbf{a}_1 = \frac{a}{2}(\mathbf{y}_0 + \mathbf{z}_0 - \mathbf{x}_0), \quad \mathbf{a}_2 = \frac{a}{2}(\mathbf{z}_0 + \mathbf{x}_0 - \mathbf{y}_0), \quad \mathbf{a}_3 = \frac{a}{2}(\mathbf{x}_0 + \mathbf{y}_0 - \mathbf{z}_0) \quad (1.19)$$

The *face-centered cubic lattice* (fcc), see Fig. 1.6(c): To construct the fcc lattice we add to the simple cubic lattice an additional lattice point in the center of each square face. The fcc lattice is of great importance, since an enormous variety of solids crystallize in this form with an atom (or ion) at each lattice site. A symmetric



**Fig. 1.7** Face-centered cubic lattice with two-atom basis forming either the so-called diamond (when the two atoms are identical) or the zincblende (when the two atoms in the basis are different) structure

set of primitive vectors for the fcc lattice is

$$\mathbf{a}_1 = \frac{a}{2}(\mathbf{y}_0 + \mathbf{z}_0), \quad \mathbf{a}_2 = \frac{a}{2}(\mathbf{z}_0 + \mathbf{x}_0), \quad \mathbf{a}_3 = \frac{a}{2}(\mathbf{x}_0 + \mathbf{y}_0) \quad (1.20)$$

Essentially all semiconductors of interest for electronics and optoelectronics have the fcc structure. However, they have two atoms per basis. The coordinates of the two basis atoms are (000) (the grey atom) and  $(a/4)(111)$  (white), indicated in Fig. 1.7 by two tilted arrows. If the two atoms of the basis are identical, the structure is called the diamond structure. Semiconductors such as silicon, germanium and carbon (also refer to Fig. 1.5) fall into this category. If the two atoms are different, for example, GaAs, AlAs, CdS, the structure is called zincblende. The structure can be viewed as a stack, left side of Fig. 1.7, of four regular tetrahedrons shown in Fig. 1.5.

Semiconductors with the diamond structure are often called elemental semiconductors, while the zincblende semiconductors are usually called compound semiconductors. The compound semiconductors are also denoted by the positions of the atoms in the periodic table, for example, GaAs, AlAs and InP are called III–V semiconductors while CdS, CdSe and CdTe are called II–VI semiconductors.

Many of the properties of crystals and many of the theoretical techniques used to describe crystals derive from the periodicity of crystalline structures. This suggests the use of Fourier analysis as an analytical tool. In the analysis of periodic time varying fields (for example, the acoustic signal analysis and radio signal analysis) we often do much of the analytical work in the frequency domain rather than in the time domain. In analogy with the time-frequency duality, there is a corresponding *real space-reciprocal space* or *wave vector space* duality for crystal-related discussions. Many concepts are best understood in terms of functions of the *wave vector*. We describe a plane wave with wavelength  $\lambda$  equivalently as a plane wave with

wave vector  $\mathbf{k}$  of magnitude  $2\pi/\lambda$  and propagation direction perpendicular to the wave front. The space of the wave vectors is called the reciprocal space, the analogue of the frequency domain for the time problem.

A simple transformation is carried out to map the real space lattice into the reciprocal space ( $\mathbf{k}$ -space)

$$\mathbf{b}_1 = 2\pi \frac{\mathbf{a}_2 \times \mathbf{a}_3}{\mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3}, \quad \mathbf{b}_2 = 2\pi \frac{\mathbf{a}_3 \times \mathbf{a}_1}{\mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3}, \quad \mathbf{b}_3 = 2\pi \frac{\mathbf{a}_1 \times \mathbf{a}_2}{\mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3} \quad (1.21)$$

by which it is easy to find that the simple cubic Bravais lattice, with cubic primitive cell of lattice constant  $a$ , and primitive vectors Eq. (1.18), has a simple cubic reciprocal lattice

$$\mathbf{a}_1 = ax_0, \quad \mathbf{a}_2 = ay_0, \quad \mathbf{a}_3 = az_0 \quad (1.22)$$

has a simple cubic reciprocal lattice with cubic primitive cell of side  $2\pi/a$ , and primitive vectors

$$\mathbf{b}_1 = \frac{2\pi}{a}x_0, \quad \mathbf{b}_2 = \frac{2\pi}{a}y_0, \quad \mathbf{b}_3 = \frac{2\pi}{a}z_0 \quad (1.23)$$

Similarly, the primitive vectors of the reciprocal lattice of the fcc Bravais lattice with primitive vectors Eq. (1.20) are

$$\mathbf{b}_1 = \frac{2\pi}{a}(\mathbf{y}_0 + \mathbf{z}_0 - \mathbf{x}_0), \quad \mathbf{b}_2 = \frac{2\pi}{a}(\mathbf{z}_0 + \mathbf{x}_0 - \mathbf{y}_0), \quad \mathbf{b}_3 = \frac{2\pi}{a}(\mathbf{x}_0 + \mathbf{y}_0 - \mathbf{z}_0) \quad (1.24)$$

A general vector

$$\mathbf{G} = m'_1 \mathbf{b}_1 + m'_2 \mathbf{b}_2 + m'_3 \mathbf{b}_3 \quad (1.25)$$

is called a reciprocal lattice vector, where the  $m'_1$ ,  $m'_2$  and  $m'_3$  are three integers (either positive or negative).

It is easy to see that by Eq. (1.21),

$$\mathbf{b}_i \cdot \mathbf{a}_j = 2\pi \delta_{ij} \quad (1.26)$$

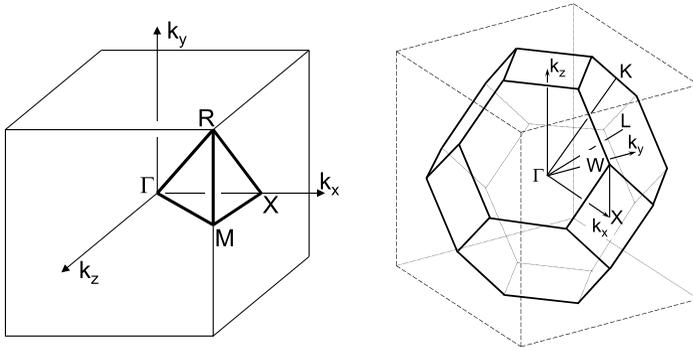
which resulting in the following special relationship

$$e^{i\mathbf{G} \cdot \mathbf{R}} = e^{i2\pi(m'_1 m_1 + m'_2 m_2 + m'_3 m_3)} = 1 \quad (1.27)$$

where  $\mathbf{R}$  is a lattice vector in Eq. (1.17) which is often called the direct lattice vector to distinguish it from the reciprocal lattice vector  $\mathbf{G}$ .

Because of the above relationship, two wave vectors  $\mathbf{k}$  and  $\mathbf{k}'$  satisfying

$$\mathbf{k}' = \mathbf{k} + \mathbf{G} \quad (1.28)$$



**Fig. 1.8** First Brillouin zones for simple cubic lattice (*left*) and fcc (*right*) lattice. Points and lines of symmetry are indicated

are said to be equivalent. This implies that we only need to focus on those  $\mathbf{k}$  points that lie within or on the so-called Brillouin zone, which has the property that no two interior  $\mathbf{k}$  points are equivalent. From here and throughout, we consider only the first Brillouin zone which is the region in the reciprocal space that is closer to the center of the reciprocal space than to any other reciprocal lattice point. The first Brillouin zones for the simple cubic and fcc lattices are shown in Fig. 1.8.

Most importantly we consider XR-R $\Gamma$ - $\Gamma$ X-XM in the  $\mathbf{k}$ -space (in unit of  $2\pi/a$ )

$$\Gamma = (0, 0, 0), \quad X = \left(\frac{1}{2}, 0, 0\right) \quad R = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right), \quad M = \left(\frac{1}{2}, 0, \frac{1}{2}\right) \quad (1.29)$$

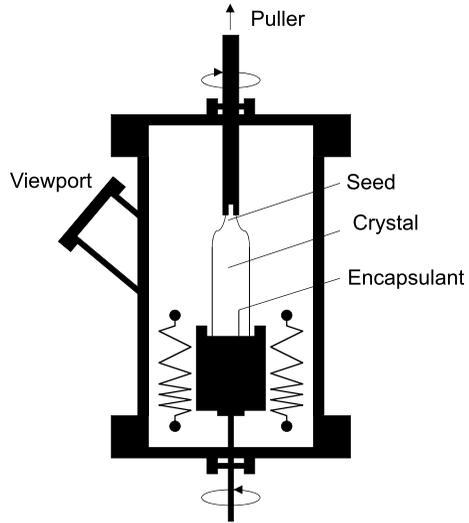
for the simple cubic lattice. For the fcc lattice, such as the diamond as well as zincblende structures (fcc lattices with bases), we mostly consider XU-UL-L $\Gamma$ - $\Gamma$ X-XW-WK in the  $\mathbf{k}$ -space (in unit of  $2\pi/a$ ),

$$\begin{aligned} \Gamma &= (0, 0, 0), & X &= (1, 0, 0), & L &= \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right), & W &= \left(1, 0, \frac{1}{2}\right) \\ U &= \left(\frac{1}{4}, \frac{1}{4}, 1\right), & K &= \left(\frac{3}{4}, 0, \frac{3}{4}\right) \end{aligned} \quad (1.30)$$

## 1.2 Bulk and Epitaxial Crystal Growth

So far we have discussed crystal structures that are present in natural semiconductors. These structures are the lowest free energy configuration of the solid state of the atoms. Since the electrical and optical properties of the semiconductors are completely determined by the crystal structures, artificial structures, e.g., hetero materials (among them the well-known superlattices have been fabricating ever since mid-1970s inspired by the pioneering work of Esaki and Tsu at IBM) grown by

**Fig. 1.9** Schematic of Czochralski-style crystal grower used to produce substrate ingots. The approach is widely used for Si, GaAs and InP



heteroepitaxial crystal growth techniques such as molecular beam epitaxy (MBE) and metal-organic chemical vapor deposition (MOCVD) have made a tremendous impact on the semiconductor physics, the semiconductor technology and the semiconductor electronic and optoelectronic device industry.

Bulk crystal growth techniques are used mainly to produce substrates on which devices are eventually fabricated. While for some semiconductors like silicon and GaAs (to some extent for InP) the bulk crystal growth techniques are highly matured; for most other semiconductors it is difficult to obtain high quality, large area substrate. The aim of the bulk crystal growth techniques is to produce single crystal boules with as large a diameter as possible and with as few defects as possible. For silicon the boule diameters have reached 30 cm with boule lengths approaching 100 cm. Large size substrates ensure low cost device production.

Any material that will crystallize can be crystallized by slow cooling from a molten mass, or by cooling a supersaturated solution of the material. The classic home experiment is that of cooling a supersaturated solution of copper sulphate in water; crystal platelets will readily form as the liquor cools. A much larger crystal can be grown if a seed crystal of copper sulphate is suspended in the solution as it cools, the growth is then onto the seed crystal. One important technique is the Czochralski (CZ) technique. In the CZ technique shown in Fig. 1.9, the melt of the charge (i.e., the high quality polycrystalline material) is held in a vertical crucible. The top surface of the melt is just barely above the melting temperature. A seed crystal is then lowered into the melt and slowly withdrawn. As the heat from the melt flows up the seed, the melt surface cools and the crystal begins to grow. The seed is rotated about its axis to produce a roughly circular cross-section crystal. The rotation inhibits the natural tendency of the crystal to grow along certain orientations to produce a faceted crystal. The resulting crystal is called a boule and may be several centimeters in diameter and a good fraction of a meter in length. Some materials, for

example GaAs, must be used very carefully. Arsenic is likely to boil off the melt, which has to be kept under pressure of an inert gas to prevent this; one may also have a layer of suitable molten glass over the melt as a further protection. Material grown by this process is referred to as bulk grown.

The CZ technique is widely employed for silicon, GaAs, and InP and produces long ingots (boules) with very good circular cross-sections. For silicon up to 100 kg ingots can be obtained. In the case of GaAs and InP the CZ technique has to face problems arising from the very high pressures of As and P at the melting temperatures of the compounds. Not only does the chamber have to withstand such pressures, also the As and P leave the melt and condense on the side walls. To avoid the second problem one can seal the melt by covering it with a molten layer of a second material (e.g., boron oxide) which floats on the surface. The technique is then referred as liquid encapsulated Czochralski, or the LEC technique.

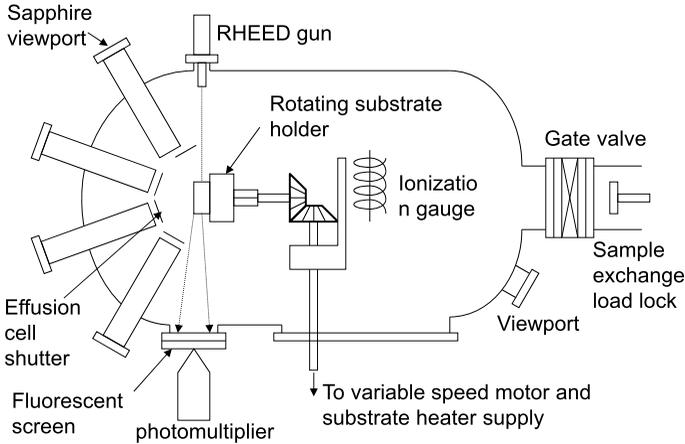
A second bulk crystal growth technique involves a charge of material loaded in a quartz container. The charge may be composed of either high quality polycrystalline material or carefully measured quantities of elements which make up a compound crystal. The container called a “boat” is heated till the charge melts and wets the seed crystal. The seed is then used to crystallize the melt by slowly lowering the boat temperature starting from the seed end. In the gradient-freeze approach the boat is pushed into a furnace (to melt the charge) and slowly pulled out. In the Bridgeman approach, the boat is kept stationary while the furnace temperature is temporally varied to form the crystal.

The easiest approach for the boat technique is to use a horizontal boat. However, the shape of the boule that is produced has a D-shaped form. To produce circular cross-sections vertical configurations have now been developed for GaAs and InP.

In addition to produce high-purity bulk crystals, the techniques discussed above are also responsible for producing crystals with specified electrical properties. This may involve high-resistivity materials along with *n*- or *p*-type materials. In silicon it is difficult to produce high resistivity substrate by bulk crystal growth and resistivities are usually less than  $10^4 \Omega \cdot \text{cm}$ . However, in compound semiconductors carrier trapping impurities such as chromium and iron can be used to produce materials with resistivities of about  $10^8 \Omega \cdot \text{cm}$ . The high resistivity or semi-insulating substrates are extremely useful in device isolation and for high-speed devices. For *n*- or *p*-type dopings carefully measured dopants are added in the melt.

The availability of high quality substrates is essential to any device technology. Other than the three materials of Si, GaAs, and InP, the substrate fabrication of semiconductors is still in its infancy. Since epitaxial growth techniques used for devices require close lattice matching between the substrate and the overlayer, non-availability of substrates can seriously hinder the progress of a material technology. This is, for example, one of the reasons of slow progress in large bandgap semiconductor technology necessary for high-power and high-temperature electronic devices and short-wavelength semiconductor lasers.

The epitaxial growth techniques have a very slow growth rate (as low as a monolayer per second for some techniques) which allow one to control very accurately the dimensions in the growth direction. In fact, in techniques like molecular beam



**Fig. 1.10** Principle of the MBE growth system. In situ monitoring capabilities are often standard components of the system

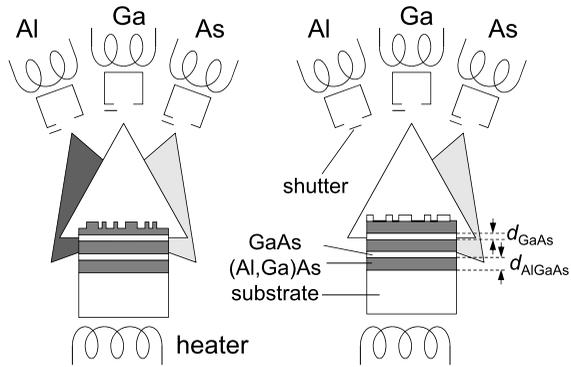
epitaxy and metal organic chemical vapor deposition, one can achieve monolayer (about  $3 \text{ \AA}$ ) control in the growth direction. This level of control is essential for the variety of heterostructure devices that have been beginning to be used in electronics and optoelectronics. The epitaxial techniques are also very useful for precise doping profiles that can be achieved.

Molecular beam epitaxy (MBE) is one of the most important epitaxial techniques as far as heterostructure physics and devices are concerned. MBE is a high vacuum technique (about  $10^{-11}$  torr vacuum when fully pumped down) in which crucibles containing a variety of elemental charges are placed in the growth chamber (Fig. 1.10). The elements contained in the crucibles make up the components of the crystal to be grown as well as the dopants that may be used. When a crucible is heated, atoms or molecules of the charge are evaporated and these travel in straight lines to impinge on a heated substrate.

The growth rate in MBE is about 0.1 monolayer per second and this slow rate coupled with shutters placed in front of the crucibles allow one to switch the composition of the growing crystal with monolayer control. However, to do so, the growth conditions have to be adjusted so that growth occurs in the monolayer by monolayer mode rather than by three dimensional island formation. This requires that atoms impinging on the substrate have enough kinetics to reach an atomically flat profile. Thus the substrate temperature has to be maintained at a point where it is high enough to provide enough surface migration to the incorporating atoms, but not so high as to cause entropy controlled defects.

Since MBE allows one to grow crystal structures with atomic control, one can change the periodicity of the crystals. This leads to the concept of superlattices where two (or even more) semiconductors  $A$  and  $B$  are grown alternately with thickness  $d_A$  and  $d_B$  respectively along the growth direction. The periodicity of the su-

**Fig. 1.11** Schematic diagram illustrating the MBE growth of an AlGaAs/GaAs superlattice sample. Deposition of (a) (Al, Ga)As, (b) GaAs



perlattice in the growth direction is then  $d_A + d_B$ . An AlGaAs/GaAs superlattice grown by MBE is illustrated in Fig. 1.11.

Because of the different sizes of atoms that compose the semiconductor materials, different semiconductor materials have different lattice constants  $a$ . For example,  $a_{\text{GaAs}} = 5.65 \text{ \AA}$ ,  $a_{\text{AlAs}} = 5.66 \text{ \AA}$ ,  $a_{\text{InAs}} = 6.06 \text{ \AA}$  (see more later in Table 1.4). Superlattices can then be placed in three general categories: (i) lattice matched such as AlGaAs/GaAs, (ii) lattice strained (InAs/GaAs), and (iii) lattice strained with intermediate substrate. We shall discuss lattice strain in hetero materials more extensively later.

Since no chemical reactions occur in MBE, the growth is the simplest of all epitaxial techniques and is quite controllable. However, since the growth involves high vacuum, leaks can be a major problem. The growth chamber walls are usually cooled by liquid  $\text{N}_2$  to ensure high vacuum and to prevent atoms/molecules to come off from the chamber walls.

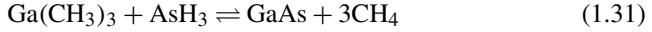
The low background pressure in MBE allows one to use electron beams to monitor the growing crystal. The reflection high-energy electron diffraction (RHEED) technique relies on electron diffraction to monitor both the quality of the growth substrate and the layer-by-layer growth mode. As each monolayer gets filled up, one can see this reflected in the RHEED intensity by the naked eye!

While MBE is a simple and elegant growth technique, it cannot be used conveniently for all semiconductors. For example, phosphides are often not grown by MBE due to the danger in handling elemental phosphorus. Also elements with very low vapor pressures are difficult to use since it is not easy to heat the crucibles beyond 1500 K. Silicon epitaxy in MBE, for example, requires an electron-beam evaporation where an electron beam is used to knock off Si atoms for growth.

In general, MBE is a relatively safe technique and has become the technique of choice for the testing of almost all new ideas on heterostructure physics.

Metal organic chemical vapor deposition (MOCVD) is another important growth technique widely used for heteroepitaxy. Like MBE, it is also capable of producing monolayer-sharp interfaces between semiconductors. Unlike in MBE, the gases that are used in MOCVD are not made of single elements, but are complex molecules which contain elements like Ga or As to form the crystal. Thus the growth depends

upon the chemical reactions occurring at the heated substrate surface. For example, in the growth of GaAs one often uses Triethyl Gallium and Arsine and the crystal growth depends upon the following reaction



One advantage of the growth occurring via a chemical reaction is that one can use lateral temperature control to carry out local area growth. Laser assisted local area growth is also possible for some materials and can be used to produce new kinds of device structures. Such local area growth is difficult in MBE.

There are several varieties of MOCVD reactors. In the atmospheric MOCVD the growth chamber is essentially at atmospheric pressure. One needs a large amount of gases for growth in this case, although one does not have the problems associated with vacuum generation. In the low-pressure MOCVD the growth chamber pressure is kept low. The growth rate is then slow as in the MBE.

The use of the MOCVD equipment requires very serious safety precautions. The gases used are highly toxic and a great many safety features have to be incorporated to avoid any deadly accidents.

In addition to MBE and MOCVD one has hybrid epitaxial techniques often called MOMBE (metal organic MBE) which try to combine the best of MBE and MOCVD. In MBE one has to open the chamber to load the charge for the materials to be grown while this is avoided in MOCVD where gas bottles can be easily replaced from outside. Additionally, in MBE one has occasional spitting of material in which small clumps of atoms are evaporated off on to the substrate. This is avoided in MOCVD and MOMBE.

### 1.3 Bloch Theorem of Electrons in Solids

We now study the properties of electrons in solids. We start with the real-space Schrödinger equation for an electron in a periodic lattice structure

$$H_0(\mathbf{r})\Psi(\mathbf{r}) = \left[ \frac{-\hbar^2 \nabla^2}{2m_0} + V(\mathbf{r}) \right] \Psi(\mathbf{r}) = E\Psi(\mathbf{r}) \quad (1.32)$$

where the first term represents the kinetic energy of the electron and  $V(\mathbf{r})$  is the potential energy of the lattice

$$V(\mathbf{r} + \mathbf{R}) = V(\mathbf{r}) \quad (1.33)$$

where  $\mathbf{R}$  is any lattice vector defined by Eq. (1.17).  $m_0$  is the free electron mass.

The Bloch theorem states that the solutions of the Schrödinger equation of Eq. (1.32) with periodic condition of Eq. (1.33), denoted by quantum numbers  $n$  and  $\mathbf{k}$ , have the following properties

$$\Psi_{n\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{N}} u_{n\mathbf{k}}(\mathbf{r}) e^{i\mathbf{k} \cdot \mathbf{r}}$$

$$u_{nk}(\mathbf{r}) = u_{nk}(\mathbf{r} + \mathbf{R}) \quad (1.34)$$

$$\int_{\text{cell}} u_{nk}^*(\mathbf{r})u_{nk}(\mathbf{r})d\mathbf{r} = 1$$

and  $E = E_n(\mathbf{k})$  is the energy dispersion relationship. Here  $N$  is the total number of unit cells in the crystal. The unit cell is defined by the primitive vectors of the lattice.  $n$  is the energy band index and  $\hbar\mathbf{k}$  is the quasi-momentum of the electron. The rigorous derivation of the Bloch theorem can be found in many solid state textbooks, while a brief understanding can be formulated as follows.

Because of the periodic condition, the physical properties at  $\mathbf{r}$  are expected to be identical to the ones at  $\mathbf{r} + \mathbf{R}$  for which we can write

$$|\Psi_{nk}(\mathbf{r})|^2 = |\Psi_{nk}(\mathbf{r} + \mathbf{R})|^2 \quad (1.35)$$

Note that the wave function itself is not directly physical from the quantum mechanical point of view, while the amplitude of the wave function represents the spatial distribution of the electron, thus resulting in the above equation due to the periodic condition. The above equation can be fulfilled when the wave function is decomposed into a part which has the periodicity as the crystal and a phase factor. The expression for the phase factor in Eq. (1.34), i.e.,  $e^{i\mathbf{k}\cdot\mathbf{r}}$ , becomes the most natural choice when we recall Eq. (1.17), i.e.,  $\mathbf{R} = m_1\mathbf{a}_1 + m_2\mathbf{a}_2 + m_3\mathbf{a}_3$ , where  $m_1, m_2$  and  $m_3$  are integers,  $\mathbf{a}_1, \mathbf{a}_2$  and  $\mathbf{a}_3$  are primitive vectors of the lattice.

We now see that the solution of the Schrödinger equation of Eq. (1.32) is to be characterized by  $\mathbf{k}$ . Furthermore, there must be many electron states because of the large numbers of electrons in the crystal for which we introduce a quantum number  $n$  to distinguish them so that the total wave function of an electron state  $n\mathbf{k}$  in the crystal is expressed as

$$\Psi_{nk}(\mathbf{r}) = Bu_{nk}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}} \quad (1.36)$$

$$u_{nk}(\mathbf{r}) = u_{nk}(\mathbf{r} + \mathbf{R})$$

with an eigen value  $E_n(\mathbf{k})$ . This is almost identical to the Bloch theorem of Eq. (1.34) except a parameter  $B$  which is to be obtained by normalizing the wave function

$$\int_{\Omega} |\Psi_{nk}(\mathbf{r})|^2 d\mathbf{r} = 1 \quad (1.37)$$

where  $\Omega$  denotes the volume of the crystal. Insert Eq. (1.36) into the above equation,

$$\int_{\Omega} |\Psi_{nk}(\mathbf{r})|^2 d\mathbf{r} = |B|^2 \int_{\Omega} |u_{nk}(\mathbf{r})|^2 d\mathbf{r} = 1 \quad (1.38)$$

Because of the periodicity of the crystal, we only need to focus on the spatial region in one unit cell in the crystal, say unit cell 1. All other unit cells, denoted as  $i = 2, 3, \dots, N$ , in the crystal can be expressed by displacing unit cell 1 by the lattice

vector  $\mathbf{R}_i - \mathbf{R}_1$ , see Eq. (1.17). Here  $N$  is the total number of unit cells in the crystal volume  $\Omega$ . Thus,

$$\begin{aligned} \int_{\Omega} |\Psi_{nk}(\mathbf{r})|^2 d\mathbf{r} &= |B|^2 \int_{\Omega} |u_{nk}(\mathbf{r})|^2 d\mathbf{r} = |B|^2 \sum_{i=1}^N \int_{\text{cell}} |u_{nk}(\mathbf{r}_i)|^2 d\mathbf{r}_i \\ &= |B|^2 \sum_{i=1}^N \int_{\text{cell}} |u_{nk}(\mathbf{r}_1 + \mathbf{R}_i - \mathbf{R}_1)|^2 d\mathbf{r}_1 \\ &= |B|^2 \sum_{i=1}^N \int_{\text{cell}} |u_{nk}(\mathbf{r}_1)|^2 d\mathbf{r}_1 = 1 \end{aligned} \quad (1.39)$$

We have inserted the second equation of Eqs. (1.36) to obtain the fourth equality. Using the third equation of Eqs. (1.34), i.e.,  $u_{nk}(\mathbf{r})$  is normalized in the unit cell, we readily obtain  $B = 1/\sqrt{N}$ , and thus the Bloch theorem of Eqs. (1.34).

There are very important consequences of the Bloch theorem about the properties of electrons in solids. One of them is the acceleration theorem, which will be studied in Sect. 2.4. When applying an external force  $\mathbf{F}$ , e.g., due to an external electromagnetic field  $(\mathbf{E}, \mathbf{B})$  on the electrons in the solid,

$$\hbar \dot{\mathbf{k}} = \mathbf{F} = -e(\mathbf{E} + \mathbf{v}_{nk} \times \mathbf{B}), \quad \mathbf{v}_{nk} = \frac{1}{\hbar} \frac{\partial E_{nk}}{\partial \mathbf{k}} \quad (1.40)$$

Here  $-e$  is the electron charge and  $\mathbf{v}_{nk}$  is the electron group velocity.  $\hbar \mathbf{k}$  is therefore commonly referred to as the quasi-momentum of the electron in the crystal.

Moreover, as a consequences of time reversal symmetry, for a crystal of

$$H\Psi_{nk}(\mathbf{r}) = E_n(\mathbf{k})\Psi_{nk}(\mathbf{r})$$

the following relationship exists

$$E_n(\mathbf{k}) = E_n(-\mathbf{k}) \quad (1.41)$$

regardless of the spatial symmetry of the system, i.e., the energy of state with a wave vector  $\mathbf{k}$  is the same as  $-\mathbf{k}$ . This is known as Kramers' theorem [33].

## 1.4 $sp^3s^*$ Tight-Binding Model

In this and the following sections we introduce two most applied energy band structure models to calculate Eq. (1.32), namely, tight-binding model and  $\mathbf{k} \cdot \mathbf{p}$  model.

As the atoms of the elements making up the semiconductors are brought together to form the crystal, the valence electronic states are perturbed by the presence of neighboring atoms. While the original atomic functions describing the valence electrons are, of course, no longer eigenstates of the problem, they can be used as a

good approximate set of basis states to describe the “crystalline” electrons. This motivates the tight-binding method. For most semiconductor materials of interest, the atomic functions  $|\alpha, j\rangle$  (centered at atom  $j$ ) required to describe the outermost valence electrons are the  $s$ ,  $p_x$ ,  $p_y$ , and  $p_z$  types, see Table 1.2. Moreover, since there are more than one atom per unit cell, the Bloch function is in the form of

$$\Psi_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{R}_i} \sum_{\alpha} \sum_j C_{\alpha,j}(\mathbf{k}) |\alpha, j, \mathbf{r} - \mathbf{r}_j - \mathbf{R}_i\rangle e^{i\mathbf{k} \cdot \mathbf{R}_i} \quad (1.42)$$

where the sum over  $\mathbf{R}_i$  runs over all unit cells,  $\alpha$  is the index of the different atomic functions  $|\alpha, j\rangle$  used in the basis, and  $\mathbf{r}_j$  denotes the spatial position of atom  $j$  in unit cell  $\mathbf{R}_i$ .

Once the expansion set for the crystal states has been chosen, the coefficients  $C_{\alpha,j}$  remain to be determined. To this end, the Schrödinger equation is in the form of a secular determinant

$$|\langle \alpha', j', \mathbf{r} - \mathbf{r}_{j'} - \mathbf{R}_{i'} | H - E | \Psi_{\mathbf{k}}(\mathbf{r}) \rangle| = 0 \quad (1.43)$$

where  $H$  is the Hamiltonian of the system under investigation.

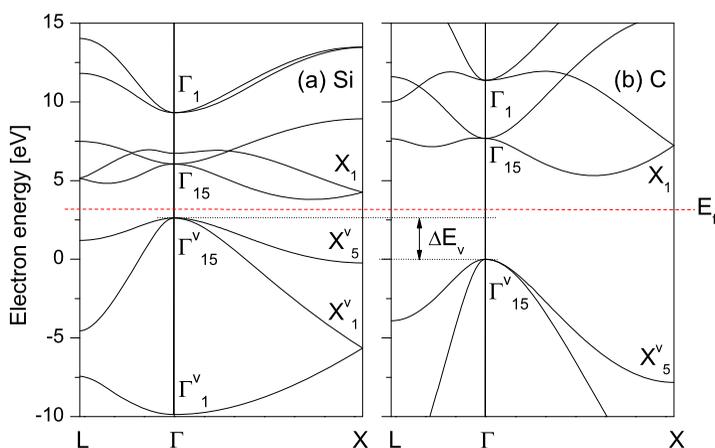
In theory, one can calculate the matrix elements in the secular determinant, Eq. (1.43), by first determining the crystal potential. This however is very difficult because of the complexity of the problem. Slater and Koster were the first to advocate the use of the tight-binding method as an empirical technique. In their formalism, the matrix elements of the secular determinant are treated as disposable constants. Energy levels in the band structure can be obtained and fitted with the measurement data by adjusting the disposable constants.

In 1983 Vogl, Hjalmarson and Dow published their results of a  $sp^3s^*$  nearest-neighbor semi-empirical tight-binding theory of energy bands in zincblende and diamond structure materials [34]. The theory was developed from the  $sp^3$  tight-binding model of Harrison [35]. Here we introduce five Löwdin orbitals,  $|s\rangle$ ,  $|p_x\rangle$ ,  $|p_y\rangle$ ,  $|p_z\rangle$ , and  $|s^*\rangle$ , at each atomic site  $\mathbf{R}_i$ . The Hamiltonian matrix element is denoted as  $h(\alpha\beta, ij)$  between the  $\alpha$ th orbital on the  $i$ th atomic site  $|\alpha, i\rangle$  and the  $\beta$ th orbital on  $j$ th atomic site  $|\beta, j\rangle$ , where either  $i = j$  or  $i$  is a nearest neighbor of  $j$ . The values of these matrix elements are listed in Table 1.3 for crystal Si, C, Ge, AlAs, InAs, and GaAs [34]. In Table 1.3 the diagonal elements are denoted as  $E$  (orbital energies), and the off-diagonal elements are  $V$  (interaction energies). For both orbital and interaction energies,  $s$ ,  $p$  and  $s^*$  denote  $s$ ,  $p$  and  $s^*$  orbitals,  $a$  denotes atomic site anion, and  $c$  the cation.  $V(x, x)$  represents  $V(p_x a, p_x c)$ ,  $V(p_y a, p_y c)$  and  $V(p_z a, p_z c)$ .  $V(x, y)$  represents  $V(p_x a, p_y c)$ , while other parameters can be derived from the ones listed in Table 1.3 after proper considerations of orbital symmetries.

Figure 1.12 shows the energy band structures of bulk silicon and carbon calculated by the  $sp^3s^*$  tight-binding model. As schematically shown in Fig. 1.4, because of the large number of atoms in the bulk material, energy levels form bands. Mathematically, the second equation in Eq. (1.34) actually implies  $N \rightarrow \infty$ . In reality,  $N$  is always finite in which case the second equation in Eq. (1.34) assumes that the effects

**Table 1.3** Energy band structure parameters [eV] for  $sp^3s^*$  tight-binding band calculation [34]

	Si	C	Ge	AlAs	InAs	GaAs
$E(s, a)$	-4.2000	-4.5450	-5.8800	-7.5273	-9.5381	-8.3431
$E(p, a)$	1.7150	3.8400	1.6100	0.9833	0.9099	1.0414
$E(s^*, a)$	6.6850	11.3700	6.3900	7.4833	7.4099	8.5914
$E(s, c)$	-4.2000	-4.5450	-5.8800	-1.1627	-2.7219	-2.6569
$E(p, c)$	1.7150	3.8400	1.6100	3.5867	3.7201	3.6686
$E(s^*, c)$	6.6850	11.3700	6.3900	6.7267	6.7401	6.7386
$V(s, s)$	-8.3000	-22.7250	-6.7800	-6.6642	-5.6052	-6.4513
$V(x, x)$	1.7150	3.8400	1.6100	1.8780	1.8398	1.9546
$V(x, y)$	4.5750	11.6700	4.900	4.2919	4.4693	5.0779
$V(sa, pc)$	5.7292	15.2206	5.4649	5.1106	3.0354	4.4800
$V(sc, pa)$	5.7292	15.2206	5.4649	5.4965	5.4389	5.7839
$V(s^*a, pc)$	5.3749	8.2109	5.2191	4.5216	3.3744	4.8422
$V(pa, s^*c)$	5.3749	8.2109	5.2191	4.9950	3.9097	4.8077
$\Delta E_v$	2.63	0.0		0.4	0.1	0.0

**Fig. 1.12** Energy band structures of diamond-structure (a) silicon and (b) carbon calculated by the  $sp^3s^*$  tight-binding model. The horizontal dashed line marked as  $E_f$  is referred to as the Fermi level for pure material at zero temperature

of the real boundaries of the solids with a finite number of unit cells on the electron states under investigation in the solid is negligibly small since  $N$  is very large.

Knowing the energy band structure, we then fill the energy levels by the available valence electrons in the solid. At zero temperature and for pure semiconductor, all energy levels below the horizontal dashed line  $E_f$  in Fig. 1.12 are fully occupied, Energy levels above  $E_f$  are completely empty.  $E_f$  is known as the Fermi level. The

highest occupied energy band is the valence band, and the lowest unoccupied band is the conduction band. For almost all semiconductor materials of interest, the top of the valence band locates at  $\mathbf{k} = 0$  ( $\Gamma$  point). The lowest conduction band states in Si and C locate somewhere between  $\Gamma$  and X points, see Fig. 1.12. Because of the different  $\mathbf{k}$  values of the valence and conduction band optimal points (the materials are referred as indirect-band materials), the optical properties of Si and C are rather bad. However, the lowest conduction band states of III–V materials locate at the  $\Gamma$  points so that the optical properties of III–V materials are extremely good. We will discuss the optical properties of semiconductor materials late.

Now we apply the  $sp^3s^*$  tight-binding theory to study  $\text{Si}_{1-x}\text{C}_x$  alloy as a means to further elaborate the tight-binding theory. The realization of many kinds of electronic and optoelectronic devices in strained layer  $\text{Si}_{1-x}\text{Ge}_x/\text{Si}$  heterostructures has stimulated a great interest in investigating IV–IV binary and ternary alloys [36, 37]. However, the strained epitaxial  $\text{Si}_{1-x}\text{Ge}_x$  layers without misfit dislocations can be grown on a Si substrate only by a low-temperature growth technique [36]. Moreover, the application of  $\text{Si}_{1-x}\text{Ge}_x/\text{Si}$  materials is restricted by the strain in the epitaxial layers. To compensate the strain, C atoms with an atomic diameter smaller than the ones of both Si and Ge atoms are introduced into the Si-Ge system to form  $\text{Si}_{1-x-y}\text{Ge}_x\text{C}_y$  alloys. The substitutional C atoms in  $\text{Si}_{1-y}\text{C}_y$  and related alloys also offer an additional parameter for tailoring the energy band structure [37]. The investigation on  $\text{Si}_{1-y}\text{C}_y$  and related alloys is thus of great importance to understand the bandgap engineering for Si-based semiconductor materials.

In Ref. [34], the top of the valence band,  $\Gamma_{15}^v = 0$  is referred as the reference energy for every individual material. Referring to the energy band of C, an energy band offset between C and Si,  $\Delta E_v(\text{Si}) = \Gamma_{15}^v(\text{Si}) - \Gamma_{15}^v(\text{C})$  is to be added to the Si orbital energies, see Fig. 1.12. From available values of the electron affinity (the electron affinity of silicon is 4.05 eV, whereas for diamond it is much substrate-orientation-dependent [38], a value of 2.2 eV is obtained for (001)-orientation) it is easy to obtain the absolute positions of valence bandedges ( $\Gamma_{15}^v$ ) below the vacuum level: 5.17 eV for Si and 7.8 eV for C. Thus,  $\Delta E_v(\text{Si}) = 2.63$  eV.

For  $\text{Si}_{1-y}\text{C}_y$  alloy, the interaction elements are obtained by the well-known  $d^{-2}$  scaling rule [35], where  $d$  is the spatial distance between the interacting atoms.

The eigen function of the system is expressed by

$$\sum_{\alpha, i} C(\alpha, i) |\alpha, i\rangle$$

The coefficients in the above linear combination satisfy the eigenvalue equation

$$\sum_{\beta, j} h(\alpha\beta, ij) C(\beta, j) = EC(\alpha, i) \quad (1.44)$$

For crystal system with translational system (i.e., unit cells are periodically positioned in space), we apply the Fourier transformation

$$C(\alpha, \mathbf{k}) = \frac{1}{\sqrt{N}} \sum_i C(\alpha, i) e^{i\mathbf{k}\cdot\mathbf{r}_i}$$

$$H(\alpha\beta, \mathbf{k}\mathbf{q}) = \frac{1}{N} \sum_{ij} h(\alpha\beta, ij) e^{i(\mathbf{k}\cdot\mathbf{r}_i - \mathbf{q}\cdot\mathbf{r}_j)} \quad (1.45)$$

so that Eq. (1.44) reduces to

$$\sum_{\beta, \mathbf{q}} H(\alpha\beta, \mathbf{k}\mathbf{q}) C(\beta, \mathbf{q}) = E(\mathbf{k}) C(\alpha, \mathbf{k}) \quad (1.46)$$

Here  $N$  is the number of unit cells in the system.

Because of the translational symmetry,  $h(\alpha\beta, ij) = h(\alpha\beta, \mathbf{r}_i - \mathbf{r}_j)$ ,

$$H(\alpha\beta, \mathbf{k}\mathbf{q}) = \delta_{\mathbf{k}, \mathbf{q}} H(\alpha\beta, \mathbf{k}) = \delta_{\mathbf{k}, \mathbf{q}} \sum_{ij} h(\alpha\beta, ij) e^{i\mathbf{k}\cdot(\mathbf{r}_i - \mathbf{r}_j)} \quad (1.47)$$

$$\sum_{\beta} H(\alpha\beta, \mathbf{k}) C(\beta, \mathbf{k}) = E(\mathbf{k}) C(\alpha, \mathbf{k})$$

In a  $\text{Si}_{1-y}\text{C}_y$  alloy,  $h(\alpha\beta, ij)$  is not invariant with respect to lattice translations, Eqs. (1.47) are thus not valid. However, let us approximate  $h(\alpha\beta, ij)$  by its mean value in the limit of long wavelength (small  $|\mathbf{k}|$  and  $|\mathbf{q}|$ ). In this case, the term  $e^{i(\mathbf{k}-\mathbf{q})\cdot\mathbf{r}_j}$  in Eq. (1.45)

$$e^{i(\mathbf{k}\cdot\mathbf{r}_i - \mathbf{q}\cdot\mathbf{r}_j)} = e^{i\mathbf{k}\cdot(\mathbf{r}_i - \mathbf{r}_j)} e^{i(\mathbf{k}-\mathbf{q})\cdot\mathbf{r}_j} \quad (1.48)$$

can be approximated as constant over a large area of the structure (which is normally known as the effective medium approximation). Let

$$h'(\alpha\beta, ij) = \frac{\sum_{mn} h(\alpha\beta, mn) \delta_{\mathbf{r}_m - \mathbf{r}_n, \mathbf{r}_i - \mathbf{r}_j}}{\sum_{mn} \delta_{\mathbf{r}_m - \mathbf{r}_n, \mathbf{r}_i - \mathbf{r}_j}} \quad (1.49)$$

which is invariant under lattice translations, we then have

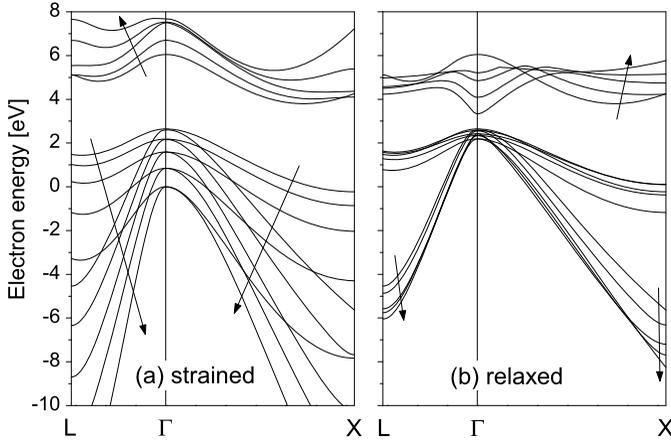
$$H'(\alpha\beta, \mathbf{k}\mathbf{q}) \approx \frac{1}{N} \sum_{ij} h'(\alpha\beta, ij) e^{i(\mathbf{k}\cdot\mathbf{r}_i - \mathbf{q}\cdot\mathbf{r}_j)} \quad (1.50)$$

and

$$H'(\alpha\beta, \mathbf{k}\mathbf{q}) = \delta_{\mathbf{k}, \mathbf{q}} H'(\alpha\beta, \mathbf{k}) = \delta_{\mathbf{k}, \mathbf{q}} \sum_{ij} h'(\alpha\beta, ij) e^{i\mathbf{k}\cdot(\mathbf{r}_i - \mathbf{r}_j)} \quad (1.51)$$

$$\sum_{\beta} H'(\alpha\beta, \mathbf{k}) C(\beta, \mathbf{k}) = E(\mathbf{k}) C(\alpha, \mathbf{k})$$

The above equations are mathematically identical to Eqs. (1.47). It must be reminded that the above solutions are valid only when  $|\mathbf{k}|$  and  $|\mathbf{q}|$  are small. The



**Fig. 1.13** The energy dispersion relations of relaxed and strained  $\text{Si}_{1-y}\text{C}_y$  alloys.  $y = 0, 0.25, 0.5, 0.75, 1.0$  (Y. Fu, M. Willander, P. Han, T. Matsuura, and J. Murota, Si-C atomic bond and electronic band structure of a cubic  $\text{Si}_{1-y}\text{C}_y$  alloy, Phys. Rev. B, vol. 58, pp. 7717–7722, 1998)

approximation is generally acceptable for the valence band top at  $\Gamma_{15}^v$ . For the conduction band of an indirect bandgap material like Si and C, a similar but modified scheme can be performed for conduction band states close to the bandedge. Let  $\mathbf{k}_0$  be the wave vector of the conduction bandedge state, the interaction in Eq. (1.45) between two conduction band states close to  $\mathbf{k}_0$ , i.e.,  $\mathbf{k} + \mathbf{k}_0$  and  $\mathbf{q} + \mathbf{k}_0$ , where  $|\mathbf{k}|$  and  $|\mathbf{q}|$  are small, can be expressed as

$$H(\alpha\beta, \mathbf{k} + \mathbf{k}_0, \mathbf{q} + \mathbf{k}_0) = \frac{1}{N} \sum_{ij} h(\alpha\beta, ij) e^{i\mathbf{k}_0 \cdot (\mathbf{r}_i - \mathbf{r}_j)} e^{i(\mathbf{k} \cdot \mathbf{r}_i - \mathbf{q} \cdot \mathbf{r}_j)} \quad (1.52)$$

Writing  $h(\alpha\beta, ij) e^{i\mathbf{k}_0 \cdot (\mathbf{r}_i - \mathbf{r}_j)}$  as the new  $h(\alpha\beta, ij)$ , we are then able to use the above numerical approximation.

Figure 1.13 shows the energy dispersion relations of relaxed and strained  $\text{Si}_{1-y}\text{C}_y$  alloys as functions of the C mole fraction  $y$ , where arrows indicate the increase of the C mole fraction from 0 to 1.0. Note that the complete dispersion relations are not monotonous functions of the C mole fraction, only parts with arrows in Fig. 1.13 are. Here we consider two cases. When the  $\text{Si}_{1-y}\text{C}_y$  alloy is grown on a Si substrate, it is strained when the layer is thin. The atomic bond length is uniform and fixed by the substrate Si; The atomic bonds become relaxed when the layer becomes thicker. In the effective medium approximation, the lattice constant of the relaxed  $\text{Si}_{1-y}\text{C}_y$  alloy is obtained by linear interpolation between bond lengths of C and SiC when  $y > 0.5$ , it is obtained from SiC and Si when  $y < 0.5$ .

For strained  $\text{Si}_{1-y}\text{C}_y$  alloy, both the valence and conduction bands are vertically shifted along the energy axis following the increase of the C mole fraction. The bandgap is indirect and increases monotonically with the C mole fraction. On the other hand, the valence band of a relaxed  $\text{Si}_{1-y}\text{C}_y$  alloy is not much affected by

the C mole fraction. The bandgap of relaxed  $\text{Si}_{1-y}\text{C}_y$  alloy is indirect and increases with increasing  $y$  when  $y < 0.35$ . However it becomes a direct-bandgap material when  $y > 0.35$ . The energy bandgap increases with increasing  $y$  from 0 to 0.35, then decreases when increasing  $y$  from 0.35 to 0.5. After that, the energy bandgap increases again with  $y$ .

Let us check the validity of the conclusion that the energy bandgap increases with increasing but small  $y$ . It is easy to see that Eq. (1.49) can be rewritten as

$$h' = (1 - y)h_{\text{Si}} + yh_{\text{SiC}} \quad (1.53)$$

for small  $y$  value, where  $h_{\text{Si}}$  and  $h_{\text{SiC}}$  are interaction elements in crystal Si and SiC, respectively. Since the energy bandgap is proportional to the interaction elements, the above equation indicate an increasing energy bandgap of  $\text{Si}_{1-y}\text{C}_y$  as a function of the C content (the bandgaps of C and SiC are wider than the Si one). It is thus observed that the conclusion of increasing bandgap with  $y$  is very general, even we are working with the relatively simple  $sp^3s^*$  tight binding model.

## 1.5 Bandedge States

Our goal is essentially to solve the Schrödinger equation to get the eigenenergies and their associated eigenvectors, i.e., wave functions. The wave functions usually have too high frequencies to be feasible to calculate explicitly using numerical methods on computers as an inordinately high number of grid points would be necessary to capture an acceptable numerical representation of the wave functions. A solution is to separate the wave functions into an oscillatory part at unit cell scale and a modulating part which is of the same scale as the solid. This is the basic idea of the envelope function approximation—the modulating part is called the envelope function. We write the envelope function as a Bloch function:

$$\Psi_{nk}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{nk}(\mathbf{r}) \quad (1.54)$$

where  $n$  is the state index and  $\mathbf{k}$  is the wave vector, see Eq. (1.34). The Schrödinger equation for this wave function is simply

$$\left[ \frac{\mathbf{p}^2}{2m_0} + V(\mathbf{r}) \right] \Psi_{nk}(\mathbf{r}) = E_n(\mathbf{k}) \Psi_{nk}(\mathbf{r}) \quad (1.55)$$

Substituting the factorization into the equation requires some care in evaluating the product of the momentum operator  $\mathbf{p}$  and the two parts of the wave function. With  $\mathbf{p} = -i\hbar\nabla$  and knowing from vector calculus that  $\nabla e^{i\mathbf{k}\cdot\mathbf{r}} = i\mathbf{k}e^{i\mathbf{k}\cdot\mathbf{r}}$ , the product of the momentum operator and the wave function becomes

$$-i\hbar\nabla[e^{i\mathbf{k}\cdot\mathbf{r}} u_{nk}(\mathbf{r})] = -i\hbar e^{i\mathbf{k}\cdot\mathbf{r}} (\nabla + i\mathbf{k}) u_{nk}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} (\mathbf{p} + \hbar\mathbf{k}) u_{nk}(\mathbf{r}) \quad (1.56)$$

Applying the momentum operator a second time gives  $e^{i\mathbf{k}\cdot\mathbf{r}}(\mathbf{p} + \hbar\mathbf{k})^2 u_{n\mathbf{k}}(\mathbf{r})$ , so the Schrödinger equation can be written in the following way that the oscillatory part will cancel out:

$$e^{i\mathbf{k}\cdot\mathbf{r}} \left[ \frac{(\mathbf{p} + \hbar\mathbf{k})^2}{2m_0} + V(\mathbf{r}) \right] u_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} E_n(\mathbf{k}) u_{n\mathbf{k}}(\mathbf{r}) \quad (1.57)$$

Expanding the  $(\mathbf{p} + \hbar\mathbf{k})^2$  term gives

$$\left[ \frac{\mathbf{p}^2}{2m_0} + V(\mathbf{r}) + \frac{\hbar\mathbf{k} \cdot \mathbf{p}}{m_0} + \frac{\hbar^2 k^2}{2m_0} \right] u_{n\mathbf{k}}(\mathbf{r}) = E_n(\mathbf{k}) u_{n\mathbf{k}}(\mathbf{r}) \quad (1.58)$$

The first two terms are identical to the original Hamiltonian, so if the two other terms are treated as two small perturbations the Hamiltonian can be expressed as

$$(H_0 + H_1 + H_2) u_{n\mathbf{k}}(\mathbf{r}) = E_n(\mathbf{k}) u_{n\mathbf{k}}(\mathbf{r}) \quad (1.59)$$

where

$$H_1 = \frac{\hbar}{m_0} \mathbf{k} \cdot \mathbf{p}, \quad H_2 = \frac{\hbar^2 k^2}{2m_0}$$

are the first-order and second-order perturbations, respectively.

If the equation is solved for  $\mathbf{k} = 0$  with only  $H_0$  remaining nonzero, the result is a set of eigenvectors  $u_{n0}(\mathbf{r})$ , typically at the optimal points such as the  $\Gamma$  point of the valence band structures of Si and C bulk materials in Fig. 1.12.

An expression of the Hamiltonian similar to Eq. (1.59) can be formulated for the electron states in the vicinity of the conduction bandedges of Si and C which do not locate at  $\Gamma$  but with a finite  $\mathbf{k}_0$ . For this, we express the wave vector of the modulation envelope function in terms of  $\mathbf{k}_0$ , i.e.,  $\mathbf{k} + \mathbf{k}_0$  so that  $|\mathbf{k}|$  is small. Equation (1.58) becomes

$$\left[ \frac{(\mathbf{p} + \hbar\mathbf{k}_0)^2}{2m_0} + V(\mathbf{r}) + \frac{\hbar\mathbf{k} \cdot (\mathbf{p} + \hbar\mathbf{k}_0)}{m_0} + \frac{\hbar^2 k^2}{2m_0} \right] u_{n,\mathbf{k}+\mathbf{k}_0}(\mathbf{r}) = E_n(\mathbf{k} + \mathbf{k}_0) u_{n,\mathbf{k}+\mathbf{k}_0}(\mathbf{r}) \quad (1.60)$$

For the following discussion we express the eigenfunction using Dirac notation as  $|m\rangle$ , and eigenvalues  $E_m$ . The eigenfunction for a given  $\mathbf{k}$ ,  $|n\mathbf{k}\rangle$ , is a linear combination of the basis functions:

$$u_{n\mathbf{k}}(\mathbf{r}) = \sum_m c_{nm}(\mathbf{k}) |m\rangle \quad (1.61)$$

so the objective is now to find the coefficients  $c_{nm}(\mathbf{k})$  that form our envelope parts together with the basis functions at  $\Gamma$ .

If we insert this linear combination into the Schrödinger equation, we get

$$H \sum_m c_{nm}(\mathbf{k}) |m\rangle = \sum_m H |m\rangle c_{nm}(\mathbf{k}) = E_n(\mathbf{k}) \sum_m c_{nm}(\mathbf{k}) |m\rangle \quad (1.62)$$

Multiply this on the left with the conjugate of any, say  $|\ell\rangle$ , of the basis functions, we obtain

$$\sum_m \langle \ell | H | m \rangle c_{nm}(\mathbf{k}) = E_n(\mathbf{k}) \sum_m c_{nm}(\mathbf{k}) \langle \ell | m \rangle = E_n(\mathbf{k}) c_{n\ell}(\mathbf{k}) \quad (1.63)$$

The right-hand part is the result of wave function's orthonormal property  $\langle \ell | m \rangle = \delta_{\ell m}$ .

Inserting the expanded Hamiltonian equation (1.59), multiplying with the conjugate and integrating over an unit cell, give

$$E_n(\mathbf{k}) c_{n\ell}(\mathbf{k}) = \left( E_\ell + \frac{\hbar^2 k^2}{2m_0} \right) c_{n\ell}(\mathbf{k}) + \sum_m \frac{\hbar}{m_0} \langle \ell | \mathbf{k} \cdot \mathbf{p} | m \rangle c_{nm}(\mathbf{k}) \quad (1.64)$$

We first set  $c_{n\ell}(\mathbf{k}) = \delta_{n\ell}$  on the right side of the above Eq. (1.64), thus to neglect the wave function corrections from which we obtain the first-order correction to the energy of state  $|n\mathbf{k}\rangle$

$$E_n(\mathbf{k}) c_{nn}(\mathbf{k}) = \left( E_n + \frac{\hbar^2 k^2}{2m_0} + \frac{\hbar}{m_0} \langle n | \mathbf{k} \cdot \mathbf{p} | n \rangle \right) c_{nn}(\mathbf{k}) \quad (1.65)$$

which is

$$E_n(\mathbf{k}) = E_n + \frac{\hbar^2 k^2}{2m_0}, c_{nn}(\mathbf{k}) = 1 \quad (1.66)$$

since  $\langle n | \mathbf{p} | n \rangle = 0$ . This is normally referred to as the first-order approximation.

Similarly, for  $\ell \neq n$ , Eq. (1.64)

$$E_n(\mathbf{k}) c_{n\ell}(\mathbf{k}) = \left( E_\ell + \frac{\hbar^2 k^2}{2m_0} \right) c_{n\ell}(\mathbf{k}) + \frac{\hbar}{m_0} \langle \ell | \mathbf{k} \cdot \mathbf{p} | n \rangle \quad (1.67)$$

so that the first-order correction to the wave function is

$$c_{n\ell}(\mathbf{k}) = \frac{\hbar}{m_0} \frac{\langle \ell | \mathbf{k} \cdot \mathbf{p} | n \rangle}{E_n - E_\ell} \quad (1.68)$$

by using Eq. (1.66). Note that in Eq. (1.67) we have implicitly assumed that both  $\hbar \langle \ell | \mathbf{k} \cdot \mathbf{p} | m \rangle / m_0$  and  $c_{nm}$  ( $m \neq n$ ) are small so that their products on the right side are neglected, so that only the product of  $\hbar \langle \ell | \mathbf{k} \cdot \mathbf{p} | n \rangle / m_0$  and  $c_{nn}$  ( $c_{nn} = 1$ ) remains.

By inserting the above expression back into Eq. (1.64), we have obtained the energy of state  $u_{n\mathbf{k}}(\mathbf{r})$  at the second-order approximation

$$E_n(\mathbf{k}) = E_n + \frac{\hbar^2 k^2}{2m_0} + \frac{\hbar}{m_0} \sum_{\ell \neq n} \frac{|\langle \ell | \mathbf{k} \cdot \mathbf{p} | n \rangle|^2}{E_n - E_\ell} \quad (1.69)$$

The result can be expressed in terms of an effective mass  $m^*$ :

$$E_n(\mathbf{k}) = E_n + \sum_{i,j} \frac{\hbar^2}{2m_{ij}^*} k_i k_j \quad (1.70)$$

where  $i, j = x, y, z$ , and

$$\frac{m_0}{m_{ij}^*} = \delta_{ij} + \frac{2}{m_0} \sum_{\ell \neq n} \frac{\langle n | p_i | \ell \rangle \langle \ell | p_j | n \rangle}{E_n - E_\ell} \quad (1.71)$$

Note here that the effective mass can be anisotropic which can be utilized for optical coupling in quantum well photodetection [39].

It can be easily seen that a narrow bandgap, which leads to two states being close to each other, i.e.,  $E_n - E_\ell$  is small in the above equation, gives a small effective mass. This agrees very well with the experimental data which indicates that InSb has both the smallest bandgap and the lowest effective mass.

For semiconductors of device application interest, we concentrate on the conduction and valence bands of cubic semiconductors with both diamond (silicon and germanium) and zincblende symmetries (III–V group).

Refer to Fig. 1.12, the conduction band generally consists of three sets of band minima located at the  $\Gamma_{15}$ -point at  $\mathbf{k} = 0$ , the L-points at  $\mathbf{k} = (\pi/a, \pi/a, \pi/a)$ , and along the  $\Delta$  lines from  $(0, 0, 0)$  to  $(\pi/a, 0, 0)$ , from  $(0, 0, 0)$  to  $(0, \pi/a, 0)$ , and from  $(0, 0, 0)$  to  $(0, 0, \pi/a)$ , where  $a$  is the lattice constant. The valence band tops are located at  $\Gamma_{15}^v$ . Two bands are normally degenerate at this point, which are heavy-hole and light-hole bands; the third one is the spin-split-off band due to the spin-orbital interaction.

In close proximity to an energy minimum at  $\mathbf{k}_0$  in the conduction band, the energy dispersion relationship  $E_{\mathbf{k}}$  can be expressed as

$$\begin{aligned} E(\mathbf{k}) = E(\mathbf{k}_0) + \sum_i \left. \frac{\partial E(\mathbf{k})}{\partial k_i} \right|_{\mathbf{k}=\mathbf{k}_0} (k_i - k_{0,i}) \\ + \sum_{ij} \left. \frac{\partial^2 E(\mathbf{k})}{\partial k_i \partial k_j} \right|_{\mathbf{k}=\mathbf{k}_0} (k_i - k_{0,i})(k_j - k_{0,j}) + \dots \end{aligned} \quad (1.72)$$

where  $i, j = x, y, z$ . The linear terms vanish because of the spatial invariance under translation of  $\mathbf{k} \rightarrow -\mathbf{k}$ . In the region around  $\mathbf{k}_0$  where the higher orders can be neglected, the energy dispersion  $E(\mathbf{k})$  is approximated by a quadratic function of  $\mathbf{k}$ :

$$\begin{aligned} E(\mathbf{k}) = E(\mathbf{k}_0) + \sum_{ij} \frac{\hbar^2}{2} \frac{1}{m_{ij}^*} (k_i - k_{0,i})(k_j - k_{0,j}) \\ \frac{1}{m_{ij}^*} = \frac{1}{\hbar^2} \left. \frac{\partial^2 E_{\mathbf{k}}}{\partial k_i \partial k_j} \right|_{\mathbf{k}=\mathbf{k}_0} \end{aligned} \quad (1.73)$$

Here  $1/m_{ij}^*$  is equivalent to the definition of Eq. (1.71).  $[1/m_{ij}^*]$  for  $i, j = (x, y, z)$  forms a so-called inverse effective-mass tensor  $\bar{w}$ , which will have a profound effect on the optical properties of semiconductor materials to be discussed in Sect. 5.2.

Two typical band structures are:

1. Spherical band:

$$E(\mathbf{k}) = E(\mathbf{k}_0) + \frac{\hbar^2(\mathbf{k} - \mathbf{k}_0)^2}{2m^*} \quad (1.74)$$

The conduction bands of III–V materials are well described by the above expression, see Table 1.4.

2. Ellipsoidal band:

$$E(\mathbf{k}) = E(\mathbf{k}_0) + \frac{\hbar^2}{2} \left[ \frac{(k_\ell - k_{0,\ell})^2}{m_\ell^*} + \frac{(\mathbf{k}_t - \mathbf{k}_{0,t})^2}{m_t^*} \right] \quad (1.75)$$

where  $k_\ell$  and  $\mathbf{k}_t$  are longitudinal and transverse components of wave vector  $\mathbf{k}$ ,  $m_\ell^*$  and  $m_t^*$  are longitudinal and transverse effective masses. The conduction band of Si consists of six ellipsoids described by a longitudinal effective mass  $m_\ell^* = 0.9163$  and a transverse effective mass  $m_t^* = 0.1905$  expressed in the unit of free electron mass  $m_0$  [40]. The six ellipsoids become divided into groups. (1)  $m_z^* = m_\ell^*$ ,  $m_x^* = m_y^* = m_t^*$ , degeneracy  $g = 2$ ; (2)  $m_z^* = m_t^*$ ,  $m_x^* = m_t^*$ ,  $m_y^* = m_\ell^*$ ,  $g = 2$ . (3)  $m_z^* = m_t^*$ ,  $m_x^* = m_\ell^*$ ,  $m_y^* = m_t^*$ ,  $g = 2$ . Here  $m_x^*$ ,  $m_y^*$  and  $m_z^*$  are the effective masses in the  $x$ ,  $y$  and  $z$  directions, respectively.

There are many different effective-mass concepts defined by various physical properties. When a wave is subjected to an external force  $\mathbf{F}$ , the acceleration is given by

$$\frac{dv_i}{dt} = \frac{d}{dt} \left( \frac{1}{\hbar} \frac{\partial E_{\mathbf{k}}}{\partial k_i} \right) = \sum_j \frac{1}{\hbar} \frac{\partial^2 E_{\mathbf{k}}}{\partial k_i \partial k_j} \dot{k}_j \quad (1.76)$$

where  $i, j = x, y, z$ . The above equation can be written as

$$\frac{dv_i}{dt} = \sum_j \frac{1}{m_{ij}^*} F_j \quad (1.77)$$

by Eq. (1.40). Here the effective mass  $m^*$  is defined as the *acceleration effective mass*. The inverse effective-mass tensor by Eq. (1.73) is the one evaluated at the band minimal points.

The *conductivity effective mass*,  $m_c^*$ , is defined as the ratio of the electron momentum to its group velocity

$$\frac{\hbar \mathbf{k}}{m_c^*} = \frac{1}{\hbar} \frac{\partial E_{\mathbf{k}}}{\partial \mathbf{k}} \quad (1.78)$$

For a parabolic band,  $m_c^* = m^*$ , and for an ellipsoidal band,

$$\frac{3}{m_c^*} = \frac{1}{m_\ell^*} + \frac{2}{m_t^*} \quad (1.79)$$

where  $m_\ell^*$  and  $m_t^*$  are longitudinal and transverse effective masses.

Each electronic state  $E(\mathbf{k})$ , characterized by  $\mathbf{k}$ , can be occupied by two electrons, one spin up and the other spin down. The occupation probability of state  $E(\mathbf{k})$  is given by Fermi distribution function  $f[E(\mathbf{k}), E_f]$ , where  $E_f$  is the Fermi energy. The electron density  $n$  is given by

$$n = \int f[E(\mathbf{k}), E_f] \frac{2d\mathbf{k}}{(2\pi)^3} \quad (1.80)$$

where the integration is restricted within the first Brillouin zone. See more discussions below in Sect. 1.9 about the density of states. For a simple band of Eq. (1.74),

$$n = \int_{E(\mathbf{k}_0)} f(E, E_f) N_3(E) dE \quad (1.81)$$

where  $N_3(E)$  is the three-dimensional density of states

$$N_3(E) = \frac{1}{2\pi^2} \left( \frac{2m^*}{\hbar^2} \right)^{3/2} \sqrt{E - E(\mathbf{k}_0)} \quad (1.82)$$

For a complicated band, Eq. (1.81) can still hold with a more elaborate expression for  $N_3(E)$ , where the *density-of-states effective mass*  $m_d^*$  is defined in place of  $m^*$ . Note that  $m_d^*$  can depend on  $E$ . For a parabolic band,  $m_d^* = m^*$ , and for an ellipsoidal band,

$$m_d^* = [m_\ell^* (m_t^*)^2]^{1/3} \quad (1.83)$$

For simple band of Eq. (1.74) with a density of state  $N_3(E)$  in Eq. (1.82), the electron density at temperature  $T$  is

$$n = \frac{1}{2\pi^2} \left( \frac{2m^*}{\hbar^2} \right)^{3/2} \int \frac{E^{1/2} dE}{1 + \exp[(E - E_f)/k_B T]} \quad (1.84)$$

where  $k_B$  is the Boltzmann constant, from which the *carrier-concentration effective mass*  $m_{cc}^*$  can be introduced for complicated energy bands so that the above equation remains intact

$$(m_{cc}^*)^{3/2} = \frac{1}{(k_B T)^{3/2} F_{1/2}(E_f/k_B T)} \int \frac{(m_d^*)^{3/2} E^{1/2} dE}{1 + \exp[(E - E_f)/k_B T]} \quad (1.85)$$

where

$$F_{1/2}(x) = \int \frac{y^{1/2} dy}{1 + \exp(x - y)} \quad (1.86)$$

is called Fermi integral of order 1/2.

For more complicated non-parabolic conduction band, a simple analytical way is usually applied

$$E(\mathbf{k})[1 + \alpha E(\mathbf{k})] = \gamma(\mathbf{k}) \quad (1.87)$$