

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)

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

1	Semiconductor Materials	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
2	Electron Transport	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
3	Optical Properties of Semiconductors	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
4	Electronic Quantum Devices	185
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	δ -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
5	Nanostructured Optoelectronics	271
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 δ 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- μm gate length metal-oxide-semiconductor field-effect transistor (MOSFET) was reported in 1974 by Dennard et al. [1]; 0.1- μm 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- μm 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 μm (1 Gb) in 2001, 0.13 μm (4 Gb) in 2004, 0.10 μm (16 Gb) in 2007, and 0.07 μm (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 ϵ_0 is the permittivity of free space. From Newton's second law we have

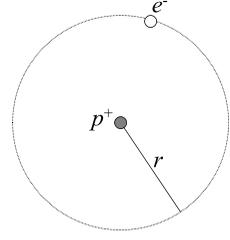
$$\frac{m_0 v^2}{r} = \frac{e^2}{4\pi\epsilon_0 r^2} \quad (1.1)$$

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

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

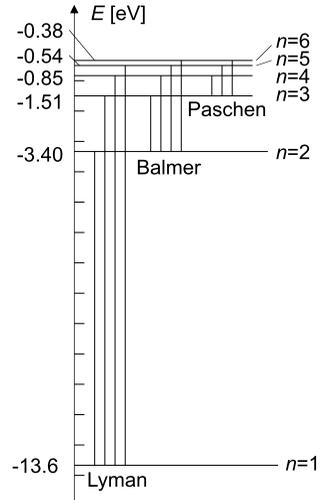
is the Rydberg constant.

Bohr’s theory may be applied to other single electron systems such as He^+ or 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	ℓ	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 ∞ , as shown by Eq. (1.7). The magnitude of the orbital angular momentum, L , of a state is determined by the orbit quantum number ℓ ,

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

where the maximum value of ℓ 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 orbit 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 ℓ 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

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

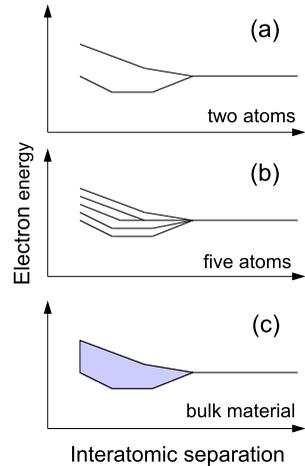
where the spin magnetic quantum number, $m_s = \pm 1/2$. The introduction of spin doubles the number of states allowed for each value of n .

Now, four quantum numbers, n , ℓ , m_ℓ , 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 , ℓ , m_ℓ , and m_s .

With the aid of the exclusion principle one can see how electrons fill shells (n) and subshells (ℓ). For each value of ℓ there are $(2\ell + 1)$ values of m_ℓ 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	$1s^2$	$2s^2 2p^3$
Ga	$1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10}$	$4s^2 4p^1$	P	$1s^2 2s^2 2p^6$	$3s^2 3p^3$
In	[Kr] $4d^{10}$	$5s^2 5p^1$	As	$1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10}$	$4s^2 4p^3$
			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	$1s^2$	$2s^2 2p^4$
Cd	[Kr] $4d^{10}$	$5s^2$	S	$1s^2 2s^2 2p^6$	$3s^2 3p^4$
Hg	[Xe] $4f^{14} 5d^{10}$	$6s^2$	Se	$1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10}$	$4s^2 4p^4$
			Te	[Kr] $4d^{10}$	$5s^2 5p^4$

atoms/m³: 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-

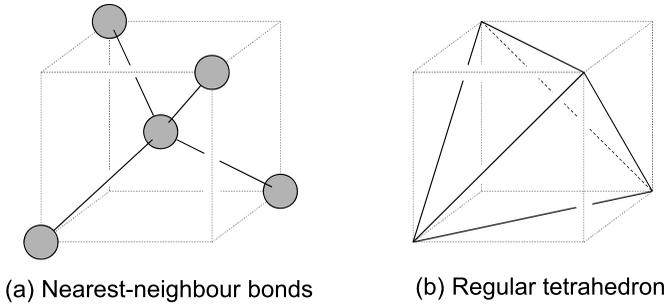


Fig. 1.5 (a) Si atom interacts with four neighboring Si atoms positioned at the vertices of (b) a regular tetrahedron by sharing its four valence electrons ($3s^23p^2$) with the four neighboring Si atoms

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.

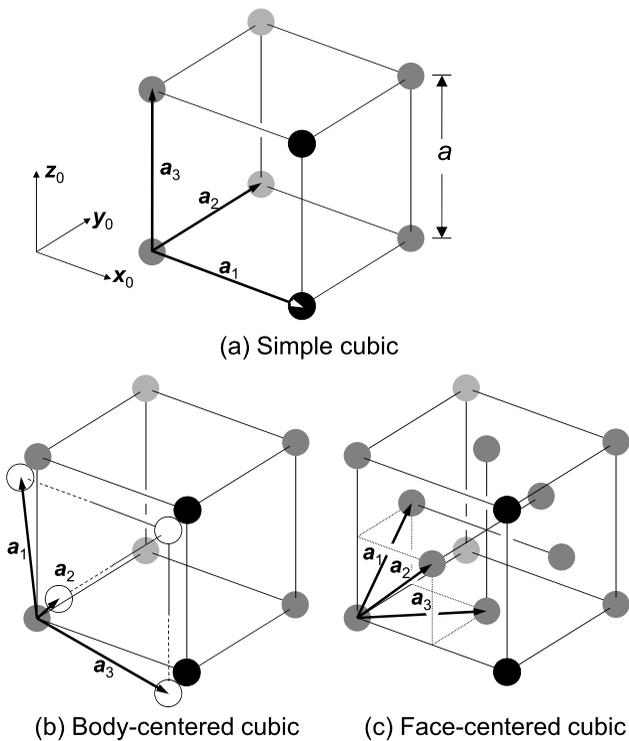


Fig. 1.6 (a) Simple cubic, (b) body-centered cubic, and (c) face-centered cubic lattices. a is the lattice constant

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

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

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

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

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

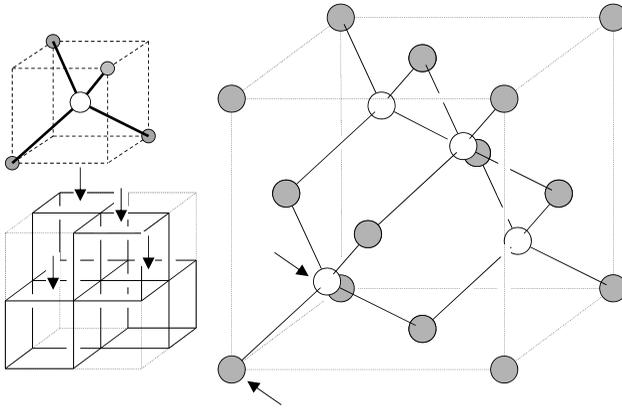


Fig. 1.7 Face-centered cubic lattice with two-atom basis forming either the so-called diamond (when the two atoms are identical) or the zincblende (when the two atoms in the basis are different) structure

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 λ 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

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

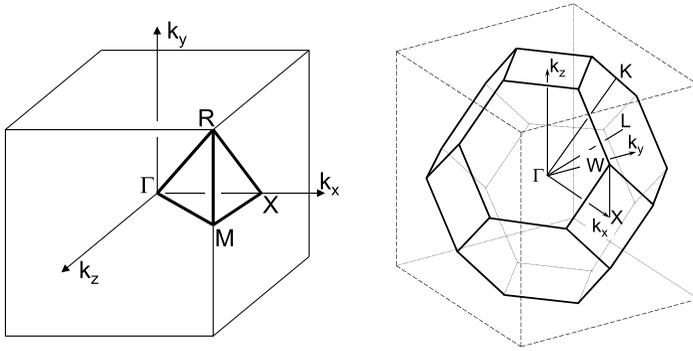


Fig. 1.8 First Brillouin zones for simple cubic lattice (*left*) and fcc (*right*) lattice. Points and lines of symmetry are indicated

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 Γ - Γ X-XM in the \mathbf{k} -space (in unit of $2\pi/a$)

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

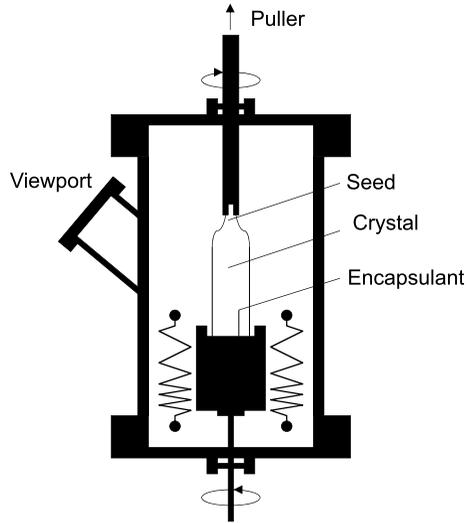
for the simple cubic lattice. For the fcc lattice, such as the diamond as well as zincblende structures (fcc lattices with bases), we mostly consider XU-UL-L Γ - Γ X-XW-WK in the \mathbf{k} -space (in unit of $2\pi/a$),

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

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

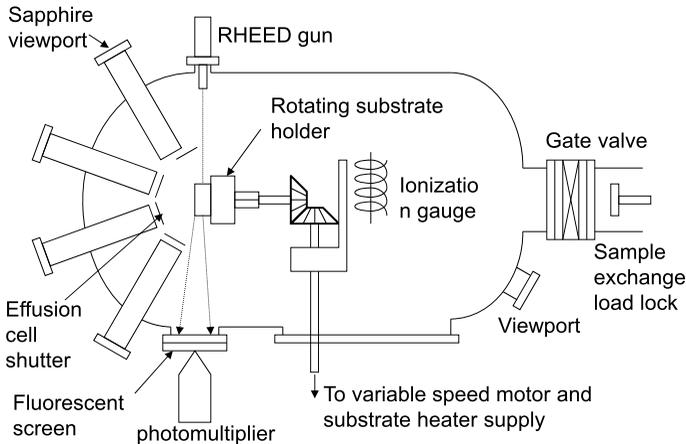


Fig. 1.10 Principle of the MBE growth system. In situ monitoring capabilities are often standard components of the system

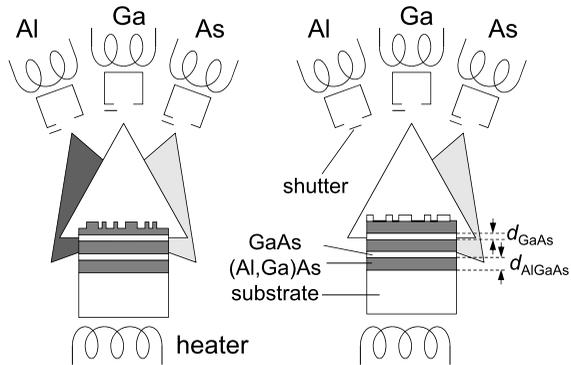
epitaxy and metal organic chemical vapor deposition, one can achieve monolayer (about 3 \AA) control in the growth direction. This level of control is essential for the variety of heterostructure devices that have been beginning to be used in electronics and optoelectronics. The epitaxial techniques 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 N_2 to ensure high vacuum and to prevent atoms/molecules to come off from the chamber walls.

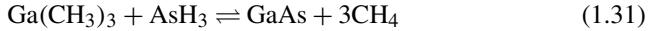
The low background pressure in MBE allows one to use electron beams to monitor the growing crystal. The reflection high-energy electron diffraction (RHEED) technique relies on electron diffraction to monitor both the quality of the growth substrate and the layer-by-layer growth mode. As each monolayer 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

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

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

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

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

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

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

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

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

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

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

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

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

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 Ω . 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, α 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.

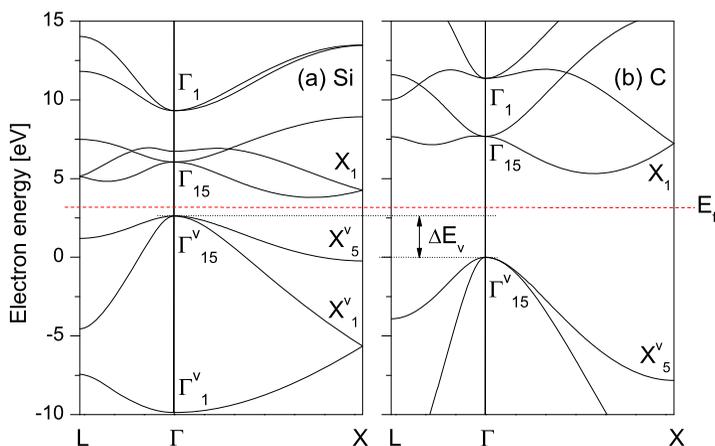
In theory, one can calculate the matrix elements in the secular determinant, Eq. (1.43), by first determining the crystal potential. This however is very difficult because of the complexity of the problem. Slater and Koster were the first to advocate the use of the tight-binding method as an empirical 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 α th orbital on the i th atomic site $|\alpha, i\rangle$ and the β 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
ΔE_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 $\mathbf{k} = 0$ (Γ point). The lowest conduction band states in Si and C locate somewhere between Γ 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 Γ 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 (Γ_{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

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

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

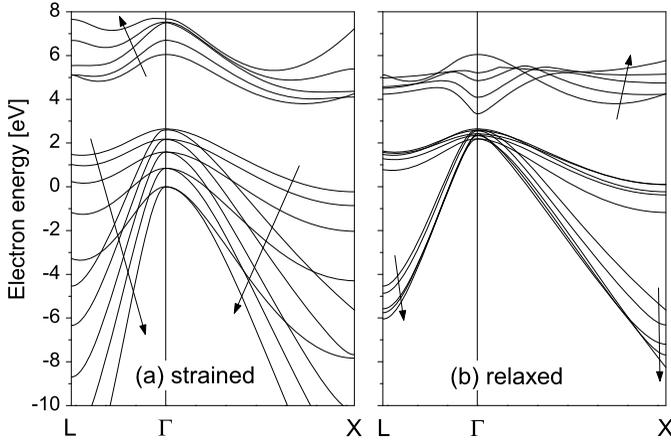


Fig. 1.13 The energy dispersion relations of relaxed and strained $\text{Si}_{1-y}\text{C}_y$ alloys. $y = 0, 0.25, 0.5, 0.75, 1.0$ (Y. Fu, M. Willander, P. Han, T. Matsuura, and J. Murota, Si-C atomic bond and electronic band structure of a cubic $\text{Si}_{1-y}\text{C}_y$ alloy, Phys. Rev. B, vol.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

approximation is generally acceptable for the valence band top at Γ_{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_{Si} and h_{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 Γ 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 Γ 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 Γ .

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 Γ_{15} -point at $\mathbf{k} = 0$, the L-points at $\mathbf{k} = (\pi/a, \pi/a, \pi/a)$, and along the Δ 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 Γ_{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_ℓ and \mathbf{k}_t are longitudinal and transverse components of wave vector \mathbf{k} , m_ℓ^* 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_ℓ^* 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

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

where $\gamma(\mathbf{k})$ is given by one of the right-hand sides of Eqs. (1.74) and (1.75). α is a nonparabolicity parameter, which in some cases can be used as a fitting parameter in the interpretation of the electron transport data. The group velocity associated with state \mathbf{k} and the conductivity effective mass are proven to be

$$\begin{aligned} \mathbf{v}(\mathbf{k}) &= \frac{1}{\hbar} \frac{\partial E(\mathbf{k})}{\partial \mathbf{k}} = \frac{\hbar \mathbf{k}}{m^* [1 + 2\alpha E(\mathbf{k})]} \\ m_c^* &= m^* [1 + \alpha E(\mathbf{k})] \end{aligned} \quad (1.88)$$

We can easily conclude from the above equations that the non-parabolicity effect is to reduce the group velocity and to increase the density of states (larger effective mass). Since the scattering rate is proportional to the density of states, the second effect enhances the scattering rate. In other words, it reduces the mobility of electrons at high temperature and high fields when high-energy states are occupied.

For valence bands, non-parabolicity cannot be parameterized in this simple form because of the correlation among the heavy hole, light hole and spin-split-off band. The following $\mathbf{k} \cdot \mathbf{p}$ model is thus essential when quantitative descriptions of the valence band states are required.

1.6 Eight-Band $\mathbf{k} \cdot \mathbf{p}$ Model

The well-established eight-band $\mathbf{k} \cdot \mathbf{p}$ model for III–V semiconductor materials was derived rigorously for bulk materials [41–45]. And by a series of experiments the parameters for the $\mathbf{k} \cdot \mathbf{p}$ Hamiltonian can be determined, and the final result is a matrix with many material-dependent parameters and no arbitrary parameters to adjust. The model has been extensively applied to many low-dimensional systems, e.g., see Ref. [46], and the review of Ref. [47], as it gives a good approximation of the states close to the Γ point ($\mathbf{k} = 0$), which is the most relevant region for many device applications. The approximation is given with good accuracy, while in many nanostructure calculations, this is the only feasible numerical model that can be implemented (even though it is still much restricted due to the computational capability).

For the $\mathbf{k} \cdot \mathbf{p}$ model, a common approach of choosing the basis functions $|n\rangle$ at the Γ point is to follow Kane [44] and define a set of eight states, $|m\rangle$, i.e., $u_{n\mathbf{k}}(\mathbf{r})$ in Eq. (1.54) at Γ ($\mathbf{k} = 0$),

$$\begin{aligned} |m\rangle &= (|S \uparrow\rangle, |X \uparrow\rangle, |Y \uparrow\rangle, |Z \uparrow\rangle, |S \downarrow\rangle, |X \downarrow\rangle, |Y \downarrow\rangle, |Z \downarrow\rangle) \\ H_0|m\rangle &= E_m|m\rangle \end{aligned} \quad (1.89)$$

where the arrows indicate spin up and down. $|X\rangle$, $|Y\rangle$ and $|Z\rangle$ denote degenerate valence bands while $|S\rangle$ denotes the conduction band. Other remote bands are not included unless specifically required.

By following the notation and phase conventions used by Enders et al. [45], the Hamiltonian of the eight-band $\mathbf{k} \cdot \mathbf{p}$ model is written explicitly as

$$H = H_0 + H_{\text{so}} + D + H_1 + H_2 \quad (1.90)$$

H_0 represents the energies of the basis states, H_{so} the spin-orbit interaction, D the deformation potential, and H_1 and H_2 the $\mathbf{k} \cdot \mathbf{p}$ interactions of first and second order, respectively:

$$H_0 = \begin{bmatrix} E_g & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\Delta'_{\text{so}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\Delta'_{\text{so}} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\Delta'_{\text{so}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & E_g & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\Delta'_{\text{so}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\Delta'_{\text{so}} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\Delta'_{\text{so}} \end{bmatrix} \quad (1.91)$$

E_g is the fundamental bandgap of the unstrained material and $\Delta'_{\text{so}} = \Delta_{\text{so}}/3$, Δ_{so} is the spin-orbit interaction energy.

$$H_{\text{so}} = \Delta'_{\text{so}} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -i & 0 & 0 & 0 & 0 & 1 \\ 0 & i & 0 & 0 & 0 & 0 & 0 & -i \\ 0 & 0 & 0 & 0 & 0 & -1 & i & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & i & 0 \\ 0 & 0 & 0 & -i & 0 & -i & 0 & 0 \\ 0 & 1 & i & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (1.92)$$

The Hamilton matrix of the deformation potential interaction reads

$$D = \begin{bmatrix} a_c(\epsilon_{xx} + \epsilon_{yy} + \epsilon_{zz}) & 0 & 0 & 0 \\ 0 & H' & 0 & 0 \\ 0 & 0 & a_c(\epsilon_{xx} + \epsilon_{yy} + \epsilon_{zz}) & 0 \\ 0 & 0 & 0 & H' \end{bmatrix} \quad (1.93)$$

where

$$\begin{vmatrix} \epsilon_{xx} & \epsilon_{xy} & \epsilon_{xz} \\ \epsilon_{yx} & \epsilon_{yy} & \epsilon_{yz} \\ \epsilon_{zx} & \epsilon_{zy} & \epsilon_{zz} \end{vmatrix} \quad (1.94)$$

is the strain tensor (see more discussions in the following section). H' is the 3×3 matrix in the following expression

$$H' = \begin{vmatrix} \ell\epsilon_{xx} + m(\epsilon_{yy} + \epsilon_{zz}) & n\epsilon_{xy} & n\epsilon_{xz} \\ n\epsilon_{yx} & \ell\epsilon_{yy} + m(\epsilon_{zz} + \epsilon_{xx}) & n\epsilon_{yz} \\ n\epsilon_{zx} & n\epsilon_{zy} & \ell\epsilon_{zz} + m(\epsilon_{xx} + \epsilon_{yy}) \end{vmatrix} \quad (1.95)$$

parameters ℓ , m , and n are defined by the deformation potential parameters a_v , b , and d as:

$$a_v = \frac{\ell + 2m}{3}, \quad b = \frac{\ell - m}{3}, \quad d = \frac{n}{\sqrt{3}} \quad (1.96)$$

$$H_1 = \begin{bmatrix} H'_1 & 0 \\ 0 & H'_1 \end{bmatrix} \quad (1.97)$$

with

$$H'_1 = \begin{bmatrix} 0 & iPk_{x\epsilon} & iPk_{y\epsilon} & iPk_{z\epsilon} \\ -iPk_{x\epsilon} & 0 & 0 & 0 \\ -iPk_{y\epsilon} & 0 & 0 & 0 \\ -iPk_{z\epsilon} & 0 & 0 & 0 \end{bmatrix} \quad (1.98)$$

where

$$P = \frac{\hbar p_x}{2m_0}, \quad p_x = \langle S | i\hbar \nabla_x | X \rangle \quad (1.99)$$

For zincblende lattices

$$p_x = p_y \equiv \langle S | i\hbar \nabla_y | Y \rangle = p_z \equiv \langle S | i\hbar \nabla_z | Z \rangle \quad (1.100)$$

so that in the above expression, P is not direction-dependent.

Note that in the eight-band $\mathbf{k} \cdot \mathbf{p}$ model, $|X\rangle$, $|Y\rangle$ and $|Z\rangle$ denote degenerate valence bands while $|S\rangle$ denotes the conduction band, $\mathbf{p}_{cv} = (p_x, p_y, p_z)$ is actually the optical transition matrix between the conduction band and the valence band.

Back to the expression (1.98),

$$k_{\alpha\epsilon} = k_\alpha - \sum_{\beta=x,y,z} \epsilon_{\alpha\beta} k_\beta \quad (1.101)$$

where k_x , k_y and k_z are wave vector components in x , y and z directions. The free-electron and second-order $\mathbf{k} \cdot \mathbf{p}$ interaction contributions are contained in the matrix

$$H_2 = \begin{bmatrix} Ak_\epsilon^2 & 0 & 0 & 0 \\ 0 & S(\mathbf{k}_\epsilon) & 0 & 0 \\ 0 & 0 & Ak_\epsilon^2 & 0 \\ 0 & 0 & 0 & S(\mathbf{k}_\epsilon) \end{bmatrix} \quad (1.102)$$

where $k_\epsilon^2 = k_{x\epsilon}^2 + k_{y\epsilon}^2 + k_{z\epsilon}^2$. $S(\mathbf{k}_\epsilon)$ is the Shockley matrix of Eq. (1.103)

$$S(\mathbf{k}_\epsilon) = \begin{bmatrix} H_{\mathbf{k}_\epsilon} & 0_{3 \times 3} \\ 0_{3 \times 3} & H_{\mathbf{k}_\epsilon} \end{bmatrix} \quad (1.103)$$

$$H_{\mathbf{k}_\epsilon} = \begin{bmatrix} Lk_x^2 + Mk_{yz}^2 & Nk_x k_y & Nk_x k_z \\ Nk_y k_x & Lk_y^2 + Mk_{zx}^2 & Nk_y k_z \\ Nk_z k_x & Nk_z k_y & Lk_z^2 + Mk_{xy}^2 \end{bmatrix}$$

where $k_{ij}^2 \equiv k_i^2 + k_j^2$, $0_{3 \times 3}$ is the 3×3 zero matrix, L , M , N are band parameters given by

$$L = \frac{-\hbar^2}{m_0}(\gamma_1 + 4\gamma_2), \quad M = \frac{-\hbar^2}{m_0}(\gamma_1 - 2\gamma_2), \quad N = \frac{-\hbar^2}{m_0}(6\gamma_3) \quad (1.104)$$

where γ_1 , γ_2 and γ_3 are Luttinger parameters.

In Eq. (1.102), $A = \hbar^2/2m_e^*$, where m_e^* is normally referred to as the effective mass of electrons in the conduction band given by

$$\frac{m_0}{m_e^*} = 1 + 2F + \frac{E_p(E_g + \frac{2}{3}\Delta_{so})}{E_g(E_g + \Delta_{so})} \quad (1.105)$$

In the above equation, $E_p = p_{cv}^2/2m_0$ is the energy parameter related to the momentum matrix p_{cv} , F is the Kane parameter.

It is important to account for the temperature dependence of the bandgap energies through the Varshni approximation

$$E_g(T) = E_g(0) - \frac{\alpha T^2}{T + \beta} \quad (1.106)$$

where α and β are the empirical Varshni parameters and T is the absolute temperature.

To determine the confinement potential we can apply the method where the top of the valence band of every binary III–V compound is measured with respect to a common reference energy level through a valence band offset (VBO) value [48]. The VBO values of the ternary alloys are obtained from the binaries using expressions similar to Eq. (1.110).

Principal band structure parameters of the common binary compounds are listed in Table 1.4 [47]. The permittivities ϵ are obtained from Ref. [49] (which will be used in studying quantum devices). Moreover, the permittivity of InN is derived from its refractive index of 2.56 listed in Ref. [49]. Parameters for relaxed bulk ternary and quaternary alloys are obtained by interpolations similar to Eq. (1.110). Bowing effects are included whenever the values of the bowing parameters are available in Ref. [47] and references therein.

As mentioned before, Table 1.4 lists principal band structure parameters of binary compounds, whereas ternary alloys of these binary compounds are commonly used for energy band structure engineering, as alloying of two materials is one of the oldest techniques to modify properties of materials. When two semiconductors A and B are mixed via an appropriate growth technique, one must have the following information regarding the structure of the alloy:

1. The crystalline structure of the alloy lattice.
2. The ordering of atoms. If an alloy of $A_x B_{1-x}$ is formed, where x is the mole fraction of material A ,
 - All of A atoms are localized in one region while B atoms localized in another region. Such alloys are called phase separated.

Table 1.4 Band structure parameters for zinc-blende GaAs, InP, InAs, GaN, AlN, and InN [47, 49]

Parameters	GaAs	InP	InAs	GaN	AlN	InN
a [Å]	5.65325	5.8697	6.0583	4.50	4.38	4.98
E_g^Γ [eV]	1.519	1.4236	0.417	3.299	4.9	1.94
$\alpha(\Gamma)$ [meV/K]	0.5405	0.363	0.276	0.593	0.593	0.245
$\beta(\Gamma)$ [K]	204	162	93	600	600	624
Δ_{so} [eV]	0.341	0.108	0.39	0.017	0.019	0.006
m_e^*	0.067	0.0795	0.026	0.15	0.25	0.12
γ_1	6.98	5.08	20.0	2.67	1.92	3.72
γ_2	2.06	1.60	8.5	0.75	0.47	1.26
γ_3	2.93	2.10	9.2	1.10	0.85	1.63
F	-1.94	-1.31	-2.90	-0.92	0.76	-0.92
VBO [eV]	-0.80	-0.94	-0.59	-2.64	-3.44	-2.38
a_c [eV]	-7.17	-6.0	-5.08	-2.2	-6.0	-1.85
a_v [eV]	1.16	0.6	1.0	-5.2	-3.4	-1.5
b [eV]	-2.0	-2.0	-1.8	-2.2	-1.9	-1.2
d [eV]	-4.8	-5.0	-3.6	-3.4	-10	-9.3
C_{11} [GPa]	1221	1011	832.9	293	304	187
C_{12} [GPa]	566	561	452.6	159	160	125
E_p [eV]	28.8	20.7	21.5	25.0	27.1	25.0
ϵ [49]	12.40	12.5	15.15	10.4	9.14	6.55

- The probability that an atom next to an A -type atom is A is x and the probability that it is B is $(1 - x)$. Such are called random alloys.
- The A and B atoms form a well-ordered periodic structure, leading to a superlattice.

Most semiconductor alloys used in the electronics and optoelectronics industry are grown with the intention of making perfectly random alloys. Usually, deviations from this randomness lead to deterioration in the device performance.

For materials having the same lattice structure, the lattice constant of an alloy obeys Vegard's law

$$a_{\text{alloy}} = xa_A + (1 - x)a_B \quad (1.107)$$

where a_{alloy} , a_A and a_B are lattice constants of the alloy, material A and B , respectively. In the virtual crystal approximation the random-alloy potential is replaced by an average periodic potential

$$V_{\text{alloy}}(\mathbf{r}) = xV_A(\mathbf{r}) + (1 - x)V_B(\mathbf{r}) \quad (1.108)$$

where $V_A(\mathbf{r})$ and $V_B(\mathbf{r})$ are periodic potential energies of material A and B , in the term of Eq. (1.32).

For example, an implementation of this approach in the tight-binding method involves taking the weighted average of the matrix elements. For direct-bandgap materials,

$$E_{g,\text{alloy}} = xE_{g,A} + (1-x)E_{g,B} \quad (1.109)$$

$$\frac{1}{m_{\text{alloy}}^*} = \frac{x}{m_A^*} + \frac{1-x}{m_B^*}$$

where E_g is the energy bandgap.

When calculating the bandgap energies of the constituent ternaries we need to account for the bandgap bowing effect. The bandgap energy of a relaxed $\text{In}_x\text{Ga}_{1-x}\text{As}$ alloy is given by:

$$E_g(\text{In}_x\text{Ga}_{1-x}\text{As}) = xE_g(\text{InAs}) + (1-x)E_g(\text{GaAs}) - x(1-x)C_{\text{InGaAs}} \quad (1.110)$$

where C is the bowing parameter. The recommended value for C_{InGaAs} is 0.477 eV [47]. Similar expressions are used for the other relaxed ternary alloys with $C_{\text{InGaIn}} = 2.2$ eV for $\text{In}_x\text{Ga}_{1-x}\text{N}$ when $x = 0.36$ [50], $C_{\text{GaInAs}} = 20$ eV for $\text{GaN}_y\text{As}_{1-y}$ when $x = 0.01$ and 11 eV when $x = 0.03$ [47, 51], and for $\text{InN}_x\text{As}_{1-x}$, the recommended bowing factor (C_{InNAs}) is 4.22 eV [47, 52].

We have thus far discussed mostly III–V materials. Si and Ge are indirect bandgap materials where the lowest conduction bandedges are not located at the Γ points where the top valence bandedges locate. Therefore, the conduction-band electron states and valence-band hole states can well be separately described. For example, 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], see Eq. (1.75). When we consider a Si thin film embedded in SiO_2 which is extended in the xy plane, the translational symmetry is broken down along the z direction so that the six ellipsoids become divided into groups. (1) $m_z^* = m_\ell^*$, $m_x^* = m_y^* = m_t^*$, degeneracy $g = 2$; (2) $m_z^* = m_t^*$, $m_x^* = m_y^* = m_\ell^*$, $g = 4$. Here m_x^* , m_y^* and m_z^* are the effective masses in the x , y and z directions, respectively. In the Schrödinger equation, V_c is the potential energy for the conduction band electron. For undoped Si thin film embedded in SiO_2 , V_c is zero in Si and about 3.5 eV in SiO_2 , which is the conduction-band offset between Si and SiO_2 [53]. Coulombic potential of conduction-band electrons and dopants will be added when the sample becomes doped (see more discussions late).

The valence bands of Si and Ge are described by the 6-band $\mathbf{k} \cdot \mathbf{p}$ perturbation Hamiltonians, and the values of band parameters are listed in Table 1.5 [41, 42].

Equation (1.103) includes only terms in $k_i k_j$. For high hole energy states (high \mathbf{k} values), terms of higher order in k_i must be included. In the first order approximation we assume that $E_{xxxx}^{\alpha 4} = Q$ is independent of $|\alpha\rangle$ and $E_{ijmn}^{\alpha 4} = 0$ if i, j, m and n are

Table 1.5 Valence band parameters of the 6-band $\mathbf{k} \cdot \mathbf{p}$ theory for Si and Ge

Parameters	Unit	Si	Ge
L	$\text{eV} \cdot \text{\AA}^2$	-25.51	-143.32
M	$\text{eV} \cdot \text{\AA}^2$	-15.17	-22.90
N	$\text{eV} \cdot \text{\AA}^2$	-38.10	-161.22
Q	$\text{eV} \cdot \text{\AA}^4$	-125.0 [54]	
Δ	eV	60.044	0.282
a	eV	2.1	2.0
b	eV	-1.5	-2.2
d	eV	-3.4	-4.4
a_0	\AA	5.4309	5.6561
C_{11}	10^{11} dyn/cm ²	16.56	12.853
C_{12}	10^{11} dyn/cm ²	6.39	4.826

not the same so that

$$H(\mathbf{k}) = \begin{vmatrix} Lk_x^2 + Mk_{yz}^2 + Qk_x^4 & Nk_xk_y & Nk_xk_z \\ Nk_yk_x & Lk_y^2 + Mk_{zx}^2 + Qk_y^4 & Nk_yk_z \\ Nk_zk_x & Nk_zk_y & Lk_z^2 + Mk_{xy}^2 + Qk_z^4 \end{vmatrix} \quad (1.111)$$

The value of Q for Si was obtained by fitting the calculated Fermi levels at different doping levels with measurement data [54].

Figure 1.14 shows the valence band structures of bulk Si and $\text{Si}_{0.7}\text{Ge}_{0.3}$ alloys grown, respectively on Si and $\text{Si}_{0.7}\text{Ge}_{0.3}$ substrates. Here we have taken the valence bandedges of the alloy materials as zero energy reference (excluding the spin-orbital interaction). In the unstrained $\text{Si}_{0.7}\text{Ge}_{0.3}$ and Si materials, the heavy-hole and light-hole bands are degenerate at $k_z = 0$. However, in the tensile (compressional) strained Si ($\text{Si}_{0.7}\text{Ge}_{0.3}$) the light-hole band moves upwards (downwards), whereas the heavy-hole band is not much affected by the strain.

As discussed earlier, a valence band offset is expected at a heterostructure interface. Considering the (001) orientation and referring ΔE_v as the discontinuity in the top of the valence bands at Γ_{15}^v , van de Walle and Martin showed that ΔE_v varies almost linearly with the lattice constant a_{xy} [55]

$$\Delta E_v = 0.84 - 2.41(a_{xy} - 5.43) \quad (1.112)$$

for a heterostructure of $\text{Si}_{1-x}\text{Ge}_x$ grown on (001) Si, where a_{xy} is in \AA and ΔE_v in eV.

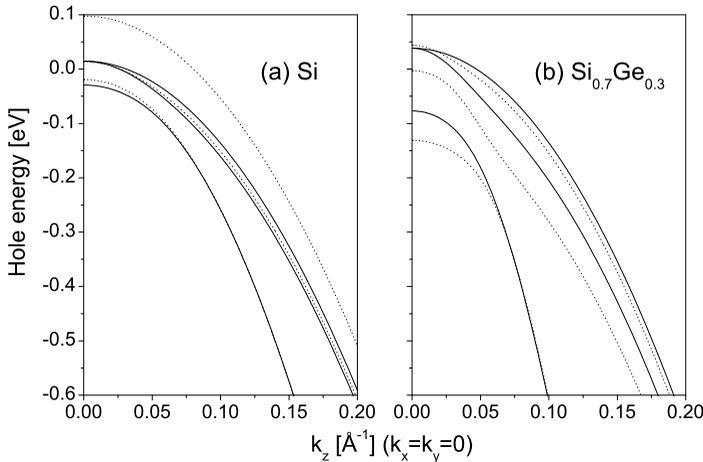


Fig. 1.14 Si and $\text{Si}_{0.7}\text{Ge}_{0.3}$ valence band structures. *Solid lines*: unstrained materials. *Dotted lines*: Si grown on $\text{Si}_{0.7}\text{Ge}_{0.3}$ substrate and $\text{Si}_{0.7}\text{Ge}_{0.3}$ on Si

1.7 Strain Field in Nanostructures

The proposal and realization of the concept “semiconductor heterostructures” was revolutionary in electronics and optoelectronics and paved the way to the fast development of information and communication technology [56]. The epitaxy methods of heterostructures, i.e., molecular beam epitaxy (MBE) and metal-organic vapor phase epitaxy (MOVPE), have already become a common technology for semiconductor industry. A critical issue is the lattice mismatch between the epitaxial layer and the substrate. When a thin epitaxial layer is grown onto a substrate with a different lattice constant, the atoms in the epitaxial layer will be forced to align with the atoms in the substrate so that the epitaxial layer is strained. The strain force will gradually accumulate during epitaxial growth, it will eventually start to relieving by forming three-dimensional (3D) islands on the growth surface, and/or misfit dislocations inside the crystal. Strain relaxation often generates lattice defects which degrade device performances, it can also be utilized. The most significant example is the self-assembled quantum dots formed by nucleation and coalescence of the 3D islands via the Stranski-Krastanov growth mode [57].

As discussed in the previous section, strain also affect the energy band structure of the epitaxial layer, and thus provides a means to tailor the energy band structure and carrier transport. In this section, we discuss the theories of the elasticity and microscopic strain field.

Refer to Fig. 1.15, we first discuss the strain in terms of displacements in two dimensions induced by tensional stresses P_1 , P_2 , and shear stress S_3 . We suppose that a naturally plane body (or sheet) is strained in such a way that all the particles remain of the body in one plane after the strain. It is necessary to refer all the displacements to a pair of axes fixed relatively to some particles of the body. Let the origin O be situated at one of the particles of the body, and if that particle moves

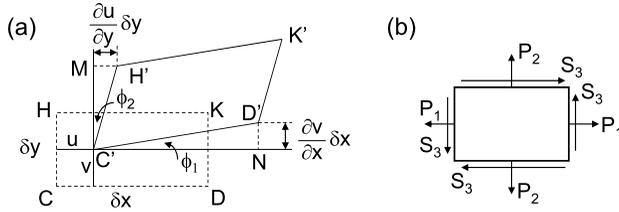


Fig. 1.15 Displacements, strain, stress, and external forces

O is supposed to move with it; and let the axis OX always pass through one other given particle of the body. The axis OY is always perpendicular to OX and in the plane of the particles.

Let the particle which was at (x, y) before strain be situated at $(x + u, y + v)$ after strain. Both u and v are functions of x and y since each will vary with both x and y . We have to investigate the change of size and shape of an element, which, in the natural state, is a rectangle with sides δx and δy .

The rectangle CDKH, Fig. 1.15(a) is displaced relatively to the axes into the position C'D'K'H'. The displacement of C has components u and v . Now suppose

$$u = f(x, y) \tag{1.113}$$

Then, denoting the displacement of D parallel to OX by $(u + \delta u)$, we get, since the coordinates of D are $(x + \delta x, y)$,

$$u + \delta u = f(x + \delta x, y) \tag{1.114}$$

Therefore

$$\delta u = f(x + \delta x, y) - f(x, y) = \frac{\partial f(x, y)}{\partial x} \delta x = \frac{\partial u}{\partial x} \delta x \tag{1.115}$$

But δu is the displacement of D' relative to C' and is therefore the excess of C'N over CD. Now the length of CD is δx , and the horizontal projection of C'D' is greater than CD by $\frac{\partial u}{\partial x} \delta x$. Hence the extensional strain in the direction of OX is

$$\frac{C'N - CD}{CD} = \frac{\frac{\partial u}{\partial x} \delta x}{\delta x} = \frac{\partial u}{\partial x} \tag{1.116}$$

Likewise the extensional strain in the direction of OY is $\partial v / \partial y$.

The shear strain for the lines C'D' and C'H' is, by definition, the whole change of angle at C'; that is, the shear strain is $(\phi_1 + \phi_2)$. But

$$\phi_1 \approx \frac{ND'}{C'N} = \frac{\partial v}{\partial x} \tag{1.117}$$

and $\phi_2 = \partial u / \partial y$. Therefore the shear strain is

$$\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \tag{1.118}$$

In three dimensions, we add tensional stress P_3 along the z axis and shear stresses S_1 and S_2 in the yz and xz planes. Let a particle originally at (x, y, z) move to $(x + u, y + v, z + w)$. It follows then that the three dimensional strains are $\partial u/\partial x$, $\partial v/\partial y$, and $\partial w/\partial z$. Likewise the three component shear strains are

$$\left(\frac{\partial v}{\partial z} + \frac{\partial w}{\partial y}\right), \quad \left(\frac{\partial w}{\partial x} + \frac{\partial u}{\partial z}\right), \quad \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right) \quad (1.119)$$

We shall denote the extensional strains by α , β , γ , respectively, and the shear strains by a , b , c , respectively, that is

$$\begin{aligned} \alpha &= \frac{\partial u}{\partial x}, & \beta &= \frac{\partial v}{\partial y}, & \gamma &= \frac{\partial w}{\partial z}, \\ a &= \frac{\partial v}{\partial z} + \frac{\partial w}{\partial y}, & b &= \frac{\partial w}{\partial x} + \frac{\partial u}{\partial z}, & c &= \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \end{aligned} \quad (1.120)$$

In contemporary physics, $\alpha = \varepsilon_{xx}$, $\beta = \varepsilon_{yy}$, $\gamma = \varepsilon_{zz}$, $a/2 = \varepsilon_{xy}$, $b/2 = \varepsilon_{yz}$, $c/2 = \varepsilon_{xz}$, $P_1 = \sigma_{xx}$, $P_2 = \sigma_{yy}$, $P_3 = \sigma_{zz}$, $S_3 = \sigma_{xy}$, $S_2 = \sigma_{xz}$, and $S_1 = \sigma_{yz}$, where $[\sigma_{ij}]$ is denoted as the three-dimensional Cauchy stress tensor and $[\varepsilon_{ij}]$ the Cauchy's infinitesimal strain tensor.

When working with a three-dimensional stress state, a fourth order tensor $[\lambda_{ijkl}]$ containing 81 elastic coefficients must be defined to link the stress tensor and the strain tensor

$$\sigma_{ij} = \sum_{k\ell} \lambda_{ijkl} \varepsilon_{k\ell} \quad (1.121)$$

$i, j, k, \ell = x, y, z$. Due to the symmetry of the stress, strain and stiffness tensors, only 21 elastic coefficients are independent. In crystals with cubic symmetry, there are only three linearly independent constants: $\lambda_{xxxx} = C_{11}$, $\lambda_{xxyy} = C_{12}$, $\lambda_{xyxy} = C_{44}$. For wurtzite structure, there are five linearly independent elastic constant: $\lambda_{xxxx} = C_{11}$, $\lambda_{zzzz} = C_{33}$, $\lambda_{xxyy} = C_{12}$, $\lambda_{xxzz} = C_{13}$, $\lambda_{xzxz} = C_{44}$ [47].

Figure 1.16 shows the schematic Stranski-Krastanov growth mode of InAs QDs on GaAs substrate. Denote a_{GaAs} is the lattice constant of substrate GaAs, and a_{InAs} is the lattice constant of the bulk InAs material. As the InAs atoms in the first epitaxial layer will align with the GaAs atoms in the xy plane so that the lattice constants of the thin wetting layer are the same as a_{GaAs} , the displacements of InAs atoms in the xy plane are

$$u = v = a_{\text{InAs}} - a_{\text{GaAs}}$$

and the wetting layer is therefore biaxially strained in the xy -plane by an amount:

$$\varepsilon_{xx} = \varepsilon_{yy} = \frac{a_{\text{InAs}}}{a_{\text{GaAs}}} - 1 \quad (1.122)$$

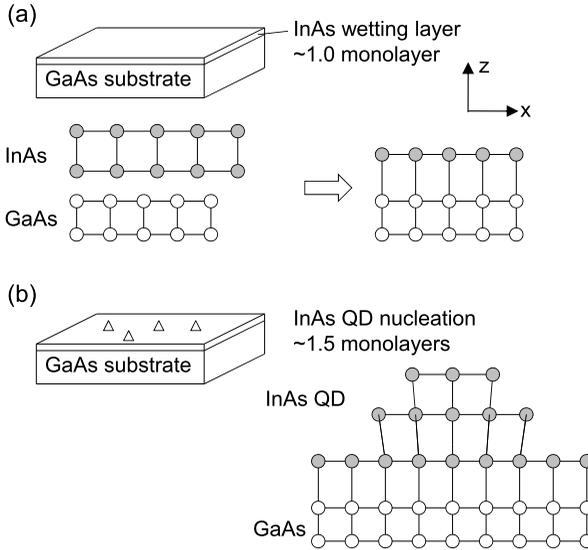


Fig. 1.16 Schematic self-assembled InAs QD by Stranski-Krastanov growth mode

where we have assumed $\partial u / \partial x = u / a_{\text{GaAs}}$ is constant. By minimizing the strain energy, the epitaxial layer is uniaxially strained in the z direction by an amount:

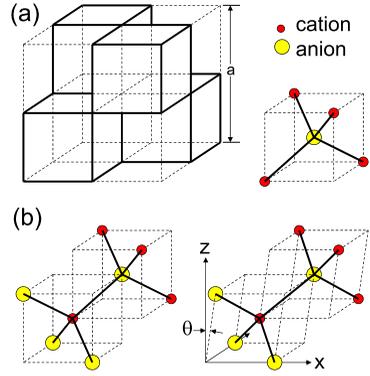
$$\varepsilon_{zz} = -\frac{2C_{12}}{C_{11}}\varepsilon_{xx} \quad (1.123)$$

where C_{12} and C_{11} are elastic constants of InAs. Other strain elements are zero, i.e., $\varepsilon_{ij} = 0$ if $i \neq j$. Note that we will discuss more carefully the above two equations at the end of this section.

The calculation of the strain in the three dimensional InAs QD of Fig. 1.16(b) is complicated. The geometric shape of the QD is determined by many factors including growth conditions and inter-diffusion between QD material and background material. By a hybrid computational approach that combines continuum calculations of strain energy with first-principles results, relatively small effects associated with surface stress were obtained for InAs/GaAs system [58], while it was theoretically demonstrated to be an important factor in Ge/Si QD formation [59]. Wise et al. demonstrated the evolution of the QD formation under the strong influence of surface energy by cubic elastic equilibrium equations [60]. Kinetic Monte Carlo study of metal organic chemical vapor deposition (MOCVD) growth mechanism of GaSb QDs in GaAs showed that the strain induced by the lattice mismatch between the epitaxial material and the substrate is directly responsible for the QD formation and the transition from two-dimensional to three-dimensional growth mode [61].

In the following we introduce the short-range valence-force-field (VFF) approach to describe inter-atomic forces in the QD systems in terms of bond stretching and bending [62, 63]. The model is microscopic and has been widely applied in bulk and

Fig. 1.17 Microscopic strain field in zincblende crystal lattice structure: the valence force field (VFF) approach



alloys [64–69], as well as low-dimensional systems [70–72]. In the VFF model, the deformation of a lattice structure is completely specified when the location of every atom in the system is given [62], and the elastic energy of a bond is minimal in its three-dimensional relaxed bulk lattice structure.

Refer to Fig. 1.17 for the bond configuration in the zincblende crystal lattice structure, the bond between two atoms is denoted by its bond length r_{i0} and spatial orientation Ω_{i0} at equilibrium, where i is the bond index. When the spatial positions of the atoms change, so change the relevant bond parameters to (r_i, Ω_i) . The elastic energy of this bond (by setting the elastic energy at equilibrium as the zero reference) can be written in the harmonic form

$$E_i = K_{ir}\delta r_i^2 + K_{i\Omega}r_{i0}^2 \sum_{j=1}^6 (1 - \cos \Omega_{ij}) + K_{irr}\delta r_i \sum_{j=1}^6 \delta r_j \quad (1.124)$$

with $j = 1, 2, \dots$ denotes the nearest-neighbor bonds. Note that there are six nearest-neighbor bonds in zincblende structure, see Fig. 1.17(b). δr_i is the variation of the length of bond i , and Ω_{ij} is the variation of the angle between the i th and the j th bond. The total elastic energy is the sum of all bond energies

$$E_{\text{elastic}} = \sum_i E_i \quad (1.125)$$

The numerical values of K 's for VFF bonds of zincblende bulk materials are easily obtained from elastic coefficients C_{11} , C_{12} , and C_{44} listed in [47]. We consider a bulk semiconductor under a hydrostatic pressure and a shear stress along the x direction. The bulk semiconductor is composed of N unit cells in x , y and z direction so that its lengths are all Na and volume $(Na)^3$. Under a hydrostatic pressure, the lengths are modified by δ_x , δ_y and δ_z so that

$$u = \frac{x}{Na}\delta_x, \quad v = \frac{y}{Na}\delta_y, \quad w = \frac{z}{Na}\delta_z \quad (1.126)$$

Table 1.6 Lattice constant a , elastic coefficients C 's [47, 49, 73] and K 's of zincblende InAs, InP, GaAs, Si, and Ge bulk materials at 100 K

	InAs	InP	GaAs	Ge	Si
a [Å]	6.05	5.86	5.65	5.657	5.432
C_{11} [GPa]	83.29	101.1	122.1	128.9	165.7
C_{12} [GPa]	45.26	56.1	56.6	48.3	63.9
C_{44} [GPa]	39.59	45.6	60.0	67.1	79.6
K_r [GPa Å]	170.13	198.57	215.09	212.4	256.1
K_{Ω} [GPa Å]	9.59	10.99	15.41	18.9	23.1
K_{rr} [GPa Å]	59.32	71.14	74.85	73.2	93.2

$$\varepsilon_{xx} = \frac{\partial u}{\partial x} = \frac{\delta_x}{Na}, \quad \varepsilon_{yy} = \frac{\delta_y}{Na}, \quad \varepsilon_{zz} = \frac{\delta_z}{Na}$$

For zinc blende structures,

$$\begin{aligned} \sigma_{xx} &= C_{11}\varepsilon_{xx} + C_{12}\varepsilon_{yy} + C_{12}\varepsilon_{zz} \\ \sigma_{yy} &= C_{21}\varepsilon_{xx} + C_{11}\varepsilon_{yy} + C_{12}\varepsilon_{zz} \\ \sigma_{zz} &= C_{21}\varepsilon_{xx} + C_{12}\varepsilon_{yy} + C_{11}\varepsilon_{zz} \end{aligned} \quad (1.127)$$

Other components of ε and σ are zero. The strain energy is

$$E_{\text{elastic}} = \frac{1}{2} [C_{11}(\varepsilon_{xx}^2 + \varepsilon_{yy}^2 + \varepsilon_{zz}^2) + 2C_{12}(\varepsilon_{xx}\varepsilon_{yy} + \varepsilon_{xx}\varepsilon_{zz} + \varepsilon_{yy}\varepsilon_{zz})] (Na)^3 \quad (1.128)$$

For the shear stress along the x direction of Fig. 1.17(b), $u = z \tan \theta$, $v = 0$ and $w = 0$, so that

$$\begin{aligned} \varepsilon_{xz} &= \frac{1}{2} \tan \theta, \quad \sigma_{xz} = \frac{1}{2} C_{44} \tan \theta \\ E_{\text{elastic}} &= \frac{1}{8} C_{44} \tan^2 \theta (Na)^3 \end{aligned} \quad (1.129)$$

We can further consider another shear stress along the y direction so that we totally have three independent linear equations for K_{ir} , $K_{i\Omega}$, and K_{irr} . Note that K values depend on the temperature of the material due to the temperature dependence of the lattice constant. The dependence, however, is found to be very weak. The K values listed in Table 1.6 are obtained for bulk materials at 100 K.

The local lattice structure of the InAs QD is determined by adjusting spatial coordinates of lattice sites in order to minimize the total strain energy. The adjustment of the lattice sites can be performed by a Monte Carlo scheme in such a way that we generate one set of random walkings for all lattice sites, check the total elastic energy. If the total elastic energy is reduced, the set of the random walkings is allowed and the lattice sites are then upgraded.

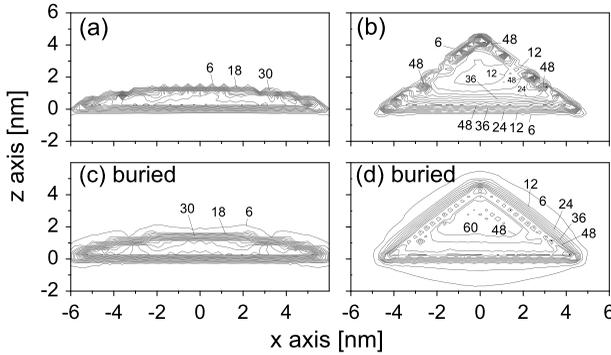


Fig. 1.18 Spatial distributions of strain energies around InAs QDs (volume = 91.3 nm^3) in GaAs matrices. $y = 0$. Linear scale applies and values associated with contour lines are in the unit of meV. (a) Uncapped hemisphere and (b) pyramid InAs QDs on GaAs substrates, (c) hemisphere and (d) pyramid InAs QDs buried in GaAs matrices (X.-F. Yang, K. Fu, W. Lu, W.-L. Xu, and Y. Fu, Strain effect in determining the geometric shape of self-assembled quantum dot, *J. Phys. D: Appl. Phys.*, vol. 42, p. 125414(8), 2009)

Spatial distributions of the strain energies around InAs/GaAs QDs are presented in Fig. 1.18. Two QDs are considered here, one is a hemisphere with a base diameter of 12 nm and a height 1.5 nm, the other QD is a pyramid with a base size 8 nm and a height 4 nm. The strain energy of the QD pyramid is significantly small when the QD is uncapped due to the lattice relaxation at the QD surface, whereas atoms in the QD hemisphere are not able to relax much since all of them are very close to the substrate and therefore are significantly strained. This demonstrates clearly the general understanding of the strain-driven self-assembling formation of QDs that the lattice-mismatch-induced strain is largely released when the QD forms on a substrate.

The strain field distribution is readily calculated which are presented in Fig. 1.19. The strain fields of the QD hemisphere can be very well approximated by Eqs. (1.122), (1.123) since the horizontal dimension (diameter = 12 nm) is much larger than the vertical one (height = 1.5 nm). The strain field distribution in the GaAs background is also very significant, especially perpendicular to the substrate surface. The strain distribution in the QD pyramid is rather complicated. We observe significant strain relaxation inside the QD in the top vicinity. The surrounding background material becomes also largely strained. The major reason is the ratio between the numbers of the surface and total QD atoms. In other words, the main difference between the hemisphere and pyramid QDs under discussion is the aspect ratio. For the hemisphere QD with a small aspect ratio, surface atoms dominate which are either very close to or in direct contact with the GaAs substrate, whereas atoms in the pyramid QD can be well catalogued into strained surface atoms and much-relaxed bulk ones.

Now we discuss a bit further about the strain field distribution of Eqs. (1.122) and (1.123). It can be noticed that the two equations have been also applied to quantum well structures formed by alloys such as $\text{Al}_x\text{Ga}_{1-x}\text{As}$ quantum well embedded in

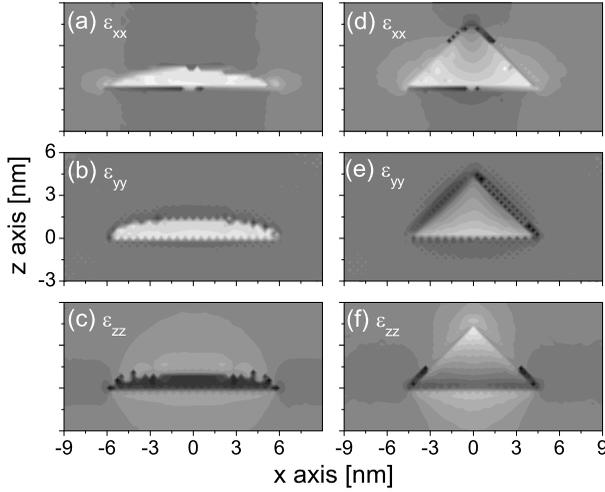


Fig. 1.19 Strain field distributions around InAs QDs embedded in GaAs matrices. $y = 0$. Linear scale from white (-0.04) to black (0.04) applies. (a–c) Hemisphere QD, (d–f) pyramid QD (X.-F. Yang, K. Fu, W. Lu, W.-L. Xu, and Y. Fu, Strain effect in determining the geometric shape of self-assembled quantum dot, J. Phys. D: Appl. Phys., vol. 42, p. 125414(8), 2009)

AlAs, or $\text{InAs}_x\text{P}_{1-x}$ quantum well in InP. In the latter case, the strain fields in the $\text{InAs}_x\text{P}_{1-x}$ quantum well are described as

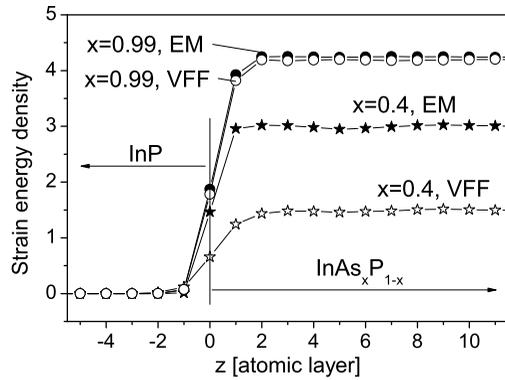
$$\varepsilon_{xx} = \varepsilon_{yy} = \frac{a_{\text{InAs}_x\text{P}_{1-x}}}{a_{\text{InP}}} - 1, \quad \varepsilon_{zz} = -\frac{2C_{12}}{C_{11}}\varepsilon_{xx} \quad (1.130)$$

where we apply Vegard's law, Eq. (1.107) to obtain the lattice constant of an alloy, and similar relationships for elastic constants

$$\begin{aligned} a_{\text{InAs}_x\text{P}_{1-x}} &= xa_{\text{InAs}} + (1-x)a_{\text{InP}} \\ C_{12} &= xC_{12,\text{InAs}} + (1-x)C_{12,\text{InP}} \\ C_{11} &= xC_{11,\text{InAs}} + (1-x)C_{11,\text{InP}} \end{aligned} \quad (1.131)$$

To validate the strain fields of Eq. (1.130), let us do two simulations about a [001] $\text{InAs}_x\text{P}_{1-x}$ quantum well grown on an InP substrate. We denote [001] as the z direction. The first simulation is to use the VFF model in the following way. The quantum well growth starts at the surface of the InP substrate which is denoted as $z = 0$. For $z \geq 0$, In atoms deposit to the cation sites of the lattice, while As and P atoms deposit to anion sites according to the mole fraction x of As atoms. The growth mode is two dimensional, i.e., atomic layer-by-layer growth along the z direction. For each growing atomic layer, we allow the atoms on the growing layer as well as atoms in three atomic layers beneath the growing layer to adjust their spatial positions in order to minimize the strain energy. Strain energy per atom, i.e., the strain energy density, along the z direction, thus obtained, is denoted as VFF

Fig. 1.20 Strain energy density (per atom) in $\text{InAs}_x\text{P}_{1-x}$ quantum well grown on InP substrate along the [001] direction. EM: effective medium approximation; VFF: valence-force field model



and plotted in Fig. 1.20. In the second simulation, the lattice constants in the x and y directions of the atoms in the $\text{InAs}_x\text{P}_{1-x}$ quantum well assume the value of InP, while the lattice constant along the z direction is obtained from Eqs. (1.130), which is denoted as the effective medium approximation (EM). The strain energy density thus calculated is plotted in Fig. 1.20 for comparison with the VFF results. Here it is observed that the VFF result is smaller than the EM model. Only when the As mole fraction reaches close to 1.0 give the two models the same strain energy.

By performing similar simulations for x ranging from 0.0 to 1.0, we calculate the ratio between the VFF and EM strain energy density, which is shown in Fig. 1.21. For all x smaller than 1.0, the VFF strain energy is always lower than the EM approximation. The reason is rather straightforward: As and P atoms in $\text{InAs}_x\text{P}_{1-x}$ can adjust their spatial positions locally so that the local strain energies, and thus

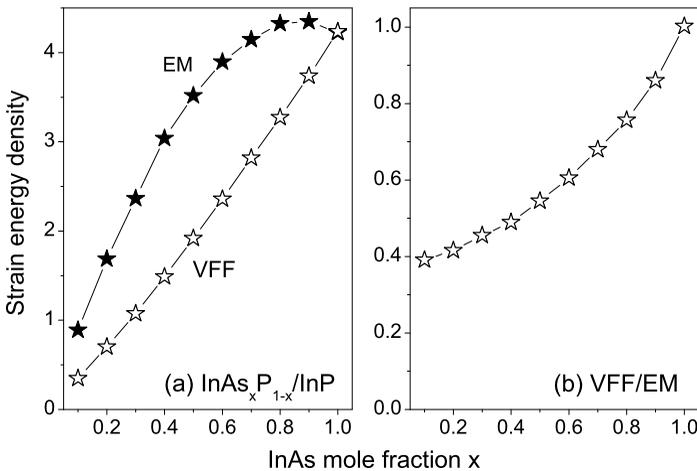


Fig. 1.21 Strain energy density (strain energy per atom) along the direction perpendicular to the two-dimensional [001] $\text{InAs}_x\text{P}_{1-x}$ quantum well embedded in InP. (a) Strain energy densities obtained by VFF and EM models, (b) ratio between VFF and EM strain energy densities

the total strain energy, become small. For large x , especially InAs on InP, the freedom for the local spatial position adjustments is small, so that the results of the VFF and the EM approximation converge.

It is thus concluded that the microscopic structures are critically important in determining the macroscopic properties of semiconductor nanostructures.

1.8 Heterostructure Material and Envelope Function

Among all possible combinations of semiconductors, the lattice-matched (or at least closely matched) combinations have attracted most interest. The reason is, of course, the possibility of ideal heterointerfaces if the lattice constants of the constituents are the same. It is interesting to note that perfect lattice match does not guarantee an ideal interface, as can be seen from the results for the GaAs-Ge system [74]. In addition to the lattice match condition, we also have to require that both systems be binary (or ternary, quaternary) in order to avoid, for instant, the site allocation problem [75].

A picture of the effects of the heterojunctions on the band structure can be obtained by employing the effective-mass theorem. The simplest approach is to view the semiconductor as homogeneous with an additional potential superimposed on the crystal potential. Then the effective-mass theory tells us that for electrons at band minima or maxima we can replace the Hamiltonian

$$\frac{-\hbar^2 \nabla^2}{2m_0} + V_0(\mathbf{r}) + V(\mathbf{r}) \quad (1.132)$$

by

$$\frac{-\hbar^2 \nabla^2}{2m^*} + V(\mathbf{r}) \quad (1.133)$$

and therefore transform away the periodic crystal potential $V_0(\mathbf{r})$, provided that $V(\mathbf{r})$ and its Fourier components fulfill certain conditions. In other words we have reduced the heterojunction problem to a well-defined potential [$V(\mathbf{r})$] problem for a quasi-free electron with effective mass m^* .

To complete the effective-mass theory we need to know the potential $V(\mathbf{r})$. Although it is difficult to determine $V(\mathbf{r})$ with respect to the vacuum level, a calculation of $V(\mathbf{r})$ relative to the bandedge of one component of the heterostructures seems to be easier. Harrison et al. proposed [35] that $V(\mathbf{r})$ can be determined by the linear-combination-of-atomic-orbitals method which gives the valence bandedge E_v as

$$E_v = \frac{E_a + E_c}{2} - \left[\left(\frac{E_a + E_c}{2} \right)^2 + V_{xx}^2 \right]^{1/2} \quad (1.134)$$

where E_a (E_c) are the p -orbital energies of the anion (cation) atoms, and V_{xx} is the interatomic matrix element of the crystal Hamiltonian between p -orbitals of

adjacent atoms. The valence band offset ΔE_v for hole of, e.g. an AlAs-GaAs heterojunction is therefore

$$\Delta E_v = E_v(\text{AlAs}) - E_v(\text{GaAs}) \quad (1.135)$$

and the conduction band discontinuity ΔE_c is

$$\Delta E_c = E_g(\text{AlAs}) - E_g(\text{GaAs}) - \Delta E_v \quad (1.136)$$

Here E_g is the energy bandgap of AlAs or GaAs. A comparison of these values with experimental results has been given by Margaritondo et al. [76], who found an overall agreement between theory and experiments.

We now seek to solve the wave equation

$$[H_0 + V(\mathbf{r})]\Psi = E\Psi \quad (1.137)$$

where

$$H_0 = \frac{p^2}{2m_0} + V_0(\mathbf{r})$$

is the Hamiltonian of a perfect crystal, $V(\mathbf{r})$ is a perturbing potential due to heterojunctions in the heterostructure system. m_0 is the free electron mass.

For the perfect crystal, the Bloch theorem says

$$H_0 u_{nk} e^{i\mathbf{k}\cdot\mathbf{r}} = E_n(\mathbf{k}) u_{nk} e^{i\mathbf{k}\cdot\mathbf{r}} \quad (1.138)$$

where $u_{nk}(\mathbf{r}) = u_{nk}(\mathbf{r} + \mathbf{R}_i)$, \mathbf{R}_i is the position of unit cell i . $n = 1, 2, \dots, \mu$ is the band index.

When we investigate a crystal whose component composition is not uniform, it is more suitable to use the Wannier function

$$a_n(\mathbf{r} - \mathbf{R}_i) = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} u_{nk}(\mathbf{r} - \mathbf{R}_i) e^{i\mathbf{k}\cdot(\mathbf{r} - \mathbf{R}_i)} \quad (1.139)$$

Here we assume that the Bloch functions are normalized for a volume Ω containing N unit cells.

Before we move on, let us study the characteristics of the above Wannier function.

$$\begin{aligned} & \langle a_n(\mathbf{r} - \mathbf{R}_i) | a_n(\mathbf{r} - \mathbf{R}_j) \rangle \\ &= \frac{1}{N} \sum_{\mathbf{k}q} \int u_{nk}^*(\mathbf{r} - \mathbf{R}_i) u_{nq}(\mathbf{r} - \mathbf{R}_j) e^{-i\mathbf{k}\cdot(\mathbf{r} - \mathbf{R}_i) + i\mathbf{q}\cdot(\mathbf{r} - \mathbf{R}_j)} d\mathbf{r} \quad (1.140) \end{aligned}$$

It is easy to see that the integration over \mathbf{r}

$$\begin{aligned}
A &\equiv \int u_{nk}^*(\mathbf{r} - \mathbf{R}_i) u_{nq}(\mathbf{r} - \mathbf{R}_j) e^{-ik \cdot (\mathbf{r} - \mathbf{R}_i) + iq \cdot (\mathbf{r} - \mathbf{R}_j)} d\mathbf{r} \\
&= e^{i(k \cdot \mathbf{R}_i - q \cdot \mathbf{R}_j)} \int u_{nk}^*(\mathbf{r} - \mathbf{R}_i) u_{nq}(\mathbf{r} - \mathbf{R}_j) e^{-i(k-q) \cdot \mathbf{r}} d\mathbf{r} \quad (1.141)
\end{aligned}$$

Because $u_{nk}(\mathbf{r} - \mathbf{R}_i)$ and $u_{nq}(\mathbf{r} - \mathbf{R}_j)$ are periodic with respect to the unit cell, so becomes $u_{nk}^*(\mathbf{r} - \mathbf{R}_i) u_{nq}(\mathbf{r} - \mathbf{R}_j)$. The above integration can therefore be performed by integrating $u_{nk}^*(\mathbf{r} - \mathbf{R}_i) u_{nq}(\mathbf{r} - \mathbf{R}_j)$ over one unit cell, then summing over all unit cells by which we obtain that A is nonzero only when $\mathbf{q} = \mathbf{k}$. The summation over \mathbf{k} in Eq. (1.140) is nonzero when $\mathbf{R}_i = \mathbf{R}_j$. In other words, $\langle a_n(\mathbf{r} - \mathbf{R}_i) | a_n(\mathbf{r} - \mathbf{R}_j) \rangle = \delta_{ij}$, i.e., Wannier function $a_n(\mathbf{r} - \mathbf{R}_i)$ is localized around the i th unit cell.

We now express the perturbed wave function of Eq. (1.137) using Wannier functions in the form of

$$\Psi = \sum_{i,n} \psi_n(\mathbf{R}_i) a_n(\mathbf{r} - \mathbf{R}_i) \quad (1.142)$$

In order to obtain the equations for $\psi_n(\mathbf{R}_i)$ we substitute the wave function given by Eq. (1.142) into Eq. (1.137), multiply by $a_n^*(\mathbf{r} - \mathbf{R}_j)$ and integrate over the volume Ω . In this way we obtain the following equations

$$\sum_{i,n} [h_{n'}(\mathbf{R}_j - \mathbf{R}_i) \delta_{n'n} + V_{n'nji}] \psi_n(\mathbf{R}_i) = E \psi_{n'}(\mathbf{R}_j) \quad (1.143)$$

where

$$h_{n'}(\mathbf{R}_j - \mathbf{R}_i) \delta_{n'n} = \langle a_{n'}(\mathbf{r} - \mathbf{R}_j) | H_0 | a_n(\mathbf{r} - \mathbf{R}_i) \rangle$$

because of the orthogonal property of wave functions from different bands and periodic property of the operator H_0 ,

$$V_{n'nji} \equiv \langle a_{n'}(\mathbf{r} - \mathbf{R}_j) | V(\mathbf{r}) | a_n(\mathbf{r} - \mathbf{R}_i) \rangle \quad (1.144)$$

$V_{n'nji}$ will be appreciable when there is considerable overlap between $a_{n'}(\mathbf{r} - \mathbf{R}_j)$ and $a_n(\mathbf{r} - \mathbf{R}_i)$. This will be so when $|\mathbf{R}_j - \mathbf{R}_i|$ is small. If $V(\mathbf{r})$ is a slowly varying function we may regard it as constant over the region where the product $\langle a_{n'}(\mathbf{r} - \mathbf{R}_j) | a_n(\mathbf{r} - \mathbf{R}_i) \rangle$ is appreciable and take it outside the integral for $V_{n'nji}$, which is equal to zero when $i \neq j$. Thus,

$$V_{n'nji} \approx V_{n'n}(\mathbf{R}_j) \delta_{ji} \quad (1.145)$$

$V(\mathbf{r})$ may not be a slowly varying function in situations like interfaces between two different materials when regarding individual atomic potentials. But as a consequence of the surface kinetic processes, a certain degree of intermixing of the two materials is inevitable. Such intermixing is clearly indicated in the GaAs/AlGaAs superlattice by the fine structure emission lines in the photoluminescence spectra corresponding well to that calculated for one monolayer change in the well width.

Therefore, it takes several atomic layers for the averaged atomic potential to change from one material over to the other. Since the Wannier function is quite localized, especially when the electronic states under investigation are close to the bandedges, the approximation of $V_{n'nji} \approx V_{n'n}(\mathbf{R}_j)\delta_{ji}$ is reasonable.

Equation (1.143) is thus reduced to the form

$$\sum_i h_{n'}(\mathbf{R}_i)\psi_{n'}(\mathbf{R}_j - \mathbf{R}_i) + \sum_n V_{n'n}(\mathbf{R}_j)\psi_n(\mathbf{R}_j) = E\psi_{n'}(\mathbf{R}_j) \quad (1.146)$$

On the other hand, writing the energy $E_{n'}(\mathbf{k})$ as

$$E_{n'}(\mathbf{k}) = \langle u_{n'}e^{i\mathbf{k}\cdot\mathbf{r}} | H_0 | u_{n'}e^{i\mathbf{k}\cdot\mathbf{r}} \rangle \quad (1.147)$$

and substituting the Bloch functions in terms of Wannier functions, we obtain

$$E_{n'}(\mathbf{k}) = \frac{1}{N} \sum_{ij} h_{n'}(\mathbf{R}_j - \mathbf{R}_i)e^{-i\mathbf{k}\cdot(\mathbf{R}_j - \mathbf{R}_i)} = \sum_i h_{n'}(\mathbf{R}_i)e^{-i\mathbf{k}\cdot\mathbf{R}_i} \quad (1.148)$$

Replacing \mathbf{k} by $-i\nabla$ then applying to $\psi_{n'}(\mathbf{R}_j)$

$$E_{n'}(-i\nabla)\psi_{n'}(\mathbf{R}_j) = \sum_i h_{n'}(\mathbf{R}_i)e^{-\nabla\cdot\mathbf{R}_i}\psi_{n'}(\mathbf{R}_j) = \sum_i h_{n'}(\mathbf{R}_i)\psi_{n'}(\mathbf{R}_j - \mathbf{R}_i) \quad (1.149)$$

where we have used the relation

$$e^{-\nabla\cdot\mathbf{R}_i}\psi_{n'}(\mathbf{r}) = \psi_{n'}(\mathbf{r} - \mathbf{R}_i) \quad (1.150)$$

Equation (1.146) can thus be re-formatted as

$$E_{n'}(-i\nabla)\psi_{n'}(\mathbf{R}_j) + \sum_n V_{n'n}(\mathbf{R}_j)\psi_n(\mathbf{R}_j) = E\psi_{n'}(\mathbf{R}_j) \quad (1.151)$$

By approximating \mathbf{R}_i by \mathbf{r} , the equation

$$[E_{n'}(-i\nabla) - E]\psi_{n'}(\mathbf{r}) + \sum_n V_{n'n}(\mathbf{r})\psi_n(\mathbf{r}) = 0 \quad (1.152)$$

is therefore seen to be equivalent to Eq. (1.146). Equation (1.152) is the differential equation for envelope function ψ_n .

To understand $\psi_n(\mathbf{r})$ more clearly, let us write the total wave function given by Eq. (1.142) as

$$\Psi = \sum_n \Psi_n$$

then expand Ψ_n in terms of Bloch functions, i.e., insert Eq. (1.139) into Eq. (1.142),

$$\Psi_n = \frac{1}{N} \sum_i \sum_k \psi_n(\mathbf{R}_i)e^{-i\mathbf{k}\cdot\mathbf{R}_i}u_{nk}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}} = \frac{1}{\sqrt{N}} \sum_k G_v(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{r}}u_{nk}(\mathbf{r}) \quad (1.153)$$

where

$$G_n(\mathbf{k}) = \frac{1}{\sqrt{N}} \sum_i \psi_n(\mathbf{R}_i) e^{-i\mathbf{k} \cdot \mathbf{R}_i}$$

Replacing \mathbf{R}_i by \mathbf{r} ,

$$G_n(\mathbf{k}) = \frac{1}{\sqrt{N}} \int \psi_n(\mathbf{r}) e^{-i\mathbf{k} \cdot \mathbf{r}} d\mathbf{r}$$

we see that $G_n(\mathbf{k})$ is the Fourier transform of $\psi_n(\mathbf{r})$.

Assuming that $u_{n\mathbf{k}}(\mathbf{r})$ does not vary much with \mathbf{k} , Eq. (1.153) becomes

$$\Psi_n \approx \frac{1}{\sqrt{N}} u_n(\mathbf{r}) \sum_{\mathbf{k}} G_n(\mathbf{k}) e^{i\mathbf{k} \cdot \mathbf{r}} = u_n(\mathbf{r}) \psi_n(\mathbf{r}) \quad (1.154)$$

where u_n is the Bloch function at the optimum point of interest. The function $\psi_n(\mathbf{r})$ therefore acts as a modulation function for the function u_n which may vary rapidly in a single cell, and is usually called the envelope function.

We now apply the envelope function theory to study the conduction band of a heterostructure composed of material A and B which are described by energy dispersion relations of $E_n(\mathbf{k})$ in the form of

$$E_i(\mathbf{k}) = E_{i0} + \frac{\hbar^2 k_{xy}^2}{2m_{i,t}^*} + \frac{\hbar^2 k_z^2}{2m_{i,\ell}^*} \quad (1.155)$$

where $m_{i,t}^*$ and $m_{i,\ell}^*$ are transverse and longitudinal effective masses in the xy plane and along the z axis, respectively. E_{i0} is the conduction band edge of material i ($i = A$ and B). Replacing \mathbf{k} by $-i\nabla$, we immediately come up to the well-known effective mass approximation from Eq. (1.152)

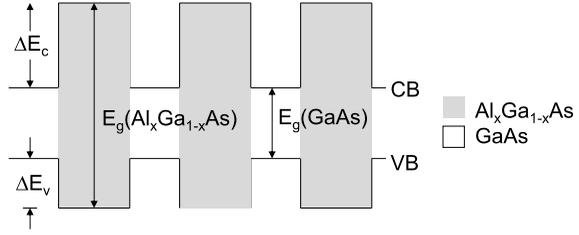
$$\left[-\frac{\hbar^2}{2m_{i,t}^*} \left(\frac{d^2}{dx^2} + \frac{d^2}{dy^2} \right) - \frac{\hbar^2}{2m_{i,\ell}^*} \frac{d^2}{dz^2} + E_{i0}(\mathbf{r}) \right] \psi(\mathbf{r}) = E \psi(\mathbf{r}) \quad (1.156)$$

where we have neglected the coupling of the conduction band with other bands (most importantly the valence bands) in the above equation, i.e., $V_{n'n} = 0$ in Eq. (1.152), which is valid for many applications.

More explicitly let us concentrate on a one-dimensional heterostructure AB grown along the z direction with the heterointerface at $z = 0$. Assume $z < 0$ is material A and $z \geq 0$ material B . Moreover, $E_{A0} < E_{B0}$, i.e., material A is the quantum well and B is the barrier, and take E_{A0} as the energy reference. For energy $0 < E \leq \Delta E_c = E_{B0} - E_{A0}$,

$$\left[-\frac{\hbar^2}{2m_{A,t}^*} \left(\frac{d^2}{dx^2} + \frac{d^2}{dy^2} \right) - \frac{\hbar^2}{2m_{A,\ell}^*} \frac{d^2}{dz^2} + V(\mathbf{r}) \right] \psi(\mathbf{r}) = E \psi(\mathbf{r}) \quad (1.157)$$

Fig. 1.22 Schematic illustration of one-dimensional square-well barrier model for conduction-band (CB) electrons and valence-band (VB) holes in GaAs/ $\text{Al}_x\text{Ga}_{1-x}\text{As}$ multiple quantum well system



where $V(\mathbf{r}) = \Delta E_c$ for $z \geq 0$ and $V(\mathbf{r}) = 0$ otherwise. Since $V(\mathbf{r}) = V(z)$ is xy -independent,

$$\begin{aligned}\psi(\mathbf{r}) &= e^{i\boldsymbol{\rho}\cdot\mathbf{k}_{xy}}\psi(z) \\ E &= E_z + \frac{\hbar^2 k_{xy}^2}{2m_{A,t}^*} \\ \left[-\frac{\hbar^2}{2m_{A,\ell}^*} \frac{d^2}{dz^2} + V(z) \right] \psi(z) &= E_z \psi(z)\end{aligned}\quad (1.158)$$

where $\boldsymbol{\rho} = (x, y)$. Note that the total wave function is $\psi(\mathbf{r})u(\mathbf{r})$, where $u(\mathbf{r})$ is the Bloch function.

In Eq. (1.157), for $0 < E \leq \Delta E_c$ the effective mass of composite material A , $m_{A,\ell}^*$ applies across the whole system, including the spatial region of material B , since the effective mass of material B , $m_{B,\ell}^*$ is only defined for $E \geq \Delta E_c$.

For $E \geq \Delta E_c$, the effective mass is spatial-position dependent, $m^*(z) = m_{B,\ell}^*$ for $z \geq 0$ and $m^*(z) = m_{A,\ell}^*$. Extended discussions about the effective mass approximation were reported [77–83] when the effective mass m^* is spatial-position dependent. It is generally accepted that

$$\left\{ -\frac{\hbar^2}{2} \frac{d}{dz} \left[\frac{1}{m^*(z)} \frac{d}{dz} \right] + V(z) \right\} \psi(z) = E \psi(z) \quad (1.159)$$

is the most reasonable choice, where

$$\psi(z) \quad \text{and} \quad \frac{1}{m^*(z)} \frac{d\psi(z)}{dz} \quad (1.160)$$

are continuous across the system.

Conventionally the III–V heterostructure is referred as type-I as the quantum wells for both conduction-band electrons and valence-band holes locate at the same spatial regions, as schematically illustrated in Fig. 1.22. A band-offset coefficient is generally defined as

$$\frac{\Delta E_c}{\Delta E_c + \Delta E_v} = \frac{\Delta E_c}{\Delta E_g} \quad (1.161)$$

where

$$\Delta E_g = E_g(\text{Al}_x\text{Ga}_{1-x}\text{As}) - E_g(\text{GaAs})$$

is the difference between energy bandgaps of $\text{Al}_x\text{Ga}_{1-x}\text{As}$ and GaAs. For III–V heterostructures, this coefficient is generally accepted as 0.65. The valence band offset between GaAs and $\text{Al}_x\text{Ga}_{1-x}\text{As}$ is therefore $0.35\Delta E_g$. This expression is normally valid for direct band-gap materials. It becomes complicated when indirect-band-gap materials are involved.

We have presented the concept of valence band offset when discussing the $\mathbf{k} \cdot \mathbf{p}$ model in Sect. 1.6. Almost identical results are obtained using the two different band offset schemes, as can be expected for III–V materials [47]. The energy values in the following are all measured with respect to the top of the valence band of bulk GaAs.

For GaAs/ $\text{Al}_x\text{Ga}_{1-x}\text{As}$ hetero materials, we have to consider the Γ -X-mixing effect when the Al mole fraction larger than 0.35 [84]. First of all, a valence band offset of $0.5x$ eV was obtained by the sp^3s^* tight-binding model which fits very well with experiments [85]. The band offset between Γ bands of GaAs and $\text{Al}_x\text{Ga}_{1-x}\text{As}$ is

$$1.247x + 1.147(x - 0.45)^2 - 0.5x \text{ eV}$$

when $x > 0.45$, since the Γ bandgap of $\text{Al}_x\text{Ga}_{1-x}\text{As}$ is [86]

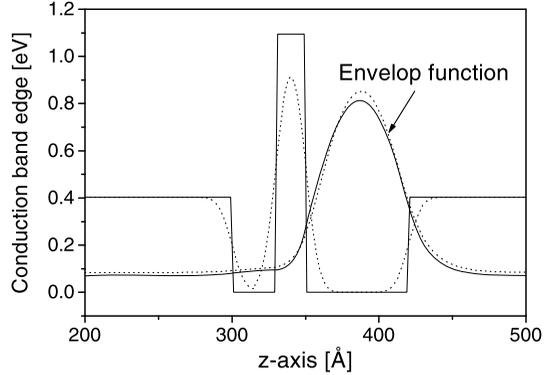
$$1.424 + 1.247x + 1.147(x - 0.45)^2 \text{ eV}$$

A good example of the one-dimensional heterostructure material is a GaAs/AlGaAs asymmetrically-coupled quantum well system. Grown on the [001] oriented semi-insulated GaAs substrate, the sample consists of a 500-nm-GaAs buffer layer, a 50-nm- $\text{Al}_{0.53}\text{Ga}_{0.47}\text{As}$ barrier, a 3-nm-GaAs well, a 2-nm-AlAs barrier, a 7-nm-GaAs well, a 50-nm- $\text{Al}_{0.53}\text{Ga}_{0.47}\text{As}$ barrier, and finally a 20-nm-GaAs cap layer.

Ion-implantations have been performed to enhance the inter-diffusion of Al atoms and consequently modify the quantum potential energy for various electronics and optoelectronics applications. After the ion-implantation, the quantum well is changed from a square well to an error-function-like well. Taking the center of the AlAs barrier (denoted as $\text{Al}_{x_2}\text{Ga}_{1-x_2}\text{As}$, $x_2 = 1.0$) as $z = 0$, the AlAs barrier width as $2h$, the two heterointerfaces between GaAs quantum wells and the AlAs barrier locate at $z = -h$ and $z = h$, $h = 10 \text{ \AA}$. The other two heterointerfaces between GaAs quantum wells and $\text{Al}_{x_1}\text{Ga}_{1-x_1}\text{As}$ barriers ($x_1 = 0.53$) locate then at $z_1 = -40 \text{ \AA}$ and $z_2 = 80 \text{ \AA}$. The degree of the Al diffusion across heterointerfaces is given by the diffusion equation of [87–90]

$$x(z) = \frac{x_1}{2} \left[2 - \operatorname{erf}\left(\frac{z - z_1}{2L}\right) + \operatorname{erf}\left(\frac{z - z_2}{2L}\right) \right] + \frac{x_2}{2} \left[\operatorname{erf}\left(\frac{z + h}{2L}\right) - \operatorname{erf}\left(\frac{z - h}{2L}\right) \right] \quad (1.162)$$

Fig. 1.23 Conduction bandedge and envelope function (vertically shifted up by the energy position of the ground sublevel) of a GaAs/AlGaAs asymmetrically-coupled quantum-well system. *Solid lines:* $L = 0.1 \text{ \AA}$; *dotted lines:* $L = 5.1 \text{ \AA}$



where L is the diffusion length, and erf the error function. Here the same diffusion length is assumed at the GaAs/AlGaAs and GaAs/AlAs heterointerfaces.

$$\frac{1}{m^*(z)} = \frac{x(z)}{m_{\text{AlAs}}^*} + \frac{1-x(z)}{m_{\text{GaAs}}^*} \quad (1.163)$$

in the virtual crystal approximation and

$$V(z) = 0.65 \times 1.247x(z) \text{ eV} \quad (1.164)$$

taking the conduction bandedge of GaAs as zero reference energy. Figure 1.23 shows the envelope function and energy position of the ground sublevel in the conduction band. By the position-dependent-effective-mass equation of Eq. (1.163), the ground sublevel in the conduction band is 59.1 meV when $L = 0.1 \text{ \AA}$, it becomes 75.2 meV when $L = 5.1 \text{ \AA}$. However, applying the effective mass of GaAs across the heterostructure, the ground sublevel becomes 70.0 and 83.5 meV, respectively, for $L = 0.1$ and 5.1 \AA , whereas the envelope function is only slightly modified.

1.9 Dimensionality and Density of States

In bulk material, the wave functions of electron states in the conduction band and hole states in the valence band can be expressed as:

$$\Psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}}(\mathbf{r}), \quad u_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r} + \mathbf{R}_n) \quad (1.165)$$

due to the periodic lattice potential $V(\mathbf{r}) = V(\mathbf{r} + \mathbf{R}_n)$, where \mathbf{k} is the wave vector and \mathbf{R} is the spatial position of the n th lattice site in the crystal. Note that the periodic Bloch function $u_{\mathbf{k}}(\mathbf{r})$ depends on \mathbf{k} .

The wave function is normalized in the spatial region Ω of the crystal. In the case of the bulk material, we have

$$\int_{\Omega} \Psi_{\mathbf{k}}^*(\mathbf{r}) \Psi_{\mathbf{k}}(\mathbf{r}) d\mathbf{r} = \int_{\Omega} u_{\mathbf{k}}^*(\mathbf{r}) u_{\mathbf{k}}(\mathbf{r}) d\mathbf{r} \quad (1.166)$$

by using the first equation of Eqs. (1.165). Since $u_{\mathbf{k}}(\mathbf{r})$ is periodic with respect to the unit cell of the lattice, the above integration becomes

$$\sum_n \int_{\text{cell}} |u_{\mathbf{k}}(\mathbf{r} + \mathbf{R}_n)|^2 d\mathbf{r} = \sum_n \int_{\text{cell}} |u_{\mathbf{k}}(\mathbf{r})|^2 d\mathbf{r} = N \quad (1.167)$$

by using the second equation of Eqs. (1.165), where \int_{cell} denotes the spatial integration within a unit cell denoted by \mathbf{R}_n , N is the number of unit cells in Ω . Here we assume that the Bloch function is normalized within the unit cell

$$\int_{\text{cell}} |u_{\mathbf{k}}(\mathbf{r})|^2 d\mathbf{r} = 1 \quad (1.168)$$

The above expression can also be expressed as $\langle u_{\mathbf{k}}(\mathbf{r}) | u_{\mathbf{k}}(\mathbf{r}) \rangle = 1$. By Eq. (1.168), we note that the physical unit of $u_{\mathbf{k}}(\mathbf{r})$ is [meter] $^{-3/2}$ in MKS. Note that within this book, we adopt MKS unit.

The normalized total wave function of one electron state in a bulk material is thus

$$\Psi_{\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{N}} e^{i\mathbf{k} \cdot \mathbf{r}} u_{\mathbf{k}}(\mathbf{r}) \quad (1.169)$$

as in many standard textbooks of solid state physics.

Now we consider the wave functions of electrons in nanostructures. The nanostructures are characterized by the quantum confinements of electrons. In practice, the size of the nanostructures can go down to tens of nanometers, in one (quantum wells, QWs), two (quantum wires, QWRs), or three dimensions (quantum dots, QDs), which actually is still very large in terms of the unit cells of common semiconductors with lattice constants of about 0.5 nm. The effective mass theory and the envelope function approach apply very well, despite the fact that it is not strictly derived both mathematically and physically. By the envelope function approach, the wave functions of the electrons can be approximately expressed as

$$\Psi_{i\mathbf{k}}(\mathbf{r}) = \psi_i(\mathbf{r}) u_{\mathbf{k}}(\mathbf{r}) \quad (1.170)$$

where $\psi_i(\mathbf{r})$ is called the envelope function, which is a slowly-varying function in space \mathbf{r} within one unit cell, while the spatial variation within the unit cell at atomic level is accounted for by the Bloch function $u_{\mathbf{k}}(\mathbf{r})$. Here i is the quantum number (can be a set of quantum numbers, see more discussions below) that denotes the state under investigation.

For a nanostructure which is confined along the z axis and extended in the xy plan, such as QWs, the envelope function is in the form of

$$\psi_i(z) e^{i(k_x x + k_y y)} \quad (1.171)$$

where $\psi_i(z)$ represents the quantum confinement along the z direction with a quantum index i , and the extension in the xy plane is represented by wave numbers k_x and k_y . The first expression of Eqs. (1.165) is easily retrieved when the quantum confinement along the z direction is lifted up so that $\psi_{k_z}(z) = e^{ik_z z}$.

In general, $\psi_i(\mathbf{r})$ is normalized within the spatial extension Ω of the electron in the nanostructure (equivalently, the spatial extension of the quantum confinement) by the envelope-function Schrödinger equation, i.e.,

$$\int_{\Omega} |\psi_i(\mathbf{r})|^2 d\mathbf{r} = 1 \quad (1.172)$$

Under the circumstance, $\psi_i(\mathbf{r})$ has the same physical unit as $u_{\mathbf{k}}(\mathbf{r})$, i.e., $[\text{meter}]^{-3/2}$, while $e^{i\mathbf{k}\cdot\mathbf{r}}$ in Eqs. (1.165) is unit-less.

For the sake of continuous discussion and presentation, we adopt the normalization of the Bloch function within one unit cell, i.e., Eq. (1.168), as well as the normalization of the envelope function in the spatial extension Ω of the quantum confinement, i.e., Eq. (1.172). We then need to find the normalization factor of the total wave function of Eq. (1.170) by calculating the following integration

$$\int_{\Omega} |\Psi_{i\mathbf{k}}(\mathbf{r})|^2 d\mathbf{r} = \int_{\Omega} |\psi_i(\mathbf{r})|^2 |u_{\mathbf{k}}(\mathbf{r})|^2 d\mathbf{r} \quad (1.173)$$

As mentioned before, $\psi_i(\mathbf{r})$ varies slowly in space so that we can use its value at \mathbf{R}_n as an average value for the whole unit cell centered at \mathbf{R}_n . Thus, the above integration becomes

$$\sum_n |\psi_i(\mathbf{R}_n)|^2 \int_{\text{cell}} |u_{\mathbf{k}}(\mathbf{r})|^2 d\mathbf{r} = \sum_n |\psi_i(\mathbf{R}_n)|^2 \quad (1.174)$$

and the last equality comes from the normalization of Bloch function, i.e., Eq. (1.168). Here $\mathbf{R}_n \in \Omega$. Let Ω_{cell} be the volume of the unit cell,

$$\sum_n |\psi_i(\mathbf{R}_n)|^2 = \frac{1}{\Omega_{\text{cell}}} \sum_n |\psi_i(\mathbf{R}_n)|^2 \Omega_{\text{cell}} = \frac{1}{\Omega_{\text{cell}}} \int_{\Omega} |\psi_i(\mathbf{r})|^2 d\mathbf{r} = \frac{1}{\Omega_{\text{cell}}} \quad (1.175)$$

The last equality comes from Eq. (1.172).

Thus, the normalized total wave function is in the form of

$$\Psi_{i\mathbf{k}}(\mathbf{r}) = \sqrt{\Omega_{\text{cell}}} \psi_i(\mathbf{r}) u_{\mathbf{k}}(\mathbf{r}) \quad (1.176)$$

where

$$\int_{\text{cell}} |u_{\mathbf{k}}(\mathbf{r})|^2 d\mathbf{r} = 1 \quad \text{and} \quad \int_{\Omega} |\psi_i(\mathbf{r})|^2 d\mathbf{r} = 1 \quad (1.177)$$

Let us consider the following four cases.

Bulk material For bulk material, the envelope function $\psi_i(\mathbf{r})$ is in the form of a plane wave, i.e., $\psi_{\mathbf{k}}(\mathbf{r}) = C e^{i\mathbf{k}\cdot\mathbf{r}}$, where C is the normalization factor

$$\int_{\Omega} |\psi_{\mathbf{k}}(\mathbf{r})|^2 d\mathbf{r} = \int_{\Omega} |C|^2 d\mathbf{r} = |C|^2 \int_{\Omega} d\mathbf{r} = |C|^2 \Omega \quad (1.178)$$

so that $C = 1/\sqrt{\Omega}$. Thus, the normalized envelope function in a bulk material is

$$\psi_{\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{\Omega}} e^{i\mathbf{k}\cdot\mathbf{r}} \quad (1.179)$$

By inserting this into Eq. (1.176), we retrieve Eq. (1.169), where $\Omega/\Omega_{\text{cell}} = N$.

Quantum well For a QW which is confined along the z direction by an effective confinement length L_{QW} , it is extended in the xy plane so that the envelope function can be expressed as

$$C\psi_i(z)e^{i(k_x x + k_y y)} \quad (1.180)$$

Here C is the normalization factor, and

$$\int_{L_{\text{QW}}} |\psi_i(z)|^2 dz = 1 \quad (1.181)$$

see the second equation in Eqs. (1.177). The normalization of Eq. (1.180) is

$$|C|^2 \int_{L_{\text{QW}}} |\psi_i(z)|^2 dz \int_{xy} dx dy = |C|^2 A_{\text{QW}} \quad (1.182)$$

where A_{QW} is the extension area of the QW in the xy plane. Thus, $C = 1/\sqrt{A_{\text{QW}}}$ and the normalized envelope function in a QW is

$$\psi_{i,\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{A_{\text{QW}}}} \psi_i(z) e^{i(k_x x + k_y y)} \quad (1.183)$$

Quantum wire Similarly we can discuss that the normalized envelope function in a QWR, with a confinement area A_{QWR} in the xy plane and an extension length L_{QWR} in the z direction, is

$$\int_{A_{\text{QWR}}} |\psi_i(x, y)|^2 dx dy = 1, \quad \psi_{i,k_z}(\mathbf{r}) = \frac{1}{\sqrt{L_{\text{QWR}}}} \psi_i(x, y) e^{ik_z z} \quad (1.184)$$

Quantum dot For a QD confined three dimensionally within an effective volume Ω_{QD}

$$\int_{\Omega_{\text{QD}}} |\psi_i(\mathbf{r})|^2 d\mathbf{r} = 1 \quad (1.185)$$

Now we discuss the density of states which is the number of available electronic states per unit volume per unit energy at energy E . Quantum mechanically, each electronic state can be occupied by two electrons (spin up and down). Here we exclude the magnetic field. In a three-dimensional (3D) bulk material, each electronic state is represented by its wave vector \mathbf{k} (Bloch theorem):

$$\Psi_{n\mathbf{k}}^{3\text{D}}(\mathbf{r}) = \frac{1}{\sqrt{N}} u_{n\mathbf{k}}(\mathbf{r}) e^{i\mathbf{k}\cdot\mathbf{r}} \quad (1.186)$$

where n is the band index and $u_{n\mathbf{k}}(\mathbf{r})$ is the Bloch function.

We consider a 3D lattice which extends along the x axis by a length of L_x . It also extends L_y and L_z along the y and z directions, respectively. We adopt the following periodic boundary of the lattice,

$$\Psi_{nk}^{3D}(x, y, z) = \Psi_{nk}^{3D}(x + L_x, y, z) = \Psi_{nk}^{3D}(x, y + L_y, z) = \Psi_{nk}^{3D}(x, y, z + L_z) \quad (1.187)$$

which requires

$$e^{ik_i L_i} = 1$$

for $i = x, y, z$. We have the following expression for the wave vector k_i

$$k_i = \frac{2m_i \pi}{L_i} \quad (1.188)$$

where m_i is an integer. When L_i is large, the spacing between adjacent allowed \mathbf{k} values, i.e., $k_{i, m_i+1} - k_{i, m_i} = 2\pi/L_i$ is very small so that we are able to discuss the volume in the \mathbf{k} space that each electronic state occupies

$$\frac{(2\pi)^3}{L_x L_y L_z} = \frac{(2\pi)^3}{\Omega} \quad (1.189)$$

where $\Omega = L_x L_y L_z$ is the volume of the crystal. The total number of electronic states within $d\mathbf{k}$ in the \mathbf{k} space is

$$\frac{2d\mathbf{k}}{(2\pi)^3} = \frac{2\Omega d\mathbf{k}}{(2\pi)^3} \quad (1.190)$$

and the density of states per unit volume is thus given by

$$N_3(E)dE = \int_E^{E+dE} \frac{2d\mathbf{k}}{(2\pi)^3} \quad (1.191)$$

for a three-dimensionally-extended system.

For a spherical-parabolic conduction band described by

$$E = E_c + \frac{\hbar^2 k^2}{2m^*} \quad (1.192)$$

where E_c is the conduction band edge and m^* is the effective mass of the conduction-band electron (we concentrate on the conduction band, whereas the extension to the valence band is straightforward),

$$\frac{2d\mathbf{k}}{(2\pi)^3} = \frac{2 \times 4\pi k^2 dk}{8\pi^3} \quad (1.193)$$

Since $k^2 = 2m^*(E - E_c)/\hbar^2$ by Eq. (1.192), the three-dimensional density of states $N_3(E)$ is

$$N_3(E) = \frac{1}{2\pi^2} \left(\frac{2m^*}{\hbar^2} \right)^{3/2} \sqrt{E - E_c} \theta(E - E_c) \quad (1.194)$$

$\theta(E)$ is the step function, i.e., $\theta(E) = 1$ if $E \geq 0$, $\theta(E) = 0$ when $E < 0$.

When the system becomes confined along, say the z axis, such as the one-dimensional GaAs/AlGaAs quantum well in Fig. 1.23, the translational symmetry in this direction is broken so that the corresponding wave vector k_z is replaced by the index i of discrete energy levels in this direction. The sublevel i is now described by

$$\begin{aligned} \psi_{ik}^{2D}(\boldsymbol{\rho}, z) &= \psi_i(z) e^{i\mathbf{k}\cdot\boldsymbol{\rho}} \\ E_i(\mathbf{k}) &= E_c + E_i + \frac{\hbar^2 k^2}{2m^*} \end{aligned} \quad (1.195)$$

where \mathbf{k} and $\boldsymbol{\rho}$ are wave vector and spatial coordinate in the xy plane. To make the equations simple, here we have approximated the energy band structure of the semiconductor as spherical and parabolic. The approach is valid when the energies of states under discussion are close to the bandedges.

We assume that the system is extended by L_x and L_y in the xy plane. It is easy to obtain that the two-dimensional density of states $N_2(E)$ is

$$N_2(E) = \frac{m^*}{\pi \hbar^2 L_z} \theta(E - E_c) \quad (1.196)$$

where L_z is the effective width of the system in the quantum confinement direction (it is reminded that the density of states is defined as per unit volume).

When the dimensionality of the system is further decreased in the y -direction, we come up with a one dimensional system or a quantum wire,

$$N_1(E) = \frac{1}{2\pi S_{yz}} \sqrt{\frac{2m^*}{\hbar^2}} \frac{1}{\sqrt{E - E_c}} \theta(E - E_c) \quad (1.197)$$

where S_{yz} is the effective area of the system in the yz plane.

Finally we have the “zero” dimensional system, either an artificial quantum dot or a natural atom consisting of a set of discrete energy states E_i ,

$$N_0(E) = \frac{1}{\Omega_{\text{QD}}} \sum_i \delta(E - E_i) \quad (1.198)$$

where Ω_{QD} is the effective volume of the quantum dot. Including the relaxation processes, the δ function is replaced by

$$\delta(E - E_i) \rightarrow \frac{\Gamma}{(E - E_i)^2 + \Gamma^2} \quad (1.199)$$

where Γ is the relaxation energy.

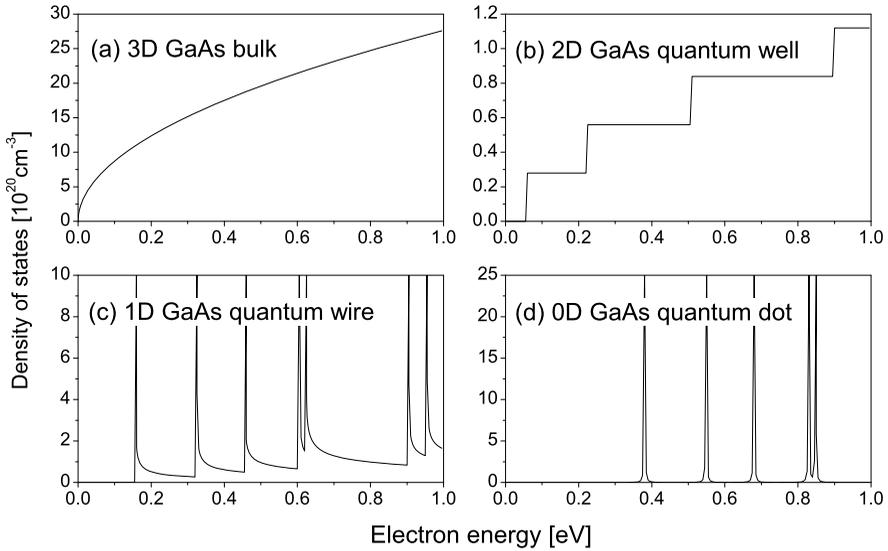


Fig. 1.24 Dimensionality of energy density of electron states in the conduction band (we have set $E_c = 0$). (a) 3D extended: GaAs bulk material. (b) 2D extended: GaAs quantum well with a well thickness 10 nm. (c) 1D extended: GaAs quantum wire with a rectangular cross section of $10 \times 7.5 \text{ nm}^2$. (d) 0D extended: GaAs quantum dot of $10 \times 7.5 \times 5.0 \text{ nm}^3$. $\Gamma = 1 \text{ meV}$

Figure 1.24 shows the energy densities of conduction-band states of GaAs materials embedded in a barrier material with infinitely high barrier. We note that as the dimensionality of the system changes, the energy dependence of the density of states also changes. For a spherical-parabolic band, we have a \sqrt{E} -dependence for a three dimensionally extended system ($E_c = 0$); for a two-dimensional system, the density of states is constant; and for a one dimensional system we have an $1/\sqrt{E}$ -dependence. And finally we have the δ -function form for a zero-dimensional system.

References

1. Dennard RH, Gaensslen FH, Yu HN, Rideout VL, Bassous E, LeBlanc AR (1974) Design of ion-implanted MOSFET's with very small physical dimensions. *IEEE J Solid-State Circuits* sc9:256–268
2. Sai-Halasz GA, Wordeman MR, Kern DP, Ganin E, Rishton S, Zicherman DS, Schmidt H, Polcari MR, Ng HY, Restle PJ, Chang THP, Dennard RH (1987) Design and experimental technology for 0.1- μm gate-length low-temperature operation FETs. *IEEE Electron Device Lett* 8:463–466
3. Hashimoto T, Sudoh Y, Kurino H, Narai A, Yokoyama S, Horiike Y, Koyanagi M (1992) 3 V operation of 70 nm gate length MOSFET with new double punchthrough stopper structure. In: Ext abs 1992, int conf solid state devices materials, Tokyo Japan, pp 490–492
4. Ono M, Saito M, Yoshitomi T, Fiegna C, Ohguro T, Iwai H (1995) A 40 nm gate length n -MOSFET. *IEEE Trans Electron Devices* 42:1822–1830

5. Kawaura H, Sakamoto T, Baba T, Ochiai Y, Fujita J, Matsui S, Sone J (1998) Transistor operation of 30 nm gate length EJ-MOSFETs. *IEEE Electron Device Lett* 19:74–76
6. Langen W, Hardtdegen H, Luth H, Kordos P (1995) Selectively grown vertical sub-100 nm dual-gate GaAs FETs. In: *Compound semiconductors 1994, proceedings of 21st international symposium*. IOP Publishing, Bristol, pp 183–188
7. Behammer D, Vescan L, Loo R, Moers J, Mück A, Lüth H, Grabolla T (1996) Selectively grown vertical Si-p MOS transistor with short channel lengths 130 nm. *Electron Lett* 32:406–407
8. Langen W, Vescan L, Loo R, Luth H, Kordos P (1996) Vertical 100 nm Si-p channel JFET grown by selective epitaxy. *Appl Surf Sci* 102:252–254
9. Klaes D, Moers J, Tonnesmann A, Wickenhauser S, Vescan L, Marso M, Grabolla T, Grimm M, Luth H (1998) Selectively grown vertical Si MOS transistor with reduced overlap capacitances. *Thin Solid Films* 336:306–308
10. Moers J, Klaes D, Tonnesmann A, Vescan L, Wickenhauser S, Marso M, Kordos P, Luth H (1999) 19 GHz vertical Si p-channel MOSFET. *Electron Lett* 35:239–240
11. Moers J, Klaes D, Tonnesmann A, Vescan V, Wickenhauser S, Grabolla T, Marso M, Kordos P, Luth H (1999) Vertical p-MOSFETs with gate oxide deposition before selective epitaxial growth. *Solid-State Electron* 43:529–535
12. Murotani T, Naritake I, Matano T, Ohtsuki T, Kasai N, Koga H, Koyama K, Nakajima K, Yamaguchi H, Watanabe H, Okuda T (1997) A 4-level storage 4 Gb DRAM. In: *1997 IEEE int solid-state circuits conf dig*, San Francisco, CA, p 74
13. Yasuda H (1997) Multimedia impact on devices in the 21st century. In: *1997 IEEE int solid-state circuits conf dig*, San Francisco, CA, p 28
14. Mitin VV, Gribnikov ZS, Korshak AN (1998) Quantum heterostructures for micro- and optoelectronics. In: *Physics of semiconductor devices*, vol 1. Narosa Publishing House, Delhi, pp 134–141; 2 vol, xxviii+1322 pp
15. Appenzeller J, Schroer CH, Schapers TH, van der Hart A, Forster A, Lengeler B, Luth H (1996) Electron interference in a T-shaped quantum transistor based on Schottky-gate technology. *Phys Rev B* 53:9959–9963
16. Smith CG, Pepper M, Newbury R, Ahmed H, Hasko DG, Peacock DC, Frost JEP, Ritchie DA, Jones GAC, Hill G (1989) One-dimensional quantised ballistics resistors in parallel configuration. *J Phys Condens Matter* 1:6763–6770
17. Avishai Y, Kaveh M, Shatz S, Band YB (1989) Ballistic electronic conductance of two parallel channels. *J Phys Condens Matter* 1:6907–6912
18. Castaño E, Kirczenow G (1990) Theory of the conductance of parallel ballistic constrictions. *Phys Rev B* 41:5055–5060
19. Baie X, Colinge JP, Bayot V, Grivei E (1995) Quantum-wire effects in thin and narrow SOI MOSFETs. In: *IEEE international SOI conference proceedings*. IEEE Press, New York, pp 66–67; xiv+183 pp (cat no 95CH35763)
20. Colinge JP, Baie X, Bayot V, Grivei E (1996) A silicon-on-insulator quantum wire. *Solid-State Electron* 39:49–51
21. Kirczenow G, Sachrajda AS, Feng Y, Taylor RP, Henning L, Wang J, Zawadzki P, Coleridge PT (1994) Artificial impurities in quantum wires: from classical to quantum behaviour. *Phys Rev Lett* 72:2069–2072
22. Kirczenow G (1994) Scattering models of conduction around an antidot in a magnetic field. *Phys Rev B* 50:1649–1655 and references therein
23. Sun JP, Haddad GI, Mazumder P, Shulman JN (1998) Resonant tunneling diodes: models and properties. *Proc IEEE* 86:641–661
24. Mazumder P, Kulkarni S, Bhattacharya M, Sun JP, Haddad GI (1998) Digital circuit applications of resonant tunneling diodes. *Proc IEEE* 86:664–686
25. Andersson JY, Alverbro J, Borglind J, Helander P, Martin H, Ostlund M (1997) 320 × 240 pixels quantum well infrared photodetector (QWIP) array for thermal imaging: fabrication and evaluation. *Proc SPIE* 3061:740–748

26. Brill B, Sarusi G (1997) QWIP research and development of 320*256 QWIP arrays in EL-OP. Proc SPIE 3061:781–788
27. Allee DA, Broers AN, Pease RFW (1991) Limits of nano-gate fabrication. Proc IEEE 79:1093–1105
28. Ahmed H (1991) Nanostructure fabrication. Proc IEEE 79:1140–1148
29. Brittain S, Paul K, Zhao X-M, Whitesides G (1998) Soft lithography and microfabrication. Phys World 11:31–36
30. Luth H (1996) Tunneling in semiconductor nanostructures: physics and devices. Acta Phys Pol A 90:667–679
31. Luth H (1998) Semiconductor nanostructures for device applications. In: Physics of semiconductor devices, vol 1. Narosa Publishing House, Delhi, pp 18–25; 2 vol, xxviii+1322
32. Luth H (1998) Semiconductor nanostructures: a new impact on electronics. Appl Surf Sci 130–132:855–865
33. Callaway J (1974) Quantum theory of the solid state. Academic Press, New York, p 233
34. Vogl P, Hjalmarson HP, Dow JD (1983) A semi-empirical tight-binding theory of the electronic structure of semiconductors. J Phys Chem Solids 44:365–378
35. Harrison WA (1980) Electronic structure and the properties of solids. Freeman, San Francisco
36. Bean JC (1992) Silicon-based semiconductor heterostructures: column IV bandgap engineering. Proc IEEE 80:571
37. Soref RA (1993) Silicon-based optoelectronics. Proc IEEE 81:1687
38. Geis MW, Gregory JA, Pate BB (1991) Capacitance-voltage measurements on metal-SiO₂-diamond structures fabricated with (100)- and (111)-oriented substrates. IEEE Trans Electron Devices 38:619
39. Xu W, Fu Y, Willander M, Shen SC (1994) Theory of normal-incidence absorption for the intersubband transition in *n*-type indirect-gap semiconductor quantum wells. Phys Rev B 49:13760–13766
40. Hensel JC, Hasegawa H, Nakayama M (1965) Cyclotron resonance in uniaxially stressed silicon. II. Nature of the covalent bond. Phys Rev 138:A225–A238
41. Dresselhaus G, Kip AF, Kittel C (1955) Cyclotron resonance of electrons and holes in silicon and germanium crystals. Phys Rev 98:368–384
42. Tiersten M (1961) Acoustic-mode scattering of holes. IBM J Res Dev 5:122–131
43. Bir GL, Pikus GE (1960) Theory of the deformation potential for semiconductors with a complex band structure. Sov Phys, Solid State 2:2039
44. Kane EO (1966) The *k* · *p* method. In: Willardson RK, Beer AC (eds) Semiconductors and semimetals. Physics of III–V compound, vol 1. Academic Press, New York, pp 75–100, Chap. 3
45. Enders P, Bärwolff A, Woerner M, Suisky D (1995) *k* · *p* theory of energy bands, wave functions, and optical selection rules in strained tetrahedral semiconductors. Phys Rev B 51:16695–16704
46. Stier O, Grundmann M, Bimberg D (1999) Electronic and optical properties of strained quantum dots modeled by 8-band *k* · *p* theory. Phys Rev B 59:5688–5701
47. Vurgaftman I, Meyer JR, Ram-Mohan LR (2001) Band parameters for III–V compound semiconductors and their alloys. J Appl Phys 89:5815–5875
48. Yu ET, McCaldin JO, McGill TC (1992) Band offsets in semiconductor heterojunctions. Solid State Phys 46:1–146
49. Madelung O (ed) (1991) Semiconductors group IV elements and III–V compounds. Springer, Berlin
50. Kent PRC, Hart GLW, Zunger A (2002) Biaxial strain-modified valence and conduction band offsets of zincblende GaN, GaP, GaAs, InN, InP, and InAs, and optical bowing of strained epitaxial InGaN alloys. Appl Phys Lett 81:4377–4379
51. Egorov AYu, Bernklau D, Borchert B, Illek S, Livshits D, Rucki A, Schuster M, Kaschner A, Hoffmann A, Dumitras Gh, Amann MC, Riechert H (2001) Growth of high quality InGaAsN heterostructures and their laser application. J Cryst Growth 227–228:545–552

52. Yang T, Nakajima S, Sakai S (1997) Tight-binding calculation of electronic structures of In-NAs ordered alloys. *Jpn J Appl Phys* 36:L320–L322
53. Muller DA, Sorsch T, Baumann FH, Evans-Lutterodt K, Timp G (1999) The electronic structure at the atomic scale of ultrathin gate oxides. *Nature* 399:758–761
54. Fu Y, Willander M (1997) Valence-band structure of heavily doped silicon. *Phys Lett A* 234:483–487
55. Van de Walle CG, Martin RM (1986) Theoretical calculations of heterojunction discontinuities in the Si/Ge system. *Phys Rev B* 34:5621–5634
56. Alferov ZI (2001) Nobel lecture: the double heterostructure concept and its applications in physics, electronics, and technology. *Rev Mod Phys* 73:767–782
57. Stranski IN, Krastanov LV (1939) *Abhandlungen der Mathematisch-Naturwissenschaftlichen Klasse. Akad Wiss Lit (Mainz)* 146:797
58. Moll N, Scheffler M, Pehlke E (1998) Influence of surface stress on the equilibrium shape of strained quantum dots. *Phys Rev B* 58:4566–4571
59. Shklyaeu OE, Beck MJ, Asta M, Miksis MJ, Voorhees PW (2005) Role of strain-dependent surface energies in Ge/Si(100) island formation. *Phys Rev Lett* 94:176102(4)
60. Wise SM, Lowengrub JS, Kim JS, Thornton K, Voorhees PW (2005) Quantum dot formation on a strain-patterned epitaxial thin film. *Appl Phys Lett* 87:133102(3)
61. Fu K, Fu Y (2008) Kinetic Monte Carlo study of metal organic chemical vapor deposition growth mechanism of GaSb quantum dots. *Appl Phys Lett* 93:101906
62. Birman JL (1958) Theory of the piezoelectric effect in the zincblende structure. *Phys Rev* 111:1510–1514
63. Musgrave MJP, Pople JA (1962) A general valence force field for diamond. *Proc R Soc Lond A* 268:474–484
64. Nusimovici MA, Birman JL (1967) Lattice dynamics of wurtzite: CdS. *Phys Rev* 156:925–938
65. Martin RM (1970) Elastic properties of ZnS structure semiconductors. *Phys Rev B* 1:4005–4011
66. Ramani R, Mani KK, Singh RP (1976) Valence force fields and the lattice dynamics of beryllium oxide. *Phys Rev B* 14:2659–2663
67. Saito T, Arakawa Y (1999) Atomic structure and phase stability of $\text{In}_x\text{Ga}_{1-x}\text{N}$ random alloys calculated using a valence-force-field method. *Phys Rev B* 60:1701–1706
68. Takayama T, Yuri M, Itoh K, Baba T, Harris JS Jr (2000) Theoretical analysis of unstable two-phase region and microscopic structure in wurtzite and zinc-blende InGaN using modified valence force field model. *J Appl Phys* 88:1104–1110
69. Nabetani Y, Matsumoto T, Sasikala G, Suemune I (2005) Theory of strain states in InAs quantum dots and dependence on their capping layers. *J Appl Phys* 98:63502(7)
70. Jiang H, Singh J (1997) Strain distribution and electronic spectra of InAs/GaAs self-assembled dots: an eight-band study. *Phys Rev B* 56:4696–4701
71. Stier O, Grundmann M, Bimberg D (1999) Electronic and optical properties of strained quantum dots modeled by 8-band $k \cdot p$ theory. *Phys Rev B* 59:5688–5701
72. Bernard JE, Zunger A (1991) Strain energy and stability of Si-Ge compounds, alloys, and superlattices. *Phys Rev B* 44:1663–1681
73. Ito T, Khor KE, Das Sarma S (1989) Empirical potential-based Si-Ge interatomic potential and its application to superlattice stability. *Phys Rev B* 40:9715–9722
74. Kroemer H, Polasko KJ, Wight SC (1980) On the (110) orientation as the preferred orientation for the molecular beam epitaxial growth of GaAs or Ge, GaP on Si, and similar zincblende-on-diamond systems. *Appl Phys Lett* 36:763–765
75. Casey HC Jr, Panish MB (1978) *Heterostructure lasers, Part B: materials and operating characteristics*. Academic Press, New York
76. Margaritondo G, Katnani AD, Stoffel NG, Daniels RR, Zhao T-X (1982) Nature of the band discontinuities at semiconductor heterojunction interfaces. *Solid State Commun* 43:163–166
77. White SR, Sham JL (1981) Electronic properties of flat-band semiconductor heterostructures. *Phys Rev Lett* 47:879–882

78. Ando T, Mori S (1982) Effective-mass theory of semiconductor heterojunctions and superlattices. *Surf Sci* 113:124–130
79. Zhu QG, Kroemer H (1983) Interface connection rules for effective-mass wave functions at an abrupt heterojunction between two different semiconductors. *Phys Rev B* 27:3519–3527
80. Kahen KB, Leburton JP (1986) Optical constants of GaAs-Al_xGa_{1-x}As superlattices and multiple quantum wells. *Phys Rev B* 33:5465–5472
81. Morrow RA (1987) Establishment of an effective-mass Hamiltonian for abrupt heterojunctions. *Phys Rev B* 35:8074–8079
82. Fu Y, Chao KA (1989) Subband structures of GaAs/Al_xGa_{1-x}As multiple quantum wells. *Phys Rev B* 40:8349–8356
83. Fu Y, Chao KA (1991) Band offset in GaAs/Al_xGa_{1-x}As multiple quantum wells calculated with the sp^3s^* tight-binding model. *Phys Rev B* 43:4119–4124
84. Fu Y, Willander M, Ivchenko EL, Kiselev AA (1993) Valley mixing in GaAs/AlAs multilayer structures in the effective mass method. *Phys Rev B* 47:13498–13507
85. Fu Y, Chao KA (1991) Band-offset in GaAs/AlGaAs multiple quantum wells calculated with the sp^3s^* tight-binding model. *Phys Rev B* 43:4119–4124
86. Adachi S (1985) GaAs, AlAs, and Al_xGa_{1-x}As: material parameters for use in research and device applications. *J Appl Phys* 58:R1–R29
87. Crank J (1956) *The mathematics of diffusion*. Clarendon, Oxford, pp 9–14
88. Redinbo GF, Craighead HG, Hong JM (1993) Proton implantation intermixing of GaAs/AlGaAs quantum well. *J Appl Phys* 74:3099–3102
89. Feng W, Chen F, Cheng WQ, Huang Q, Zhou JM (1997) Influence of growth conditions on Al-Ga interdiffusion in low-temperature grown AlGaAs/GaAs quantum wells. *Appl Phys Lett* 71:1676–1678
90. Li N, Li N, Lu W, Liu XQ, Yuan XZ, Li ZF, Dou HF, Shen SC, Fu Y, Willander M, Fu L, Tan HH, Jagadish C, Johnston MB, Gal M (1999) Proton implantation and rapid thermal annealing effects on GaAs/AlGaAs quantum well infrared photodetector. *Superlattices Microstruct* 26:317–324

Chapter 2

Electron Transport

Abstract We introduce basic quantum mechanical concepts of electron wave packet transport, the scattering theory about electron transition from one steady state to another that eventually form a net nonzero charge flow, i.e., current. Boltzmann equation is discussed about carrier transport in terms of drift and diffusion versus ballistic transport. Major scattering processes in semiconductors are introduced and discussed.

2.1 Quantum Mechanical Wave Transport

In semiclassical transport theory the electrons are treated as classical point-like particles during free flights between scattering events. The scattering processes are however treated quantum mechanically. During the free flights, two things must be consistent with each other:

1. A wavepacket representing the point-like electron with well-defined position and momentum;
2. The quantum uncertainty principle.

The semiclassical treatment of the electron transport driven by an external electric field E is normally referred to as the Drude drift model. Here it is assumed that the electron speed v is zero after a collision with a scattering center. The electron will be driven by the external electric field. By the classical Newton mechanics $F = m^*a$, where F is the force asserted on the electron, and a the acceleration, we have

$$-eE = m^* \frac{dv}{dt} \tag{2.1}$$

Let τ be the mean time between two successive collisions, we obtain the average speed v_{average} and thereafter the mobility μ are

$$v_{\text{average}} = \frac{-e\tau}{m^*} E, \quad \mu = \frac{e\tau}{m^*} \tag{2.2}$$

The mean free path $\ell = v_{\text{average}}\tau$. On the one hand, well-defined position and momentum indicate

$$\Delta p \ll p, \quad \Delta x \ll \ell \quad (2.3)$$

while the quantum uncertainty principle states

$$\hbar \approx \Delta p \Delta x \ll p\ell = \frac{p^2}{m^*}\tau \quad (2.4)$$

Since at thermal equilibrium,

$$\frac{p^2}{2m^*} \approx k_B T$$

where k_B is the Boltzmann constant and T the temperature, the validity of the semiclassical approach is insured when

$$\tau \gg \frac{\hbar}{k_B T} \quad (2.5)$$

Otherwise, quantum treatment of transport during the free flights is necessary where the wave nature of electron dynamics is to be fully accounted for.

Quantum mechanically, we denote the solution of a Schrödinger equation as a wave function $\Psi(\mathbf{r}, t)$, which in general is space \mathbf{r} and time t dependent

$$i\hbar \frac{\partial \Psi(\mathbf{r}, t)}{\partial t} = H(\mathbf{r}, t)\Psi(\mathbf{r}, t) \quad (2.6)$$

When the Hamiltonian $H(\mathbf{r}, t)$ in the above Schrödinger equation describes an electron in an electron system, $|\Psi(\mathbf{r}, t)|^2 d\mathbf{r}$ represents the probability to find the electron in a space volume $d\mathbf{r}$ centered at \mathbf{r} at time t .

In the previous chapter we have discussed the general features of electron eigen states in semiconductor, which are characterized as Bloch states denoted by wave functions Ψ_{nk} and energies $E_n(\mathbf{k})$. Here n is the energy band index and \mathbf{k} is the wave vector of the Bloch state. For the devices discussed in this book, the magnetic field is negligible so that the electron spin up and down states are degenerate. In other words, each Bloch state actually consists of two states, one spin up and one spin down.

Knowing the electron states in the semiconductor, we then fill these states by the available electrons. At zero temperature, starting from the electron state of the lowest energy we fill each electron state by two electrons (Pauli exclusion principle). For intrinsic bulk semiconductor, all electron states up to the top of the valence band are completely filled, and the electron states from the bottom of the conduction band are totally empty. Same situation happens for insulators. Actually the only difference between intrinsic semiconductors and insulators is in their energy bandgaps (energy separations between the valence band top and the conduction band bottom). Insulators have much larger energy bandgaps than semiconductors. The filling of the electron states in metals are rather different: the conduction bands of metals are

partially filled by electrons. The different occupation features of energy bands of semiconductors and metals are responsible for the different electric conductances of the materials. We will discuss it shortly.

At finite temperature T , the occupation of electron state with eigen value E_i (for Bloch state, $i = n\mathbf{k}$) is given by the Fermi Dirac distribution

$$f(E_i, E_f) = \frac{1}{e^{(E_i - E_f)/k_B T} + 1} \quad (2.7)$$

where E_f is called Fermi level which is determined by

$$N = \sum_i \frac{1}{e^{(E_i - E_f)/k_B T} + 1} \quad (2.8)$$

where N is the total number of electrons. Note that an energy state can contain either 0 or 1 electron. The mean number of electrons in the energy state is therefore numerically equal to the probability of its being occupied.

The spatial distribution of an electron occupying state denoted by $\Psi_i(\mathbf{r}, t)$ and E_i is $|\Psi_i(\mathbf{r}, t)|^2 f(E_i, E_f)$, and the spatial distribution $n(\mathbf{r}, t)$ of the total electrons is

$$n(\mathbf{r}, t) = \sum_i |\Psi_i(\mathbf{r}, t)|^2 f(E_i, E_f) \quad (2.9)$$

Note that in the above equation, the time-dependence is explicit. At steady state, $\Psi_i(\mathbf{r}, t) = e^{-E_i t/\hbar} \Psi_i(\mathbf{r})$ and the carrier distribution n is time-independent. We will return to this issue from time to time when we study electronic devices.

The motion of the electron is described by its probability flow density

$$\mathbf{j}(\mathbf{r}, t) = \frac{\hbar}{2im_0} [\Psi^*(\mathbf{r}, t) \nabla \Psi(\mathbf{r}, t) - \Psi(\mathbf{r}, t) \nabla \Psi^*(\mathbf{r}, t)] \quad (2.10)$$

In the absence of a potential energy, the Hamiltonian of the electron consists of only $-\hbar^2 \nabla^2 / 2m_0$ and the solution of the corresponding Schrödinger equation is the plane wave form $\Psi(\mathbf{r}) = e^{i\mathbf{k} \cdot \mathbf{r}}$ with a flow density

$$\mathbf{j} = \frac{\hbar \mathbf{k}}{m_0} \quad (2.11)$$

by inserting $\Psi(\mathbf{r}) = e^{i\mathbf{k} \cdot \mathbf{r}}$ into Eq. (2.10).

It is easy to show that the probability flow density \mathbf{j} satisfies the continuity equation

$$\frac{\partial(\Psi^* \Psi)}{\partial t} = -\nabla \cdot \mathbf{j} \quad (2.12)$$

Let us have a look about the probability flow density of a steady-state Bloch electron in the semiconductor. Inserting the Bloch wave function

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

into Eq. (2.10)

$$\mathbf{j}_{nk}(\mathbf{r}) = \frac{\hbar\mathbf{k}}{Nm_0}u_{nk}^*u_{nk} + \frac{\hbar}{2im_0}(u_{nk}^*\nabla u_{nk} - u_{nk}\nabla u_{nk}^*) \quad (2.13)$$

where N is the number of unit cells contained in the semiconductor.

We recall that the periodic Bloch function $u_{nk}(\mathbf{r}, t)$ is periodic with respect to unit cell, it is also normalized within one unit cell. The probability flow density of the Bloch electron averaged in the semiconductor thus becomes

$$\langle \mathbf{j}_{nk}(\mathbf{r}) \rangle = \int \frac{\hbar\mathbf{k}}{Nm_0}u_{nk}^*u_{nk}d\mathbf{r} + \int \frac{\hbar}{2im_0}(u_{nk}^*\nabla u_{nk} - u_{nk}\nabla u_{nk}^*)d\mathbf{r} \quad (2.14)$$

Since

$$\int_{\text{cell}} u_{nk}^*\nabla u_{nk}d\mathbf{r} = 0 \quad (2.15)$$

we have

$$\langle \mathbf{j}_{nk}(\mathbf{r}) \rangle = \frac{\hbar\mathbf{k}}{m_0} \quad (2.16)$$

This is very much alike the probability flow density of a free electron Eq. (2.11). It indicates that a Bloch electron $\Psi_{nk}(\mathbf{r}, t)$ can be viewed as a plane wave with a wave vector \mathbf{k} .

At equilibrium, the occupation of Bloch state Ψ_{nk} is determined by its energy $E_n(\mathbf{k})$. In other words, Ψ_{nk} and $\Psi_{n-\mathbf{k}}$ are equally occupied, since $E_n(\mathbf{k}) = E_n(-\mathbf{k})$, i.e., Eq. (1.41). Since the two Bloch states have the same probability flow density with opposite directions, there is no net electron flow at equilibrium.

We have thus far discussed the probability flow density of a Bloch electron, i.e., an eigen function of the electron Hamiltonian. In practice, the electron under investigation is injected into the device through contacts which is not initially an eigen state of the electron Hamiltonian in the device. Let us denote the initial state of injecting electron as $\psi(\mathbf{r}, 0)$ at $t = 0$. The motion of the electron state is governed by the time-dependent Schrödinger equation

$$i\hbar\frac{\partial\Psi(\mathbf{r}, t)}{\partial t} = H\Psi(\mathbf{r}, t) \quad (2.17)$$

Discretizing the time t by step δt so that $t = n\delta t$, where n is a positive integer. The wave function of the electron becomes $\Psi(\mathbf{r}, n\delta t)$ which is further denoted as $\Psi^n(\mathbf{r})$, and the above time-dependent Schrödinger equation can be written as

$$i\hbar\frac{\Psi^{n+1}(\mathbf{r}) - \Psi^n(\mathbf{r})}{\delta t} = H\Psi^n(\mathbf{r}) \rightarrow \Psi^{n+1}(\mathbf{r}) = \left(1 - \frac{i\delta t}{\hbar}H\right)\Psi^n(\mathbf{r}) \quad (2.18)$$

This equation has an intuitive form such that a future wave function Ψ^{n+1} is given explicitly in terms of its preceding wave function Ψ^n , hence it is known as the explicit forward-time-centered-space (FTCS) scheme [1]. However this solution is

unstable due to boundless growth of errors through successive time integration. Alternatively we can try the implicit form

$$\Psi^n(\mathbf{r}) = \left(1 + \frac{i\delta t}{\hbar}H\right)\Psi^{n+1}(\mathbf{r}) \quad (2.19)$$

and though this form is stable, it is fairly inaccurate. Thus Goldberg, Schey, and Schwartz have employed the Cayley form [2]

$$\left(1 + \frac{i\delta t}{2\hbar}H\right)\Psi^{n+1}(\mathbf{r}) = \left(1 - \frac{i\delta t}{2\hbar}H\right)\Psi^n(\mathbf{r}) \quad (2.20)$$

It was indicated [2] that unitarity is the characteristics of the original Schrödinger equation which ensures that the normalization of the wave functions does not change in time. When H is Hermitian, $e^{\pm i\delta t H/\hbar}$ are clearly unitary, while their first-order approximations $1 \pm i\delta t H/\hbar$ are not unitary. On the other hand, the Cayley form Eq. (2.20) is unitary, which has a further desirable property of being precise to order δt^2 .

The value of δt in Eq. (2.20) is determined in such a way that $\delta t E \leq \hbar$ in order to convert the Schrödinger equation (2.17) into the Cayley form of Eq. (2.20).

In the Goldberg scheme, the initial wave packet is described by a single-momentum Gaussian wave packet [2]

$$\left(\frac{\sigma^2}{\pi}\right)^{3/4} e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}_0)-|\mathbf{r}-\mathbf{r}_0|^2/2\sigma^2} \quad (2.21)$$

centered at \mathbf{r}_0 , where \mathbf{k} is the momentum and σ describes the spatial extension of the wave packet. As an example, let us study the wave packet transport of one conduction-band electron with $m_x^* = 0.1905$, $m_y^* = 0.1905$, and $m_z^* = 0.9163$, in a Si quantum wire embedded in SiO₂ with a cross section $(-L_x : L_x) \times (-L_y : L_y)$ extended along the z axis. In the following numerical simulations, $L_x = L_y = 2.7$ nm, $L_z = (-500 : 500)$ nm. We assume that the potential barrier provided by the surrounding SiO₂ is infinitely high. Furthermore, we introduce a negative point charge at the center of the quantum wire. The initial Gaussian wave packet can be expressed as

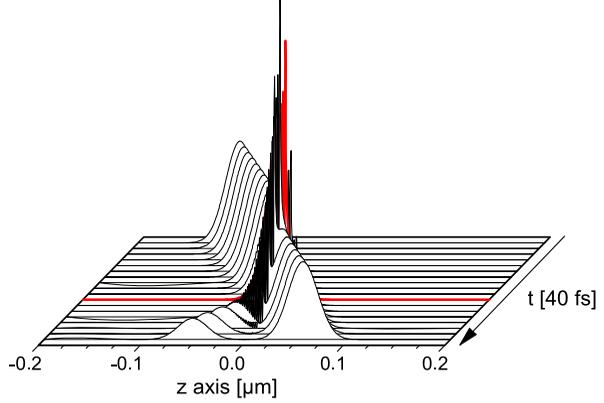
$$\psi_{i,k}(\mathbf{r}, 0) = \psi_i(x, y) \left(\frac{\sigma^2}{\pi}\right)^{1/4} e^{ik(z-z_0)-(z-z_0)^2/2\sigma^2} \quad (2.22)$$

where $\psi_i(x, y)$ is the i th eigen function in the xy plane. Let $z_0 = -100$ nm, $\sigma = 10$ nm, and

$$\psi_i(x, y) = \sin\left[\frac{\pi(x + L_x)}{2L_x}\right] \sin\left[\frac{\pi(y + L_y)}{2L_y}\right] \quad (2.23)$$

which is the ground state in the xy plane. Denote $E_k = \hbar^2 k^2 / 2m_z^*$ as the kinetic energy of the electron along the z axis, it is just straightforward to calculate the

Fig. 2.1 Wave packet transport $a(z, t)$ initially centered at $z = -0.1 \mu\text{m}$. $E_k = 150 \text{ meV}$. The red curve is the wave packet profile when the wave packet center reaches $z = 0$



wave packet transport $\psi(\mathbf{r}, t)$ as a function of E_k by Eq. (2.20), the flow density $\mathbf{j}(\mathbf{r}, t)$ by Eq. (2.10). Physically, the current flow is nonzero only along the z axis, so that we present the wave packet transport in the form of

$$a(z, t) = \iint |\psi(\mathbf{r}, t)|^2 dx dy \quad (2.24)$$

Figure 2.1 shows the temporal development of a Gaussian wave packet initially centered at $z = -0.1 \mu\text{m}$ with a kinetic energy $E_k = 150 \text{ meV}$ along the z axis. It is easy to estimate that the speed of such a wave packet is $2.4 \times 10^5 \text{ m/s}$ (effective mass $m_z^* = 0.9163$) and the time to move from $z = -0.1 \mu\text{m}$ to $z = 0$ is 0.417 ps , as marked by the red profile in Fig. 2.1. Here it is observed that the wave packet splits into two by the negative point charge, one is reflected back (moves along the $-z$ axis) and the other tunnels through the potential barrier induced by the negative point charge then transports further along the z axis.

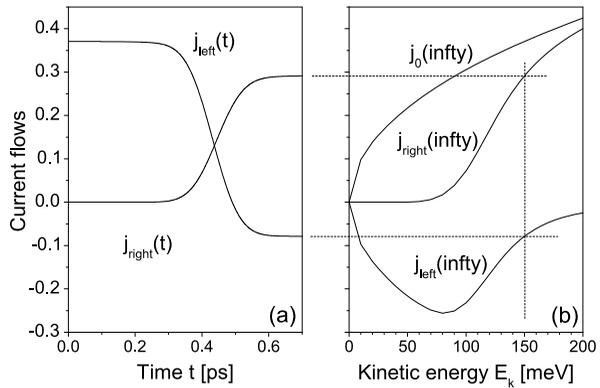
In order to see the temporal development of the wave packet more clearly, we split the current flow into two parts, one for $z < 0$ and the other $z \geq 0$

$$\begin{aligned} \mathbf{j}_{\text{left}}(t) &= \int_{-\infty}^0 \left[\iint \mathbf{j}(\mathbf{r}, t) dx dy \right] dz \\ \mathbf{j}_{\text{right}}(t) &= \int_0^{\infty} \left[\iint \mathbf{j}(\mathbf{r}, t) dx dy \right] dz \end{aligned} \quad (2.25)$$

Figure 2.2(a) shows the temporal developments of the above two current flows when $E_k = 150 \text{ meV}$. Here we see that after the wave packet reaches the point charge, the tunneling part starts to emerge on the right side of the point charge, thus a gradual increase of $\mathbf{j}_{\text{right}}(t)$ (> 0 , i.e., along the z axis), the reflected part move along the $-z$ axis, thus $\mathbf{j}_{\text{left}}(t) < 0$.

If we put two electrodes at $z = \pm\infty$, two currents, $\mathbf{j}_{\text{left}}(\infty)$ and $\mathbf{j}_{\text{right}}(\infty)$, will be eventually detected when $t \rightarrow \infty$, which are shown in Fig. 2.2(b). For the sake of comparison, the transport of the wave packet through a perfect Si quantum wire

Fig. 2.2 (a) Current flows $j_{\text{left}}(t)$ and $j_{\text{right}}(t)$ as functions of time t . $E_k = 150$ meV. (b) $j_{\text{left}}(\infty)$ and $j_{\text{right}}(\infty)$ as functions of E_k . $j_0(\infty)$ denotes the current flow through a perfect Si quantum wire



(without the point charge at the quantum wire center), $j_0(\infty)$, is also presented which is exactly the relationship of Eq. (2.16) as expected. Figure 2.2(b) shows that the wave packet is easily scattered when its kinetic energy is small. When the kinetic energy is larger than 200 meV, the wave packet overcomes the point charge so that the scattering becomes negligible.

The wave packet transport is extremely important in nano-scale electronic and photonic devices. See more in Chaps. 4 and 5.

2.2 Scattering Theory

By the above section we understand that at equilibrium, the electrons remain in their initial states in the semiconductor and there will be no net carrier flow. External forces are therefore needed in order to obtain any responses from the electrons. Before we formulate the electron transport theory as well as discuss all other properties of semiconductor materials and devices, we repeat here the basic quantum mechanical description about the response of electrons under external perturbation.

We shall use the quantum theory in the following to discuss various transition processes of electron. We first formulate the quantum perturbation theory in its simplest form. The electron is described by the Hamiltonian H_0

$$\begin{aligned}
 i\hbar \frac{\partial \Psi_i(\mathbf{r}, t)}{\partial t} &= H_0 \Psi_i(\mathbf{r}, t) \\
 H_0 \Psi_i(\mathbf{r}) &= E_i \Psi_i(\mathbf{r}) \\
 \Psi_i(\mathbf{r}, t) &= e^{-iE_i t/\hbar} \Psi_i(\mathbf{r})
 \end{aligned} \tag{2.26}$$

At $t = 0$ a perturbation $V' e^{-i\omega t}$ is switched on, where V' is time independent. The wave function $\Psi(\mathbf{r}, t)$ of the new Schrödinger equation

$$i\hbar \frac{\partial \Psi(\mathbf{r}, t)}{\partial t} = (H_0 + V' e^{-i\omega t}) \Psi(\mathbf{r}, t) \tag{2.27}$$

is expanded in the basis of $\{\Psi_i, E_i\}$ in the form of

$$\Psi(\mathbf{r}, t) = \sum_i a_i(t) \Psi_i(\mathbf{r}) e^{-iE_i t/\hbar} \quad (2.28)$$

so that

$$i\hbar \frac{da_m(t)}{dt} = \sum_i \langle \Psi_m(\mathbf{r}) | V' | \Psi_i(\mathbf{r}) \rangle e^{i(E_m - E_i - \hbar\omega)t/\hbar} a_i(t) \quad (2.29)$$

Suppose that before $t = 0$, the electron under investigation occupies state Ψ_n , i.e., $a_n(t < 0) = 1$ and $a_i(t < 0) = 0$ if $i \neq n$. By the first-order perturbation theory,

$$i\hbar \frac{da_m(t)}{dt} = \langle \Psi_m(\mathbf{r}) | V' | \Psi_n(\mathbf{r}) \rangle e^{i(E_m - E_n - \hbar\omega)t/\hbar} \quad (2.30)$$

It must be emphasized here that the above equation is actually valid for a very short time duration of $[t, t + dt]$. After dt , the occupations of all the electron states will all be modified and eventually we need to work with Eq. (2.29) along the time axis, which will be done in the next section. However, Eq. (2.30) itself is very rich in physics if being treated carefully. The most important aspect to be taken into account for is the fact that during the transition from Ψ_n to Ψ_m under the influence of $V' e^{-i\omega t}$, the occupation $a_n(t)$ of the initial state Ψ_n will be gradually dissipated.

Let

$$\langle \Psi_m | V' | \Psi_n \rangle = V'_{mn}$$

We introduce the dissipation process phenomenologically by the concept of a lifetime, \hbar/Γ_n , of the electron occupying state Ψ_n . In other words, due to the transitions from Ψ_n to Ψ_m , the occupation a_n gradually decreases. Hence, we can simulate that the total wave function of the initial state Ψ_n is

$$\Psi_n(\mathbf{r}) e^{iE_n t/\hbar - \Gamma_n t/\hbar}$$

Equation (2.30) is now replaced by

$$i\hbar \frac{da_m(t)}{dt} = V'_{mn} e^{i(E_m - E_n - \hbar\omega)t/\hbar - \Gamma_n t/\hbar} \quad (2.31)$$

The transition probability is calculated by integrating from 0 to t

$$a_m(t) = \frac{1}{i\hbar} \int_0^t V'_{mn} e^{i(E_m - E_n - \hbar\omega)t/\hbar - \Gamma_n t/\hbar} dt = \frac{V'_{mn} [1 - e^{i(E_m - E_n - \hbar\omega)t/\hbar - \Gamma_n t/\hbar}]}{E_m - E_n - \hbar\omega + i\Gamma_n} \quad (2.32)$$

Let $t \rightarrow \infty$,

$$a_m(\infty) = \frac{V'_{mn}}{E_m - E_n - \hbar\omega + i\Gamma_n} \quad (2.33)$$

so that

$$|a_m(\infty)|^2 = \frac{|V'_{mn}|^2}{(E_m - E_n - \hbar\omega)^2 + \Gamma_n^2} \quad (2.34)$$

And the averaged transition probability per unit time, i.e., the transition rate, is obtained by dividing $|a_m(\infty)|^2$ by lifetime \hbar/Γ_n of occupation of state Ψ_n

$$p_m \equiv \frac{|a_m(\infty)|^2}{\hbar/\Gamma_n} = \frac{|V'_{mn}|^2}{\hbar} \frac{\Gamma_n}{(E_m - E_n - \hbar\omega)^2 + \Gamma_n^2} \quad (2.35)$$

In the following, we try to derive the scattering theory in a rigorous way which will be the starting point of the principles of almost all semiconductor nanostructured devices to be discussed in this book.

By purely and formally writing

$$H = H_0 + V' \quad (2.36)$$

where H_0 represents the Hamiltonian operator of the electron at equilibrium, V' is the interaction of the electron with an external field. By the Bloch theorem, the eigen solution of H_0 is

$$\begin{aligned} i\hbar \frac{\partial \Psi_{nk}(\mathbf{r}, t)}{\partial t} &= H_0 \Psi_{nk}(\mathbf{r}, t) \\ \Psi_{nk}(\mathbf{r}, t) &= e^{-iE_{nk}t/\hbar} \Psi_{nk}(\mathbf{r}) \\ H_0 \Psi_{nk}(\mathbf{r}) &= E_{nk} \Psi_{nk}(\mathbf{r}, t) \end{aligned} \quad (2.37)$$

where n is the band index and \mathbf{k} the wave vector. In the following we use short notations to simply mathematical expressions: \mathbf{k} represents quantum numbers $n\mathbf{k}$ for Bloch state, and $|\mathbf{k}\rangle = \Psi_{nk}(\mathbf{r}, t) = e^{-iE_{nk}t/\hbar} \Psi_{nk}(\mathbf{r})$. And between two states \mathbf{k} and \mathbf{q} , $\langle \mathbf{q} | \mathbf{k} \rangle = \delta_{\mathbf{q}, \mathbf{k}}$, i.e., the wave functions are orthogonal. Note that $\delta_{\mathbf{q}, \mathbf{k}} = 0$ if $\mathbf{q} \neq \mathbf{k}$, and $\delta_{\mathbf{q}, \mathbf{k}} = 1$ if $\mathbf{q} = \mathbf{k}$.

Assume that the electron occupies state $|\mathbf{k}\rangle$ before we switch on interaction V' . We like to solve the equation of

$$(H_0 + V')|\Psi\rangle = E_{\mathbf{k}}|\Psi\rangle \quad (2.38)$$

i.e., $|\Psi\rangle$ is a state of Hamiltonian $H_0 + V'$. The above equation can be re-formatted as $(E_{\mathbf{k}} - H_0)|\Psi\rangle = V'|\Psi\rangle$, which is equivalent to

$$(E_{\mathbf{k}} - H_0)|\Psi\rangle = (E_{\mathbf{k}} - H_0)|\mathbf{k}\rangle + V'|\Psi\rangle$$

since $(E_{\mathbf{k}} - H_0)|\mathbf{k}\rangle = 0$. We obtain the following equation by dividing the above expression by $(E_{\mathbf{k}} - H_0)$

$$|\Psi\rangle = |\mathbf{k}\rangle + \frac{1}{E_{\mathbf{k}} - H_0} V'|\Psi\rangle \quad (2.39)$$

We obtain the formal formula of scattering wave function by adding a correction $i\Gamma$

$$|\Psi\rangle = |\mathbf{k}\rangle + \frac{1}{E_{\mathbf{k}} - H_0 + i\Gamma} V' |\Psi\rangle = |\mathbf{k}\rangle + G_{\mathbf{k}} V' |\Psi\rangle \quad (2.40)$$

which is known as the Lippmann-Schwinger equation. We will discuss Γ more in the next section. In the above equation,

$$G_{\mathbf{k}} = \frac{1}{E_{\mathbf{k}} - H_0 + i\Gamma} = \sum_{\mathbf{q}} \frac{|\mathbf{q}\rangle \langle \mathbf{q}|}{E_{\mathbf{k}} - E_{\mathbf{q}} + i\Gamma} \quad (2.41)$$

In the last equality, a representation of unity as a sum over eigenstates of H_0 is inserted to make the expression more explicitly.

We can write the solution of Eq. (2.40) as

$$|\Psi\rangle = \frac{1}{1 - G_{\mathbf{k}} V'} |\mathbf{k}\rangle \quad (2.42)$$

and expand it

$$|\Psi\rangle = |\mathbf{k}\rangle + G_{\mathbf{k}} V' |\mathbf{k}\rangle + G_{\mathbf{k}} V' G_{\mathbf{k}} V' |\mathbf{k}\rangle + G_{\mathbf{k}} V' G_{\mathbf{k}} V' G_{\mathbf{k}} V' |\mathbf{k}\rangle + \dots \quad (2.43)$$

The n th approximation, also known as the n th Born approximation, to the scattering state $|\Psi\rangle$ consists of terminating the above expansion after n terms.

We define a transition matrix $T_{\mathbf{k}}$ as

$$\begin{aligned} T_{\mathbf{k}} &= V' + V' G_{\mathbf{k}} V' + V' G_{\mathbf{k}} V' G_{\mathbf{k}} V' + V' G_{\mathbf{k}} V' G_{\mathbf{k}} V' G_{\mathbf{k}} V' + \dots \\ &= V' + V' \frac{1}{E_{\mathbf{k}} - H_0 + i\Gamma} V' + \dots \end{aligned} \quad (2.44)$$

i.e., $V' |\Psi\rangle = T_{\mathbf{k}} |\mathbf{k}\rangle$, so that the scattering state can be written

$$|\Psi\rangle = |\mathbf{k}\rangle + G_{\mathbf{k}} T_{\mathbf{k}} |\mathbf{k}\rangle \quad (2.45)$$

The probability that the electron initially at state $|\mathbf{k}\rangle$ ends up at state $|\mathbf{q}\rangle$ is

$$\begin{aligned} \langle \mathbf{q} | \Psi \rangle &= \langle \mathbf{q} | \mathbf{k} \rangle + \frac{1}{E_{\mathbf{k}} - E_{\mathbf{q}} + i\Gamma} \left(\langle \mathbf{q} | V' | \mathbf{k} \rangle + \langle \mathbf{q} | V' \frac{1}{E_{\mathbf{k}} - H_0 + i\Gamma} V' | \mathbf{k} \rangle + \dots \right) \\ &= \delta_{\mathbf{q}, \mathbf{k}} + \frac{1}{E_{\mathbf{k}} - E_{\mathbf{q}} + i\Gamma} \left(\langle \mathbf{q} | V' | \mathbf{k} \rangle + \langle \mathbf{q} | V' \frac{1}{E_{\mathbf{k}} - H_0 + i\Gamma} V' | \mathbf{k} \rangle + \dots \right) \end{aligned} \quad (2.46)$$

We have thus far introduced the formal scattering theory about the probability of an electron at initial state $|\mathbf{k}\rangle$ being scattered to state $|\mathbf{q}\rangle$ due to an external perturbation V' . Now we apply the theory to calculate transition amplitudes between the relevant unperturbed eigenstates due to the appearance of the external perturbation V' which is switched on at $t = 0$.

Let H_0 be time-independent and all solutions of the eigenvalue problem $H_0|\mathbf{k}\rangle = E_{\mathbf{k}}|\mathbf{k}\rangle$ are known, it is straightforward to construct the solution of the time-dependent Schrödinger equation

$$i\hbar \frac{\partial |\Psi(t)\rangle}{\partial t} = H_0 |\Psi(t)\rangle \quad (2.47)$$

which coincides at $t = 0$ with a given initial state $|\Psi(0)\rangle$. Under these circumstances, the state

$$|\Psi(t)\rangle = e^{-iH_0t/\hbar} |\Psi(0)\rangle = \sum_{\mathbf{k}} e^{-iE_{\mathbf{k}}t/\hbar} |\mathbf{k}\rangle \langle \mathbf{k} | \Psi(0)\rangle \quad (2.48)$$

represents the general solution of Eq. (2.47).

When the external perturbation V' is switched on at $t = 0$, the Hamiltonian becomes $H_0 + V'$ and the Schrödinger equation becomes

$$i\hbar \frac{\partial |\Psi(t)\rangle}{\partial t} = (H_0 + V') |\Psi(t)\rangle \quad (2.49)$$

The above expression is normally referred to as the Schrödinger equation in the Schrödinger picture. We transform the above equation into the so-called interaction picture by a time-dependent unitary operator

$$|\hat{\Psi}(t)\rangle = e^{iH_0t/\hbar} |\Psi(t)\rangle \quad (2.50)$$

In the interaction picture, the Schrödinger equation becomes

$$i\hbar \frac{\partial |\hat{\Psi}(t)\rangle}{\partial t} = \hat{V}'(t) |\hat{\Psi}(t)\rangle \quad (2.51)$$

where the new interaction operator is given by

$$\hat{V}'(t) = e^{iH_0t/\hbar} V' e^{-iH_0t/\hbar} \quad (2.52)$$

Next we introduce the time development operator $\hat{T}(t)$ such that

$$|\hat{\Psi}(t)\rangle = \hat{T}(t) |\hat{\Psi}(0)\rangle \quad (2.53)$$

Insert the above expression into the Schrödinger equation in the interaction picture, i.e., Eq. (2.51), we can see that the time development operator satisfies the integral equation

$$\hat{T}(t) = 1 - \frac{i}{\hbar} \int_0^t \hat{V}'(t') \hat{T}(t') dt' \quad (2.54)$$

A direct calculation of the above time development operator from eigenstate $|\mathbf{k}\rangle$ to $|\mathbf{q}\rangle$ is

$$\langle \mathbf{q} | \hat{T}(t) | \mathbf{k} \rangle = \delta_{\mathbf{q},\mathbf{k}} - \frac{i}{\hbar} \int_0^t \langle \mathbf{q} | \hat{V}'(t') \hat{T}(t') | \mathbf{k} \rangle dt'$$

$$\begin{aligned}
&= \delta_{\mathbf{q},\mathbf{k}} - \frac{i}{\hbar} \sum_{\mathbf{k}'} \int_0^t \langle \mathbf{q} | \hat{V}'(t') | \mathbf{k}' \rangle \langle \mathbf{k}' | \hat{T}(t') | \mathbf{k} \rangle dt' \\
&= \delta_{\mathbf{q},\mathbf{k}} - \frac{i}{\hbar} \sum_{\mathbf{k}'} \int_0^t e^{i(E_{\mathbf{q}} - E_{\mathbf{k}'})t'/\hbar} \langle \mathbf{q} | V'(t') | \mathbf{k}' \rangle \langle \mathbf{k}' | \hat{T}(t') | \mathbf{k} \rangle dt'
\end{aligned} \tag{2.55}$$

On the other hand, when V' is small, Eq. (2.54) can be solved by a power series in terms of \hat{V}' such that the zero-order solution

$$\hat{T}^{(0)}(t) = 1 \tag{2.56}$$

which is the solution when $\hat{V}'(t') = 0$. The first-order solution is

$$\hat{T}^{(1)}(t) = 1 - \frac{i}{\hbar} \int_0^t \hat{V}'(t') \hat{T}^{(0)}(t') dt' = 1 - \frac{i}{\hbar} \int_0^t \hat{V}'(t') dt' \tag{2.57}$$

Iterating the mathematical process we obtain a formal solution to Eq. (2.54)

$$\hat{T}(t) = 1 - \frac{i}{\hbar} \int_0^t \hat{V}'(t') dt' - \frac{1}{\hbar^2} \int_0^t \hat{V}'(t') \left[\int_0^{t'} \hat{V}'(t'') dt'' \right] dt' + \dots \tag{2.58}$$

When $\mathbf{k} \neq \mathbf{q}$, the transition amplitude in the first-order perturbation theory is then given by the simple and much-used formula

$$\langle \mathbf{q} | \hat{T}^{(1)}(t) | \mathbf{k} \rangle = -\frac{i}{\hbar} \int_0^t e^{i(E_{\mathbf{q}} - E_{\mathbf{k}})t'/\hbar} \langle \mathbf{q} | V'(t') | \mathbf{k} \rangle dt' \tag{2.59}$$

Notice that $t = 0$ is a time preceding the onset of the perturbation V' .

If V' does not depend on the time, or

$$V'(t) = V_0^* e^{i\omega t} + V_0 e^{-i\omega t}$$

where V_0 and ω are time-independent, the time integration of Eq. (2.59) is trivial. We first focus on the first case, i.e., V' is time-independent, while the second case will be studied in the next section.

Starting with an initial state $|\mathbf{k}\rangle$ at $t = 0$, Eq. (2.59) shows that after a time interval t , the electron will evolve from its initial state $|\mathbf{k}\rangle$ into state $|\mathbf{q}\rangle$ due to perturbation V' switched on at $t = 0$. In other words, the probability that the electron will stay in its initial state will be decreased. Let us now study the probability that the electron dwells in its initial state. Back to Eq. (2.55) and note that V' is weak and time-independent,

$$\langle \mathbf{q} | \hat{T}(t) | \mathbf{k} \rangle = -\frac{i}{\hbar} \langle \mathbf{q} | V' | \mathbf{k} \rangle \int_0^t e^{i(E_{\mathbf{q}} - E_{\mathbf{k}})t'/\hbar} \langle \mathbf{k} | \hat{T}(t') | \mathbf{k} \rangle dt' \tag{2.60}$$

for $\mathbf{q} \neq \mathbf{k}$. Here the transitions via intermediate states $\mathbf{k}' \neq \mathbf{k}$ are neglected since under the first-order perturbation assumption, the possibility of a two-step process, i.e., from \mathbf{k} to \mathbf{k}' ($\mathbf{k}' \neq \mathbf{k}$), then from \mathbf{k}' to \mathbf{q} ($\mathbf{q} \neq \mathbf{k}$), is expected to be very small.

By Eq. (2.55), the equation of motion for $\langle \mathbf{k} | \hat{T}(t) | \mathbf{k} \rangle$ is, rigorously,

$$\frac{\partial}{\partial t} \langle \mathbf{k} | \hat{T}(t) | \mathbf{k} \rangle = -\frac{i}{\hbar} \sum_{\mathbf{q} \neq \mathbf{k}} e^{i(E_{\mathbf{k}} - E_{\mathbf{q}})t/\hbar} \langle \mathbf{k} | V' | \mathbf{q} \rangle \langle \mathbf{q} | \hat{T}(t) | \mathbf{k} \rangle - \frac{i}{\hbar} \langle \mathbf{k} | V' | \mathbf{k} \rangle \langle \mathbf{k} | \hat{T}(t) | \mathbf{k} \rangle \quad (2.61)$$

If Eq. (2.60) is substituted into (2.61), we obtain a differential-integral equation for the probability amplitude that the electron will dwell in the initial state $|\mathbf{k}\rangle$ at time t after switching on V' at time 0

$$\begin{aligned} \frac{\partial}{\partial t} \langle \mathbf{k} | \hat{T}(t) | \mathbf{k} \rangle = & -\frac{1}{\hbar^2} \sum_{\mathbf{q} \neq \mathbf{k}} |\langle \mathbf{q} | V' | \mathbf{k} \rangle|^2 \int_0^t \langle \mathbf{k} | \hat{T}(t') | \mathbf{k} \rangle e^{i(E_{\mathbf{q}} - E_{\mathbf{k}})(t' - t)/\hbar} dt' \\ & - \frac{i}{\hbar} \langle \mathbf{k} | V' | \mathbf{k} \rangle \langle \mathbf{k} | \hat{T}(t) | \mathbf{k} \rangle \end{aligned} \quad (2.62)$$

The solution of this equation demands care. We are interested in a long time which implies rapid oscillations of the exponential factor in the integrand as a function of the final state energy $E_{\mathbf{q}}$. The slowly varying amplitude $\langle \mathbf{k} | \hat{T}(t') | \mathbf{k} \rangle$ can therefore be removed from the t' integrand. The remaining integral can be evaluated

$$\int_0^t e^{i(E_{\mathbf{q}} - E_{\mathbf{k}})(t' - t)/\hbar} dt' = \frac{\hbar [1 - e^{-i(E_{\mathbf{q}} - E_{\mathbf{k}})t/\hbar}]}{i(E_{\mathbf{q}} - E_{\mathbf{k}})} \quad (2.63)$$

which oscillates between 0 and $-i\hbar/(E_{\mathbf{q}} - E_{\mathbf{k}})$ as a function of t . Moreover, it becomes singular when $E_{\mathbf{q}} = E_{\mathbf{k}}$. Such a result is not physical. One critical factor we missed so far is that the transition process from initial state $|\mathbf{k}\rangle$ to other state $|\mathbf{q}\rangle$ will diminish gradually the probability that the electron will stay at its initial state $|\mathbf{k}\rangle$. This is normally referred to as the lifetime of state $|\mathbf{k}\rangle$ under perturbation V' . In other words, $|\mathbf{k}\rangle$, which is an eigen state of Hamiltonian H_0 , is no long an eigen state of Hamiltonian $H_0 + V'$.

We thus introduce a lifetime related parameter $\Gamma_{\mathbf{k}}$ to initial state $|\mathbf{k}\rangle$, such that the eigen value is now $E_{\mathbf{k}} + i\Gamma_{\mathbf{k}}$. In other words, we now express the total wave function of state $|\mathbf{k}\rangle$ as

$$e^{-i(E_{\mathbf{k}} + i\Gamma_{\mathbf{k}})t/\hbar} \Psi_{\mathbf{k}}(\mathbf{r}) \quad (2.64)$$

in the presence of V' , see Eq. (2.37). $\Gamma_{\mathbf{k}}$ will be obtained self-consistently shortly. The self-consistency is more clear in Sect. 3.7 when we discuss the excitation of an exciton in a quantum dot.

Integration (2.63) now becomes

$$\int_0^t e^{i(E_q - E_k - i\Gamma_k)(t' - t)/\hbar} dt' = \frac{\hbar[1 - e^{-i(E_q - E_k - i\Gamma_k)t/\hbar}]}{i(E_q - E_k - i\Gamma_k)}$$

$$\implies \frac{\hbar}{i(E_q - E_k - i\Gamma_k)} \quad (2.65)$$

when $t \gg \Gamma_k/\hbar$. The resulting differential equation is

$$\frac{\partial}{\partial t} \langle \mathbf{k} | \hat{T}(t) | \mathbf{k} \rangle = \left(\frac{i}{\hbar} \sum_{q \neq \mathbf{k}} \frac{|\langle \mathbf{q} | V' | \mathbf{k} \rangle|^2}{E_q - E_k - i\Gamma_k} - \frac{i}{\hbar} \langle \mathbf{k} | V' | \mathbf{k} \rangle \right) \langle \mathbf{k} | \hat{T}(t) | \mathbf{k} \rangle \quad (2.66)$$

and the solution is

$$\langle \mathbf{k} | \hat{T}(t) | \mathbf{k} \rangle = \exp \left[\left(\frac{i}{\hbar} \sum_{q \neq \mathbf{k}} \frac{|\langle \mathbf{q} | V' | \mathbf{k} \rangle|^2}{E_q - E_k - i\Gamma_k} - \frac{i}{\hbar} \langle \mathbf{k} | V' | \mathbf{k} \rangle \right) t \right] \quad (2.67)$$

since $\langle \mathbf{k} | \hat{T}(0) | \mathbf{k} \rangle = 1$.

Just like (2.64), the above expression describes the exponential decay of the electron occupation in the initial state $|\mathbf{k}\rangle$. Therefore,

$$\langle \mathbf{k} | \hat{T}(t) | \mathbf{k} \rangle = \exp \left(-\frac{\Gamma_k}{\hbar} t - \frac{i\Delta E_k}{\hbar} t \right) \quad (2.68)$$

where

$$\Gamma_k = \sum_{q \neq \mathbf{k}} |\langle \mathbf{q} | V' | \mathbf{k} \rangle|^2 \frac{\Gamma_k}{(E_k - E_q)^2 + \Gamma_k^2} \quad (2.69)$$

normally referred to as the relaxation energy of state $|\mathbf{k}\rangle$ under the presence of V' , describes how long the electron will stay in its initial state $|\mathbf{k}\rangle$. \hbar/Γ_k is therefore denoted as the lifetime. And the eigen value of state $|\mathbf{k}\rangle$ is shifted, up to second order in V' , by

$$\Delta E_k = \langle \mathbf{k} | V' | \mathbf{k} \rangle + \sum_{q \neq \mathbf{k}} |\langle \mathbf{q} | V' | \mathbf{k} \rangle|^2 \frac{E_k - E_q}{(E_k - E_q)^2 + \Gamma_k^2} \quad (2.70)$$

Equation (2.69) shows that Γ_k is proportional to V' , which is natural. Without V' , the electron will stay at its initial state forever. Switching on perturbation V' makes $|\mathbf{k}\rangle$ no longer an eigen state of the new Hamiltonian $H_0 + V'$ so that the probability of finding the electron at $|\mathbf{k}\rangle$ will decrease.

When perturbation $\langle \mathbf{q} | V' | \mathbf{k} \rangle$ is very small, Γ_k will also be very small. In this case, we retrieve common textbook expressions about the relaxation energy Γ_k and energy shift ΔE_k

$$\Gamma_k = \pi \sum_{q \neq k} |\langle \mathbf{q} | V' | \mathbf{k} \rangle|^2 \delta(E_k - E_q) \quad (2.71)$$

$$\Delta E_k = \langle \mathbf{k} | V' | \mathbf{k} \rangle + \sum_{q \neq k} \frac{|\langle \mathbf{q} | V' | \mathbf{k} \rangle|^2}{E_k - E_q}$$

since

$$\lim_{\Gamma \rightarrow 0} \frac{\Gamma}{E^2 + \Gamma^2} = \pi \delta(E) \quad (2.72)$$

We now understand the fundamental constrain about the validity of the above expressions: Γ_k must be very small, which, however, is not always guaranteed by a small $\langle \mathbf{q} | V' | \mathbf{k} \rangle$. We will discuss in more detail in the next section.

When Eq. (2.68) is substituted back into Eq. (2.60), the integration can be carried out and we obtain

$$\langle \mathbf{q} | \hat{T}(t) | \mathbf{k} \rangle = \langle \mathbf{q} | V' | \mathbf{k} \rangle \frac{1 - e^{-\Gamma_k t / \hbar} e^{i(E_q - E_k - \Delta E_k)t / \hbar}}{E_q - E_k - \Delta E_k + i\Gamma_k} \quad (2.73)$$

for $t \geq 0$, and hence, the probability that the electron has evolved from its initial state $|\mathbf{k}\rangle$ into state $|\mathbf{q}\rangle$ is

$$\begin{aligned} P_{q \leftarrow k}(t) &= |\langle \mathbf{q} | \hat{T}(t) | \mathbf{k} \rangle|^2 \\ &= |\langle \mathbf{q} | V' | \mathbf{k} \rangle|^2 \frac{1 - 2e^{-\Gamma_k t / \hbar} \cos[(E_q - E_k - \Delta E_k)t / \hbar] + e^{-2\Gamma_k t / \hbar}}{(E_q - E_k - \Delta E_k)^2 + \Gamma_k^2} \end{aligned} \quad (2.74)$$

After a time that is very long compared with the lifetime \hbar/Γ_k that the electron stays in eigenstate $|\mathbf{k}\rangle$, i.e., $t \rightarrow \infty$, the probability that the electron will be found in state $|\mathbf{q}\rangle$ is

$$P_{q \leftarrow k}(\infty) = \frac{|\langle \mathbf{q} | V' | \mathbf{k} \rangle|^2}{(E_q - E_k - \Delta E_k)^2 + \Gamma_k^2} \quad (2.75)$$

exhibiting the typical bell-shaped resonance behavior with a peak at $E_q = E_k + \Delta E_k$ and a width Γ_k . And the transition rate, i.e., the transition probability per unit time, is obtained by dividing $P_{q \leftarrow k}(\infty)$ by lifetime \hbar/Γ_k

$$p_{q \leftarrow k} = \frac{|\langle \mathbf{q} | V' | \mathbf{k} \rangle|^2}{\hbar} \frac{\Gamma_k}{(E_q - E_k - \Delta E_k)^2 + \Gamma_k^2} \quad (2.76)$$

This is almost the same as we obtained before, see Eq. (2.35), where the lifetime was introduced phenomenologically. Note that in deriving Eq. (2.35), the effect of V' on the eigenvalue of state $|\mathbf{k}\rangle$, i.e., ΔE_k of Eq. (2.70), was neglected.

Let us have a close look at Γ_k . By Eq. (2.71), and when $\langle \mathbf{q} | V' | \mathbf{k} \rangle$ is constant in a small energy range ΔE in the neighborhood of E_q ,

$$\begin{aligned} \Gamma_k &= \pi \sum_{\mathbf{q} \neq \mathbf{k}} |\langle \mathbf{q} | V' | \mathbf{k} \rangle|^2 \delta(E_q - E_k) \\ &\approx \pi |\langle \mathbf{q} | V' | \mathbf{k} \rangle|^2 \int_{\mathbf{q} \neq \mathbf{k}} \delta(E_q - E_k) N_{\text{DOS}}(E_q) dE_q \\ &= \pi |\langle \mathbf{q}_k | V' | \mathbf{k} \rangle|^2 N_{\text{DOS}}(E_{q_k}) \end{aligned} \quad (2.77)$$

$N_{\text{DOS}}(E_q)$ is the density of states. Here $|\mathbf{q}_k\rangle$ is such a representative state that its eigenvalue equals E_k , $N_{\text{DOS}}(E_{q_k})$ is the density of these states. Notice that $|\mathbf{q}_k\rangle \neq |\mathbf{k}\rangle$!

2.3 Time-Dependent Perturbation

We now discuss the scattering processes involving energy dissipation. The processes include both the absorption and emission of energy by the electron to and from its surrounding environment.

We consider an external perturbation in the form of

$$V'(t) = 2V' \cos(\omega t) = V' e^{i\omega t} + V' e^{-i\omega t} \quad (2.78)$$

where V' and ω are time-independent. Rather importantly, V' and ω are physical quantities so that they are real.

By Eq. (2.55), and for $\mathbf{q} \neq \mathbf{k}$,

$$\langle \mathbf{q} | \hat{T}(t) | \mathbf{k} \rangle = -\frac{i}{\hbar} \langle \mathbf{q} | V' | \mathbf{k} \rangle \int_0^t [e^{i(E_q - E_k - \hbar\omega)t'/\hbar} + e^{i(E_q - E_k + \hbar\omega)t'/\hbar}] \langle \mathbf{k} | \hat{T}(t') | \mathbf{k} \rangle dt' \quad (2.79)$$

And the probability that the electron remains in its initial state,

$$\begin{aligned} \frac{\partial}{\partial t} \langle \mathbf{k} | \hat{T}(t) | \mathbf{k} \rangle &= -\frac{i}{\hbar} \sum_{\mathbf{q} \neq \mathbf{k}} [e^{i(E_k - E_q - \hbar\omega)t/\hbar} + e^{i(E_k - E_q + \hbar\omega)t/\hbar}] \langle \mathbf{k} | V' | \mathbf{q} \rangle \langle \mathbf{q} | \hat{T}(t) | \mathbf{k} \rangle \\ &\quad - \frac{2i}{\hbar} \cos(\omega t) \langle \mathbf{k} | V' | \mathbf{k} \rangle \langle \mathbf{k} | \hat{T}(t) | \mathbf{k} \rangle \end{aligned} \quad (2.80)$$

By inserting Eq. (2.79) into the above equation and following the same considerations when solving Eq. (2.62), we expect to have

$$\int_0^t e^{i(E_q - E_k \pm \hbar\omega)(t'-t)/\hbar} dt' \implies \frac{\hbar}{i(E_q - E_k - i\Gamma_k \pm \hbar\omega)} \quad (2.81)$$

which will be most significant when $E_q - E_k \pm \hbar\omega = 0$. The physics is straightforward that the electron, initially occupying state $|\mathbf{k}\rangle$, will transit to another state $|\mathbf{q}\rangle$

after absorbing a quasi particle $\hbar\omega$ of the perturbation field described by Eq. (2.78), when $E_q - E_k - \hbar\omega = 0$, whereas for $E_q - E_k + \hbar\omega = 0$, the electron will emit $\hbar\omega$ while transiting from $|\mathbf{k}\rangle$ to $|\mathbf{q}\rangle$. The probabilities of other transition processes are weak. There is another term

$$\int_0^t e^{i(E_q - E_k)(t' - t)/\hbar} \cos[\omega(t' + t)] dt' \rightarrow 0 \quad (2.82)$$

since (1) we now study the case of $\omega \neq 0$ so that $E_q \neq E_k$ when expression (2.81) is significant, and (2) $\cos[\omega(t' + t)]$ oscillates rapidly.

Furthermore, in practical systems and applications, the effective range of $\hbar\omega$ is at least in the order of eV (photodetections and light-emitting devices in the visible optical range) so that $1/\omega$ is less than 10^{-12} s, while the time scale of most first-order perturbation processes is in the range of 10^{-15} . In other words, we consider the time duration of $\omega t \rightarrow 0$ so that $\cos(\omega t) = 1$ in Eq. (2.80). Thus

$$\frac{\partial}{\partial t} \langle \mathbf{k} | \hat{T}(t) | \mathbf{k} \rangle = - \left(\frac{i}{\hbar} \sum_{q \neq k} \frac{|\langle \mathbf{q} | V' | \mathbf{k} \rangle|^2}{E_q - E_k \pm \hbar\omega - i\Gamma_k} - \frac{2i}{\hbar} \langle \mathbf{k} | V' | \mathbf{k} \rangle \right) \langle \mathbf{k} | \hat{T}(t) | \mathbf{k} \rangle \quad (2.83)$$

which leads to the solution of

$$\langle \mathbf{k} | \hat{T}(t) | \mathbf{k} \rangle = \exp \left(-\frac{\Gamma_k}{\hbar} t - \frac{i \Delta E_k}{\hbar} t \right) \quad (2.84)$$

exactly the same as Eq. (2.68) but different parameters

$$\begin{aligned} \Gamma_k &= \sum_{q \neq k} |\langle \mathbf{q} | V' | \mathbf{k} \rangle|^2 \frac{\Gamma_k}{(E_k - E_q \pm \hbar\omega)^2 + \Gamma_k^2} \\ \Delta E_k &= \langle \mathbf{k} | V' | \mathbf{k} \rangle + \sum_{q \neq k} |\langle \mathbf{q} | V' | \mathbf{k} \rangle|^2 \frac{E_k - E_q \pm \hbar\omega}{(E_k - E_q \pm \hbar\omega)^2 + \Gamma_k^2} \end{aligned} \quad (2.85)$$

It is easy to see that Eqs. (2.69) and (2.70) are retrieved when $\omega \rightarrow 0$. Note that when $\omega = 0$, the external perturbation expression of Eq. (2.78) is $2V'$, while it is V' for Eqs. (2.69) and (2.70).

By Eq. (2.79),

$$\langle \mathbf{q} | \hat{T}(t) | \mathbf{k} \rangle = \langle \mathbf{q} | V' | \mathbf{k} \rangle \frac{1 - e^{i(E_q - E_k - \Delta E_k \pm \hbar\omega + i\Gamma_k)t/\hbar}}{E_q - E_k - \Delta E_k \pm \hbar\omega + i\Gamma_k} \quad (2.86)$$

so that when $t \rightarrow \infty$,

$$\langle \mathbf{q} | \hat{T}(\infty) | \mathbf{k} \rangle = \frac{\langle \mathbf{q} | V' | \mathbf{k} \rangle}{E_q - E_k - \Delta E_k \pm \hbar\omega + i\Gamma_k} \quad (2.87)$$

Finally we obtain the total transition probability

$$P_{q \leftarrow k} = |\langle \mathbf{q} | \hat{T}(\infty) | \mathbf{k} \rangle|^2 = \left| \frac{\langle \mathbf{q} | V' | \mathbf{k} \rangle}{E_q - E_k - \Delta E_k \pm \hbar\omega + i\Gamma_k} \right|^2 \quad (2.88)$$

and the transition rate $p_{q \leftarrow k} = P_{q \leftarrow k} / (\hbar / \Gamma_k)$

$$p_{q \leftarrow k} = \frac{|\langle \mathbf{q} | V' | \mathbf{k} \rangle|^2}{\hbar} \frac{\Gamma_k}{(E_q - E_k - \Delta E_k \pm \hbar\omega)^2 + \Gamma_k^2} \quad (2.89)$$

Equations (2.85) and (2.89) will be extensively used in this book. The “−” sign in “±” in the above equation is normally referred as the absorption of a quasi particle $\hbar\omega$, and the “+” sign as the emission of $\hbar\omega$.

2.4 Acceleration Theorems

We now take a close look at the properties of electrons transporting in a semiconductor crystal. As we learned before, the state of an electron in a periodic crystal is described the Bloch theorem. By substituting the Bloch solution of

$$\Psi_{nk}(\mathbf{r}) = e^{i\mathbf{k} \cdot \mathbf{r}} u_{nk}(\mathbf{r}) \quad (2.90)$$

into the Hamiltonian

$$H_0 \Psi(\mathbf{r}) = \left[\frac{-\hbar^2 \nabla^2}{2m_0} + V(\mathbf{r}) \right] \Psi(\mathbf{r}) = E \Psi(\mathbf{r}) \quad (2.91)$$

for an electron in the periodic lattice potential

$$V(\mathbf{r} + \mathbf{R}) = V(\mathbf{r}) \quad (2.92)$$

where \mathbf{R} is any lattice vector, we find that $u_{nk}(\mathbf{r})$ is determined by the following eigenvalue equation

$$\left[\frac{\hbar^2}{2m_0} \left(\frac{1}{i} \nabla + \mathbf{k} \right)^2 + V(\mathbf{r}) \right] u_{nk}(\mathbf{r}) = E_{nk} u_{nk}(\mathbf{r}) \quad (2.93)$$

with boundary condition

$$u_{nk}(\mathbf{r}) = u_{nk}(\mathbf{r} + \mathbf{R}) \quad (2.94)$$

For the sake of the following discussion, we define

$$H_k \equiv \frac{\hbar^2}{2m_0} \left(\frac{1}{i} \nabla + \mathbf{k} \right)^2 + V(\mathbf{r}) \quad (2.95)$$

Because of the periodic boundary condition we can regard Eq. (2.93) as an eigenvalue problem restricted to a single cell of the crystal.

We expand the energy dispersion relation $E_{n\mathbf{k}+\mathbf{q}}$ of the Bloch electron around \mathbf{k}

$$E_{n\mathbf{k}+\mathbf{q}} = E_{n\mathbf{k}} + \sum_i \frac{\partial E_{n\mathbf{k}}}{\partial k_i} q_i + \frac{1}{2} \sum_{ij} \frac{\partial^2 E_{n\mathbf{k}}}{\partial k_i \partial k_j} q_i q_j + O(q^3) \quad (2.96)$$

On the other hand, since $E_{n\mathbf{k}+\mathbf{q}}$ is the eigenvalue of $H_{\mathbf{k}+\mathbf{q}}$, we write

$$\begin{aligned} H_{\mathbf{k}+\mathbf{q}} &= \frac{\hbar^2}{2m_0} \left(\frac{1}{i} \nabla + \mathbf{k} + \mathbf{q} \right)^2 + V(\mathbf{r}) \\ &= H_{\mathbf{k}} + \frac{\hbar^2}{m_0} \mathbf{q} \cdot \left(\frac{1}{i} \nabla + \mathbf{k} \right) + \frac{\hbar^2 q^2}{2m_0} \\ &= H_{\mathbf{k}} + V' + \frac{\hbar^2 q^2}{2m_0} \end{aligned} \quad (2.97)$$

where

$$V' = \frac{\hbar^2}{m_0} \mathbf{q} \cdot \left(\frac{1}{i} \nabla + \mathbf{k} \right) \quad (2.98)$$

By the first-order perturbation theory, i.e., Eq. (2.70), we find that the energy correction due to \mathbf{q} in the Hamiltonian Eq. (2.97) is

$$\langle u_{n\mathbf{k}} | V' | u_{n\mathbf{k}} \rangle = \int u_{n\mathbf{k}}^* \frac{\hbar^2}{m_0} \mathbf{q} \cdot \left(\frac{1}{i} \nabla + \mathbf{k} \right) u_{n\mathbf{k}} \mathbf{d}\mathbf{r} \quad (2.99)$$

Here we have only kept the term linear in \mathbf{q} , while the q^2 -term in Eq. (2.97) is related to the effective mass (see Chap. 1). We expect that the above term is equivalent to the second term on the right side of Eq. (2.96), i.e.,

$$\sum_i \frac{\partial E_{n\mathbf{k}}}{\partial k_i} q_i = \int u_{n\mathbf{k}}^* \frac{\hbar^2}{m_0} \mathbf{q} \cdot \left(\frac{1}{i} \nabla + \mathbf{k} \right) u_{n\mathbf{k}} \mathbf{d}\mathbf{r} \quad (2.100)$$

It is easy to see that the above equation can be re-formatted as

$$\frac{\partial E_{n\mathbf{k}}}{\partial \mathbf{k}} = \frac{\hbar^2}{m_0} \int u_{n\mathbf{k}}^* \left(\frac{1}{i} \nabla + \mathbf{k} \right) u_{n\mathbf{k}} \mathbf{d}\mathbf{r} \quad (2.101)$$

Expressing the above equation in terms of the Bloch functions $\Psi_{n\mathbf{k}}(\mathbf{r})$, we have

$$\frac{1}{\hbar} \frac{\partial E_{n\mathbf{k}}}{\partial \mathbf{k}} = - \int \Psi_{n\mathbf{k}}^* \frac{i\hbar \nabla}{m_0} \Psi_{n\mathbf{k}} \mathbf{d}\mathbf{r} \quad (2.102)$$

Since $-i\hbar \nabla / m_0$ is the velocity operator

$$\mathbf{v} = \frac{\mathbf{p}}{m_0} = \frac{-i\hbar \nabla}{m_0}$$

it is thus established by Eq. (2.102),

$$\mathbf{v}_{nk} = \frac{1}{\hbar} \frac{\partial E_{nk}}{\partial \mathbf{k}} \quad (2.103)$$

is the mean velocity of an electron in Bloch state $\Psi_{nk}(\mathbf{r})$. This is the first acceleration theorem about the electron in a Bloch state. It indicates that an electron in a Bloch state has a non-vanishing mean velocity, the electron moves forever without degradation of its mean velocity, in spite of the interaction of the electron with the fixed lattice of ions.

We now introduce the second acceleration theorem which states that in a steady applied electric field, \mathbf{E} , the acceleration of an electron in a periodic lattice $V(\mathbf{r} + \mathbf{R}) = V(\mathbf{r})$ is described by

$$\hbar \dot{\mathbf{k}} = -e\mathbf{E} \quad (2.104)$$

We write the total Hamiltonian of the electron as $H = H_0 + V'(\mathbf{r})$, where

$$H_0 = \frac{\mathbf{p}^2}{2m_0} + V(\mathbf{r}), \quad V'(\mathbf{r}) = -\mathbf{F} \cdot \mathbf{r} \quad (2.105)$$

m_0 is the free electron mass, H_0 is the original Hamiltonian of the electron in the lattice for which we have the Bloch solution

$$H_0 \Psi_{nk}(\mathbf{r}) = E_{nk} \Psi_{nk}(\mathbf{r}), \quad \Psi_{nk}(\mathbf{r}) = \frac{1}{\sqrt{N}} e^{i\mathbf{k} \cdot \mathbf{r}} u_{nk}(\mathbf{r})$$

where N is the number of unit cells. $\mathbf{F} = -e\mathbf{E}$ is the force on an electron due to the applied electric field \mathbf{E} which induces $V'(\mathbf{r})$ the extra potential energy in the above equation. Note that $\mathbf{E}(\mathbf{r}) = -\nabla U(\mathbf{r})$ so that $U(\mathbf{r}) = -\mathbf{E} \cdot \mathbf{r}$ when \mathbf{E} is constant. Here $U(\mathbf{r})$ is the electric potential. The potential energy $V'(\mathbf{r})$ of an electron in an electric potential $U(\mathbf{r})$ is $-eU(\mathbf{r})$, resulting in the expression of $V'(\mathbf{r})$ in Eqs. (2.105).

Since

$$\begin{aligned} \nabla_{\mathbf{k}} \Psi_{nk}(\mathbf{r}) &= i\mathbf{r} e^{i\mathbf{k} \cdot \mathbf{r}} u_{nk}(\mathbf{r}) + e^{i\mathbf{k} \cdot \mathbf{r}} \nabla_{\mathbf{k}} u_{nk}(\mathbf{r}) \\ &= i\mathbf{r} \Psi_{nk}(\mathbf{r}) + e^{i\mathbf{k} \cdot \mathbf{r}} \nabla_{\mathbf{k}} [e^{-i\mathbf{k} \cdot \mathbf{r}} \Psi_{nk}(\mathbf{r})] \end{aligned} \quad (2.106)$$

Multiple the above expression by $i\mathbf{F}$,

$$i\mathbf{F} \cdot \nabla_{\mathbf{k}} \Psi_{nk}(\mathbf{r}) = -\mathbf{F} \cdot \mathbf{r} \Psi_{nk}(\mathbf{r}) + i e^{i\mathbf{k} \cdot \mathbf{r}} \mathbf{F} \cdot \nabla_{\mathbf{k}} e^{-i\mathbf{k} \cdot \mathbf{r}} \Psi_{nk}(\mathbf{r}) \quad (2.107)$$

i.e.,

$$i\mathbf{F} \cdot \nabla_{\mathbf{k}} = -\mathbf{F} \cdot \mathbf{r} + i e^{i\mathbf{k} \cdot \mathbf{r}} \mathbf{F} \cdot \nabla_{\mathbf{k}} e^{-i\mathbf{k} \cdot \mathbf{r}} \quad (2.108)$$

by which we rewrite the total Hamiltonian of the electron as

$$H = H_0 + H_{\mathbf{F}} + i\mathbf{F} \cdot \nabla_{\mathbf{k}} \quad (2.109)$$

where

$$H_F = -ie^{i\mathbf{k}\cdot\mathbf{r}} \mathbf{F} \cdot \nabla_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{r}} \quad (2.110)$$

We first have a look at H_F by calculating the transition rate from an initial Bloch state $|n\mathbf{k}\rangle$ to another Bloch state $|m\mathbf{q}\rangle$ using the perturbation theory in Sect. 2.2. By Eq. (2.76), we first calculate the transition matrix

$$\begin{aligned} A_{mqnk} &= \langle m\mathbf{q} | H_F | n\mathbf{k} \rangle = \langle m\mathbf{q} | -ie^{i\mathbf{k}\cdot\mathbf{r}} \mathbf{F} \cdot \nabla_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{r}} | n\mathbf{k} \rangle \\ &= -\frac{i}{N} \int e^{i(\mathbf{k}-\mathbf{q})\cdot\mathbf{r}} u_{m\mathbf{q}}^*(\mathbf{r}) \mathbf{F} \cdot \nabla_{\mathbf{k}} u_{n\mathbf{k}}(\mathbf{r}) d\mathbf{r} \end{aligned} \quad (2.111)$$

Since $u_{n\mathbf{k}}(\mathbf{r})$ is periodic with respect to unit cells, $\nabla_{\mathbf{k}} u_{n\mathbf{k}}(\mathbf{r})$ is also periodic. The integration over \mathbf{r} in the above equation can therefore be separated into an integration over \mathbf{r} within one unit cell and a summation over all unit cells involved in the semiconductor

$$A_{mqnk} = -\frac{i}{N} \sum_{\ell} e^{i(\mathbf{k}-\mathbf{q})\cdot\mathbf{R}_{\ell}} \int_{\text{cell}} u_{m\mathbf{q}}^*(\mathbf{R}_{\ell} + \mathbf{r}) \mathbf{F} \cdot \nabla_{\mathbf{k}} u_{n\mathbf{k}}(\mathbf{R}_{\ell} + \mathbf{r}) d\mathbf{r} \quad (2.112)$$

where \mathbf{R}_{ℓ} is the lattice vector of unit cell ℓ . A few further mathematical manipulations lead to

$$\begin{aligned} A_{mqnk} &= -\frac{i}{N} \sum_{\ell} e^{i(\mathbf{k}-\mathbf{q})\cdot\mathbf{R}_{\ell}} \mathbf{F} \cdot \nabla_{\mathbf{k}} \left[\int_{\text{cell}} u_{m\mathbf{q}}^*(\mathbf{r}) u_{n\mathbf{k}}(\mathbf{r}) d\mathbf{r} \right] \\ &= -\frac{i}{N} \mathbf{F} \cdot \nabla_{\mathbf{k}} \left[\int_{\text{cell}} u_{m\mathbf{q}}^*(\mathbf{r}) u_{n\mathbf{k}}(\mathbf{r}) d\mathbf{r} \right] \sum_{\ell} e^{i(\mathbf{k}-\mathbf{q})\cdot\mathbf{R}_{\ell}} \\ &= -i \mathbf{F} \cdot \nabla_{\mathbf{k}} \left[\int_{\text{cell}} u_{m\mathbf{q}}^*(\mathbf{r}) u_{n\mathbf{k}}(\mathbf{r}) d\mathbf{r} \right] \delta_{\mathbf{q},\mathbf{k}} \end{aligned} \quad (2.113)$$

In other words, H_F does not mix electron states of different \mathbf{k} . Furthermore, it is easy to see that

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

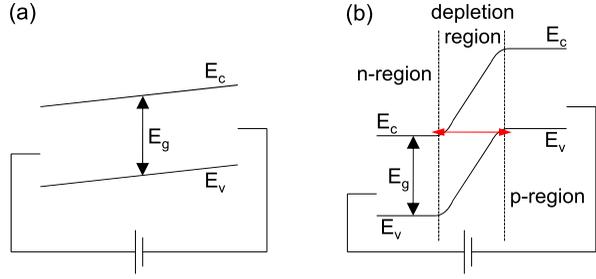
so that A_{mqnk} is nonzero only when $m \neq n$, which indicates that H_F mixes only states with the same \mathbf{k} but from different bands. In other words, H_F induces inter-band mixing.

Note that

$$\int_{\text{cell}} u_{m\mathbf{k}}^*(\mathbf{r}) u_{n\mathbf{k}}(\mathbf{r}) d\mathbf{r}$$

is \mathbf{k} dependent when $m \neq n$. See optical properties of semiconductor materials in late chapters.

Fig. 2.3 (a) Schematic band structure of bulk semiconductor under a weak electric field. (b) Electric-field-induced interband transitions in the depletion region of a $p-n$ junction



By Eqs. (2.76), (2.77), the transition of an electron from its initial valence band state (v) to a final conduction band state (c) induced by H_F

$$p_{ck \leftarrow vk} = \frac{|A_{ckvk}|^2}{\hbar} \frac{\Gamma_{vk}}{(E_{ck} - E_{vk} - \Delta E_{vk})^2 + \Gamma_{vk}^2} \quad (2.114)$$

$$\Gamma_{vk} \approx \pi \int |A_{ckvk}|^2 \delta(E_{ck} - E_{vk}) N_{\text{DOS}}(E_{ck}) dE_{ck}$$

Since the conduction band is separated from the valence band by the energy bandgap, $E_{ck} - E_{vk} \neq 0$, $\Gamma_{vk} \rightarrow 0$ and thus $p_{ck \leftarrow vk} = 0$.

The above conclusion is only valid for bulk semiconductor under weak electric field, see Fig. 2.3(a). Things will be totally different when the applied field is very strong. For example, the electric field can induce interband transitions in the depletion region of a pn junction, see Fig. 2.3(b).

For semiconductor bulk materials, we can therefore neglect H_F and define a set of functions $\xi_{\mathbf{k}}(\mathbf{r}, t)$ as solutions to the time-dependent Schrödinger equation of our electron in electric field \mathbf{E}

$$i\hbar \frac{\partial \xi_{\mathbf{k}}(\mathbf{r}, t)}{\partial t} = (H_0 + i\mathbf{F} \cdot \nabla_{\mathbf{k}}) \xi_{\mathbf{k}}(\mathbf{r}, t) \quad (2.115)$$

We do a perturbation analysis by assuming a solution with $\xi_{\mathbf{k}}(\mathbf{r}, t)$ confined to one band, say band n , with a zero-order eigen energy $E_{n\mathbf{k}}$

$$\xi_{n\mathbf{k}}(\mathbf{r}, t) = e^{-iE_{n\mathbf{k}}t/\hbar} \tilde{\xi}_{n\mathbf{k}}(\mathbf{r}) \quad (2.116)$$

differentiate it with respect to t then multiply with $i\hbar$

$$i\hbar \frac{\partial \tilde{\xi}_{n\mathbf{k}}(\mathbf{r}, t)}{\partial t} = \left(E_{n\mathbf{k}} + i\hbar \frac{\partial \mathbf{k}}{\partial t} \cdot \nabla_{\mathbf{k}} \right) \tilde{\xi}_{n\mathbf{k}}(\mathbf{r}, t) \quad (2.117)$$

Comparing Eq. (2.115) with Eq. (2.117) we see

$$\hbar \frac{\partial \mathbf{k}}{\partial t} = \hbar \dot{\mathbf{k}} = \mathbf{F} \quad (2.118)$$

which is the second acceleration theorem. It is just what can be expected in classical mechanics for a particle of charge $-e$ and momentum $\hbar\mathbf{k}$.

We note that the above discussion is performed in the absence of magnetic fields. It is more difficult to treat rigorously the motion of a lattice electron in a magnetic field. For electrons in non-degenerate bands and not-too-strong magnetic fields the result of the detailed calculations is that the equation of motion may be generalized to

$$\mathbf{F} = -e(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \quad (2.119)$$

where \mathbf{E} and \mathbf{B} are electric and magnetic fields.

We have discussed the energy band structure of the electrons in pure and bulk semiconductor in Chap. 1 and understand that at zero temperature (no thermal excitation. In reality, the temperature is not important in terms of thermal excitation between valence band and conduction since under almost all device function temperature, $k_B T$ is far less than the energy bandgap.) and in darkness (no optical excitation) the energy bands up to the valence band are completely filled while the energy bands from the conduction band are totally empty. An energy bandgap exists between the valence band and the conduction band. A consequence of the acceleration theorems is that semiconductor in its pure and bulk form (in darkness) is not conducting, for two major reasons. First, the interband transition induced by an electric field is negligibly small, see Eqs. (2.114) and Fig. 2.3(a). The change of the electron in the valence band from one Bloch state to another, Eq. (2.118), is actually suppressed by the Pauli exclusion principle since all Bloch states in the valence band are occupied. As discussed in Sect. 2.1, a complete occupied band does not contribute any net carrier transport current.

Metals, on the other hand, are totally different. The conduction band of a metal is partially filled. Electrons, especially those close to the Fermi level, can gain momentum from the electric field by Eq. (2.118) then transit to other unoccupied Bloch states above the Fermi level, resulting in a conducting carrier transport current.

2.5 Impurities and Fermi Level of Doped Semiconductor

By Sect. 2.4 we know that pure semiconductors have rarely been used for device applications by themselves. Semiconductors become useful when one uses the concept of doping to alter, in a controllable manner, the density of carriers that can carry current.

To increase the free carrier density, impurities known as dopants are introduced. The dopants are chosen from the periodic table so that they either have an extra electron in their outer shell compared to the host semiconductor, or have one less electron. The resulting dopant is called a donor or acceptor. The impurities of key interest in semiconductors are the point ones which create a local disturbance in the crystal structure. The effect of this crystal disturbance can be of two kinds:

1. The disturbance may create a potential profile which differs from the periodic lattice potential only over one or a few unit cells. This potential is deep and localized and the impurity is called a deep-level impurity.

2. The disturbance may create a long range potential disturbance which may extend over tens or more unit cells. Such impurities are called shallow-level impurities.

If a slice of semiconductor (take silicon for example) is placed in a furnace with a dense vapor of an impurity (e.g., boron), then it is found that the impurity atoms will tend to diffuse into the solid slice. The movement of the impurity atoms is described by Fick's law, which gives the relationship between the concentration $c(x)$, at a spatial position x , with the flux of impurities (number crossing unit area per unit time) $F(x)$ by

$$F(x) = -D \frac{\partial c(x)}{\partial x} \quad (2.120)$$

D is the diffusion coefficient which typically varies with temperature as

$$D = D_0 \exp\left(-\frac{eV_a}{k_B T}\right) \quad (2.121)$$

The significance of V_a is that it represents an energy that is required to displace a crystal atom by the impurity atom. For boron in silicon this energy appears to be close to 5 eV so that, even though $D_0 \sim 10^8$ m²/h, it still takes around ten hours for boron to diffuse a few micrometers into silicon at 1100 °C [3].

It is common to deposit the impurities for a short interval, then “turn off” the supply and continue the diffusion process to drive in the impurities.

By using oxide masking the process of diffusion can be carried out in selected areas. Silicon dioxide will prevent many standard impurities from diffusing into the semiconductor beneath. Thus SiO₂ masking can determine areas for diffusion. However, the diffusion proceeds in general as far laterally under the oxide as it goes down into the semiconductor.

When impurities are inserted into a layer of semiconductor by diffusion, the profile (the change of impurity density with depth) is limited to a Gaussian or error function. It is difficult to make extremely abrupt changes of concentrations. A proton-enhanced diffusion method fires energetic H⁺ ions into the semiconductor which damages the crystal lattice. It is found that this damage is closely confined to a well-defined plane below the surface of the material and that the depth of this plane can be controlled by the proton energy. In silicon, one needs about 100 keV for a depth of 1 μm and proportional accelerating voltages for other depths. The damage to the lattice permits impurities already present in the material to diffuse at lower temperatures than is usual (i.e., 800 °C for Si rather than above 1000 °C). Impurities can be moved into relatively well-defined regions with negligible diffusion of the conventional type occurring. More abrupt changes of concentration can be achieved and other types of profiling may be possible. The damage done to the crystal lattice by the proton beam has to be annealed out by heating for a short time-but not too long enough to cause significant diffusion.

Ion implantation, depicted schematically in Fig. 2.4, is a technique (in a high-vacuum system) closely linked to proton-enhanced diffusion but here the impurities are fired into the semiconductor in the form of a stream of ionized atoms- an ion

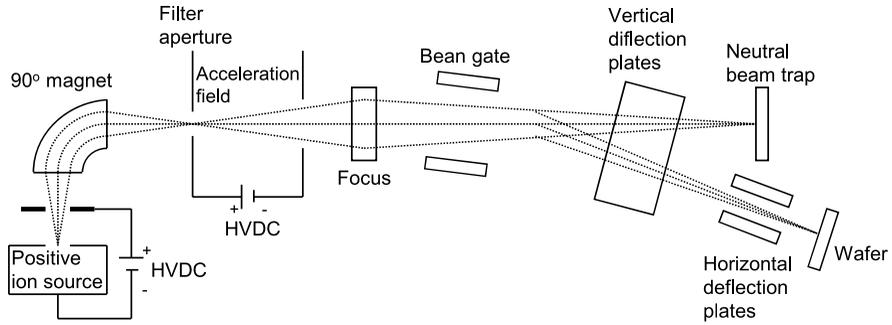


Fig. 2.4 Principle of an ion implanter

beam. This is formed by first creating a gas discharge which contains the required impurity atoms. In the discharge the atoms are stripped of their valence electrons (ionized), then drawn off from the gas discharge by suitable apertures and focusing electrodes to form an ion beam. If a gas which contains the desired impurity atoms cannot be found, one uses some inert gas in the discharge. The energetic ions in this gas can then knock out, or sputter, the required impurity atoms from some solid material which does contain them. The sputtered ions can be drawn off along with other ions. The required impurity ions are selected by passing the beam through a magnetic field which gives different curvatures in the trajectory of particles with different charge/mass ratios. The selected impurities can then pass through an aperture or slit to be focused and accelerated into the semiconductor material.

The ions, when implanted, cease to be ions and become impurity atoms but they are by no means necessarily correctly placed in substitutional sites. They usually have created considerable damage to the crystal structure on their way into the crystal. To activate the impurities in the desired substitutional sites and to reduce the damage one anneals the crystal at certain temperatures. At these temperatures, negligible diffusion of the impurity atoms occurs, but the dislocations of the atoms and the interstitial impurities can re-adjust under the action of thermal agitation. Annealing, like diffusion, may take several hours to achieve the desired result. In summary, impurities can in principle be implanted at required depth by different acceleration potentials. Special impurity profiles, especially shallow and sharp ones, can be made and are likely to find commercial use in the manufacture of specialized devices.

Associated with the impurities are new electronic states that are called impurity levels. The impurity levels can be produced in the regions of allowed bands (i.e., conduction or valence band) in which case their effects are minimal. However, the new levels could be produced in the bandgap region (the so-called bandgap levels) in which case they can greatly alter the electronic and optical properties of the semiconductors.

The wave functions of bandgap levels are no longer in the Bloch form. In general, the impurity level can be expressed in terms of the Bloch states. The deep level is highly localized so that its wave function is made up of a large number of Bloch

states. The wave function of a shallow long range impurity is extended in the real space and is made up of only a few Bloch states arising from the bandedge states.

Normally, the deep level impurities are to be avoided as much as possible in semiconductors, since an electron in such a level is “trapped” near the impurity. This trapped electron cannot participate in current flow very easily. However, in the case of indirect bandgap material, deep level impurities are often purposely introduced to increase the optical response of the material.

We are now concerned with the theory of the shallow donor and acceptor states associated with impurities in semiconductors. The ionization energies of these impurities are of the order of 40 meV in silicon, 10 meV in germanium and 6 meV in GaAs. Such energies are much less than the energy bandgap; thus it is reasonable to expect the impurity states to be formed from one-particle states of the appropriate band.

We treat a simplified model of a substitutional impurity in the semiconductor with a single spherical energy band by applying the envelope function theory, see Sect. 1.8. Let us be more specific that we substitute a Ga atom in GaAs by a Si atom. The single-particle Hamiltonian of the electron in such a system can be written as $H_0 + V'$, where H_0 is the single-particle Hamiltonian in perfect GaAs, and the modification V' of the lattice potential due to the substitution is

$$V'(\mathbf{r}) = \frac{Z_{\text{Si}}}{4\pi\epsilon_0|\mathbf{r} - \mathbf{R}_\ell|} - \frac{Z_{\text{Ga}}}{4\pi\epsilon_0|\mathbf{r} - \mathbf{R}_\ell|} = \frac{1}{4\pi\epsilon_0|\mathbf{r} - \mathbf{R}_\ell|} \quad (2.122)$$

where $Z_{\text{Si}} = 4$ and $Z_{\text{Ga}} = 3$ are the numbers of valence electrons in Si and Ga atoms, respectively. Here we assume that the substitutional Si atom locates at unit cell \mathbf{R}_ℓ . The above equation is written down when neglecting other valence electrons in the system. The correction to the above expression due to the valence electrons is normally approached by introducing a relative dielectric constant ϵ_r so that

$$V'(\mathbf{r}) = \frac{1}{4\pi\epsilon_0\epsilon_r|\mathbf{r} - \mathbf{R}_\ell|} \quad (2.123)$$

By Eqs. (1.139) and (1.144),

$$V'_{n'nji} = \left\langle a_{n'}(\mathbf{r} - \mathbf{R}_j) \left| \frac{1}{4\pi\epsilon_0\epsilon_r|\mathbf{r} - \mathbf{R}_\ell|} \right| a_n(\mathbf{r} - \mathbf{R}_i) \right\rangle \quad (2.124)$$

Since V' is most prominent in the proximity of $\mathbf{r} = \mathbf{R}_\ell$ while the Wannier function $a_n(\mathbf{r} - \mathbf{R}_i)$ is localized at unit cell \mathbf{R}_i , $V'_{n'nji}$ is significant only when $\mathbf{R}_j = \mathbf{R}_i = \mathbf{R}_\ell$. Furthermore, the integration of the Wannier functions at the same unit cell is nonzero when the Wannier functions are from the same energy band. Finally,

$$V'_{n'nji} \approx \frac{1}{4\pi\epsilon_0\epsilon_r|\mathbf{r} - \mathbf{R}_\ell|} \delta_{n'n} \delta_{ji,\ell} \quad (2.125)$$

Thus, by Eq. (1.152), we have the Schrödinger equation for the envelope function

$$\left[E_n(-i\nabla) - E \right] \psi_n(\mathbf{r}) + \frac{1}{4\pi\epsilon_0\epsilon_r|\mathbf{r}|} \psi_n(\mathbf{r}) = 0 \quad (2.126)$$

Consider a spherical conduction band

$$E_{ck} = E_c + \frac{\hbar^2 k^2}{2m^*}, \quad E_c(-i\nabla) = E_c - \frac{\hbar^2 \nabla^2}{2m^*} \quad (2.127)$$

Equation (2.126) becomes

$$\left(-\frac{\hbar^2 \nabla^2}{2m^*} - \frac{e^2}{4\pi\epsilon_0\epsilon_r r} \right) \psi(\mathbf{r}) = (E - E_c) \psi(\mathbf{r}) \quad (2.128)$$

This is just a hydrogen-like Schrödinger equation (1.8), and a series of energy levels are produced,

$$E = E_c - E_b^n = E_c - \frac{13.6}{n^2\epsilon_r^2} \frac{m^*}{m_0} \text{ eV} \quad (2.129)$$

where $n = 1, 2, \dots$. For silicon with ellipsoidal energy surface, the effective mass in the above equation is to be replaced by

$$\frac{3}{m^*} = \frac{2}{m_t^*} + \frac{1}{m_\ell^*} \quad (2.130)$$

where m_ℓ and m_t are longitudinal and transverse effective masses.

According to this simple picture of the donor impurity, the donor energy levels depend only upon the host crystal and not on the nature of the dopant. This is actually due to the fact that Eqs. (2.125), (2.126) are derived under the condition that V' does not vary much within one unit cell. This is only true for shallow impurities. More accurate theories for the donor levels must include the short-range correction within one unit cell related to the dopant impurity atom for a better agreement with the experiments.

Table 2.1 lists E_b^1 of shallow impurities (donors and acceptors) in common semiconductors. E_b^1 ($n = 1$) is normally referred to as the ionization or binding energy of the impurity. Here we observe that the impurity states are hydrogen-like, and loosely bound, largely because the dielectric constant of the semiconductor is high (about 10 for commonly used semiconductors). The small effective mass of carriers in the semiconductor further reduces the binding energy.

We now discuss the distribution of the extra electrons (missing electrons or holes) from donor (acceptor) impurities. As usual, the following assumptions are made:

1. Electrons *in one state* cannot be distinguished from one another.
2. Each state can be occupied by not more than two electrons with opposite spin, i.e., the Pauli exclusion principle.
3. Each impurity state can be occupied by just one electron. Here we only consider the donor impurities. Similar analysis can be made for the acceptor impurities.

Consider that we have n_j electrons in a state e_j whose density of states is g_j ($j = 1, 2, \dots$), the total number of distinguishable distributions W of these electrons

Table 2.1 Shallow impurity energy levels E_b^1 [meV] in semiconductors

Material	Donor	E_b^1	Acceptor	E_b^1
GaAs	Si	5.8	C	26
	Ge	6.0	Be	28
	S	6.0	Mg	28
	Sn	6.0	Si	35
Si	Li	33	B	45
	Sb	39	Al	67
	P	45	Ga	72
	As	54	In	160
Ge	Li	9.3	B	10
	Sb	9.6	Al	10
	P	12.0	Ga	11
	As	13.0	In	11

is

$$W = \prod_j \frac{g_j}{n_j!(g_j - n_j)!} \quad (2.131)$$

Now consider that the system contains N_D donors, whose energy state is e_D , of which N_{DX} are occupied by one electron each. It is reminded here that a donor can be occupied by only one electron. For these donors, we must know if they are correlated or not in order to calculate the number of distinguishable distributions. There are three different situations concerning the impurity correlation:

1. The density of impurities in the system is so low that the electron wave functions of the impurity state e_D are well localized around the donors and the overlappings of the wave functions at different donor sites are negligible.
2. The density of impurities is increased so that wave functions of e_D at different donor sites overlap with each other, forming impurity band.
3. The impurity band is widened due to the increase of the impurity concentration and eventually it merges with the conduction band (Mott transition).

Statistically however, the last two situations are equivalent. In other words, we only need to know whether or not the wave functions of the impurity electrons are localized. If the wave functions are localized, we can distinguish the electron at donor site i from the one at j , even their state energies are the same. But if the wave functions become delocalized, there is then no way to distinguish the electron at donor i from the one at j because the electrons are identical and they are in the same energy state.

First, let us assume that the electron wave functions of the N_D impurities in the system overlap with each other. Among N_D impurities, N_{DX} of them are neutral (they are occupied by electrons). The corresponding number of distinguishable

states, W_e , is

$$W_e = g_D^{N_{DX}} \frac{N_D!}{N_{DX}!(N_D - N_{DX})!} \quad (2.132)$$

Here we add a subscript index e to the number of states indicating that the corresponding impurity electron wave functions are “extended”.

The total number of distinguishable distributions of the system is given by

$$W_t = W_i W_e \quad (2.133)$$

By adopting the method of Lagrange undetermined multipliers and introducing the donor state, e_D , the sum

$$\Sigma = \ln W_t + \alpha \left(n - N_{DX} - \sum_j n_j \right) + \beta \left(U - N_{DX} e_D - \sum_j n_j e_j \right) \quad (2.134)$$

has the following restrains to fulfill

$$0 = \frac{d\Sigma}{dN_{DX}} = \frac{d\Sigma}{dn_j} \quad (2.135)$$

where n is the total number of electrons and U the total energy of the system, $j = 1, 2, \dots$

It is easy to derive that

$$\alpha = -\frac{E_f}{k_B T}, \quad \beta = \frac{1}{k_B T} \quad (2.136)$$

where E_f is the Fermi level. From the restrains of Eq. (2.135) the Fermi distribution function for the band electrons is obtained:

$$n_j = \frac{g_j}{\exp\left(\frac{e_j - E_f}{k_B T}\right) + 1} \quad (2.137)$$

It must be emphasized that the differentiation, $d(\ln N!)/dN$, which appears in Eq. (2.135), is approximated as

$$\frac{d(\ln N!)}{dN} \approx \ln N \quad (2.138)$$

in order to obtain Eq. (2.137). And Eq. (2.138) is valid only when N is a very large number where statistical analysis is meaningful.

It is usually always valid by saying that N_D and N_{DX} are large number so that Eq. (2.138) remains true, it should be however very cautious to use Eq. (2.137), because these results depend very much on another condition: that the wave functions of those electrons to neutralize impurities (impurity electrons) overlap with each other.

If the wave functions of the impurity electrons are localized and do not overlap with each other, and if there are n_m electrons at impurity site m , $m = 1, 2, \dots, N_D$, the numbers of states of the neutral impurities and of the whole system become

$$W_\ell = \prod_m g_D^{n_m}, \quad W_t = W_i W_\ell \quad (2.139)$$

where the subscript index ℓ indicating “localized”. For each impurity, n_m can be either 0 or 1, as mentioned earlier. Σ , which is defined in Eq. (2.134), now becomes

$$\Sigma = \ln W_t + \alpha \left(n - \sum_m n_m - \sum_j n_j \right) + \beta \left(U - \sum_m n_m e_D - \sum_j n_j e_j \right) \quad (2.140)$$

Since n_j are still large numbers, Eq. (2.140) still has the following restrains to fulfill

$$0 = \frac{d\Sigma}{dn_j} \quad (2.141)$$

so that Eq. (2.137) is obtained again. However, for n_m we can only calculate the change of Σ following the change in n_m from 0 to 1 and from 1 to 0:

$$\Delta \Sigma_{0 \rightarrow 1} = -\frac{E_f - e_D}{k_B T}, \quad \Delta \Sigma_{1 \rightarrow 0} = \frac{E_f - e_D}{k_B T} \quad (2.142)$$

Here the statistics tells us that if E_f is higher than e_D , it is better to occupy the impurity state, from $n_m = 0$ to $n_m = 1$, while for E_f lower than e_D , it is better to empty the impurity state, from $n_m = 1$ to $n_m = 0$ in order to minimize Σ .

We thus see that if the wave functions of the impurity electrons are localized and well separated from each other spatially, the Fermi level of the system is completely determined by the band electrons n_j , Eq. (2.137). The occupation of the impurity states depends then on the relative positions of the Fermi level E_f and the impurity state energy e_D . For shallow impurities, we can determine the critical impurity concentration, N_D^c where the wave functions of the impurity electrons begin to overlap with each other. By the hydrogen-like model, the effective Bohr radius of the shallow impurity state is

$$a = \frac{\epsilon_r}{m^*/m_0} a_0 \quad (2.143)$$

where $a_0 = 0.529 \text{ \AA}$ is the Bohr radius. On the other hand, the averaged distance between two adjacent impurity atoms is

$$a_D = \frac{1}{N_D^{1/3}} \quad (2.144)$$

For n -type GaAs, $\epsilon_r = 13.1$ and $m^* = 0.067m_0$, so that $a = 10.34 \text{ nm}$. It is easy to obtain the critical impurity concentration, $N_D^c = 9.05 \times 10^{17} \text{ cm}^{-3}$ when $a = a_D$.

For n -type Si, $\epsilon_r = 11.9$, $m^* = 0.43m_0$, so that $a = 1.46$ nm and $N_D^c = 3.28 \times 10^{20}$ cm $^{-3}$.

The treatment in the above consideration is of course over simplified. For example, the hydrogen-like model may not apply to the n -type Si because the resulting impurity Bohr radius is comparable with its lattice constant (0.543 nm).

2.6 Boltzmann Equation

Carrier transport can be characterized in terms of either classical or quantum physics, while for most cases, the generalized classical approach described by the Boltzmann transport equation is adequate enough.

In the formalism of the Boltzmann transport equations, the electron system is described by the distribution function $f_n(\mathbf{r}, \mathbf{k}, t)$, which gives the occupation probability of a state characterized by a band index n , the \mathbf{k} vector and the space vector \mathbf{r} . To calculate the distribution function $f_n(\mathbf{r}, \mathbf{k}, t)$ in given external fields, we examine its temporal behavior. We consider a group of electrons in the volume element $d\mathbf{r}d\mathbf{k}$ in the phase space. The total differential-quotient is caused by the scattering of electrons from one state to the other and vice versa,

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \dot{\mathbf{k}} \cdot \nabla_{\mathbf{k}} f + \dot{\mathbf{r}} \cdot \nabla_{\mathbf{r}} f = \left(\frac{\partial f}{\partial t} \right)_{\text{coll}} \quad (2.145)$$

In the steady state the local differential quotient is zero and we are left with the usual form of the Boltzmann transport equation for the electron system which allows the distribution function to be calculated for given external fields and known scattering mechanisms:

$$\dot{\mathbf{k}} \cdot \nabla_{\mathbf{k}} f + \dot{\mathbf{r}} \cdot \nabla_{\mathbf{r}} f = \left(\frac{\partial f}{\partial t} \right)_{\text{coll}} \quad (2.146)$$

On the left side of Eq. (2.146),

$$\dot{\mathbf{k}} = -\frac{e\mathbf{E}}{\hbar} \quad (2.147)$$

is the external force by the acceleration theorem, here we only consider the transport under the influence of an electric field \mathbf{E} . And

$$\dot{\mathbf{r}} = \frac{1}{\hbar} \nabla_{\mathbf{k}} E_{nk} \quad (2.148)$$

is the group velocity of the Bloch electron, where E_{nk} is the energy dispersion of the system under consideration.

If the scattering is induced by a quasi-particle, e.g., the phonon state, it is generally assumed that the quasi-particle system sets up an equilibrium state so quickly that we can ignore disturbances in the quasi-particle system. In this approximation

the form of the collision term on the right of Eq. (2.146) can be written in the simpler form

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = \int \frac{2d\mathbf{q}}{(2\pi)^3} \{P_{q,k}[1-f(\mathbf{k})]f(\mathbf{q}) - P_{k,q}[1-f(\mathbf{q})]f(\mathbf{k})\} \quad (2.149)$$

where $P_{k,q}$ is the probability of scattering from state \mathbf{k} to \mathbf{q} , see Eq. (2.88), and we have applied the Pauli exclusion principle.

Furthermore, physically collisions cannot change f_0 which implies that

$$\int \frac{2d\mathbf{q}}{(2\pi)^3} \{P_{q,k}[1-f_0(\mathbf{k})]f_0(\mathbf{q}) - P_{k,q}[1-f_0(\mathbf{q})]f_0(\mathbf{k})\} = 0 \quad (2.150)$$

It follows that

$$P_{q,k}[1-f_0(\mathbf{k})]f_0(\mathbf{q}) = P_{k,q}[1-f_0(\mathbf{q})]f_0(\mathbf{k}) \quad (2.151)$$

i.e.,

$$P_{q,k}e^{-E_k/k_B T} = P_{k,q}e^{-E_q/k_B T} \quad (2.152)$$

For elastic scattering processes, e.g., alloy and impurity scatterings, so is the acoustic phonon scattering because of the negligible acoustic phonon energy, the electron does not lose or gain energy,

$$P_{k,q} = P_{q,k} \quad (2.153)$$

Therefore Eq. (2.149) becomes much simplified:

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = \int \frac{2d\mathbf{q}}{(2\pi)^3} P_{q,k}[f(\mathbf{q}) - f(\mathbf{k})] \quad (2.154)$$

Equation (2.154) looks quite odd at the first sight. At equilibrium, the principle of detailed balance says

$$P_{q,k}[f_0(\mathbf{q}) - f_0(\mathbf{k})] = 0 \quad (2.155)$$

which means that $f_0(\mathbf{q}) = f_0(\mathbf{k})$ if $P_{k,q} \neq 0$. However, we know that there is a factor of $\delta[E_k - E_q]$ in $P_{k,q}$ (energy conservation) and $f_0(\mathbf{k}) = f_0(E_k)$. Therefore, Eq. (2.155) and thus Eq. (2.154) are perfectly alright.

For the optical phonon scattering, $P_{k,q}$ and $P_{q,k}$ are two different processes, one is the phonon-absorption while the other emits a phonon:

$$P_{k,q} = AN_{\text{op}}\delta(E_k - E_q - \hbar\omega_0)$$

for emitting a phonon when transiting from state \mathbf{k} to \mathbf{q} , where A is a constant, N_{op} is the phonon density of states, $\hbar\omega_0$ is the phonon energy, and

$$P_{q,k} = A(N_{\text{op}} + 1)\delta(E_q - E_k + \hbar\omega_0)$$

to absorb a phonon from state \mathbf{q} to \mathbf{k} .

Denoting s as the phonon wave vector and $P_- = AN_{\text{op}}$ and $P_+ = A(N_{\text{op}} + 1)$, Eq. (2.149) in full detail is now given by

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = \int \frac{2ds}{(2\pi)^3} \{ P_- [1 - f(\mathbf{k} - s)] f(\mathbf{k}) + P_+ [1 - f(\mathbf{k} + s)] f(\mathbf{k}) - P_- [1 - f(\mathbf{k})] f(\mathbf{k} + s) - P_+ [1 - f(\mathbf{k})] f(\mathbf{k} - s) \} \quad (2.156)$$

Therefore Eqs. (2.154)–(2.155) are derived again if we treat P_- and P_+ as two different scattering processes and Eq. (2.154) remains valid. Other inelastic processes can be treated similarly.

By Eq. (2.154) we then see that if there are several principal scattering mechanisms which are independent from each other, we can write

$$P_{\mathbf{k},\mathbf{q}} = \sum_i P_{i,\mathbf{k},\mathbf{q}} \quad (2.157)$$

If we are only interested in the linear response to the external force, the collision integral Eq. (2.149) can be approached by the relaxation-time approximation

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = -\frac{f - f_0}{\tau} \quad (2.158)$$

where τ is the relaxation time and f_0 is the distribution function at equilibrium. It is easy to see that if Eq. (2.157) holds true, we obtain the well-known Mathiessen rule:

$$\frac{1}{\tau} = \sum_i \frac{1}{\tau_i} \quad (2.159)$$

where τ_i is the relaxation time corresponding to scattering process $P_{i,\mathbf{k},\mathbf{q}}$.

For weak external fields, the distribution function can be approximated by

$$f(\mathbf{k}) \approx f_0(\mathbf{k}) + f_1(\mathbf{k}) \quad (2.160)$$

where $f_0(\mathbf{k}) \gg f_1(\mathbf{k})$. By substituting the above equation into Eq. (2.154) together with Eq. (2.158) we obtain the relaxation-time expression:

$$-\frac{f(\mathbf{k}) - f_0(\mathbf{k})}{\tau} = \int P_{\mathbf{q},\mathbf{k}} [f_1(\mathbf{q}) - f_1(\mathbf{k})] \frac{2d\mathbf{q}}{(2\pi)^3} \quad (2.161)$$

As a first-order approximation we can neglect $f_1(\mathbf{q})$ in the right side of the above equation to obtain the usual relaxation-time expression:

$$\frac{1}{\tau(\mathbf{k})} = \int P_{\mathbf{q},\mathbf{k}} \frac{2d\mathbf{q}}{(2\pi)^3} \quad (2.162)$$

Equation (2.162) is the usual definition of the total scattering rate.

2.7 Drift, Diffusion and Ballistic Transport

At a phenomenological level, electron transport is described by

$$j_i(\mathbf{r}) = -e \left[n(\mathbf{r}) v_{d,i} - \sum_j D_{ij} \frac{\partial n(\mathbf{r})}{\partial x_j} \right] \quad (2.163)$$

where $i, j = x, y, z$, $n(\mathbf{r})$ and \mathbf{j} are carrier density and current density, respectively, D_{ij} is the diffusion coefficient tensor. v_d is the drift velocity induced by an electric field.

For an dc electric field \mathbf{E} ,

$$\nabla_{\mathbf{r}} f = 0, \quad \frac{\partial f}{\partial t} = 0 \quad (2.164)$$

the Boltzmann transport equation becomes

$$-\frac{e\mathbf{E}}{\hbar} \cdot \nabla_{\mathbf{k}} f_0 = -\frac{f - f_0}{\tau} \quad (2.165)$$

together Eq. (1.40), we obtain

$$f = f_0 + \frac{e\mathbf{E} \cdot \nabla_{\mathbf{k}} f_0}{\hbar} \tau \quad (2.166)$$

The drift velocity is given as

$$\mathbf{v}_d \equiv \frac{1}{n} \int \mathbf{v}_{\mathbf{k}} f \frac{2d\mathbf{k}}{(2\pi)^3} = \frac{1}{n} \int \mathbf{v}_{\mathbf{k}} \frac{e\mathbf{E} \cdot \nabla_{\mathbf{k}} f_0}{\hbar} \tau \frac{2d\mathbf{k}}{(2\pi)^3} \quad (2.167)$$

where

$$\int \mathbf{v}_{\mathbf{k}} f_0 \frac{2d\mathbf{k}}{(2\pi)^3} = 0$$

since the group velocity of the system is zero at equilibrium. And

$$n = \int f_0 \frac{2d\mathbf{k}}{(2\pi)^3}$$

is the carrier concentration.

We can also formulate the drift velocity into the concept of carrier mobility such

$$\mathbf{v}_{d,i} = \sum_j \mu_{ij} E_j \quad (2.168)$$

where $\mathbf{E} = (E_x, E_y, E_z)$ is the electric field.

$$\mathbf{v}_d = \frac{1}{n} \int \mathbf{v}_{nk} \frac{e\mathbf{E} \cdot \nabla_{\mathbf{k}} E_{nk}}{\hbar} \frac{\partial f_0}{\partial E_{nk}} \tau \frac{2d\mathbf{k}}{(2\pi)^3} = \frac{1}{n} \int \mathbf{v}_{nk} e\mathbf{E} \cdot \mathbf{v}_{nk} \frac{\partial f_0}{\partial E_{nk}} \tau \frac{2d\mathbf{k}}{(2\pi)^3} \quad (2.169)$$

see Eq. (1.40) about the expression of the group velocity \mathbf{v}_{nk} of Bloch state $n\mathbf{k}$. More explicitly,

$$v_{d,x} = \frac{e}{n} \int (v_{nk,x}^2 E_x + v_{nk,x} v_{nk,y} E_y + v_{nk,x} v_{nk,z} E_z) \frac{\partial f_0}{\partial E_{nk}} \tau \frac{2d\mathbf{k}}{(2\pi)^3} \quad (2.170)$$

Similar expressions can be written down for $v_{d,y}$ and $v_{d,z}$.

For common semiconductors, the integration over $v_{nk,i} v_{nk,j}$ is zero when $i \neq j$ (symmetry consideration). Thus,

$$v_{d,i} = \frac{eE_i}{n} \int v_{nk,i}^2 \frac{\partial f_0}{\partial E_{nk}} \tau \frac{2d\mathbf{k}}{(2\pi)^3} \quad (2.171)$$

Comparing with Eq. (2.168), we obtain the expression for calculating the mobility

$$\mu_i = \frac{e}{n} \int v_{nk,i}^2 \frac{\partial f_0}{\partial E_{nk}} \tau \frac{2d\mathbf{k}}{(2\pi)^3} \quad (2.172)$$

For weak electric-fields, D and μ are electric field independent and verify the Einstein relation

$$D = \frac{\mu k_B T}{e} \quad (2.173)$$

At high electric fields, the Einstein relation fails and the diffusion coefficient becomes $D_{ij}(E)$.

Let us consider a material having a geometry of length L and a cross-section $w \times w$, see Fig. 2.5(a). When L and w are far larger than the mean free path ℓ , it is known by the Ohm's law that the resistance of such a material is

$$R = \frac{L}{w^2} \rho \quad (2.174)$$

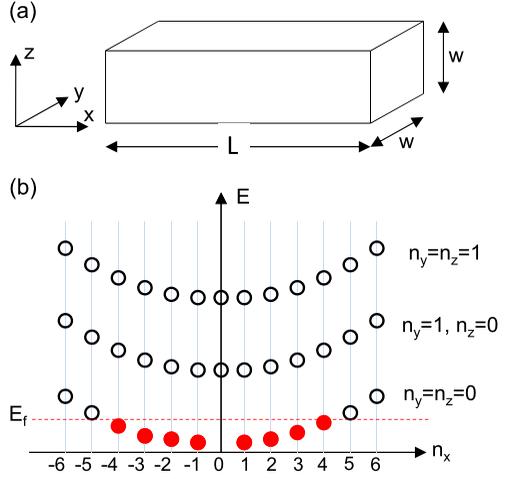
where ρ is the resistivity. The electron transport is characterized as drift-diffusion. Denote n as the density of the electrons in the electron system, the current density j and the conductivity σ are

$$j = -nev_{\text{average}} = \frac{ne^2\tau}{m^*} E, \quad \sigma = \frac{ne^2\tau}{m^*} \quad (2.175)$$

from which we can easily obtain the relationship between ρ and σ .

Now we reduce the dimension of the material so that both L and w become much smaller than ℓ . In this case, the electron will not experience any scattering and its transport is just a free flight. The situation is known as ballistic transport. Assume

Fig. 2.5 Electron transport.
(a) Geometric structure.
(b) Energy band diagram, where E_f denotes the Fermi level below which the states (solid dots) are occupied and above it the states (hollow) are empty



$L \gg w$ and the external electric field applies along the x axis. Consider the electron states within the material

$$k_x = \frac{n_x \pi}{L}, \quad k_y = \frac{n_y \pi}{w}, \quad k_z = \frac{n_z \pi}{w} \quad (2.176)$$

where n_x , n_y and n_z are integers. And the energy of the state is

$$E = \frac{\hbar^2}{2m^*} (k_x^2 + k_y^2 + k_z^2) = \frac{\hbar^2 \pi^2}{2m^*} \left(\frac{n_x^2}{L^2} + \frac{n_y^2}{w^2} + \frac{n_z^2}{w^2} \right) \quad (2.177)$$

which is plotted schematically in Fig. 2.5(b). Fill the system with electrons up to the Fermi level E_f . For simplicity we consider only the single band of $n_y = n_z = 0$. The time that an electron with charge $-e$ across length L at a speed of v is L/v , resulting in a current of

$$\frac{e}{L/v} = \frac{ev}{L} \quad (2.178)$$

The total current of all electrons is

$$I_{\text{left} \rightarrow \text{right}} = \sum_{\text{occupied } n_x > 0} \frac{2ev}{L} = \frac{2e}{L} \sum_{\text{occupied } n_x > 0} \frac{\hbar k_x}{m^*} = \frac{2\pi e \hbar}{m^* L^2} \sum_{\text{occupied } n_x > 0} n_x \quad (2.179)$$

In the above equation, a factor of “2” is introduced to account for the spin degeneracy.

Since L is relatively large so that we can convert the sum into integration by inserting $\delta n_x = dn_x = 1$

$$I_{\text{left} \rightarrow \text{right}} = \frac{2\pi e \hbar}{m^* L^2} \int_{\text{occupied}} n_x dn_x \quad (2.180)$$

Because of Eq. (2.177),

$$n_x dn_x = \frac{m^*}{\hbar^2(\pi/L)^2} dE \quad (2.181)$$

so that

$$I_{\text{left} \rightarrow \text{right}} = \frac{2e}{h} \int_{\text{occupied}} dE \quad (2.182)$$

where $h = 2\pi\hbar$.

The current from the right to the left can be similarly obtained. To induce a net nonzero current along the x direction, an external bias V is applied, which causes a potential energy change eV . Assume that the left end is grounded and the right end is biased V , the potential energy at the right end becomes $-eV$, and the net current

$$I_{\text{total}} = \frac{2e}{h} \left(\int dE_{\text{left} \rightarrow \text{right}} - \int dE_{\text{right} \rightarrow \text{left}} \right) = \frac{2e}{h} [E_f - (E_f - eV)] = \frac{2e^2}{h} V \quad (2.183)$$

where E_f is the Fermi level. The resistance of the ballistic transport is

$$R_{\text{ballistic}} = \frac{h}{2e^2} = 12.5 \text{ k}\Omega \quad (2.184)$$

which is a constant.

2.8 Carrier Scatterings

Finally we discuss three type of scattering mechanisms in the rest of this chapter, namely, the phonon scattering where the electron interacts with the lattice; two electrons interact with each other; and, an electron is scattered by an external force, e.g., a photon, or an impurity.

2.8.1 Phonon Scattering

When an ion in a compound semiconductor lattice structure, or an atom in an elemental semiconductor, is displaced by an amount of \mathbf{Q} from its lattice position \mathbf{R} ,

$$\mathbf{Q} = \sum_q \sqrt{\frac{\hbar}{2\rho V \omega_q}} (a_q + a_{-q}^+) e^{i\mathbf{q}\cdot\mathbf{r}} \boldsymbol{\Xi} \quad (2.185)$$

a deformation potential is induced

$$H' = \epsilon \frac{\partial \mathbf{Q}}{\partial \mathbf{r}} = \sum_{\ell j, \mathbf{q}} \sqrt{\frac{\hbar}{2\rho V \omega_q}} (a_q + a_{-q}^+) e^{i\mathbf{q}\cdot\mathbf{r}} \epsilon_{\ell j q \ell} \boldsymbol{\Xi}_j \quad (2.186)$$

where ρ is the material density, ω_q is the frequency of the ion displacement and \mathbf{q} is the wave vector, $\boldsymbol{\Xi}$ is the polarization of the ion displacement, ϵ is a tensor describing the shift of the electron band per unit deformation. a_q^+ is the creation operator of such a collective lattice displacement, which is called the phonon. a_q is the corresponding annihilation operator.

We consider a state which consists of one electron and the phonon background in the form of

$$\psi(\mathbf{k})\Phi(N_q)$$

where $\psi(\mathbf{k})$ is the electron Bloch function and $\Phi(N_q)$ is the phonon state of \mathbf{q} , N_q is the phonon density.

To calculate the transition probability from electron state \mathbf{k} to state \mathbf{k}' , we need to perform the integration

$$\frac{1}{N} u_{\mathbf{k}'}^* u_{\mathbf{k}} e^{i(\mathbf{k}-\mathbf{k}'\pm\mathbf{q})\cdot\mathbf{r}}$$

over the space, where N is the number of the unit cells. The integration can be divided into a sum over the lattice vector \mathbf{R} and an integration in the unit cell because $u_{\mathbf{k}}$ is translationally symmetric:

$$\sum_{\mathbf{R}} \frac{1}{N} e^{i(\mathbf{k}-\mathbf{k}'\pm\mathbf{q})\cdot\mathbf{R}} = \begin{cases} 1 & \text{if } \mathbf{k} - \mathbf{k}' \pm \mathbf{q} = \mathbf{G} \\ 0 & \text{otherwise} \end{cases} \quad (2.187)$$

$$p(\mathbf{k}, \mathbf{k}') = \int_{\text{cell}} u_{\mathbf{k}'}^* u_{\mathbf{k}} d\mathbf{r}$$

where \mathbf{G} is the reciprocal lattice. The terms corresponding to $\mathbf{G} = 0$ are called ‘‘normal’’ (‘‘N’’) terms, others are the ‘‘umklapp’’ (‘‘U’’) terms. $p(\mathbf{k}, \mathbf{k}')$ is the so-called overlap integral.

Since

$$a_q \Phi(N_q) = N_q \Phi(N_q - 1), \quad a_q^+ \Phi(N_q) = (N_q + 1) \Phi(N_q + 1) \quad (2.188)$$

which correspond to the emission and absorption of a phonon, the transition probability per unit time of one electron from \mathbf{k} to state \mathbf{k}' is

$$P_{\mathbf{k}, \mathbf{k}'} = \frac{\pi p^2(\mathbf{k}, \mathbf{k}')}{\rho V \omega_q} \begin{pmatrix} N_q + 1 \\ N_q \end{pmatrix} |\epsilon_{lj} q_l \Xi_j|^2 \frac{\gamma}{[\Omega \pm \omega_q]^2 + \gamma^2/4} \quad (2.189)$$

Here $\hbar\Omega = E(\mathbf{k}') - E(\mathbf{k})$. The upper and lower symbols refer to the phonon emission and absorption, respectively.

Let us discuss the multiple quantum well system grown along the z -direction. The system is based on a semiconductor material with lattice constant a . The thickness of the quantum well is ma , where m is the monolayer number of the well. The period of the system is Na . It is easy to see then $(N - m)a$ is the barrier thickness. Here we concentrate on a lattice-matched heterostructure. The electron-phonon interaction is

a very complicated quantity, which depends strongly on the electron energy as well as sample structure [4, 5]. Generally speaking, in a system whose lattice structure is defined as \mathbf{R}_j , the electron-phonon interaction Hamiltonian, H' , can be written as

$$H' = \sum_j \mathbf{Q}_j \cdot \nabla V(\mathbf{r} - \mathbf{R}_j) \quad (2.190)$$

where \mathbf{Q}_j is the displacement of the atom placed at lattice site \mathbf{R}_j , \mathbf{r} is the electron position, and V is the lattice potential. Let us consider the following phonon mode whose wave vector is \mathbf{q} . The imaginary part of \mathbf{q}

$$\text{Im}(\mathbf{q}) = \begin{cases} k_1 \mathbf{z}_0 & \text{when } z > 0 \\ -k_1 \mathbf{z}_0 & \text{otherwise} \end{cases} \quad (2.191)$$

where \mathbf{z}_0 is the unit vector in the z direction. Such a phonon mode is localized along the z direction but is extended in the xy plane. Using the periodic boundary condition in the region of $(-Na/2, Na/2)$, where N is the number of the lattice sites in one period (we assume that Na is considerably larger than the well width ma so that the boundary conditions do not affect the phonon in the well), the lattice displacement can be expressed by

$$\mathbf{Q}_j = A \sqrt{\frac{N(1 - e^{-k_1 a})}{2 - 2e^{-Nk_1 a/2}}} e^{i\mathbf{q} \cdot \mathbf{R}_j} \quad (2.192)$$

where A is a constant containing factors like the phonon polarization and normalization in the xy plane.

$$\sqrt{\frac{N(1 - e^{-k_1 a})}{2 - 2e^{-Nk_1 a/2}}}$$

is the normalization factor in the z direction. It is easy to see that

$$\sqrt{\frac{N(1 - e^{-k_1 a})}{2 - 2e^{-Nk_1 a/2}}} = 1 \quad (2.193)$$

when $k_1 = 0$, and Eq. (2.192) becomes the normal expression for the extended phonon.

The initial and final electronic states are expressed similarly

$$\psi_i = u \sqrt{\frac{ak_2}{1 - e^{-Nk_2 a}}} e^{i\mathbf{k}_i \cdot \mathbf{r}}, \quad \psi_f = u \sqrt{\frac{ak_3}{1 - e^{-Nk_3 a}}} e^{i\mathbf{k}_f \cdot \mathbf{r}} \quad (2.194)$$

where u is the Bloch function which is unit-cell periodic, and \mathbf{k}_i and \mathbf{k}_f are wave vectors of the initial and final electronic states,

$$\text{Im}(\mathbf{k}_i) = k_2 \mathbf{z}_0, \quad \text{Im}(\mathbf{k}_f) = k_3 \mathbf{z}_0$$

when $z > 0$; and for $z < 0$,

$$\text{Im}(\mathbf{k}_i) = -k_2 z_0, \quad \text{Im}(\mathbf{k}_f) = -k_3 z_0$$

The difference between the normalization factors in Eqs. (2.192), (2.194) is due to the fact that the lattice vector \mathbf{R}_i is discrete while the electronic coordinate \mathbf{r} is continuous. The transition probability between states is

$$P = \langle \psi_i | H' | \psi_f \rangle = \sqrt{\frac{Na^2 k_2 k_3 (1 - e^{-2k_1 a})}{8(1 - e^{-Nk_1 a/2})(1 - e^{-Nk_2 a})(1 - e^{-Nk_3 a})}} \times \sum_j e^{i\mathbf{q} \cdot \mathbf{R}_j} \int_{\text{cell}, j} e^{j(\mathbf{k}_i - \mathbf{k}_f) \cdot \mathbf{r}} |u|^2 \mathbf{A} \cdot \nabla V(\mathbf{r} - \mathbf{R}_j) d\mathbf{r} \quad (2.195)$$

For most applications, the doping levels in the electrodes of normal GaAs/AlGaAs systems are about $5 \times 10^{17} \text{ cm}^{-3}$, the corresponding Fermi level is about 25 meV. The doping level in the quantum well, e.g., in a double barrier resonant tunneling diode, is kept as low as possible so that the tunneling or other quantum processes will not be interfered by impurity scattering. The wave length at the Fermi level ($5 \times 10^{17} \text{ cm}^{-3}$ doping) is considerably large ($> 500 \text{ \AA}$) in a GaAs/AlGaAs system) compared with the lattice constant, e.g., 2.87 \AA for an (001) GaAs. For those electronic states whose energies are lower than the Fermi level (these are the states of interest because of the Fermi distribution factor), the wave lengths are even larger. Therefore, $\exp[i(\mathbf{k}_i - \mathbf{k}_f) \cdot \mathbf{r}]$ can be replaced by $\exp[i(\mathbf{k}_i - \mathbf{k}_f) \cdot \mathbf{R}_j]$. Since

$$\int_{\text{cell}, j} |u|^2 \mathbf{A} \cdot \nabla V(\mathbf{r} - \mathbf{R}_j) d\mathbf{r} = A \quad (2.196)$$

is the same for every unit cell, Eq. (2.195) becomes

$$P = A \sqrt{\frac{Na^2 k_2 k_3 (1 - e^{-2k_1 a})}{8(1 - e^{-Nk_1 a/2})(1 - e^{-Nk_2 a})(1 - e^{-Nk_3 a})}} \sum_j e^{i(\mathbf{q} + \mathbf{k}_i - \mathbf{k}_f) \cdot \mathbf{R}_j} \quad (2.197)$$

Due to the translational symmetry in the xy plane, the above equation is actually one-dimensional. Since

$$\sum_{n=-N/2}^{N/2} e^{in\alpha - |n\beta|} = \frac{2\beta}{\alpha^2 + \beta^2} \quad (2.198)$$

and let \mathbf{q} , \mathbf{k}_i and \mathbf{k}_f be real components of the wave vectors of the phonon, the initial and the final electronic states, respectively, in the z direction, we finally come

to a rather simple expression for the electron-phonon interaction:

$$P = A \sqrt{\frac{Nk_2k_3(1 - e^{-2k_1a})}{(1 - e^{-Nk_1a/2})(1 - e^{-Nk_2a})(1 - e^{-Nk_3a})}} \times \frac{|k_1 + k_2 - k_3|}{(\mathbf{q} + \mathbf{k}_i - \mathbf{k}_f)^2 + (k_1 + k_2 - k_3)^2} \quad (2.199)$$

When $k_1 = k_2 = k_3 = 0$,

$$\frac{|k_1 + k_2 - k_3|}{Na[(\mathbf{q} + \mathbf{k}_i - \mathbf{k}_f)^2 + (k_1 + k_2 - k_3)^2]} = \delta(\mathbf{q} + \mathbf{k}_i - \mathbf{k}_f) \quad (2.200)$$

so that Eq. (2.199) becomes

$$P = A\delta(\mathbf{q} + \mathbf{k}_i - \mathbf{k}_f) \quad (2.201)$$

which is exactly the expression of electron-phonon interaction for extended electron and phonon states [6].

Let us now discuss the following two cases. First we assume that there is one localized state involved.

By localization we mean the extension of the state $ma = 2/k$ is much smaller than Na . For the quantum well of the usual GaAs/AlGaAs double barrier resonant tunneling system, m is about 20 so that ma is about 50 Å. We have mentioned early that when the doping level in the electrodes is about $5 \times 10^{17} \text{ cm}^{-3}$, the corresponding Fermi level is about 25 meV and $1/k$ (either k_i or k_f) is longer than 500 Å. On the other hand, the most active phonons are the optical ones whose wave vector q is very small. Therefore $(\mathbf{q} + \mathbf{k}_i - \mathbf{k}_f)$ can be neglected when particles are localized.

Let $k_1 = 2/ma$ and $k_2 = k_3 = 0$. (It can be shown that the following conclusions remain intact if the values of k_1 , k_2 and k_3 are interchanged.) Since

$$1 - e^{-Nka} = Nka$$

when k is small, Eq. (2.199) becomes

$$P = \frac{A}{ak} \sqrt{\frac{1 - e^{-2ka}}{2N(1 - e^{-Nka/2})}} \quad (2.202)$$

where we have dropped off the subscripts from k . When $k = 2/ma$,

$$1 - e^{-2ka} \approx \frac{4}{m}$$

for large m and

$$1 - e^{-Nka/2} = 1 - e^{-N/m} \approx 1$$

when N/m is much larger than 1. Hence we obtain

$$P \propto A \sqrt{\frac{m}{N}} \quad (2.203)$$

The k 's are not exactly the same as $2/ma$, they depend on the state energy. When considering the electronic state, k decreases when the state energy increases (the state with higher energy penetrates more into the barrier region). Together with other constant from mathematical operation, the relation of proportionality (\propto) instead of equality ($=$) is used in Eq. (2.203) and later in Eqs. (2.205), (2.206).

It is concluded from Eq. (2.203) that when one state (either the phonon, or the initial or the final electronic state) is localized, the electron-phonon interaction is reduced when the extension of the localization, ma , is reduced. It was shown that the electron-phonon interaction is also reduced when one of the electronic states is evanescent [7]. The evanescent state is different from the localized one. It is only bounded at one end while the other end is free.

When the state becomes de-localized, i.e., when m becomes much large and is comparable with N , $(\mathbf{q} + \mathbf{k}_i - \mathbf{k}_f)$ cannot be neglected. By simple mathematical manipulation it is easy to see

$$P \propto A \quad (2.204)$$

Now we consider the situation when two or three states are localized. By similar mathematics as we used to discuss the case when one state is localized, we obtain for the case when two states are localized

$$P \propto \begin{cases} A & \text{for } |k_1 + k_2 - k_3| \neq 0 \\ A \frac{N}{m} & \text{otherwise} \end{cases} \quad (2.205)$$

For the case that all of the three states are localized, we obtain

$$P \propto \begin{cases} A \left(\frac{N}{m}\right)^{1/2} & \text{for } |k_1 + k_2 - k_3| \neq 0 \\ A \left(\frac{N}{m}\right)^{3/2} & \text{otherwise} \end{cases} \quad (2.206)$$

Quite different from the case when there is only one state localized, Eqs. (2.205), (2.206) show that the electron-phonon interaction is enhanced when more than one interacting states become localized. The strength of the electron-phonon interaction also increases when the extension of the localization is reduced.

The reader is reminded that the localization defined here is only along the z direction. The effect of two or three dimensional localization on the electron-phonon interaction can be discussed in a similar way. Such an effect could be very important for the novel quantum wire and quantum dot system. It can be expected from the above one-dimensional localization discussions that the dependence of the electron-phonon interaction on the localization extension can be much stronger.

2.8.2 Carrier-Carrier Interaction

Two main contributions to the carrier-carrier scattering can be easily identified:

- the individual carrier-carrier interaction via the screened Coulombic potential in the form of

$$V' = \frac{e^2}{4\pi\epsilon r} e^{-\beta r} \quad (2.207)$$

which accounts for two-body short-range interaction;

- the electron-plasmon interaction, which accounts for the collective long-range behavior of the electron gas.

Screened Carrier-Carrier Interaction The scattering probability of two electrons from state $(\mathbf{k}_1, \mathbf{k}_2)$ to $(\mathbf{k}_3, \mathbf{k}_4)$ is

$$\frac{|V'|^2}{\hbar^2} \frac{\gamma}{\Omega^2 + \gamma^2/4} \quad (2.208)$$

where

$$\hbar\Omega = E(\mathbf{k}_1) + E(\mathbf{k}_2) - E(\mathbf{k}_3) - E(\mathbf{k}_4) \quad (2.209)$$

The transition matrix element is

$$|V'|^2 = \frac{e^2}{\epsilon} \frac{\delta_{\mathbf{k}_1+\mathbf{k}_2, \mathbf{k}_3+\mathbf{k}_4}}{|\mathbf{k}_1 - \mathbf{k}_3|^2 + \beta^2} \quad (2.210)$$

The overlap integral is omitted here.

Electron-Plasmon Interaction The squared matrix element is calculated

$$\begin{aligned} |V'|^2 &= \frac{e^2 \hbar^3 (2\mathbf{k} \cdot \mathbf{q} + q^2)^2}{8\epsilon\omega_q q^2 m^2} \\ &\times \left[\frac{N_q \gamma}{(\Omega^+ - \omega_q)^2 + \gamma^2/4} + \frac{(N_q + 1)\gamma}{(\Omega^- + \omega_q)^2 + \gamma^2/4} \right] \quad (2.211) \end{aligned}$$

where

$$\hbar\Omega^+ = E(\mathbf{k} + \mathbf{q}) - E(\mathbf{k})$$

$$\hbar\Omega^- = E(\mathbf{k} - \mathbf{q}) - E(\mathbf{k})$$

\mathbf{q} is the wave vector of the plasmon, N_q is the equilibrium distribution population of the plasmon, ω_q is the plasmon frequency.

In the above equation, it is clear to see the term when an electron at state \mathbf{k} absorbs/emits a plasmon of state \mathbf{q} with an energy exchange $\hbar\omega_q$ and a wave vector change \mathbf{q} .

2.8.3 Impurity Scattering

- Conwell and Weisskopf approach (CW)

$$V'(r) = \begin{cases} \frac{Ze^2}{4\pi\epsilon r} & r < b \\ 0 & r > b \end{cases} \quad (2.212)$$

where $b = (3/4\pi n_i)^{1/3}$ is the mean distance between impurities, n_i the impurity concentration. Z is the number of charge units of the impurity.

- Brooks and Herring approach (BH):

$$V'(r) = \frac{Ze^2}{4\pi\epsilon r} e^{-\beta r} \quad (2.213)$$

where β^{-1} is the screening length, ϵ the dielectric constant. The scattering probability of BH approach is

$$P_{k,k'} = \frac{2\pi Z^2 n_i e^4 p^2(\mathbf{k}, \mathbf{k}')}{\hbar\epsilon^2(\beta^2 + |\mathbf{k} - \mathbf{k}'|^2)^2} \cdot \frac{\gamma}{[E(\mathbf{k}') - E(\mathbf{k})]^2/\hbar^2 + \gamma^2/4} \quad (2.214)$$

References

1. Garcia AL (1994) Numerical methods for physics. Prentice Hall, New York
2. Goldberg A, Schey HM, Shwartz JL (1967) Computer-generated motion pictures of one-dimensional quantum mechanical transmission and reflection phenomena. Am J Phys 35:177–186
3. Carroll JE (1974) Physical models for semiconductor devices. Edward Arnold, London, p 28
4. Fu Y, Xu W-L (1988) Phonon density of states calculation for superlattice systems. Chin J Infrared Res A 7:1–6
5. Mahan GD (1990) Many-particle physics, 2nd edn. Plenum, New York, p 33
6. Smith RA (1969) Wave mechanics of crystalline solids. Chapman & Hall, London, p 472
7. Fu Y, Willander M (1993) Evanescent channels in calculation of phonon-assisted tunneling spectrum of a semiconductor tunneling structure. J Appl Phys 73:1848–1852

Chapter 3

Optical Properties of Semiconductors

Abstract We first introduce the Maxwell's equations about the electromagnetic field and the Hamiltonian of electron in the electromagnetic field from which we obtain the formula for light-matter interaction which forms the base for the optical electronics. We discuss the general absorption and emission spectra of nanostructure materials. Major focus of the rest of the chapter is about electron-hole pair, i.e., exciton in nanostructures which is the base for the fast developing nanophotonics.

3.1 Electromagnetic Field

Electromagnetic wave is associated with a combination of electric and magnetic fields that can be characterized by a wave function which varies as $e^{i(\mathbf{s}\cdot\mathbf{r}-\omega_s t)}$, where

$$\omega_s = 2\pi f$$

is the angular frequency, f is the radiation frequency, \mathbf{s} ($|\mathbf{s}| = s = \omega/c$) is the wave vector, and c is the phase velocity, which in general differs from the group velocity

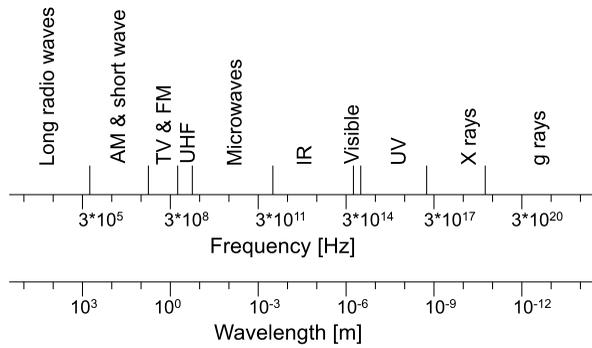
$$c_g = \frac{d\omega_s}{ds}$$

The range of the electromagnetic wave frequency spans immensely, from very-long-wavelength radio waves around 100 Hz to extremely high energy γ rays from space (around 10^{23} Hz). With the exception of the visible part of the electromagnetic spectrum shown in Fig. 3.1, the boundaries between the classifications are not sharp, which are based roughly on how the waves are generated and/or detected.

The electromagnetic wave, described in general by the Maxwell's equations in differential form

$$\begin{aligned}\nabla \times \mathbf{H} &= \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \\ \nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t}\end{aligned}\tag{3.1}$$

Fig. 3.1
Wavelength-frequency
spectrum of electromagnetic
waves (photons)



$$\nabla \cdot \mathbf{D} = \rho$$

$$\nabla \cdot \mathbf{B} = 0$$

by MKS units (meter-kilogram-second), where \mathbf{H} is the magnetic field intensity, \mathbf{J} the conduction current density, \mathbf{D} the electric flux density, \mathbf{E} the electric field intensity vector, \mathbf{B} the magnetic flux density, ρ the charge density.

These are the basics of electromagnetic theory, which were formulated in the early nineteenth century and combined by Maxwell in 1864 into such a consistent set of equations. To the best of our knowledge, Maxwell's equations correctly describe the large-scale (macroscopic) electromagnetic phenomena that occur in nature. Ample experimental evidence is available to support this view, including the classical experiments of Coulomb, Ampère, and Faraday, which provided the original motivation for Maxwell to postulate the concept of the "electromagnetic theory".

In linear, isotropic nondispersive materials (i.e., materials having field-independent, direction-independent, and frequency-independent electric and magnetic properties), we can relate \mathbf{B} to \mathbf{H} and \mathbf{D} to \mathbf{E} using simple proportions:

$$\mathbf{B} = \mu \mathbf{H}, \quad \mathbf{D} = \epsilon \mathbf{E} \quad (3.2)$$

Here μ is the magnetic permeability and ϵ is the electric permittivity. In the normal molecular, bio and semiconductor materials, we assume $\mu = \mu_0$, where μ_0 is the free-space permeability. In free space, $\epsilon = \epsilon_0$ is the free-space permittivity, while in various materials, $\epsilon = \epsilon_r \epsilon_0$, where ϵ_r is referred to be the relative permittivity, or simply dielectric constant. Now, permitting the possibility of electric losses that can dissipate electromagnetic fields in materials, we define an electric current to account for the electric loss mechanisms:

$$\mathbf{J} = \sigma \mathbf{E} \quad (3.3)$$

where σ is the electric conductivity. Between \mathbf{J} and the charge density ρ ,

$$\nabla \cdot \mathbf{J} = -\frac{\partial \rho}{\partial t} \quad (3.4)$$

due to the charge conservation law, see Eq. (2.12). This relationship is also implicit in the Maxwell equations. With these material properties, the Maxwell equations reduce to

$$\nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t} \quad (3.5)$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \epsilon \frac{\partial \mathbf{E}}{\partial t} \quad (3.6)$$

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon} \quad (3.7)$$

$$\nabla \cdot \mathbf{H} = 0 \quad (3.8)$$

By introducing a vector potential, \mathbf{A} , and a scalar potential, ϕ , the electric and magnetic fields can be obtained from the relations of

$$\mathbf{E} = -\nabla\phi - \frac{\partial \mathbf{A}}{\partial t} \quad (3.9)$$

$$\mathbf{B} = \nabla \times \mathbf{A}$$

The first and last Maxwell equations are automatically satisfied by the above definitions. By the relation of

$$\nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}$$

and in the Lorentz gauge of

$$\frac{1}{\mu} \nabla \cdot \mathbf{A} + \epsilon \frac{\partial \phi}{\partial t} = 0 \quad (3.10)$$

we have the following equations for the vector and scalar potentials

$$\nabla^2 \mathbf{A} - \epsilon\mu \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mu \mathbf{J} \quad (3.11)$$

$$\nabla^2 \phi - \epsilon\mu \frac{\partial^2 \phi}{\partial t^2} = -\frac{\rho}{\epsilon} \quad (3.12)$$

Let us examine the electric properties of the common semiconductor materials as compared with metals. Typically the free electron density of elemental metals is 10^{22} cm^{-3} , whereas it is about 10^{19} cm^{-3} for heavily doped semiconductors. The resistivity of gold is $2.01 \times 10^{-8} \Omega \cdot \text{m}$ at room temperature. For silicon at 300 K, the resistivity is about $4 \times 10^{-5} \Omega \cdot \text{m}$ when the silicon is n -type doped at 10^{19} cm^{-3} . The resistivity of Si increases drastically to $4 \times 10^{-3} \Omega \cdot \text{m}$ when the doping level becomes 10^{16} cm^{-3} [1]. For common molecular and bio materials, the charge and current densities are even smaller. We can thus neglect the effects of the carrier concentration and current density on the electromagnetic field, i.e., $\mathbf{J} = 0$

and $\phi = 0$ when discussing electromagnetic fields in semiconductor materials (even highly doped). In this case, the time-dependent solution for the vector potential becomes

$$\mathbf{A}(\mathbf{r}, t) = \mathbf{A}_0 [e^{i(\mathbf{s} \cdot \mathbf{r} - \omega_s t)} + \text{c.c.}] \quad (3.13)$$

with $s^2 = \epsilon\mu\omega_s^2$. “c.c.” in the above equation stands for “complex conjugate”. Note that in the MKS units, $\epsilon_0\mu_0 = 1/c_0^2$, where c_0 is the velocity of light in the vacuum. By the Lorentz gauge, it is observed that a constant scalar potential implies $\nabla \cdot \mathbf{A} = 0$, indicating $\mathbf{A}_0 \cdot \mathbf{s} = 0$, i.e., the vector potential is perpendicular to the propagation direction of the electromagnetic wave.

It must be noticed that the scalar potential ϕ here is devoted solely for the electromagnetic field of an optical radiation. The scalar potential exists in the electron system due to the charge distribution, the ac and/or dc biases applied on the electron system. In almost all device applications, the frequencies of ac biases are very low as compared with the frequency of the optical field, so that the vector field as well as the time-dependence of the scalar field can be neglected, with which the second equation of Eqs. (3.11) reduces to the standard steady-state Poisson equation.

The Poynting vector \mathbf{S} representing the optical power of the electromagnetic field is

$$\mathbf{S} = \mathbf{E} \times \mathbf{H} = \frac{4cs^2|\mathbf{A}_0|^2 \sin^2(\mathbf{s} \cdot \mathbf{r} - \omega_s t)}{\mu} \mathbf{s}_0 \quad (3.14)$$

where c is the velocity of light in the medium, $c = 1/\sqrt{\epsilon\mu}$, and \mathbf{s}_0 is a unit vector in the direction of \mathbf{s} . In the linear, homogeneous and isotropic medium, the phase velocity equals the group velocity, i.e., $c = c_g$.

It is easy to obtain that the time-averaged value of the optical power is

$$\langle \mathbf{S} \rangle_t = 2c\epsilon\omega_s^2 |\mathbf{A}_0|^2 \mathbf{s}_0 \quad (3.15)$$

and the optical energy density is

$$\frac{|\langle \mathbf{S} \rangle_t|}{c} = 2\epsilon\omega_s^2 |\mathbf{A}_0|^2 \quad (3.16)$$

Here we have used the energy dispersion relation of the light:

$$s = \frac{\omega_s}{c} \quad (3.17)$$

We now discuss the energy dissipation of the electromagnetic field. Going back to the Maxwell equations and noting $\mathbf{J} = \sigma \mathbf{E}$, we obtain the wave equation for the electric field

$$\nabla^2 \mathbf{E} = \epsilon\mu \frac{\partial^2 \mathbf{E}}{\partial t^2} + \sigma\mu \frac{\partial \mathbf{E}}{\partial t} \quad (3.18)$$

This represents a wave propagating with dissipation. The general solution can be chosen as

$$\mathbf{E} = \mathbf{E}_0 e^{i(\mathbf{s} \cdot \mathbf{r} - \omega_s t)} \quad (3.19)$$

so that the amplitude of the wave vector, $s = |s|$ is given by

$$s^2 = \epsilon \mu \omega_s^2 + i \sigma \mu \omega_s \quad (3.20)$$

as compared with the energy dispersion, Eq. (3.17), in the non-conductive space ($\sigma = 0$).

A complex refractive index, \tilde{n} , is normally defined by dividing c_0 by the phase velocity in the medium which can be written in terms of its real and imaginary parts so that

$$s = \tilde{n} \frac{\omega_s}{c_0} = (n + i\kappa) \frac{\omega_s}{c_0} \quad (3.21)$$

where n is often called the refractive index in connection with Snells' law of refraction, and κ is known as the extinction coefficient. Inserting the above relation into Eq. (3.19), the solution of the electric field wave equation now becomes

$$\mathbf{E} = \mathbf{E}_0 e^{i\omega_s(nz/c_0 - t)} e^{-\kappa\omega_s z/c_0} \quad (3.22)$$

Here we assume that the electromagnetic field propagates along the z axis. The velocity of the electromagnetic field is reduced from c_0 in vacuum by n to c_0/n in the medium and its amplitude is damped exponentially. The damping of the electromagnetic field is associated with the absorption of the electromagnetic energy. The absorption coefficient α is defined by the energy intensity ($\propto |\mathbf{E}|^2$) decrease by a factor of $e = 2.718282$ so that

$$\alpha = \frac{2\kappa\omega_s}{c_0} \quad (3.23)$$

We now discuss the energy dissipation of the electromagnetic field in terms of the complex dielectric constant $\epsilon' + i\epsilon''$ while neglecting σ , since this is more common for semiconductor materials. The wave equation for the electric field \mathbf{E} and its solution now become

$$\begin{aligned} \nabla^2 \mathbf{E} &= (\epsilon' + i\epsilon'') \mu_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} \\ \mathbf{E} &= \frac{\mathbf{E}_0}{2} e^{i(s \cdot \mathbf{r} - \omega_s t)} + \text{c.c.}, \quad s^2 = (\epsilon' + i\epsilon'') \mu_0 \omega_s^2 \end{aligned} \quad (3.24)$$

for monochromatic light with a frequency ω_s , an amplitude E_0 for its \mathbf{E} field. Poynting's theorem describing conservation of energy is

$$\int \mathbf{J} \cdot \mathbf{E} d\mathbf{r} = - \int \left[\nabla \cdot (\mathbf{E} \times \mathbf{H}) + \mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} + \mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t} \right] d\mathbf{r} \quad (3.25)$$

As discussed before, we can neglect the effects of conducting current and charge density of the semiconductor materials on the electromagnetic field so that the above equation simplifies to

$$\oint \mathbf{E} \times \mathbf{H} \cdot d\mathbf{A} = - \int \left(\mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} + \mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t} \right) d\mathbf{r} \quad (3.26)$$

where S denotes the surface and Ω the volume.

Write

$$\mathbf{D} = (\epsilon' + i\epsilon'')\frac{\mathbf{E}_0}{2} + \text{c.c.}$$

it is easy to obtain

$$\left\langle \mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} \right\rangle_t = -\omega_s \epsilon'' \frac{E_0^2}{2} \quad (3.27)$$

where $\langle \dots \rangle_t$ denotes time average operation. The time average of Poynting's theorem in Eq. (3.26) gives

$$\oint \langle \mathbf{E} \times \mathbf{H} \rangle_t \cdot d\mathbf{S} = \int \omega_s \epsilon'' E_0^2 d\Omega \quad (3.28)$$

Assume surface S encloses a small volume $d\Omega = \Delta\Omega$, the dissipated optical power is

$$\oint \langle \mathbf{E} \times \mathbf{H} \rangle_t \cdot d\mathbf{S} = \omega_s \epsilon'' E_0^2 \Delta\Omega \quad (3.29)$$

On the other hand, consider a thin slab of material with thickness Δz and normal surface area S for which $\Delta\Omega = S\Delta z$, and a plane wave traveling in the z direction with complex wave number $s = s' + is''$. The time average Poynting vector, assuming a single traveling wave, is

$$\langle \mathbf{E} \times \mathbf{H} \rangle_t = z_0 \left[\frac{E_0^2}{2} e^{-2s''z} \left(\sqrt{\frac{\epsilon}{\mu_0}} \right)^* + \text{c.c.} \right] \quad (3.30)$$

where z_0 is the unit vector of the z direction.

The closed surface integral of the Poynting vector, i.e., the optical power dissipated in $\Delta\Omega = S\Delta z$, the energy flow into the volume at one surface ($z = 0$) minus the energy flow out from the other surface ($z = \Delta z$) is thus

$$\oint \langle \mathbf{E} \times \mathbf{H} \rangle_t \cdot d\mathbf{S} = S \frac{E_0^2}{2} \left[(1 - e^{-2s''\Delta z}) \left(\sqrt{\frac{\epsilon}{\mu_0}} \right)^* + \text{c.c.} \right] \quad (3.31)$$

Assuming Δz very small so that $e^{-2s''\Delta z} \ll 1$,

$$\oint \langle \mathbf{E} \times \mathbf{H} \rangle_t \cdot d\mathbf{S} = S \frac{E_0^2}{2\sqrt{\mu_0}} [2s''\Delta z (\sqrt{\epsilon})^* + \text{c.c.}] \quad (3.32)$$

Since $s = \omega_s \sqrt{(\epsilon' + i\epsilon'')\mu_0}$, see Eq. (3.24),

$$s''(\sqrt{\epsilon})^* + \text{c.c.} = \frac{2s's''}{\omega_s \sqrt{\mu_0}} = \omega_s \sqrt{\mu_0} \epsilon'' \quad (3.33)$$

so that

$$\oint (\mathbf{E} \times \mathbf{H})_t \cdot d\mathbf{S} = S\omega_s \epsilon'' E_0^2 \Delta z = \omega_s \epsilon'' E_0^2 \Delta \Omega \quad (3.34)$$

as Eq. (3.29). Therefore, the energy and Poynting vector descriptions for dissipation are in agreement, as they must be from Eq. (3.26). We will apply the above expression late to derive the effective dielectric constant of exciton polariton in a quantum dot.

We now briefly introduce the concept of second quantization of the electromagnetic field, i.e., the photons. Photons are quantum mechanical particles associated with an electromagnetic radiation. They are bosons with zero mass; in free space they carry momentum $\hbar s$ and energy $\hbar\omega_s$. Unlike electrons, photons do not obey the Pauli exclusion principle: one can have as many photons with the same energy and wave vector as one can generate.

Now, taken the density of photons as n_s , the optical energy density is

$$n_s \hbar\omega_s$$

together with Eq. (3.16) we obtain the amplitude of the vector potential

$$|\mathbf{A}_0|^2 = \frac{n_s \hbar}{2\epsilon\omega_s} \quad (3.35)$$

The photon density n_s is a physically measurable quantity which according to this relation tells us the strength of the vector potential \mathbf{A} .

Let b_s^+ and b_s be creation and annihilation operators satisfying the usual Bose commutation relations. In the momentum representation, the electromagnetic field is characterized by photon occupation number N_s

$$b_s^+ |N_s\rangle = \sqrt{N_s + 1} |N_s + 1\rangle, \quad b_s |N_s\rangle = \sqrt{N_s} |N_s - 1\rangle \quad (3.36)$$

The important physical operators for a system of free photons are the total energy, the total momentum and the total number of photons

$$\hbar\omega_s b_s^+ b_s, \quad \hbar s b_s^+ b_s, \quad b_s^+ b_s \quad (3.37)$$

We have neglected the zero-point energy of $\sum_s \hbar\omega_s/2$.

Together with Eq. (3.13), and denoting \mathbf{e}_s as the unit vector of vector potential \mathbf{A} , which is normally known as the polarization of the electromagnetic field, the vector potential operator of the electromagnetic field can be expressed as

$$\mathbf{A}(\mathbf{r}, t) = \sqrt{\frac{\hbar}{2\epsilon\omega_s\Omega}} [b_s e^{i(s\cdot\mathbf{r}-\omega_s t)} + b_s^+ e^{-i(s\cdot\mathbf{r}-\omega_s t)}] \mathbf{e}_s \quad (3.38)$$

where Ω is the spatial volume where the wave function of the photon in the coordinate representation is normalized as $e^{i(s\cdot\mathbf{r}-\omega_s t)}/\sqrt{\Omega}$. And $N_s/\Omega = n_s$.

3.2 Electron in Electromagnetic Field

We first discuss the matrix element of the velocity operator

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} = \frac{\mathbf{p}}{m_0} \quad (3.39)$$

of a conduction-band Bloch electron. In the presence of an electromagnetic field, the wave function of the electron can be expressed by the following general form

$$\Psi_i(\mathbf{r}) = \sum_k F_{ik} \Psi_{ck}(\mathbf{r}) \quad (3.40)$$

where $\Psi_{ck}(\mathbf{r})$ is the conduction-band Bloch function with wave vector \mathbf{k} .

We consider the matrix element of the velocity operator in the effective mass approximation. Intuitively we would expect that in this approximation the mean velocity of the Bloch electron given by crystal momentum $\mathbf{p} = \hbar\mathbf{k}$ is

$$\frac{\mathbf{p}}{m^*} \quad (3.41)$$

where m^* is the effective mass of the conduction-band electron and the envelope wave function is

$$F_i(\mathbf{r}) = \sum_k F_{ik} e^{i\mathbf{k}\cdot\mathbf{r}} \quad (3.42)$$

In this case the velocity matrix element of two conduction-band states is given by

$$\langle F_i(\mathbf{r}) | \mathbf{v} | F_j(\mathbf{r}) \rangle = \int F_i^*(\mathbf{r}) \frac{-i\hbar\nabla}{m^*} F_j(\mathbf{r}) d\mathbf{r} \quad (3.43)$$

where we have written $\mathbf{p} = -i\hbar\nabla$.

Rigorously, the velocity matrix element between state $\Psi_i(\mathbf{r})$ and state $\Psi_j(\mathbf{r})$ of Eq. (3.40) is given by

$$\langle \Psi_i(\mathbf{r}) | \mathbf{v} | \Psi_j(\mathbf{r}) \rangle = \sum_{k,q} F_{ik}^* F_{jq} \int \Psi_{ck}^*(\mathbf{r}) \frac{-i\hbar\nabla}{m_0} \Psi_{cq}(\mathbf{r}) d\mathbf{r} \quad (3.44)$$

where the momentum operator is \mathbf{p}/m_0 . Since

$$\frac{1}{\hbar} \frac{\partial E_{ck}}{\partial \mathbf{k}} = \int \Psi_{ck}^*(\mathbf{r}) \frac{-i\hbar\nabla}{m_0} \Psi_{ck}(\mathbf{r}) d\mathbf{r} \quad (3.45)$$

i.e., Eq. (2.102), we obtain Eq. (3.41) for a parabolic conduction band of $E_{ck} = \hbar^2 k^2 / 2m^*$, which is exactly the starting point of the effective mass approximation.

The Hamiltonian describing an electron with an electronic charge $-e$ in the electromagnetic field of (\mathbf{A}, ϕ) is

$$\frac{1}{2m_0} (\mathbf{p} - e\mathbf{A})^2 - e\phi + V_0(\mathbf{r})$$

$$= \frac{p^2}{2m_0} - e\phi + V_0(\mathbf{r}) - \frac{e}{m_0} \mathbf{A} \cdot \mathbf{p} + \frac{ie\hbar}{2m_0} \nabla \cdot \mathbf{A} + \frac{e^2}{2m_0} \mathbf{A}^2 \quad (3.46)$$

where $V_0(\mathbf{r})$ is the potential energy of the electron system in the darkness, m_0 is the free electron mass, and $\mathbf{p} = -i\hbar\nabla$ is the electron momentum.

Since the scalar potential ϕ is negligible for an electromagnetic field (see last section) which results in $\nabla \cdot \mathbf{A} = 0$ in the Lorentz gauge, four terms remain on the right side of Eq. (3.46). Consider the following ratios

$$\left| \frac{e\mathbf{A} \cdot \mathbf{p}}{m_0} \right| : \left| \frac{p^2}{2m_0} \right| = \frac{2e|\mathbf{A}|}{p}, \quad \left| \frac{e^2\mathbf{A}^2}{2m_0} \right| : \left| \frac{e\mathbf{A} \cdot \mathbf{p}}{m_0} \right| = \frac{e|\mathbf{A}|}{2p} \quad (3.47)$$

The average optical power from the Sun is about 120 W/m². Let us consider an optical power of $S = 10^3$ W/m², i.e., ten times the optical power from the Sun, the photon density n_s of an optical beam in vacuum is

$$n_s = \frac{S}{c_0\hbar\omega_s} = 2.1 \times 10^{13} \text{ m}^{-3} \quad (3.48)$$

when $\hbar\omega_s = 1.0$ eV (the corresponding wavelength is 1.24 μm). By Eq. (3.35),

$$\begin{aligned} e|\mathbf{A}| &= e \sqrt{\frac{n_s\hbar}{2\epsilon\omega_s}} \\ &= 1.60219 \times 10^{-19} \times \sqrt{\frac{2.1 \times 10^{13} \times 1.05459^2 \times 10^{-68}}{2 \times 8.8541853 \times 10^{-12} \times 1.60219 \times 10^{-19}}} \\ &= 4.6 \times 10^{-32} \text{ kg} \cdot \text{m/s} \end{aligned} \quad (3.49)$$

Consider the electron speed v corresponding to the average thermal energy $k_B T$ in a semiconductor. At room temperature ($k_B T = 25$ meV) and for a conduction-band electron in GaAs ($m^* \approx 0.1m_0$), one finds

$$\begin{aligned} v &= \sqrt{\frac{2k_B T}{m^*}} = \sqrt{\frac{2 \times 25 \times 10^{-3} \times 1.60219 \times 10^{-19}}{0.1 \times 9.1096 \times 10^{-31}}} = 3 \times 10^5 \text{ m/s} \\ p &= m^* v = 0.1 \times 9.1096 \times 10^{-31} \times 3 \times 10^5 = 2.7 \times 10^{-26} \text{ kg} \cdot \text{m/s} \end{aligned} \quad (3.50)$$

and finally

$$\frac{e|\mathbf{A}|}{p} = \frac{4.6 \times 10^{-32}}{2.7 \times 10^{-26}} = 1.7 \times 10^{-6} \quad (3.51)$$

Thus, even for an optical beam carrying 10^9 W/m² (i.e., 10^7 times of the Sun), the value of $e|\mathbf{A}|/p = 1.7 \times 10^{-6}$ is still small enough so that only the term linear in \mathbf{A} is significant in Eq. (3.46). We can think about optical power from the latest femtosecond laser pulse, where the typical output power is about 0.1 W and the

pulse duration about 100 fs and pulse repetition rate 100 MHz, the peak optical power is 10^6 W. By focusing such an optical beam to an area of $10 \times 10 \mu\text{m}^2$, we reach an optical power of 10^{16} W/m² (i.e., 10^{12} W/cm²) for which the A^2 must be taken into account under this circumstance. In some cases, e.g., when the linear- A induced transitions are forbidden, the term in A^2 has to be included. We neglect the A^2 in the following discussion.

The time-dependent Schrödinger equation for the electron is now written as

$$i\hbar \frac{\partial \Psi}{\partial t} = (H_0 + V')\Psi \quad (3.52)$$

where

$$H_0 = \frac{p^2}{2m_0} + V_0(\mathbf{r}), \quad V' = \frac{ie\hbar}{m_0} \mathbf{A} \cdot \nabla \quad (3.53)$$

where $\mathbf{A}(\mathbf{r}, t)$ is given by Eq. (3.38). V' is normally referred as the light-matter interaction.

Equation (3.52) actually describes an open system, where the electromagnetic field, i.e., the photons, is treated as an external environment. The discussion of the electron dynamics in the electromagnetic field is simpler when the photons are included. In this case, the total Hamiltonian of the composite system composed of both the electron and photons is

$$H_0(\mathbf{r}) + \frac{ie\hbar}{m_0} \mathbf{A}(\mathbf{r}, t) \cdot \nabla + \sum_s \hbar\omega_s \left(b_s^+ b_s + \frac{1}{2} \right) \quad (3.54)$$

where H_0 is the Hamiltonian of the electron in darkness. b_s^+ and b_s are creation and annihilation operators of the boson field.

The form of the above Hamiltonian can be used to describe other potential fields in the form of $V'(\mathbf{r}, t) = V'(\mathbf{r}) \cos(\omega_s t)$ which can be quantized also as quasi boson particle systems. The quasi particle has an energy of $\hbar\omega_s$ and the density of the quasi particles is proportional to $\sqrt{V'(\mathbf{r})}$. Typical examples of such field quantization are lattice vibration (phonon), and the electromagnetic field (photon), which was just discussed.

The electron is scattered from an initial state to a final one by gaining or losing energy, a boson is annihilated or created at the same time. Notice, however, that the total energy of the composite system, including the electron and the boson particles, is conserved. For simplicity, we consider only one type quasi particles so that

$$\begin{cases} H_0 & t \leq 0 \\ H_0 + [V_s'^+(\mathbf{r})e^{i\omega_s t} b_s^+ + V_s'(\mathbf{r})e^{-i\omega_s t} b_s] + \hbar\omega_s b_s^+ b_s & t > 0 \end{cases} \quad (3.55)$$

and the state of the composite system as $|\mathbf{k}N_{\mathbf{k}}\rangle$, where $|\mathbf{k}\rangle$ is the electron wave function while $N_{\mathbf{k}}$ denotes the number of the boson particles associated with $|\mathbf{k}\rangle$.

Table 3.1 Temporal development of Eq. (3.57)

	$t \leq 0$	$0 < t \leq \delta t$	$\delta t < t \leq 2\delta$	$2\delta t < t \leq 3\delta t$
$N - 3$				k_3
$N - 2$			k_2	k_2
$N - 1$		k_1	k_1	k_1, k_3
N	k_0	k_0	k_0, k_2	k_0, k_2
$N + 1$		k_1	k_1	k_1, k_3
$N + 2$			k_2	k_2
$N + 3$				k_3

By Eq. (2.54), we obtain the transition matrix element between two eigenstates of the composite system

$$\begin{aligned}
& \langle \mathbf{q} N_{\mathbf{q}} | \hat{T}(t) | \mathbf{k} N_{\mathbf{k}} \rangle \\
&= \delta_{\mathbf{q}, \mathbf{k}} \delta_{N_{\mathbf{q}}, N_{\mathbf{k}}} \\
&\quad - \frac{i}{\hbar} \sum_{\mathbf{k}'} \int_0^t [\langle \mathbf{q} | V' | \mathbf{k}' \rangle e^{i(E_{\mathbf{q}} - E_{\mathbf{k}'} + \hbar\omega_s)t'/\hbar} \sqrt{N_{\mathbf{q}} + 1} \langle \mathbf{k}' N_{\mathbf{q}} + 1 | \hat{T}(t') | \mathbf{k} N_{\mathbf{k}} \rangle \\
&\quad + \langle \mathbf{q} | V'^+ | \mathbf{k}' \rangle e^{i(E_{\mathbf{q}} - E_{\mathbf{k}'} - \hbar\omega_s)t'/\hbar} \sqrt{N_{\mathbf{q}}} \langle \mathbf{k}' N_{\mathbf{q}} - 1 | \hat{T}(t') | \mathbf{k} N_{\mathbf{k}} \rangle] dt' \quad (3.56)
\end{aligned}$$

And in its differential form

$$\begin{aligned}
& \frac{d}{dt} \langle \mathbf{q} N_{\mathbf{q}} | \hat{T}(t) | \mathbf{k} N_{\mathbf{k}} \rangle \\
&= -\frac{i}{\hbar} \sum_{\mathbf{k}'} [\langle \mathbf{q} | V' | \mathbf{k}' \rangle e^{i(E_{\mathbf{q}} - E_{\mathbf{k}'} + \hbar\omega_s)t/\hbar} \sqrt{N_{\mathbf{q}} + 1} \langle \mathbf{k}' N_{\mathbf{q}} + 1 | \hat{T}(t) | \mathbf{k} N_{\mathbf{k}} \rangle \\
&\quad + \langle \mathbf{q} | V'^+ | \mathbf{k}' \rangle e^{i(E_{\mathbf{q}} - E_{\mathbf{k}'} - \hbar\omega_s)t/\hbar} \sqrt{N_{\mathbf{q}}} \langle \mathbf{k}' N_{\mathbf{q}} - 1 | \hat{T}(t) | \mathbf{k} N_{\mathbf{k}} \rangle] \quad (3.57)
\end{aligned}$$

We can follow the temporal development of Eq. (3.57) as in Table 3.1, where δt is the time interval of creating or annihilating a boson. Starting from an initial composite state $|\mathbf{k}_0 N\rangle$, the electron is scattered into state $|\mathbf{k}_1\rangle$ after δt . At the same time, the boson field is modified by either creating or annihilating a boson. $2\delta t$ later, $|\mathbf{k}_0 N\rangle \rightarrow |\mathbf{k}_1 N \pm 1\rangle$, $|\mathbf{k}_1 N \pm 1\rangle \rightarrow |\mathbf{k}_2 N + 2\rangle$, $|\mathbf{k}_2 N\rangle$ and $|\mathbf{k}_2 N - 2\rangle$. Naturally the time development is a multiple-channel process, as reflected by $\sum_{\mathbf{k}'}$ in Eq. (3.57), whereas we have followed only one channel of $|\mathbf{k}_0\rangle \rightarrow |\mathbf{k}_1\rangle$, $|\mathbf{k}_1\rangle \rightarrow |\mathbf{k}_2\rangle$ and $|\mathbf{k}_2\rangle \rightarrow |\mathbf{k}_3\rangle$ in Table 3.1.

It is noticed that the transition rate from energy level $E_{\mathbf{k}}$ to $E_{\mathbf{q}}$ is determined by V' . The wave vector of a photon is negligibly small, so is the one of the optical phonon, which results in $\langle \mathbf{q} | V' | \mathbf{k}' \rangle = \langle \mathbf{q} | V'^+ | \mathbf{k}' \rangle$ by setting $s = 0$. However, for acoustic phonon field, $\langle \mathbf{q} | V' | \mathbf{k}' \rangle$ can be significantly different from $\langle \mathbf{q} | V'^+ | \mathbf{k}' \rangle$. Furthermore, the creation and annihilation of a boson particle are progressive.

Thus far, we have not made any approximations. Equation (3.57) can be solved numerically by the standard iteration method. In order to elucidate more clearly the temporal development of the composite system, here we follow our early approximation by assuming the dominance of one-step process so that for $|\mathbf{q}\rangle \neq |\mathbf{k}\rangle$,

$$\begin{aligned} & \langle \mathbf{q} N_{\mathbf{q}} | \hat{T}(t) | \mathbf{k} N_{\mathbf{k}} \rangle \\ &= -\frac{i}{\hbar} \int_0^t [\langle \mathbf{q} | V' | \mathbf{k} \rangle e^{i(E_{\mathbf{q}} - E_{\mathbf{k}} + \hbar\omega_s)t'/\hbar} \sqrt{N_{\mathbf{q}} + 1} \langle \mathbf{k} N_{\mathbf{q}} + 1 | \hat{T}(t') | \mathbf{k} N_{\mathbf{k}} \rangle \\ & \quad + \langle \mathbf{q} | V'^+ | \mathbf{k} \rangle e^{i(E_{\mathbf{q}} - E_{\mathbf{k}} - \hbar\omega_s)t'/\hbar} \sqrt{N_{\mathbf{q}}} \langle \mathbf{k} N_{\mathbf{q}} - 1 | \hat{T}(t') | \mathbf{k} N_{\mathbf{k}} \rangle] dt' \end{aligned} \quad (3.58)$$

By inserting the above equation into Eq. (3.57), we obtain four terms:

$$\begin{aligned} & \sqrt{(N_{\mathbf{k}} + 1)(N_{\mathbf{k}} + 2)} \langle \mathbf{k} | V' | \mathbf{q} \rangle \langle \mathbf{q} | V' | \mathbf{k} \rangle e^{i(E_{\mathbf{q}} - E_{\mathbf{k}})(t' - t)/\hbar + i\omega_s(t' + t)} \\ & \quad \times \langle \mathbf{k} N_{\mathbf{k}} + 2 | \hat{T}(t') | \mathbf{k} N_{\mathbf{k}} \rangle \\ & (N_{\mathbf{k}} + 1) |\langle \mathbf{k} | V' | \mathbf{q} \rangle|^2 e^{i(E_{\mathbf{q}} - E_{\mathbf{k}} - \hbar\omega_s)(t' - t)/\hbar} \langle \mathbf{k} N_{\mathbf{k}} | \hat{T}(t') | \mathbf{k} N_{\mathbf{k}} \rangle \\ & N_{\mathbf{k}} |\langle \mathbf{k} | V' | \mathbf{q} \rangle|^2 e^{i(E_{\mathbf{q}} - E_{\mathbf{k}} + \hbar\omega_s)(t' - t)/\hbar} \langle \mathbf{k} N_{\mathbf{k}} | \hat{T}(t') | \mathbf{k} N_{\mathbf{k}} \rangle \\ & \sqrt{N_{\mathbf{k}}(N_{\mathbf{k}} - 1)} \langle \mathbf{k} | V'^+ | \mathbf{q} \rangle \langle \mathbf{q} | V'^+ | \mathbf{k} \rangle e^{i(E_{\mathbf{q}} - E_{\mathbf{k}})(t' - t)/\hbar - i\omega_s(t' + t)} \langle \mathbf{k} N_{\mathbf{k}} - 2 | \hat{T}(t') | \mathbf{k} N_{\mathbf{k}} \rangle \end{aligned}$$

We keep only the one-step process by neglecting $\langle \mathbf{k} N_{\mathbf{k}} + 2 | \hat{T}(t') | \mathbf{k} N_{\mathbf{k}} \rangle$ and $\langle \mathbf{k} N_{\mathbf{k}} - 2 | \hat{T}(t') | \mathbf{k} N_{\mathbf{k}} \rangle$ so that

$$\begin{aligned} & \frac{d}{dt} \langle \mathbf{k} N_{\mathbf{k}} | \hat{T}(t) | \mathbf{k} N_{\mathbf{k}} \rangle \\ &= -\sum_{\mathbf{q}} \frac{|\langle \mathbf{q} | V' | \mathbf{k} \rangle|^2}{\hbar^2} \int_0^t \langle \mathbf{k} N_{\mathbf{k}} | \hat{T}(t') | \mathbf{k} N_{\mathbf{k}} \rangle \\ & \quad \times [(N_{\mathbf{k}} + 1) e^{i(E_{\mathbf{q}} - E_{\mathbf{k}} - \hbar\omega_s)(t' - t)/\hbar} + N_{\mathbf{k}} e^{i(E_{\mathbf{q}} - E_{\mathbf{k}} + \hbar\omega_s)(t' - t)/\hbar}] dt' \end{aligned} \quad (3.59)$$

At steady state, i.e., $\langle \mathbf{k} N_{\mathbf{k}} | \hat{T}(t) | \mathbf{k} N_{\mathbf{k}} \rangle$ is constant over the time scale t' , we obtain

$$\begin{aligned} & \frac{d}{dt} \langle \mathbf{k} N_{\mathbf{k}} | \hat{T}(t) | \mathbf{k} N_{\mathbf{k}} \rangle = -\langle \mathbf{k} N_{\mathbf{k}} | \hat{T}(t) | \mathbf{k} N_{\mathbf{k}} \rangle \\ & \quad \times \sum_{\mathbf{q}} \frac{\pi |\langle \mathbf{q} | V' | \mathbf{k} \rangle|^2}{\hbar} [(N_{\mathbf{k}} + 1) \delta(E_{\mathbf{q}} - E_{\mathbf{k}} - \hbar\omega_s) \\ & \quad + N_{\mathbf{k}} \delta(E_{\mathbf{q}} - E_{\mathbf{k}} + \hbar\omega_s)] \end{aligned} \quad (3.60)$$

In the above equation, the sum over $|\mathbf{q}\rangle$ includes the term of $|\mathbf{q}\rangle = |\mathbf{k}\rangle$. However, the corresponding process is automatically switched off when ω_s is finite due to the δ functions, so that there is no energy level shift, i.e., $\Delta E_{\mathbf{k}} = 0$ in Eq. (2.70). The

first (second) term in the above equation indicates the transition from state $|q\rangle$ to $|k\rangle$ by emitting (absorbing) a boson.

By denoting

$$w_{kN_k} = \sum_q \frac{2\pi |\langle q|V'|k\rangle|^2}{\hbar} [(N_k + 1)\delta(E_q - E_k - \hbar\omega_s) + N_k\delta(E_q - E_k + \hbar\omega_s)] \quad (3.61)$$

we obtain

$$\langle kN_k|\hat{T}(t)|kN_k\rangle = e^{-w_{kN_k}t/2} \quad (3.62)$$

Then, for $t \rightarrow \infty$,

$$\langle qN_q|\hat{T}(+\infty)|kN_k\rangle = \frac{\langle q|V^{'+}|k\rangle\sqrt{N_k+1}\delta_{N_q,N_k+1}}{E_q - E_k - \hbar\omega_s + i\Gamma_{kN_k}} + \frac{\langle q|V'|k\rangle\sqrt{N_k}\delta_{N_q,N_k-1}}{E_q - E_k + \hbar\omega_s + i\Gamma_{kN_k}} \quad (3.63)$$

where $\Gamma_{kN_k} = \hbar w_{kN_k}/2$.

For a finite $\hbar\omega_s$, we can write explicitly that for $E_q > E_k$,

$$\langle qN_k + 1|\hat{T}(+\infty)|kN_k\rangle_{E_q > E_k} = \frac{\langle q|V^{'+}|k\rangle\sqrt{N_k+1}}{E_q - E_k - \hbar\omega_s + i\Gamma_{kN_k}} \quad (3.64)$$

In other words, only the process involving the absorption of a boson is significant when $E_q > E_k$. Similarly,

$$\langle qN_k - 1|\hat{T}(+\infty)|kN_k\rangle_{E_q < E_k} = \frac{\langle q|V'|k\rangle\sqrt{N_k}}{E_q - E_k + \hbar\omega_s + i\Gamma_{kN_k}} \quad (3.65)$$

which is an emission of a boson. And the relevant transition rates of the two processes are

$$p_{qN_q \leftarrow kN_k|E_q > E_k} = \frac{|\langle q|V^{'+}|k\rangle|^2}{\hbar} \frac{(N_k + 1)\Gamma_{kN_k}}{(E_q - E_k - \hbar\omega_s)^2 + \Gamma_{kN_k}^2} \quad (3.66)$$

$$p_{qN_q \leftarrow kN_k|E_q < E_k} = \frac{|\langle q|V'|k\rangle|^2}{\hbar} \frac{N_k\Gamma_{kN_k}}{(E_q - E_k + \hbar\omega_s)^2 + \Gamma_{kN_k}^2}$$

They are identical to Eqs. (2.85), (2.87) and (2.88) in Chap. 2, where we have assumed, implicitly, that the boson field is a reservoir $N \gg 1$ (and N is included in V'). We have neglected possible phase difference between V and V^+ , which is valid here since the electron momentum $\hbar\mathbf{k}$ is rather large compared with the photon momentum $\hbar\mathbf{s}$.

Equations (3.66) describe the absorption and emission processes of one boson during the transition of electron from state $|k\rangle$ to $|q\rangle$. This is in accordance with what

we have assumed the one-step process of $|\mathbf{k}\rangle \rightarrow |\mathbf{q}\rangle$ by neglecting $|\mathbf{k}\rangle \rightarrow |\mathbf{k}'\rangle \rightarrow |\mathbf{q}\rangle$, where $|\mathbf{k}'\rangle \neq |\mathbf{k}\rangle$.

We now study multiple-boson processes for which we have to go beyond the one-step approximation by solving Eq. (3.57) directly in the numerical ways. Instead of studying the time development operator, here we discuss the multiple-boson process in the Schrödinger picture defined by the total Hamiltonian of the electron-boson composite system Eq. (3.55). By denoting the wave function of the total system (including the electron and the quasi particles) as

$$\Psi(\mathbf{r}, t) = \sum_{qN_s} C_{qN_s}(t) |\mathbf{q}N\rangle \exp[-i(E_q + N_s \hbar \omega_s)t/\hbar] \quad (3.67)$$

where $|N_s\rangle$ denotes the boson field having N_s bosons, $C_{qN_s}(t)$ is the probability that the electron state $|\mathbf{q}\rangle$ is associated with a boson field having N_s bosons at time t . See Table 3.1.

By inserting the above expression into the time-dependent Schrödinger equation

$$i\hbar \frac{\partial \Psi(\mathbf{r}, t)}{\partial t} = H(\mathbf{r}, t) \Psi(\mathbf{r}, t) \quad (3.68)$$

it is easy to obtain the following equation for the wave function coefficient

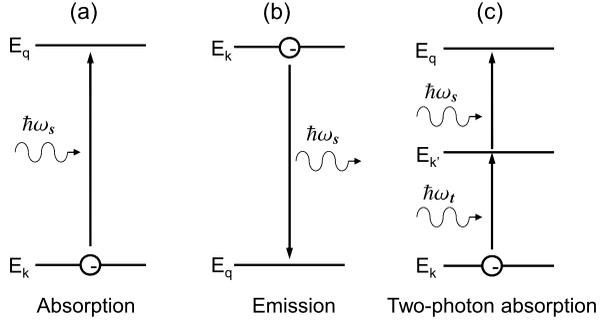
$$i\hbar \frac{dC_{qN_s}(t)}{dt} = \sum_{\mathbf{k}} [C_{kN_s-1}(t) \langle \mathbf{q} | V' | \mathbf{k} \rangle e^{i(E_q - E_k - \hbar \omega_s)t/\hbar} \sqrt{N_s} + C_{kN_s+1}(t) \langle \mathbf{q} | V' | \mathbf{k} \rangle e^{i(E_q - E_k + \hbar \omega_s)t/\hbar} \sqrt{N_s + 1}] \quad (3.69)$$

where the first terms on the right side of the equation describe the electron transition from state \mathbf{k} to \mathbf{q} accompanied by absorbing a boson, i.e., $E_k + \hbar \omega_s \rightarrow E_q$, while the second terms are about emitting a boson, $E_k - \hbar \omega_s \rightarrow E_q$. Let us discuss the electron transition induced by an electromagnetic field of Eq. (3.38). We again neglect the wave vector of the photon so that $\langle \mathbf{q} | V' | \mathbf{k} \rangle = \langle \mathbf{q} | V' | \mathbf{k} \rangle$. Moreover, we assume that the electron is initially at electron state $|\mathbf{k}\rangle$, thus $C_{\mathbf{k}} = 1$. A light $\hbar \omega_s$ is switched on at $t = 0$ and $N_s \gg 1$. By including $\Gamma_{\mathbf{k}}$ and integrating from $t = 0$ to ∞ , the first-order result of the above equation is

$$C_{\mathbf{q}} = \sqrt{N_s} \langle \mathbf{q} | V' | \mathbf{k} \rangle \left(\frac{1}{E_q - E_k - \hbar \omega_s + i\Gamma_{\mathbf{k}}} + \frac{1}{E_q - E_k + \hbar \omega_s + i\Gamma_{\mathbf{k}}} \right) \quad (3.70)$$

which has a peak when $E_q - E_k - \hbar \omega_s = 0$ or $E_q - E_k + \hbar \omega_s = 0$. In other words, if the energy of the initial electron state $|\mathbf{k}\rangle$ is lower than the energy of the final electron state $|\mathbf{q}\rangle$, $C_{\mathbf{q}}$ is significant only when a photon is absorbed. On the other hand, when $E_k > E_q$, a photon will be emitted by the electron when it transits from $|\mathbf{k}\rangle$ to $|\mathbf{q}\rangle$. Figure 3.2(a) and (b) show the schematic represents of the electron transition from $|\mathbf{k}\rangle$ to $|\mathbf{q}\rangle$ by absorbing or emitting one photon. This is exactly the same as what we have obtained via the scattering theory Eqs. (3.64) and (3.65). Note that we phenomenologically introduced $\Gamma_{\mathbf{k}}$ when solving Eq. (3.69) while $\Gamma_{\mathbf{k}}$ is automatic

Fig. 3.2 Schematic represents of electron transit from electron state $|\mathbf{k}\rangle$ to $|\mathbf{q}\rangle$ by (a) absorbing one photon $\hbar\omega_s$ and (b) emitting one photon $\hbar\omega_s$, and (c) electron transition $|\mathbf{k}\rangle \rightarrow |\mathbf{k}'\rangle \rightarrow |\mathbf{q}\rangle$ by absorbing two photons $\hbar\omega_s$ and $\hbar\omega_t$



in the scattering theory. Furthermore, the physical pictures of Figs. 3.2(a) and (b) about the photon absorption and emission are valid when $t \rightarrow \infty$.

Let us discuss the situation when several photon beams are present. For simplicity, we assume $E_k < E_q$ so that we only consider the photon absorption case

$$i\hbar \frac{dC_q(t)}{dt} = \sum_{ks} C_k(t) \sqrt{N_s} \langle \mathbf{q} | V' | \mathbf{k} \rangle e^{i(E_q - E_k - \hbar\omega_s)t/\hbar - \Gamma_k t/\hbar} \quad (3.71)$$

We write $C_q = C_q^{(0)} + C_q^{(1)} + C_q^{(2)}$ to obtain the first-order and second-order results

$$C_q^{(1)}(t) = \sum_{ks} \frac{\sqrt{N_s} \langle \mathbf{q} | V' | \mathbf{k} \rangle C_k^{(0)}}{E_q - E_k - \hbar\omega_s + i\Gamma_k} \left[1 - e^{i(E_q - E_k - \hbar\omega_s)t/\hbar - \Gamma_k t/\hbar} \right]$$

$$C_q^{(2)}(t) = \sum_{ksk't} \frac{\sqrt{N_s N_t} \langle \mathbf{q} | V' | \mathbf{k}' \rangle \langle \mathbf{k}' | V' | \mathbf{k} \rangle C_k^{(0)}}{E_{k'} - E_k - \hbar\omega_t + i\Gamma_k} \left[\frac{1 - e^{i(E_q - E_{k'} - \hbar\omega_s)t/\hbar - \Gamma_{k'} t/\hbar}}{E_q - E_{k'} - \hbar\omega_s + i\Gamma_{k'}} \right. \\ \left. - \frac{1 - e^{i(E_q - E_k - \hbar\omega_t - \hbar\omega_s)t/\hbar - (\Gamma_{k'} + \Gamma_k)t/\hbar}}{E_q - E_k - \hbar\omega_s - \hbar\omega_t + i(\Gamma_k + \Gamma_{k'})} \right] \quad (3.72)$$

When $t \rightarrow \infty$,

$$C_q^{(1)}(\infty) = \sum_{ks} \frac{\sqrt{N_s} \langle \mathbf{q} | V' | \mathbf{k} \rangle C_k^{(0)}}{E_q - E_k - \hbar\omega_s + i\Gamma_k}$$

$$C_q^{(2)}(\infty) = \sum_{ksk't} \frac{\sqrt{N_s N_t} \langle \mathbf{q} | V' | \mathbf{k}' \rangle \langle \mathbf{k}' | V' | \mathbf{k} \rangle C_k^{(0)}}{E_{k'} - E_k - \hbar\omega_t + i\Gamma_k} \left[\frac{1}{E_q - E_{k'} - \hbar\omega_s + i\Gamma_{k'}} \right. \\ \left. - \frac{1}{E_q - E_k - \hbar\omega_s - \hbar\omega_t + i(\Gamma_k + \Gamma_{k'})} \right] \quad (3.73)$$

$C_q^{(1)}(\infty)$ in the above equations is no more than the electron transition from $|\mathbf{k}\rangle$ to $|\mathbf{q}\rangle$ via one-photon absorption. $C_q^{(2)}(\infty)$ is the most interesting part: it will be

significant when $E_{k'} - E_k - \hbar\omega_t = 0$, $E_q - E_{k'} = \hbar\omega_s$, and $E_q - E_k - \hbar\omega_s - \hbar\omega_t = 0$. The process is depicted in Fig. 3.2(c), which is a two-photon absorption process when the electron transits from $|k\rangle \rightarrow |k'\rangle \rightarrow |q\rangle$.

The concept of multiphoton optical process has been widely applied to a variety of fields after the pioneer work of Goepfert-Mayer [2–5]. The multiphoton process has largely been treated theoretically by steady-state perturbation approaches, e.g., the scaling rules of multiphoton absorption by Wherrett [6] and the analysis of two-photon excitation spectroscopy of CdSe quantum dots by Schmidt et al. [7]. Here we see that Eq. (3.69) actually contains all these multiphoton processes, which will be discussed more carefully later.

3.3 Optical Spectrum of Nanostructure

After the general discussions about the energy dissipation processes, we discuss in the following sections of the chapter two special optical processes, namely, optical absorption and emission. We now know by Eqs. (3.66) that when the optical excitation is relatively low (only linear A -term is significant), the transition rate at steady-state of an electron from an initial electron state $|k\rangle$ to a final state $|q\rangle$ induced by these photons is

$$p_{q \leftarrow k}^{\text{abs}} = \frac{e^2 \hbar^2 N_s |\langle q | \mathbf{e}_s \cdot \nabla | k \rangle|^2}{2m_0^2 \epsilon \omega_s \Omega} \frac{\Gamma_k}{(E_q - E_k - \hbar\omega_s)^2 + \Gamma_k^2} \quad (3.74)$$

$$\Gamma_k = \sum_q \frac{\pi e^2 \hbar^3 N_s |\langle q | \mathbf{e}_s \cdot \nabla | k \rangle|^2}{2m_0^2 \epsilon \omega_s \Omega} \delta(E_q - E_k - \hbar\omega_s)$$

for $E_k < E_q$ and for finite $\hbar\omega_s$. The above expressions describe the photon absorption process. Superscript “abs” in $p_{q \leftarrow k}^{\text{abs}}$ stands for “absorption”. For $E_k > E_q$, the photon emission rate is

$$p_{q \leftarrow k}^{\text{em}} = \frac{e^2 \hbar^2 (N_s + 1) |\langle q | \mathbf{e}_s \cdot \nabla | k \rangle|^2}{2m_0^2 \epsilon \omega_s \Omega} \frac{\Gamma_k}{(E_q - E_k + \hbar\omega_s)^2 + \Gamma_k^2} \quad (3.75)$$

$$\Gamma_k = \sum_q \frac{\pi e^2 \hbar^3 (N_s + 1) |\langle q | \mathbf{e}_s \cdot \nabla | k \rangle|^2}{2m_0^2 \epsilon \omega_s \Omega} \delta(E_q - E_k + \hbar\omega_s)$$

Superscript “em” in $p_{q \leftarrow k}^{\text{em}}$ stands for “emission”. Here N_s is the number of monochromatic photons having an energy of $\hbar\omega_s$, \mathbf{e}_s is the polarization vector of the optical field. Note that $n_s = N_s/\Omega$ is the photon density, and Ω is the normalization volume, see Eq. (3.38).

The understanding of Eq. (3.74) is straightforward, the electron will transit from $|k\rangle$ to $|q\rangle$ by absorbing one photon from the photon field and the transition rate is proportional to the intensity N_s of the photon field. It is the fundamental mechanism

of photodetection (e.g., photodetector and solar cell). We notice that the emission term $p_{\mathbf{q}\leftarrow\mathbf{k}}^{\text{em}}$ can be rewritten as stimulated and spontaneous emission terms

$$\begin{aligned} p_{\mathbf{q}\leftarrow\mathbf{k}}^{\text{st}} &= \frac{e^2 \hbar^2 N_s |\langle \mathbf{q} | \mathbf{e}_s \cdot \nabla | \mathbf{k} \rangle|^2}{2m_0^2 \epsilon \omega_s \Omega} \frac{\Gamma_{\mathbf{k}}}{(E_{\mathbf{q}} - E_{\mathbf{k}} + \hbar\omega_s)^2 + \Gamma_{\mathbf{k}}^2} \\ p_{\mathbf{q}\leftarrow\mathbf{k}}^{\text{spon}} &= \frac{e^2 \hbar^2 |\langle \mathbf{q} | \mathbf{e}_s \cdot \nabla | \mathbf{k} \rangle|^2}{2m_0^2 \epsilon \omega_s \Omega} \frac{\Gamma_{\mathbf{k}}}{(E_{\mathbf{q}} - E_{\mathbf{k}} + \hbar\omega_s)^2 + \Gamma_{\mathbf{k}}^2} \end{aligned} \quad (3.76)$$

Superscript “st” in $p_{\mathbf{q}\leftarrow\mathbf{k}}^{\text{st}}$ stands for “stimulated”, and superscript “spon” in $p_{\mathbf{q}\leftarrow\mathbf{k}}^{\text{spon}}$ stands for “spontaneous”. $p_{\mathbf{q}\leftarrow\mathbf{k}}^{\text{st}}$ represents the transition rate from $|\mathbf{k}\rangle$ to $|\mathbf{q}\rangle$ accompanied by emitting a photon in an existing photon field N_s . This is known as the stimulated emission, which is the fundamental mechanism of lasing. $p_{\mathbf{q}\leftarrow\mathbf{k}}^{\text{spon}}$ on the other hand describe the photon emission process in the absence of any photon field, thus named as spontaneous. It is the mechanism of the light-emitting devices. These are known as the Einstein theory of photon. We will discuss more in late chapters.

We consider an electron system composed of many electrons, such as in a device. The transition rate between state $|\mathbf{k}\rangle$ and $|\mathbf{q}\rangle$ is

$$W_{\mathbf{qk}}(\hbar\omega_s) = p_{\mathbf{q}\leftarrow\mathbf{k}} f(E_{\mathbf{k}}) [1 - f(E_{\mathbf{q}})] - p_{\mathbf{k}\leftarrow\mathbf{q}} f(E_{\mathbf{q}}) [1 - f(E_{\mathbf{k}})] \quad (3.77)$$

by including the Pauli exclusion principle, i.e., an electron in an occupied state $|\mathbf{k}\rangle$ is capable to transit to an empty state $|\mathbf{q}\rangle$. At equilibrium, the occupation of electron state $|\mathbf{k}\rangle$ is expressed by the Fermi function

$$f(E_{\mathbf{k}}) = \frac{1}{1 + \exp\left(\frac{E_{\mathbf{k}} - E_f}{k_B T}\right)} \quad (3.78)$$

where E_f is the Fermi level.

Normally $N_s \gg 1$ so that $p_{\mathbf{q}\leftarrow\mathbf{k}} \approx p_{\mathbf{k}\leftarrow\mathbf{q}} \equiv p_{\mathbf{qk}}$

$$W_{\mathbf{qk}}(\hbar\omega_s) = p_{\mathbf{qk}} [f(E_{\mathbf{k}}) - f(E_{\mathbf{q}})] \quad (3.79)$$

Summing over all possible electron states in the electron system gives us the total optical transition rate

$$W(\hbar\omega_s) = \sum_{\mathbf{qk}} W_{\mathbf{qk}}(\hbar\omega_s) \quad (3.80)$$

We now relate the optical gain (lasing action) and absorption (detection action) coefficients of an optical photon beam to the optical transition rate of electrons when the optical beam travels through the electron system. We consider the photon beam traveling along the z axis, we can write the continuity equation for the photon density as

$$\frac{dN_s}{dt} = \frac{\partial N_s}{\partial t} - \frac{\partial(cN_s)}{\partial z} \quad (3.81)$$

where on the right side of the equation, the first term represents the net gain-absorption rate of photons,

$$\frac{\partial N_s}{\partial t} = W(\hbar\omega_s) \quad (3.82)$$

The second term represents the photons leaving the region between z and $(z + dz)$ due to the photon propagation. Here c is the velocity of light. In the steady state, $dN_s/dt = 0$. Moreover, in general we have

$$N_s(z) = N_s(0)e^{g(\hbar\omega_s)z} \quad (3.83)$$

which defines the optical coefficient $g(\hbar\omega_s)$

$$g(\hbar\omega_s) = \sum_{qk} \frac{e^2 \hbar^2 |\langle \mathbf{q} | \mathbf{e}_s \cdot \nabla | \mathbf{k} \rangle|^2}{2m_0^2 c \epsilon \omega_s \Omega} \frac{\Gamma_k}{(E_q - E_k \pm \hbar\omega_s)^2 + \Gamma_k^2} [f(E_k) - f(E_q)] \quad (3.84)$$

It is thus noticed by Eqs. (3.83) and (3.84) that when $g(\hbar\omega_s) > 0$, the intensity of the incident radiation becomes amplified when it travels through the electron system, optical amplification is expected, thus defined as optical gain; whereas if $g(\hbar\omega) < 0$, the radiation intensity reduces along its transmission through the electron system, i.e., radiation becomes absorbed (the absorption coefficient is normally defined as $\alpha = -g$ when $g < 0$), e.g., in photodetectors.

Furthermore, when we consider a two-level electron system such that there are only two electron states $|\mathbf{k}\rangle$ and $|\mathbf{q}\rangle$, Eq. (3.84) immediately states that a population inversion, $f(E_q) > f(E_k)$ when $E_q > E_k$, is essential for optical amplification, where at equilibrium status, $f(E_q) < f(E_k)$.

In Sect. 1.9 we have discussed the electron states in nanostructures. In the following we consider optical spectra of the four types of nanostructures.

Bulk Material We begin with the bulk material

$$\Psi_{\mathbf{k}}(\mathbf{r}) = \sqrt{\Omega_{\text{cell}}} \psi_{\mathbf{k}}(\mathbf{r}) u_{\mathbf{k}}(\mathbf{r}), \quad \psi_{\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{\Omega}} e^{i\mathbf{k}\cdot\mathbf{r}} \quad (3.85)$$

so that

$$\begin{aligned} \langle \mathbf{q} | \mathbf{e}_s \cdot \nabla | \mathbf{k} \rangle &= \langle \Psi_{\mathbf{q}}(\mathbf{r}) | \mathbf{e}_s \cdot \nabla | \Psi_{\mathbf{k}}(\mathbf{r}) \rangle \\ &= \frac{\Omega_{\text{cell}}}{\Omega} \mathbf{e}_s \cdot \left[\int_{\Omega} e^{i(\mathbf{q}-\mathbf{k})\cdot\mathbf{r}} u_{\mathbf{q}}^*(\mathbf{r}) \nabla u_{\mathbf{k}}(\mathbf{r}) d\mathbf{r} \right. \\ &\quad \left. - i\mathbf{k} \int_{\Omega} e^{i(\mathbf{q}-\mathbf{k})\cdot\mathbf{r}} u_{\mathbf{q}}^*(\mathbf{r}) u_{\mathbf{k}}(\mathbf{r}) d\mathbf{r} \right] \end{aligned} \quad (3.86)$$

It is important to note that the total wave functions of the electron states are used when calculating the optical transition matrix element. The second term on the right

side is zero because the two wave function are orthogonal ($\mathbf{q} \neq \mathbf{k}$). The first integral can be performed

$$\begin{aligned} \int_{\Omega} e^{i(\mathbf{q}-\mathbf{k})\cdot\mathbf{r}} u_{\mathbf{q}}^*(\mathbf{r}) \nabla u_{\mathbf{k}}(\mathbf{r}) d\mathbf{r} &= \sum_n e^{i(\mathbf{q}-\mathbf{k})\cdot\mathbf{R}_n} \int_{\text{cell}} u_{\mathbf{q}}^*(\mathbf{r}) \nabla u_{\mathbf{k}}(\mathbf{r}) d\mathbf{r} \\ &= \int_{\text{cell}} u_{\mathbf{q}}^*(\mathbf{r}) \nabla u_{\mathbf{k}}(\mathbf{r}) d\mathbf{r} \sum_n e^{i(\mathbf{q}-\mathbf{k})\cdot\mathbf{R}_n} \\ &= N \delta_{\mathbf{q}\mathbf{k}} \int_{\text{cell}} u_{\mathbf{q}}^*(\mathbf{r}) \nabla u_{\mathbf{k}}(\mathbf{r}) d\mathbf{r} \end{aligned} \quad (3.87)$$

For two electrons, one occupies a valence-band state and the other in a conduction-band state,

$$\int_{\text{cell}} u_{c\mathbf{k}}^*(\mathbf{r}) \nabla u_{v\mathbf{k}}(\mathbf{r}) d\mathbf{r} = \frac{\mathbf{p}_{cv}}{\hbar} \quad (3.88)$$

and $|\mathbf{p}_{cv}| = \sqrt{m_0 E_p}$, where E_p is the same energy parameter in the eight-band $\mathbf{k} \cdot \mathbf{p}$ model in Sect. 1.6. Finally

$$\int_{\Omega} e^{i(\mathbf{q}-\mathbf{k})\cdot\mathbf{r}} u_{\mathbf{q}}^*(\mathbf{r}) \nabla u_{\mathbf{k}}(\mathbf{r}) d\mathbf{r} = N \delta_{\mathbf{q}\mathbf{k}} \frac{\mathbf{p}_{cv}}{\hbar} \quad (3.89)$$

where $\delta_{\mathbf{q}\mathbf{k}} = 1$ if $\mathbf{q} = \mathbf{k}$, it is zero otherwise. N is the number of unit cells in Ω . Since $N\Omega_{\text{cell}} = \Omega$,

$$\langle \Psi_{\mathbf{q}}(\mathbf{r}) | \mathbf{e}_s \cdot \nabla | \Psi_{\mathbf{k}}(\mathbf{r}) \rangle = \delta_{\mathbf{q}\mathbf{k}} \frac{\mathbf{e}_s \cdot \mathbf{p}_{cv}}{\hbar} \quad (3.90)$$

Thus, the optical spectrum of a bulk material is

$$\begin{aligned} g(\hbar\omega_s) &= \sum_{\mathbf{q}\mathbf{k}} \frac{e^2 \hbar^2 |\langle \Psi_{\mathbf{q}}(\mathbf{r}) | \mathbf{e}_s \cdot \nabla | \Psi_{\mathbf{k}}(\mathbf{r}) \rangle|^2 \Gamma_{\mathbf{k}} [f(E_{\mathbf{k}}) - f(E_{\mathbf{q}})]}{2m_0^2 c \epsilon \omega_s \Omega [(E_{\mathbf{q}} - E_{\mathbf{k}} \pm \hbar\omega_s)^2 + \Gamma_{\mathbf{k}}^2]} \\ &= \sum_{\mathbf{k}} \frac{e^2 \hbar^2 |\langle \Psi_{c\mathbf{k}}(\mathbf{r}) | \mathbf{e}_s \cdot \nabla | \Psi_{v\mathbf{k}}(\mathbf{r}) \rangle|^2 \Gamma_{v\mathbf{k}} [f(E_{v\mathbf{k}}) - f(E_{c\mathbf{q}})]}{2m_0^2 c \epsilon \omega_s \Omega [(E_{c\mathbf{q}} - E_{v\mathbf{k}} \pm \hbar\omega_s)^2 + \Gamma_{v\mathbf{k}}^2]} \\ &= \frac{e^2 \Gamma |\mathbf{e}_s \cdot \mathbf{p}_{cv}|^2}{2m_0^2 c \epsilon \omega_s \Omega} \int \frac{f(E_{v\mathbf{k}}) - f(E_{c\mathbf{k}})}{(E_{c\mathbf{k}} - E_{v\mathbf{k}} \pm \hbar\omega_s)^2 + \Gamma^2} \frac{2d\mathbf{k}}{(2\pi)^3 / \Omega} \end{aligned} \quad (3.91)$$

Here we assume that Γ is \mathbf{k} independent. Finally,

$$g(\hbar\omega_s) = \frac{e^2 \Gamma |\mathbf{e}_s \cdot \mathbf{p}_{cv}|^2}{2m_0^2 c \epsilon \omega_s} \int \frac{f(E_{v\mathbf{k}}) - f(E_{c\mathbf{k}})}{(E_{c\mathbf{k}} - E_{v\mathbf{k}} \pm \hbar\omega_s)^2 + \Gamma^2} \frac{2d\mathbf{k}}{(2\pi)^3} \quad (3.92)$$

Quite similar to the mathematical manipulations in normalizing the components of the total wave function in Sect. 1.9, we have used the following relationships in

deriving the above equations:

$$\sum_{\mathbf{k}} g_{\mathbf{k}} = \frac{1}{\Delta \mathbf{k}} \sum_{\mathbf{k}} g_{\mathbf{k}} \Delta \mathbf{k} \quad (3.93)$$

where $\Delta \mathbf{k}$ is the minimal separation between two adjacent \mathbf{k} 's for which $\Delta \mathbf{k} = (2\pi)^3/\Omega$ because of the Bloch theorem. Replace $\Delta \mathbf{k}$ by $(2\pi)^3/\Omega$ in front of \sum and the one in \sum by $d\mathbf{k}$, we convert the summation into integration

$$\sum_{\mathbf{k}} g_{\mathbf{k}} = \frac{1}{(2\pi)^3/\Omega} \int g_{\mathbf{k}} d\mathbf{k} \quad (3.94)$$

The double summations in Eq. (3.92) are

$$\begin{aligned} \sum_{q\mathbf{k}} g_{q\mathbf{k}} \delta_{q\mathbf{k}} &= \iint g_{q\mathbf{k}} \delta_{q\mathbf{k}} \frac{d\mathbf{q}}{(2\pi)^3/\Omega} \frac{d\mathbf{k}}{(2\pi)^3/\Omega} \\ &= \int \left[\int g_{q\mathbf{k}} \delta_{q\mathbf{k}} \frac{d\mathbf{q}}{(2\pi)^3/\Omega} \right] \frac{d\mathbf{k}}{(2\pi)^3/\Omega} \\ &= \int \left[g_{\mathbf{k}\mathbf{k}} \frac{\Delta \mathbf{k}}{(2\pi)^3/\Omega} \right] \frac{d\mathbf{k}}{(2\pi)^3/\Omega} \\ &= \int \left(g_{\mathbf{k}\mathbf{k}} \frac{\Delta \mathbf{k}}{\Delta \mathbf{k}} \right) \frac{d\mathbf{k}}{(2\pi)^3/\Omega} \\ &= \int g_{\mathbf{k}\mathbf{k}} \frac{d\mathbf{k}}{(2\pi)^3/\Omega} = \sum_{\mathbf{k}} g_{\mathbf{k}\mathbf{k}} \end{aligned} \quad (3.95)$$

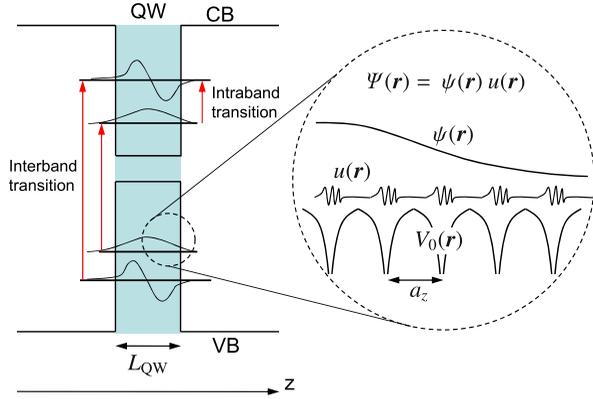
Furthermore, it is reminded that the integration in \mathbf{k} is limited within the first Brillouin zone due to the fact that the wave function in a crystal is written as a function with the periodicity of the lattice times a plane wave, while the plane wave is unique only up to one reciprocal lattice vector. It is interesting to see that the number of allowed \mathbf{k} values in the first Brillouin zone is equal to the number of unit cells in Ω :

$$\int \frac{d\mathbf{k}}{(2\pi)^3/\Omega} = N \quad (3.96)$$

Quantum Well A quantum well is characterized by an effective confinement length L_{QW} along the z direction and an extension area A_{QW} in the xy plane, and the wave function is expressed as

$$\Psi_{i\mathbf{k}}(\mathbf{r}) = \sqrt{\Omega_{\text{cell}}} \psi_{i\mathbf{k}}(\mathbf{r}) u_{\mathbf{k}}(\mathbf{r}), \quad \psi_{i\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{A_{\text{QW}}}} \psi_i(z) e^{i(k_x x + k_y y)} \quad (3.97)$$

Fig. 3.3 Schematic represents of electron's interband and intraband transitions in a quantum well (QW) along the z axis



The optical transition matrix element

$$\begin{aligned}
 \langle \Psi_{jq} | \mathbf{e}_s \cdot \nabla | \Psi_{ik} \rangle &= \frac{\Omega_{\text{cell}}}{A_{\text{QW}}} \mathbf{e}_s \cdot \int_{\Omega} e^{i[(k_x - q_x)x + (k_y - q_y)y]} \psi_j^*(z) \psi_i(z) u_{\mathbf{q}}^*(\mathbf{r}) \nabla u_{\mathbf{k}}(\mathbf{r}) d\mathbf{r} \\
 &- \frac{i(e_x k_x + e_y k_y) \Omega_{\text{cell}}}{A_{\text{QW}}} \int_{\Omega} e^{i[(k_x - q_x)x + (k_y - q_y)y]} \psi_j^*(z) \psi_i(z) u_{\mathbf{q}}^*(\mathbf{r}) u_{\mathbf{k}}(\mathbf{r}) d\mathbf{r} \\
 &+ \frac{e_z \Omega_{\text{cell}}}{A_{\text{QW}}} \int_{\Omega} e^{i[(k_x - q_x)x + (k_y - q_y)y]} \psi_j^*(z) \frac{\partial \psi_i(z)}{\partial z} u_{\mathbf{q}}^*(\mathbf{r}) u_{\mathbf{k}}(\mathbf{r}) d\mathbf{r} \quad (3.98)
 \end{aligned}$$

where $\mathbf{e}_s = (e_x, e_y, e_z)$. The second term in the above expression is zero because the wave functions are orthogonal.

Refer to Fig. 3.3, the first term represents interband transition

$$\begin{aligned}
 &\frac{\Omega_{\text{cell}}}{A_{\text{QW}}} \int_{\Omega} e^{i[(k_x - q_x)x + (k_y - q_y)y]} \psi_j^*(z) \psi_i(z) u_{\mathbf{q}}^*(\mathbf{r}) \nabla u_{\mathbf{k}}(\mathbf{r}) d\mathbf{r} \\
 &= \frac{\Omega_{\text{cell}}}{A_{\text{QW}}} \sum_{\mathbf{R}} e^{i[(k_x - q_x)R_x + (k_y - q_y)R_y]} \psi_j^*(R_z) \psi_i(R_z) \int_{\text{cell}} u_{\mathbf{q}}^*(\mathbf{r}) \nabla u_{\mathbf{k}}(\mathbf{r}) d\mathbf{r} \\
 &= \frac{\Omega_{\text{cell}}}{A_{\text{QW}}} N_{\text{QW}} \delta_{k_x q_x, k_y q_y} \frac{p_{cv}}{\hbar} \frac{1}{a_z} \sum_{R_z} \psi_j^*(R_z) \psi_i(R_z) a_z \\
 &= \frac{\Omega_{\text{cell}}}{A_{\text{QW}}} N_{\text{QW}} \delta_{k_x q_x, k_y q_y} \frac{p_{cv}}{\hbar} \frac{1}{a_z} \int_{L_{\text{QW}}} \psi_j^*(R_z) \psi_i(R_z) dz \\
 &= \delta_{k_x q_x, k_y q_y} \frac{p_{cv}}{\hbar} \int_{L_{\text{QW}}} \psi_j^*(z) \psi_i(z) dz = \delta_{k_x q_x, k_y q_y} \frac{p_{cv}}{\hbar} \langle \psi_j(z) | \psi_i(z) \rangle \quad (3.99)
 \end{aligned}$$

Here a_z is the length of the unit cell along the z direction, N_{QW} is the number of unit cells within A_{QW} . $\delta_{k_x q_x, k_y q_y} = 1$ if $k_x = q_x$ and $k_y = q_y$, it is zero otherwise.

The corresponding interband optical spectrum is

$$\begin{aligned}
 g_{\text{inter}}(\hbar\omega_s) &= \frac{e^2 \Gamma |\mathbf{e}_s \cdot \mathbf{p}_{cv}|^2}{2m_0^2 c \epsilon \omega_s L_{\text{QW}}} \sum_{ji} \iint \frac{|\langle \psi_j(z) | \psi_i(z) \rangle|^2 [f(E_{ik_x k_y}) - f(E_{jk_x k_y})]}{(E_{jk_x k_y} - E_{ik_x k_y} \pm \hbar\omega_s)^2 + \Gamma^2} \frac{2dk_x dk_y}{(2\pi)^2} \\
 & \quad (3.100)
 \end{aligned}$$

Note that $\Omega = A_{\text{QW}} L_{\text{QW}}$.

The third term in Eq. (3.98) is about the intraband transition

$$\begin{aligned}
 & \frac{e_z \Omega_{\text{cell}}}{A_{\text{QW}}} \int_{\Omega_c} e^{i[(k_x - q_x)x + (k_y - q_y)y]} \psi_j^*(z) \frac{\partial \psi_i(z)}{\partial z} u_q^*(\mathbf{r}) u_k(\mathbf{r}) d\mathbf{r} \\
 &= \frac{e_z \Omega_{\text{cell}}}{A_{\text{QW}}} \sum_{\mathbf{R}} e^{i[k_x - q_x]R_x + [k_y - q_y]R_y} \psi_j^*(R_z) \frac{\partial \psi_i(R_z)}{\partial R_z} \int_{\text{cell}} u_q^*(\mathbf{r}) u_k(\mathbf{r}) d\mathbf{r} \\
 &= \frac{e_z \Omega_{\text{cell}}}{A_{\text{QW}}} N_{\text{QW}} \delta_{k_x q_x, k_y q_y} \left[\frac{1}{a_z} \sum_{R_z} \psi_j^*(R_z) \frac{\partial \psi_i(R_z)}{\partial R_z} a_z \right] \left[\int_{\text{cell}} u_q^*(\mathbf{r}) u_k(\mathbf{r}) d\mathbf{r} \right] \\
 &= e_z \delta_{k_x q_x, k_y q_y} \left[\int_{L_{\text{QW}}} \psi_j^*(z) \frac{\partial \psi_i(z)}{\partial z} dz \right] \left[\int_{\text{cell}} u_q^*(\mathbf{r}) u_k(\mathbf{r}) d\mathbf{r} \right] \\
 &= e_z \delta_{k_x q_x, k_y q_y} \left\langle \psi_j(z) \left| \frac{\partial}{\partial z} \right| \psi_i(z) \right\rangle \quad (3.101)
 \end{aligned}$$

for states within the same band since

$$\begin{aligned}
 \int_{\text{cell}} u_{\text{CB},k}^*(\mathbf{r}) u_{\text{CB},k}(\mathbf{r}) d\mathbf{r} &= \int_{\text{cell}} u_{\text{VB},k}^*(\mathbf{r}) u_{\text{VB},k}(\mathbf{r}) d\mathbf{r} = 1 \\
 \int_{\text{cell}} u_{\text{CB},k}^*(\mathbf{r}) u_{\text{VB},k}(\mathbf{r}) d\mathbf{r} &= 0
 \end{aligned} \quad (3.102)$$

and the corresponding intraband optical spectrum is

$$\begin{aligned}
 g_{\text{intra}}(\hbar\omega_s) &= \frac{e^2 \hbar^2 \Gamma e_z^2}{2m_0^2 c \epsilon \omega_s L_{\text{QW}}} \\
 & \times \sum_{ji} \iint \frac{|\langle \psi_j(z) | \partial/\partial z | \psi_i(z) \rangle|^2 [f(E_{ik_x k_y}) - f(E_{jk_x k_y})]}{(E_{jk_x k_y} - E_{ik_x k_y} \pm \hbar\omega_s)^2 + \Gamma^2} \frac{2dk_x dk_y}{(2\pi)^2} \\
 & \quad (3.103)
 \end{aligned}$$

Quantum Wire A quantum wire is characterized by an effective confinement area A_{QWR} in the xy plane and an extension length L_{QWR} along the z direction, and

the total wave function of an electron state is expressed as

$$\Psi_{ik}(\mathbf{r}) = \sqrt{\Omega_{\text{cell}}} \psi_{ik}(\mathbf{r}) u_{\mathbf{k}}(\mathbf{r}), \quad \psi_{ik}(\mathbf{r}) = \frac{1}{\sqrt{L_{\text{QWR}}}} \psi_i(x, y) e^{ikz} \quad (3.104)$$

Similar to the derivations of the optical interband and intraband transitions in a quantum well, the interband and intraband transitions in a quantum wire are

$$g_{\text{inter}}(\hbar\omega_s) = \frac{e^2 \Gamma |\mathbf{e}_s \cdot \mathbf{p}_{cv}|^2}{2m_0^2 c \epsilon \omega_s A_{\text{QWR}}} \times \sum_{ji} \int \frac{|\langle \psi_j(x, y) | \psi_i(x, y) \rangle|^2 [f(E_{ik}) - f(E_{jk})] 2dk}{(E_{jk} - E_{ik} \pm \hbar\omega_s)^2 + \Gamma^2} \quad (3.105)$$

$$g_{\text{intra}}(\hbar\omega_s) = \frac{e^2 \hbar^2 \Gamma}{2m_0^2 c \epsilon \omega_s A_{\text{QWR}}} \sum_{ji} \int \frac{|T_{ji}(x, y)|^2 [f(E_{ik}) - f(E_{jk})] 2dk}{(E_{jk} - E_{ik} \pm \hbar\omega_s)^2 + \Gamma^2} \quad (3.106)$$

where

$$T_{ji}(x, y) = \left\langle \psi_j(x, y) \left| e_x \frac{\partial}{\partial x} + e_y \frac{\partial}{\partial y} \right| \psi_i(x, y) \right\rangle \quad (3.106)$$

Quantum Dot For a quantum dot confined three dimensionally within an effective volume Ω_{QD} ,

$$\Psi_i(\mathbf{r}) = \sqrt{\Omega_{\text{cell}}} \psi_i(\mathbf{r}) u_{\mathbf{k}}(\mathbf{r}), \quad \int_{\Omega_{\text{QD}}} |\psi_i(\mathbf{r})|^2 d\mathbf{r} = 1 \quad (3.107)$$

and the optical interband and intraband spectra are

$$g_{\text{inter}}(\hbar\omega_s) = \frac{e^2 \Gamma |\mathbf{e}_s \cdot \mathbf{p}_{cv}|^2}{2m_0^2 c \epsilon \omega_s \Omega_{\text{QD}}} \sum_{ji} \frac{|\langle \psi_j(\mathbf{r}) | \psi_i(\mathbf{r}) \rangle|^2 [f(E_i) - f(E_j)]}{(E_j - E_i \pm \hbar\omega_s)^2 + \Gamma^2} \quad (3.108)$$

$$g_{\text{intra}}(\hbar\omega_s) = \frac{e^2 \hbar^2 \Gamma}{2m_0^2 c \epsilon \omega_s \Omega_{\text{QD}}} \sum_{ji} \frac{|\langle \psi_j(\mathbf{r}) | \mathbf{e}_s \cdot \nabla | \psi_i(\mathbf{r}) \rangle|^2 [f(E_i) - f(E_j)]}{(E_j - E_i \pm \hbar\omega_s)^2 + \Gamma^2}$$

Figure 3.4 shows the optical spectra of various materials under various carrier injection conditions. Three-dimensionally extended bulk material (3D), 2D quantum well, 1D quantum wire and 0D quantum dot are considered. The extension of the quantum confinement is denoted as L , i.e., L is the quantum well width, $L \times L$ is the cross section of the quantum wire, and $L \times L \times L$ is the volume of the quantum dot. The carrier densities indicated in the figure are injected electron (hole) concentrations in the conduction (valence) bands. It is observed by Fig. 3.4 that the optical spectrum in general consists of two parts: The optical gain part in the long wavelength region, where the population inversion is achieved; and the optical absorption part in the short wavelength region, where high-energy electron states are

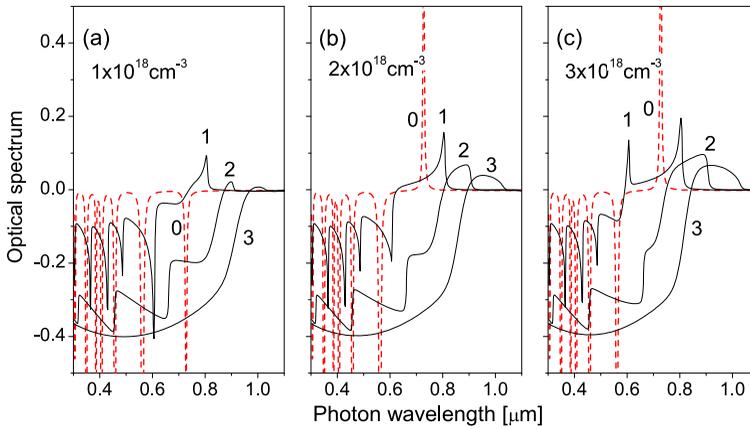


Fig. 3.4 Optical spectra $g_{\text{inter}}(\hbar\omega_s)$ of low-dimensional electron systems. $L = 10$ nm, $T = 300$ K. Integer 0 (quantum dot with a volume of $L \times L \times L$, *dashed line*), 1 (quantum wire with a cross section $L \times L$), 2 (quantum well with a thickness L), and 3 (bulk material) indicate the dimensionality of the electron system

largely unoccupied and high-energy hole states are occupied. We shall come back to discuss the optical spectra in Fig. 3.4 in Chap. 5.

3.4 Exciton and Its Optical Properties

We have thus far focused on the electric and optical properties of a single electron in semiconductor materials. However, there are many electrons in real devices. The simplification of the multiple-electron system to a single-electron system is largely justified by the envelope function approach that for almost all semiconductor device applications, we concern with the electrons that occupy bandedge states whose wave functions can be separated into the fast-varying Bloch functions and the slow-varying envelope functions. Other electrons will form approximately an effective homogeneous medium which can be described by a Poisson's equation for example (see next chapter). However, there is one critical situation that needs specific attention, namely an exciton.

Let us consider an intrinsic semiconductor bulk material for which we know that all the single-electron states up to the valence band are all occupied and all the single-electron states from the conduction band are all empty. We denote this state as the ground state Ψ_0 . As discussed in the previous section, when a photon has an energy larger than the bandgap, it can excite an electron from one of the valence-band states into one state in the conduction band. The physical problem we are dealing with now is: the conduction-band is occupied by one electron, and the valence band is almost fully occupied except one empty state. In the absence of external fields, the wave function of an ensemble of electrons in a semiconductor

material is an eigen function of the Hamiltonian

$$\sum_i H_0(\mathbf{r}_i) + \sum_{i<j} V(\mathbf{r}_i - \mathbf{r}_j) \quad (3.109)$$

where

$$H_0(\mathbf{r}_i) = -\frac{\hbar^2 \nabla_i^2}{2m_0} + V_0(\mathbf{r}_i), \quad V(\mathbf{r}_i - \mathbf{r}_j) = \frac{e^2}{4\pi\epsilon_\infty |\mathbf{r}_i - \mathbf{r}_j|} \quad (3.110)$$

\mathbf{r}_i is the spatial position of the i th electron and $V_0(\mathbf{r}_i)$ is the periodic lattice potential. $H_0(\mathbf{r})$ is the same single-electron Hamiltonian in Eq. (1.32), $V(\mathbf{r}_i - \mathbf{r}_j)$ is the Coulombic potential energy between electron \mathbf{r}_i and \mathbf{r}_j , where ϵ_∞ is the high-frequency dielectric constant of the semiconductor material. This is very similar to the Coulombic potential of a shallow impurity, see Eq. (2.123). Note that the proper treatment about the electron-electron and electron-hole interactions is rather complicated [8], while the above expression is valid when the electrons are well separated and moving relatively slowly with respect to one another [9]. When discussing an electron-hole pair (see below) that the electron and hole move relatively slowly, the electron-hole is normally referred to as the Wannier exciton. As the electron and hole approach one another, the situation becomes complicated as various effects become important, most critically the entire effective-mass formalism will eventually break down. Fortunately, these effects are not particularly important in most device applications, so that it is generally sufficient to consider the Coulombic interaction potential in the above equation.

As before, we denote $\Psi_{vk}(\mathbf{r})$ as the eigen function of $H_0(\mathbf{r})$ in the valence band, and $\Psi_{ck}(\mathbf{r})$ as the conduction-band state. In the Hartree-Fock approximation, the electronic wave function is written as a Slater determinant of single-electron functions. The wave function Ψ_0 of a state where the valence band is completely occupied and the conduction band empty, can be written as

$$\Psi_0 = \frac{1}{\sqrt{n!}} \begin{vmatrix} \Psi_{vk_1}(\mathbf{r}_1) & \Psi_{vk_2}(\mathbf{r}_1) & \cdots & \Psi_{vk_h}(\mathbf{r}_1) & \cdots \\ \Psi_{vk_1}(\mathbf{r}_2) & \Psi_{vk_2}(\mathbf{r}_2) & \cdots & \Psi_{vk_h}(\mathbf{r}_2) & \cdots \\ \dots & \dots & \dots & \dots & \dots \end{vmatrix} \quad (3.111)$$

for an n -electron system. As always, \mathbf{k}_i is limited with the first Brillouin zone. Consider an excited state of the n -electron system

$$\Psi_{k_e k_h} = \frac{1}{\sqrt{n!}} \begin{vmatrix} \Psi_{vk_1}(\mathbf{r}_1) & \Psi_{vk_2}(\mathbf{r}_1) & \cdots & \Psi_{ck_e}(\mathbf{r}_1) & \cdots \\ \Psi_{vk_1}(\mathbf{r}_2) & \Psi_{vk_2}(\mathbf{r}_2) & \cdots & \Psi_{ck_e}(\mathbf{r}_2) & \cdots \\ \dots & \dots & \dots & \dots & \dots \end{vmatrix} \quad (3.112)$$

in which a single electron has been transferred from the valence-band state Ψ_{vk_h} to a conduction-band state Ψ_{ck_e} .

The excitation energy of the excited state, $(E_{k_e k_h}, \Psi_{k_e k_h})$ expressed by Eq. (3.112), above the ground state (Ψ_0, E_0) is given by

$$E_{k_e k_h} - E_0 = E_{ck_e} - E_{vk_h} + V_{k_e k_h} \quad (3.113)$$

where E_{k_e} and E_{k_h} are given by

$$\begin{aligned}
 E_{ck_e} &= \langle \Psi_{ck_e} | H_0 | \Psi_{ck_e} \rangle + \sum_i [\langle \Psi_{ck_e}(\mathbf{r}_e) \Psi_{vk_i}(\mathbf{r}_i) | V(\mathbf{r}_e - \mathbf{r}_i) | \Psi_{ck_e}(\mathbf{r}_e) \Psi_{vk_i}(\mathbf{r}_i) \rangle \\
 &\quad - \langle \Psi_{ck_e}(\mathbf{r}_e) \Psi_{vk_i}(\mathbf{r}_i) | V(\mathbf{r}_e - \mathbf{r}_i) | \Psi_{vk_i}(\mathbf{r}_e) \Psi_{ck_e}(\mathbf{r}_i) \rangle] \\
 E_{vk_h} &= \langle \Psi_{vk_h} | H_0 | \Psi_{vk_h} \rangle + \sum_i [\langle \Psi_{vk_h}(\mathbf{r}_h) \Psi_{vk_i}(\mathbf{r}_i) | V(\mathbf{r}_h - \mathbf{r}_i) | \Psi_{vk_h}(\mathbf{r}_h) \Psi_{vk_i}(\mathbf{r}_i) \rangle \\
 &\quad - \langle \Psi_{vk_h}(\mathbf{r}_h) \Psi_{vk_i}(\mathbf{r}_i) | V(\mathbf{r}_h - \mathbf{r}_i) | \Psi_{vk_i}(\mathbf{r}_h) \Psi_{vk_h}(\mathbf{r}_i) \rangle]
 \end{aligned} \tag{3.114}$$

where the summations over i run over all occupied valence-band states, i.e., all valence-band states except $\Psi_{vk_h}(\mathbf{r}_h)$. In previous discussions, we only focused on a single electron described by Eq. (1.32) and its eigen value $\langle \Psi_{k_e} | H_0 | \Psi_{k_e} \rangle$, while neglecting all other electrons in the electron system. As shown by the above expressions, such a single-electron treatment is an approximation. On the other hand, the exact numerical calculations of the above equations, usually referred to as the first-principles calculations, are almost impossible for solid state materials because of the large number of electrons included in the electron system. A general approach is to parameterize the multiple-electron Hamiltonian, such as the sp^3s^* tight-binding model and the $\mathbf{k} \cdot \mathbf{p}$ model in terms of experimental measured energy band structures of the semiconductor materials. Such a parameterization is actually a process that has already included the interactions between the single electron and its surrounding electrons since the experiment always involve a large number of electrons. In other words, E_{ck_e} from parameterized sp^3s^* tight-binding and/or $\mathbf{k} \cdot \mathbf{p}$ models are already E_{ck_e} and E_{vk_h} expressed by Eqs. (3.114), which from now on are denoted as the single-electron conduction- and valence-band energies, they replace our early notations of $\langle \Psi_{ck_e} | H_0 | \Psi_{ck_e} \rangle$ and $\langle \Psi_{vk_h} | H_0 | \Psi_{vk_h} \rangle$.

In Eq. (3.113),

$$\begin{aligned}
 V_{k_e k_h} &= - \langle \Psi_{ck_e}(\mathbf{r}_e) \Psi_{vk_h}(\mathbf{r}_h) | V(\mathbf{r}_e - \mathbf{r}_h) | \Psi_{ck_e}(\mathbf{r}_e) \Psi_{vk_h}(\mathbf{r}_h) \rangle \\
 &\quad + \langle \Psi_{ck_e}(\mathbf{r}_e) \Psi_{vk_h}(\mathbf{r}_h) | V(\mathbf{r}_e - \mathbf{r}_h) | \Psi_{vk_h}(\mathbf{r}_e) \Psi_{ck_e}(\mathbf{r}_h) \rangle
 \end{aligned} \tag{3.115}$$

The valence band with one empty state can be treated as a particle, a ‘‘hole’’, which has a positive effective mass, see Ref. [10] about the mathematic formulation of the hole. Furthermore, when we neglect the exchange interaction term, the Coulombic interaction between the electron in the conduction band and the hole in the valence band becomes

$$V_{k_e k_h} = - \langle \Psi_{ck_e}(\mathbf{r}_e) \Psi_{vk_h}(\mathbf{r}_h) | V(\mathbf{r}_e - \mathbf{r}_h) | \Psi_{ck_e}(\mathbf{r}_e) \Psi_{vk_h}(\mathbf{r}_h) \rangle$$

In other words, we can express the interaction between the electron and the hole as

$$V(\mathbf{r}_e - \mathbf{r}_h) = - \frac{e^2}{4\pi\epsilon_\infty |\mathbf{r}_e - \mathbf{r}_h|} \tag{3.116}$$

so that the hole is positively charged. A direct consequence of such a positively charged hole in the valence band and its attractive Coulombic interaction with the negatively charged electron in the conduction band is the formation of the electron-hole pair, which is the so-called exciton.

Note that the hole is not only a concept, its presence has been demonstrated experimentally, particularly through the anomalous Hall effect. With the concept of the valence-band hole, we re-formulate the wave function expressed by Eq. (3.112) of the exciton as

$$\Psi_{ck_e}(\mathbf{r}_e)\Psi_{vk_h}(\mathbf{r}_h) \quad (3.117)$$

The momentum of such an exciton is $\mathbf{k}_e + \mathbf{k}_h$.

The physical problem of Eq. (3.113) now becomes: the conduction band is occupied by one negatively charged electron denoted as \mathbf{r}_e , and the valence band is almost fully occupied except one empty state, i.e., a positively charged hole \mathbf{r}_h . The electron and hole may be described as “free” particles with definite effective masses (the effective mass approximation). Furthermore, there is a Coulombic interaction between the electron and the hole. Because of the Coulombic interaction between the electron and the hole, Eq. (3.117) does not represent an eigenstate of the exciton system. We approximate the eigenstate by a linear combination of two-particle wave functions of Eq. (3.117)

$$\Psi(\mathbf{r}_e, \mathbf{r}_h) = \sum_{\mathbf{k}_e \mathbf{k}_h} A(\mathbf{k}_e, \mathbf{k}_h) \Psi_{ck_e}(\mathbf{r}_e) \Psi_{vk_h}(\mathbf{r}_h) \quad (3.118)$$

Equation (3.113) is now extended into a set of simultaneous equations

$$\begin{aligned} & [E - E_{ck_e} + E_{vk_h}]A(\mathbf{k}_e, \mathbf{k}_h) \\ &= \sum_{\mathbf{k}'_e \mathbf{k}'_h} \langle \Psi_{ck_e}(\mathbf{r}_e) \Psi_{vk_h}(\mathbf{r}_h) | V(\mathbf{r}_e - \mathbf{r}_h) | \Psi_{ck'_e}(\mathbf{r}_e) \Psi_{vk'_h}(\mathbf{r}_h) \rangle A(\mathbf{k}'_e, \mathbf{k}'_h) \end{aligned} \quad (3.119)$$

where E is the total energy of exciton state $\Psi(\mathbf{r}_e, \mathbf{r}_h)$, E_{ck_e} and E_{vk_h} are given by Eq. (3.114).

As we discussed before concerning Eq. (3.116), we focus on the Wannier exciton which is the most relevant in device applications. Since the spatial extension of the Wannier exciton is normally quite large, we use the Bloch functions to describe the electron and hole states

$$\Psi_{ck}(\mathbf{r}) = \frac{1}{\sqrt{N}} e^{i\mathbf{k}\cdot\mathbf{r}} u_{ck}(\mathbf{r}), \quad \Psi_{vk}(\mathbf{r}) = \frac{1}{\sqrt{N}} e^{i\mathbf{k}\cdot\mathbf{r}} u_{vk}(\mathbf{r}) \quad (3.120)$$

where N is the number of unit cells and $u_{ck}(\mathbf{r})$ and $u_{vk}(\mathbf{r})$ are the periodic Bloch functions of the conduction- and valence-band states with the period of the crystal lattice and are normalized in a unit cell of volume Ω_{cell} . $\Psi_{ck}(\mathbf{r})$ and $\Psi_{vk}(\mathbf{r})$ are normalized in the semiconductor volume $\Omega = N\Omega_{\text{cell}}$.

When $u_{ck}(\mathbf{r})$ and $u_{vk}(\mathbf{r})$ vary slowly with \mathbf{k} over the range of \mathbf{k} of interest and that this range is small compared to the dimension of the Brillouin zone, one can write

$$\begin{aligned} u_{ck_e}^*(\mathbf{r}_e)u_{ck'_e}(\mathbf{r}_e) &= \frac{1}{\Omega_{\text{cell}}} \sum_{\nu} C_{\nu}(\mathbf{k}_e, \mathbf{k}'_e) e^{i\mathbf{K}_{\nu} \cdot \mathbf{r}_e} \\ u_{vk_h}^*(\mathbf{r}_h)u_{vk'_h}(\mathbf{r}_h) &= \frac{1}{\Omega_{\text{cell}}} \sum_{\mu} C_{\mu}(\mathbf{k}_h, \mathbf{k}'_h) e^{i\mathbf{K}_{\mu} \cdot \mathbf{r}_h} \end{aligned} \quad (3.121)$$

where \mathbf{K}_{ν} and \mathbf{K}_{μ} are reciprocal lattice vectors. With these

$$\begin{aligned} &\langle \Psi_{ck_e}(\mathbf{r}_e)\Psi_{vk_h}(\mathbf{r}_h) | V(\mathbf{r}_e - \mathbf{r}_h) | \Psi_{ck'_e}(\mathbf{r}_e)\Psi_{vk'_h}(\mathbf{r}_h) \rangle \\ &= \frac{1}{\Omega^2} \sum_{\nu\mu} C_{\nu}(\mathbf{k}_e, \mathbf{k}'_e) C_{\mu}(\mathbf{k}_h, \mathbf{k}'_h) \\ &\quad \times \int \int e^{i(\mathbf{K}_{\nu} - \mathbf{q}_e) \cdot \mathbf{r}_e} V(\mathbf{r}_e - \mathbf{r}_h) e^{i(\mathbf{K}_{\mu} - \mathbf{q}_h) \cdot \mathbf{r}_h} d\mathbf{r}_e d\mathbf{r}_h \\ &= \frac{1}{\Omega} \sum_{\nu\mu} C_{\nu}(\mathbf{k}_e, \mathbf{k}'_e) C_{\mu}(\mathbf{k}_h, \mathbf{k}'_h) V(\mathbf{K}_{\nu} + \mathbf{q}_e) \delta_{\mathbf{K}_{\nu} + \mathbf{K}_{\mu} + \mathbf{q}_e + \mathbf{q}_h} \end{aligned} \quad (3.122)$$

and

$$V(\mathbf{q}) = \int e^{i\mathbf{q} \cdot \mathbf{r}} V(\mathbf{r}) d\mathbf{r} = -\frac{e^2}{\epsilon_{\infty} |\mathbf{q}|} \quad (3.123)$$

is the Fourier transform of the electron-hole interaction potential Eq. (3.116). $\mathbf{q}_e = \mathbf{k}'_e - \mathbf{k}_e$ and $\mathbf{q}_h = \mathbf{k}'_h - \mathbf{k}_h$.

Since the total exciton wave vector $\mathbf{k}_e + \mathbf{k}_h$ is restricted within the first Brillouin zone, the only possible solution of the δ function in Eq. (3.122) is

$$\mathbf{K}_{\nu} + \mathbf{K}_{\mu} = 0 \quad (3.124)$$

which means

$$\mathbf{k}_e + \mathbf{k}_h = \mathbf{k}'_e + \mathbf{k}'_h \quad (3.125)$$

indicating that the exciton momentum $\mathbf{k}_e + \mathbf{k}_h$ is a quantum number of the exciton state. Equation (3.122) can be now rewritten as

$$\begin{aligned} &\langle \Psi_{ck_e}(\mathbf{r}_e)\Psi_{vk_h}(\mathbf{r}_h) | V(\mathbf{r}_e - \mathbf{r}_h) | \Psi_{ck'_e}(\mathbf{r}_e)\Psi_{vk'_h}(\mathbf{r}_h) \rangle \\ &= \frac{1}{\Omega} \sum_{\nu} C_{\nu}(\mathbf{k}_e, \mathbf{k}'_e) C_{-\nu}(\mathbf{k}_h, \mathbf{k}'_h) V(\mathbf{K}_{\nu} + \mathbf{q}_e) \end{aligned} \quad (3.126)$$

In the above expression we denote C_{μ} as $C_{-\nu}$ under the condition of $\mathbf{K}_{\nu} + \mathbf{K}_{\mu} = 0$. For $V(\mathbf{r})$ being Eq. (3.116) so that Eq. (3.123) is true, the largest contributing term

is the one when $\mathbf{K}_v = 0$ for small \mathbf{q}_e , thus

$$\begin{aligned} & \langle \Psi_{c\mathbf{k}_e}(\mathbf{r}_e)\Psi_{vk_h}(\mathbf{r}_h) | V(\mathbf{r}_e - \mathbf{r}_h) | \Psi_{c\mathbf{k}'_e}(\mathbf{r}_e)\Psi_{vk'_h}(\mathbf{r}_h) \rangle \\ &= \frac{1}{\Omega} C_0(\mathbf{k}_e, \mathbf{k}'_e) C_0(\mathbf{k}_h, \mathbf{k}'_h) V(\mathbf{q}_e) \end{aligned} \quad (3.127)$$

As said before, the periodic Bloch functions $u_{c\mathbf{k}}$ and $u_{v\mathbf{k}}$ do not vary strongly with \mathbf{k} so that we can write

$$u_{c\mathbf{k}'_e}(\mathbf{r}_e) = u_{c\mathbf{k}_e}(\mathbf{r}_e) + \mathbf{q}_e \cdot \nabla_{\mathbf{k}_e} u_{c\mathbf{k}_e}(\mathbf{r}_e) + \dots \quad (3.128)$$

Substituting this into Eq. (3.121) and integrating over one unit cell,

$$\begin{aligned} C_0(\mathbf{k}_e, \mathbf{k}'_e) &= 1 + \mathbf{q}_e \cdot \mathbf{X}_c(\mathbf{k}_e) \dots \\ C_0(\mathbf{k}_h, \mathbf{k}'_h) &= 1 - \mathbf{q}_e \cdot \mathbf{X}_v(\mathbf{k}_h) \dots \end{aligned} \quad (3.129)$$

where

$$\mathbf{X}_c(\mathbf{k}) = \langle u_{c\mathbf{k}} | \nabla_{\mathbf{k}} | u_{c\mathbf{k}} \rangle_{\Omega_{\text{cell}}}, \quad \mathbf{X}_v(\mathbf{k}) = \langle u_{v\mathbf{k}} | \nabla_{\mathbf{k}} | u_{v\mathbf{k}} \rangle_{\Omega_{\text{cell}}} \quad (3.130)$$

With Eqs. (3.129) we obtain

$$\begin{aligned} & \langle \Psi_{c\mathbf{k}_e}(\mathbf{r}_e)\Psi_{vk_h}(\mathbf{r}_h) | V(\mathbf{r}_e - \mathbf{r}_h) | \Psi_{c\mathbf{k}'_e}(\mathbf{r}_e)\Psi_{vk'_h}(\mathbf{r}_h) \rangle \\ &= \frac{V(\mathbf{q}_e)}{\Omega} \{ 1 + \mathbf{q}_e \cdot [\mathbf{X}_c(\mathbf{k}_e) - \mathbf{X}_v(\mathbf{k}_h)] + \dots \} \end{aligned} \quad (3.131)$$

to the first order in \mathbf{q}_e .

Retaining only the first term in the above equation, Eq. (3.119) becomes

$$(E - E_{c\mathbf{k}_e} + E_{v\mathbf{k}_h}) A(\mathbf{k}_e, \mathbf{k}_h) = \frac{1}{\Omega} \sum_{\mathbf{q}} V(\mathbf{q}) A(\mathbf{k}_e + \mathbf{q}, \mathbf{k}_h - \mathbf{q}) \quad (3.132)$$

It is known mathematically that when a function $f(\mathbf{k})$ is expandable in powers of \mathbf{k} , we have the following relationships

$$\frac{1}{\Omega} \sum_{\mathbf{k}} e^{i\mathbf{k} \cdot \mathbf{r}} f(\mathbf{k}) G(\mathbf{k}) = f(-i\nabla) g(\mathbf{r}), \quad g(\mathbf{r}) = \frac{1}{\Omega} \sum_{\mathbf{k}} e^{i\mathbf{k} \cdot \mathbf{r}} G(\mathbf{k}) \quad (3.133)$$

Thus, by introducing the Fourier transform of $A(\mathbf{k}_e, \mathbf{k}_h)$:

$$\psi(\mathbf{r}_e, \mathbf{r}_h) = \frac{1}{\Omega} \sum_{\mathbf{k}_e \mathbf{k}_h} e^{i\mathbf{k}_e \cdot \mathbf{r}_e} e^{i\mathbf{k}_h \cdot \mathbf{r}_h} A(\mathbf{k}_e, \mathbf{k}_h) \quad (3.134)$$

then taking the Fourier transform of Eq. (3.132), it is easy to show that $\psi(\mathbf{r}_e, \mathbf{r}_h)$ satisfy the following differential equation

$$[E_{(-i\nabla_e)} - E_{(-i\nabla_h)} + V(\mathbf{r}_e - \mathbf{r}_h)] \psi(\mathbf{r}_e, \mathbf{r}_h) = E \psi(\mathbf{r}_e, \mathbf{r}_h) \quad (3.135)$$

where $E_{(-i\nabla_e)}$ is the expression obtained by replacing \mathbf{k}_e in the power-series expansion of $E_{c\mathbf{k}_e}$ by $-i\nabla_e$. During the derivation of the above equation we have used the fact that only small \mathbf{q} contributes significantly, see Eq. (3.123).

Consider only one exciton formed by a hole in the valence band and an electron in the conduction band, and $u_v(\mathbf{r}_h)$ and $u_c(\mathbf{r}_e)$ are independent of \mathbf{k}_h and \mathbf{k}_e , it is easy to see by Eqs. (3.118) and (3.134) that the total wave function of the exciton state is

$$\Psi(\mathbf{r}_e, \mathbf{r}_h) = \Omega \psi(\mathbf{r}_e, \mathbf{r}_h) u_c(\mathbf{r}_e) u_v(\mathbf{r}_h) \quad (3.136)$$

The exciton wave function is therefore made up of a product of the electron and hole band functions times a modulation function $\psi(\mathbf{r}_e, \mathbf{r}_h)$, which is in general referred as the exciton envelope function.

We now try to express explicitly the effective-mass equation for the exciton in a bulk semiconductor material. We write the conduction band and the valence band as

$$E_{c\mathbf{k}_e} = E_c + \frac{\hbar^2 \mathbf{k}_e^2}{2m_c^*}, \quad E_{v\mathbf{k}_h} = E_v - \frac{\hbar^2 \mathbf{k}_h^2}{2m_v^*} \quad (3.137)$$

by Eq. (1.156), where E_c and E_v are conduction- and valence-bandedges. $E_c - E_v = E_g$ is the bandgap. Inserting Eq. (3.116) into Eq. (3.135) we obtain

$$\left(\frac{\hbar^2 \mathbf{k}_e^2}{2m_c^*} + \frac{\hbar^2 \mathbf{k}_h^2}{2m_v^*} - \frac{e^2}{4\pi\epsilon_\infty |\mathbf{r}_e - \mathbf{r}_h|} \right) \psi(\mathbf{r}_e, \mathbf{r}_h) = (E - E_g) \psi(\mathbf{r}_e, \mathbf{r}_h) \quad (3.138)$$

Write

$$\mathbf{r} = \frac{m_c^* \mathbf{r}_e + m_v^* \mathbf{r}_h}{m_c^* + m_v^*}$$

as the exciton center of mass,

$$m^* = m_c^* + m_v^*$$

as the exciton mass,

$$\frac{1}{\mu} = \frac{1}{m_c^*} + \frac{1}{m_v^*}$$

where μ is the reduced mass of exciton. Moreover,

$$\psi(\mathbf{r}_e, \mathbf{r}_h) = \psi_{\mathbf{k}}(\mathbf{r}) \psi_n(\mathbf{r}_e - \mathbf{r}_h), \quad E = E_{\mathbf{k}} + E_n + E_g$$

Equation (3.138) becomes

$$\begin{aligned} -\frac{\hbar^2 \nabla_{\mathbf{r}}^2}{2m^*} \psi_{\mathbf{k}}(\mathbf{r}) &= E_{\mathbf{k}} \psi_{\mathbf{k}}(\mathbf{r}) \\ \left(-\frac{\hbar^2 \nabla_{\mathbf{r}_e - \mathbf{r}_h}^2}{2\mu} - \frac{e^2}{4\pi\epsilon_\infty |\mathbf{r}_e - \mathbf{r}_h|} \right) \psi_n(\mathbf{r}_e - \mathbf{r}_h) &= E_n \psi_n(\mathbf{r}_e - \mathbf{r}_h) \end{aligned} \quad (3.139)$$

Table 3.2 Physical parameters [11, 12] at low temperature and exciton binding energy and exciton Bohr radius

	Unit	Si	GaAs	AlAs	InAs	CdSe	ZnO
m_v^* (heavy hole)	m_0	0.537	0.45	0.409	0.35	0.45	0.59
m_c^*	m_0	1.026	0.067	0.15	0.026	0.13	0.28
E_g	eV	1.1700	1.51914	2.229	0.4180	1.751	3.2
ϵ_∞	ϵ_0	11.9	12.40	10.06	15.15	6.3	7.8
Exciton binding energy E_{ex}	meV	33.8	5.16	14.7	1.3	34.5	42.4
Exciton Bohr radius a_{Br}	nm	1.78	11.2	4.8	35.8	3.3	2.2

It is easy to see that $[E_k, \psi_k(\mathbf{r})]$ represent the motion of the exciton center of mass. In bulk semiconductor, we can write

$$\psi_k(\mathbf{r}) = e^{ik \cdot \mathbf{r}}, \quad E_k = \frac{\hbar^2 k^2}{2m^*} \quad (3.140)$$

$[E_n, \psi_n(\mathbf{r}_e - \mathbf{r}_h)]$ represent the relative motion of the electron and hole with one another, which is just hydrogen like. Refer to Eqs. (1.7), (1.9),

$$\psi_n(\mathbf{r}_e - \mathbf{r}_h) = \Psi_{n\ell m}(|\mathbf{r}_e - \mathbf{r}_h|), \quad E_n = -\frac{\mu}{m_0} \left(\frac{\epsilon_0}{\epsilon_\infty} \right)^2 \frac{R_y}{n^2} \quad (3.141)$$

Neglect the motion of the exciton center of mass ($\mathbf{k} = 0$), the exciton states (different n in the above equation) are denoted by their energy E_n below E_g . The exciton ground state ($n = 1$) is

$$\psi_{\text{ex}}(\mathbf{r}_e - \mathbf{r}_h) = e^{-|\mathbf{r}_e - \mathbf{r}_h|/a_{\text{Br}}}, \quad E_{\text{ex}} = -\frac{\mu}{m_0} \left(\frac{\epsilon_0}{\epsilon_\infty} \right)^2 R_y, \quad a_{\text{Br}} = \frac{m_0}{\mu} \frac{\epsilon_\infty}{\epsilon_0} a_0 \quad (3.142)$$

where E_{ex} is the exciton binding energy, a_{Br} is the exciton Bohr radius, a_0 is the Bohr radius, see Eq. (1.4). Table 3.2 lists the exciton binding energies and exciton Bohr radii in common semiconductor bulk materials.

We now study the effect of exciton on the optical transition. The probability that the electron makes a transition from a valence-band state in a completely filled valence band to one conduction-band state in a completely empty conduction band, i.e., from the ground state Ψ_0 to an excited exciton state $\Psi(\mathbf{r}_e, \mathbf{r}_h)$ is proportional to

$$\langle \Psi(\mathbf{r}_e, \mathbf{r}_h) | e^{i\mathbf{s} \cdot \mathbf{r}} \mathbf{e}_s \cdot \mathbf{p} | \Psi_0 \rangle \quad (3.143)$$

where \mathbf{p} is the momentum operator, \mathbf{s} and \mathbf{e}_s are respectively the wave vector and polarization vector of the photon. As discussed before, \mathbf{s} is negligible as compared with the momenta of the electron and hole. Thus the total momentum of such a

photo-generated exciton is zero. In other words, the wave function of the photogenerated exciton has the form of

$$\psi(\mathbf{r}_e, \mathbf{r}_h) = \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}_e} e^{-i\mathbf{k}\cdot\mathbf{r}_h} A(\mathbf{k}, -\mathbf{k}) \quad (3.144)$$

apart from a normalization factor, by referring to the general solution Eq. (3.134) of an exciton.

By Eq. (3.118) the optical transition probability is proportional to

$$\langle \Psi(\mathbf{r}_e, \mathbf{r}_h) | \mathbf{e}_s \cdot \mathbf{p} | \Psi_0 \rangle = \sum_{\mathbf{k}} A(\mathbf{k}, -\mathbf{k}) \langle \Psi_{c\mathbf{k}} | \mathbf{e}_s \cdot \mathbf{p} | \Psi_{v\mathbf{k}} \rangle \quad (3.145)$$

where we have neglected the contribution from all other bands except the valence- and the conduction bands. By Eq. (3.90),

$$\langle \Psi_{c\mathbf{k}} | \mathbf{e}_s \cdot \mathbf{p} | \Psi_{v\mathbf{k}} \rangle \approx \mathbf{e}_s \cdot \mathbf{p}_{cv} \quad (3.146)$$

which can be well approximated as independent of \mathbf{k} over the range q of \mathbf{k} involved so that

$$\langle \Psi(\mathbf{r}_e, \mathbf{r}_h) | \mathbf{e}_s \cdot \mathbf{p} | \Psi_0 \rangle = \mathbf{e}_s \cdot \mathbf{p}_{cv} \sum_{\mathbf{k}} A(\mathbf{k}, -\mathbf{k}) \quad (3.147)$$

By Eq. (3.144) we finally see that the transition probability is therefore proportional to

$$\langle \Psi(\mathbf{r}_e, \mathbf{r}_h) | \mathbf{e}_s \cdot \mathbf{p} | \Psi_0 \rangle = \mathbf{e}_s \cdot \mathbf{p}_{cv} \psi(\mathbf{r}_e, \mathbf{r}_h) |_{\mathbf{r}_e=\mathbf{r}_h} \quad (3.148)$$

Refer to Eq. (1.12), the radial wave functions are nonzero when $\mathbf{r}_e = \mathbf{r}_h$ only for states with zero angular momentum ($\ell = 0$). Moreover,

$$\psi(\mathbf{r}_e, \mathbf{r}_h) \propto \frac{1}{\sqrt{n^3}} e^{-|\mathbf{r}_e - \mathbf{r}_h|/na_{\text{Br}}} \quad (3.149)$$

so that the optical transition matrix element is

$$|\langle \Psi(\mathbf{r}_e, \mathbf{r}_h) | \mathbf{e}_s \cdot \mathbf{p} | \Psi_0 \rangle|^2 \propto |\psi(\mathbf{r}_e, \mathbf{r}_h) |_{\mathbf{r}_e=\mathbf{r}_h}|^2 \propto \frac{1}{n^3} \quad (3.150)$$

where n is the principal quantum number in Eq. (3.142).

The most important impact of the above equation is that electron-photon interaction, i.e., light-matter interaction, is much strong when the electron and hole form an exciton, i.e., when n is small. We will elaborate the impact of Eq. (3.150) by the end of the next section after we learn how to engineer exciton in nanostructure.

3.5 Exciton in Quantum Well

In this section we will study the binding energy of an exciton in a single GaAs quantum well. Within the effective mass approximation as discussed in the previous

sections, we need to solve the Schrödinger equation along the z axis which is also the sample growth direction. By Eq. (3.138), the Hamiltonian of an exciton has the form

$$H = -\frac{\hbar^2 \nabla_e^2}{2m_e^*} - \frac{\hbar^2 \nabla_h^2}{2m_h^*} - \frac{e^2}{4\pi\epsilon_\infty |\mathbf{r}_e - \mathbf{r}_h|} + V_e(z_e) + V_h(z_h) \quad (3.151)$$

where $V_e(z_e)$ and $V_h(z_h)$ are confinement potentials for electron and hole respectively in the z direction. ϵ_∞ is the dielectric constant of the crystal.

We build a three-dimensional trial wave function for the exciton ground state as

$$\psi(\mathbf{r}_e, \mathbf{r}_h) = \psi_e(z_e)\psi_h(z_h)\psi_{\text{ex}}(\mathbf{r}_e, \mathbf{r}_h) \quad (3.152)$$

where $\mathbf{r}_e = (\boldsymbol{\rho}_e, z_e)$ and $\mathbf{r}_h = (\boldsymbol{\rho}_h, z_h)$ are spatial coordinates of the electron and the hole, respectively. $\psi_e(z_e)$ and $\psi_h(z_h)$ are eigen functions of the electron and hole of their Schrödinger equations

$$\begin{aligned} \left[-\frac{\hbar^2}{2m_e^*} \frac{d^2}{dz_e^2} + V_e(z_e) \right] \psi_e(z_e) &= E_e \psi_e(z_e) \\ \left[-\frac{\hbar^2}{2m_h^*} \frac{d^2}{dz_h^2} + V_h(z_h) \right] \psi_h(z_h) &= E_h \psi_h(z_h) \end{aligned} \quad (3.153)$$

see inset in Fig. 3.5. The exciton envelope function $\psi_{\text{ex}}(\mathbf{r}_e, \mathbf{r}_h)$ is approximated by a three-dimensional hydrogenic-type envelope wave function

$$\psi_{\text{ex}}(\mathbf{r}_e, \mathbf{r}_h) = \exp \left[-\sqrt{\frac{|\boldsymbol{\rho}_e - \boldsymbol{\rho}_h|^2}{a_{xy}^2} + \frac{(z_e - z_h)^2}{a_z^2}} \right] \quad (3.154)$$

where a_{xy} and a_z are exciton Bohr radii in the xy plane and z direction, respectively, to be determined from minimizing the total exciton energy

$$\langle \psi(\mathbf{r}_e, \mathbf{r}_h) | H | \psi(\mathbf{r}_e, \mathbf{r}_h) \rangle = E_e + E_h - E_{\text{ex}} \quad (3.155)$$

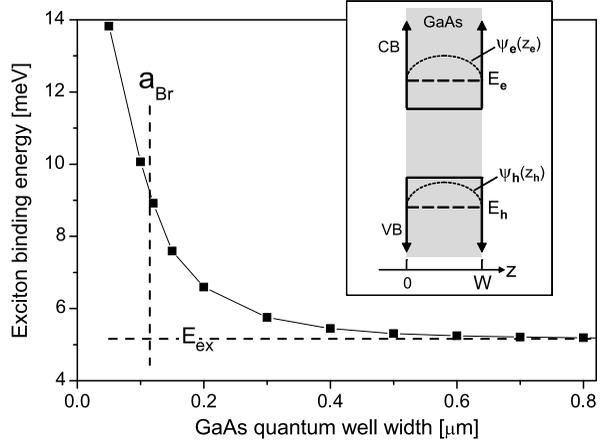
where the binding energy E_{ex} of the exciton described by Eq. (3.152) is given by

$$\begin{aligned} E_{\text{ex}} &= \int |\psi_e(z_e)\psi_h(z_h)|^2 \psi_{\text{ex}}^*(\mathbf{r}_e, \mathbf{r}_h) \\ &\quad \times \left(\frac{\hbar^2 \nabla_{\mathbf{r}_e - \mathbf{r}_h}^2}{2\mu} + \frac{e^2}{4\pi\epsilon_\infty |\mathbf{r}_e - \mathbf{r}_h|} \right) \psi_{\text{ex}}(\mathbf{r}_e, \mathbf{r}_h) d\mathbf{r}_e d\mathbf{r}_h \end{aligned} \quad (3.156)$$

Here μ is the reduced effective mass, $1/\mu = 1/m_e^* + 1/m_h^*$.

It is easy to see that Eq. (3.154) reduces to the envelope function of an exciton in a bulk material which is translationally symmetric in all the three dimensions ($a_{xy} = a_z$). In this case, the exciton Bohr radius and the binding energy of the lowest

Fig. 3.5 The binding energy of a heavy-hole-electron exciton in a single GaAs quantum well embedded in infinitely high barriers. *Inset* shows schematically the GaAs single quantum well. E_{ex} and a_{Br} are the exciton binding energy and Bohr radius of bulk GaAs



exciton state are given by Eqs. (3.142). Such hydrogenic spectra of exciton have been observed in the optical absorption that occurs below the interband threshold.

In reality, we can write down a general form for the envelope function of an exciton in a nanostructure system by introducing three different Bohr radii a_x , a_y and a_z along three spatial directions

$$\psi_{\text{ex}}(\mathbf{r}) = e^{-\alpha r}, \quad \alpha = \sqrt{\frac{x^2}{a_x^2} + \frac{y^2}{a_y^2} + \frac{z^2}{a_z^2}} \quad (3.157)$$

with which it is easy to obtain

$$\begin{aligned} \nabla^2 \psi_{\text{ex}} = \frac{e^{-\alpha r}}{\alpha^3} & \left[\left(\frac{x^2}{a_x^4} + \frac{y^2}{a_y^4} + \frac{z^2}{a_z^4} \right) \alpha - \frac{1}{a_x^2} \left(\frac{y^2}{a_y^2} + \frac{z^2}{a_z^2} \right) \right. \\ & \left. - \frac{1}{a_y^2} \left(\frac{x^2}{a_x^2} + \frac{z^2}{a_z^2} \right) - \frac{1}{a_z^2} \left(\frac{x^2}{a_x^2} + \frac{y^2}{a_y^2} \right) \right] \end{aligned} \quad (3.158)$$

which reduces to

$$\nabla^2 \psi_{\text{ex}} = \frac{e^{-r/a_{\text{Br}}}}{a_{\text{Br}} r} \left(\frac{r}{a_{\text{Br}}} - 2 \right) \quad (3.159)$$

when $a_x = a_y = a_z = a_{\text{Br}}$ for exciton in bulk material.

Back to the single GaAs quantum well, see inset in Fig. 3.5, which is grown along the z axis and extended in the xy plane. Because of the translational symmetry in the xy plane, $a_x = a_y$ so that we have two parameters a_x and a_z to describe the exciton. Actually the two parameters are not independent. Let us examine the issue from a different perspective: When the spatial coordinates of the electron and hole coincide, both the kinetic energy of the relative motion of the electron-hole pair and the Coulombic interaction between the electron and hole are infinite, but they cancel

with each other. This is easily observed for the electron-hole pair in bulk material that

$$\begin{aligned}
& \left(-\frac{\hbar^2 \psi_{\text{ex}}^* \nabla^2 \psi_{\text{ex}}}{2\mu} - \frac{e^2 |\psi_{\text{ex}}|^2}{4\pi \epsilon_\infty r} \right) \Big|_{r \rightarrow 0} \\
&= \left[-\frac{\hbar^2}{2\mu} \frac{e^{-2r/a_{\text{Br}}}}{a_{\text{Br}} r} \left(\frac{r}{a_{\text{Br}}} - 2 \right) - \frac{e^2 e^{-2r/a_{\text{Br}}}}{4\pi \epsilon_\infty r} \right] \Big|_{r \rightarrow 0} \\
&= \left(\frac{\hbar^2}{\mu a_{\text{Br}}} - \frac{e^2}{4\pi \epsilon_\infty} \right) \frac{1}{r} \Big|_{r \rightarrow 0} = 0 \times \frac{1}{r} \Big|_{r \rightarrow 0} = 0 \quad (3.160)
\end{aligned}$$

by inserting Eqs. (3.142). Therefore, for a quantum well system which is extended in the xy plane, let $x = y = z \rightarrow 0$,

$$\begin{aligned}
& \left(-\frac{\hbar^2 \psi_{\text{ex}}^* \nabla^2 \psi_{\text{ex}}}{2\mu} - \frac{e^2 |\psi_{\text{ex}}|^2}{4\pi \epsilon_\infty r} \right) \Big|_{r \rightarrow 0} \\
&= \left[\frac{\hbar^2}{\mu} \frac{a_z (2a_x^2 + a_y^2)}{a_x (a_x^2 + 2a_z^2)^{3/2}} - \frac{1}{4\sqrt{3}\pi \epsilon_\infty} \right] \frac{1}{x} \Big|_{x \rightarrow 0} \quad (3.161)
\end{aligned}$$

which is expected to be zero for any x so that

$$\frac{\sqrt{3}\beta(2 + \beta^2)}{a_x(1 + 2\beta^2)^{3/2}} = \frac{1}{a_{\text{Br}}} \quad (3.162)$$

where $\beta = a_z/a_x$.

We now calculate the binding energy of an exciton formed by the ground electron and heavy hole states in a GaAs quantum well embedded by infinitely high barriers. Refer to Table 3.2, $m_c^* = 0.067$, $m_h^* = 0.45$, $\epsilon_\infty = 12.4$ so that $E_{\text{ex}} = 5.16$ meV and $a_{\text{Br}} = 11.2$ nm in bulk GaAs. Let L_{QW} be the quantum well width so that the envelope functions of the ground electron and hole states are in the form of

$$\psi_e(z_e) = \sin\left(\frac{\pi z_e}{L_{\text{QW}}}\right), \quad \psi_h(z_h) = \sin\left(\frac{\pi z_h}{L_{\text{QW}}}\right) \quad (3.163)$$

for $(z_e, z_h) \in [0, L_{\text{QW}}]$. The binding energy of the exciton in the quantum well is the minimal value of

$$\begin{aligned}
E_{\text{ex,QW}} &= \frac{1}{C} \int_0^\infty 2\pi \rho d\rho \int_0^{L_{\text{QW}}} dz_e \int_0^{L_{\text{QW}}} dz_h \sin^2\left(\frac{\pi z_e}{L_{\text{QW}}}\right) \sin^2\left(\frac{\pi z_h}{L_{\text{QW}}}\right) e^{-2\alpha} \\
&\times \left\{ \frac{\hbar^2}{2\mu\alpha^3} \left[\left(\frac{\rho^2}{a_x^4} + \frac{(z_e - z_h)^2}{a_z^2} \right) \alpha - \frac{\rho^2}{a_x^4} - \frac{\rho^2 + 2(z_e - z_h)^2}{a_x^2 a_z^2} \right] \right. \\
&\left. + \frac{e^2}{4\pi \epsilon_\infty r} \right\} \quad (3.164)
\end{aligned}$$

in its relationship with β . Here C is the wave function normalization factor

$$C = \int_0^\infty 2\pi\rho d\rho \int_0^{L_{\text{QW}}} dz_e \int_0^{L_{\text{QW}}} dz_h \sin^2\left(\frac{\pi z_e}{L_{\text{QW}}}\right) \sin^2\left(\frac{\pi z_h}{L_{\text{QW}}}\right) e^{-2\alpha} \quad (3.165)$$

Numerical result of exciton binding energy are shown in Fig. 3.5 as a function of the width of the GaAs quantum well. The direct conclusion is straightforward: quantum confinement forces the electron and hole to stay together, resulting in a high exciton binding energy, which implies a stable exciton at high temperature.

As we have learned from the previous section, see Eq. (3.150), light-matter interaction is much strong when electron and hole stay close to form exciton. On the other hand, we know that the exciton binding energies of common semiconductor materials are very small, see Table 3.2, in terms of thermal excitation energy $k_B T = 25$ meV at room temperature. Thus room-temperature photoluminescence from bulk material is almost zero. However, in nanostructures whose sizes are comparable with exciton Bohr radii, quantum confinement strongly influences the exciton which dominate the optical absorption and the photoluminescence spectra of the nanostructures. Even at room temperature the sharp exciton lines can be detected in the absorption spectra of AlGaAs/GaAs quantum wells [13], which can hardly be observed in bulk GaAs samples. Interestingly, one of the reasons for the fast development of ZnO and its related materials is their large exciton binding energy of about 60 meV which binds the electron and hole strongly at room temperature [14]. Table 3.2 further explains that since the exciton Bohr radii of common semiconductors are in the range of nm scale, structures at nanoscale, i.e., nanostructures, are essential in order to engineer exciton. We shall discuss more the vast expanding research and technical development of nanophotonics in Chap. 5.

3.6 Colloidal Quantum Dot

To precisely engineer light-matter interactions at the nanoscale with electron-hole pairs (excitons), and their coupling with photons (excitonpolaritons), are becoming increasingly important in contemporary research in nanoscience. Materials with nanoscale features have the potential to revolutionize optoelectronics, permitting new and interesting devices and system capabilities. One key design element is the geometry, i.e. size and shape, which defines the properties of the nanostructured material and which connects to the possibility to design new and significantly improved physical, electrical and optical properties. Excitons represent one such important property that can be dictated by the size and the shape of a nanostructure in which they are confined, in addition to the material constitution of the nanostructure. Here we outline the most studied quantum dot and the optical properties of the exciton confined in the quantum dot. Figure 3.6 shows schematically a colloidal semiconductor quantum dot.

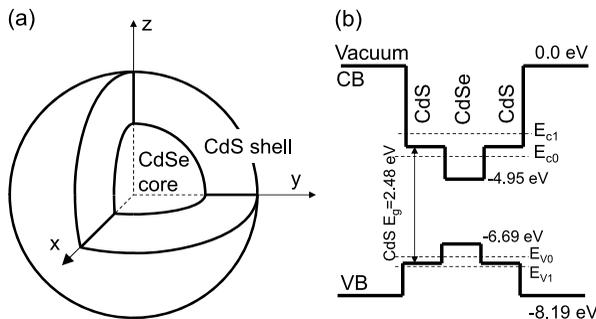


Fig. 3.6 (a) Schematic configuration of a typical colloidal quantum dot consisting of a CdSe core and a CdS shell. (b) Conduction-band (CB) and valence-band (VB) structures of the core-shell quantum dot

Optical spectroscopy deals with interactions of electromagnetic radiation with matter that occur at the ultraviolet (UV), visible (Vis), near-infrared (NIR) and infrared (IR) wavelengths. In the UV/VIS spectral regions (< 700 nm), light can penetrate only superficial tissue volumes (a few hundred microns to a millimeter in depth) due to the fact that this biological medium is highly absorbing. However, in the NIR spectral region ($700 \sim 900$ nm) tissue is generally less absorbing and, furthermore, the number of elastic scattering events of light in tissue is approximately two orders of magnitude greater than the number of absorption events. This enables the light to propagate through tissue volumes that are up to several centimeters in depth.

The success of bio application depends very much on the optical excitation and detection technologies, which are in the infrared regimes in most cases.

Because the quantum dot of interest consists of more than 10^5 atoms, we can apply the effective mass approximation, i.e., the single-particle Hamiltonian of a conduction-band electron in a quantum dot can be expressed in the form of

$$H_c = -\frac{\hbar^2 \nabla^2}{2m_c^*} + V_c(\mathbf{r}) \quad (3.166)$$

where m_c^* is the effective mass of the electron, $V_c(\mathbf{r})$ is the confinement potential

$$V_c(\mathbf{r}) = \begin{cases} E_c & |\mathbf{r}| < a \\ E_c + \Delta_c(\mathbf{r}) & \text{otherwise} \end{cases} \quad (3.167)$$

where E_c is the conduction bandedge of the quantum dot core, a is the radius of the quantum dot, and Δ_c is the band offset between the quantum dot and its surrounding medium. When referring to only quantum dot and vacuum such as a CdSe quantum dot (without CdS shell in Fig. 3.6), $\Delta_c = \chi$, which is the electron affinity of CdSe. Similar expressions can be written down for the hole in the valence band. Relevant energy band structure parameters are listed in Table 3.3.

Table 3.3 Room-temperature energy band structure parameters of II–VI materials [11, 16–25]. Data are obtained from [25] if not otherwise stated

Parameter	Wurtzite CdS	Hexagonal CdSe	Zincblende CdTe
E_g [eV]	2.485	1.751	1.49
Δ_{so} [eV]	0.062	0.416	0.80
m_c^*	0.18 [15], 0.2, 0.20 [11]	0.12, 0.13 [11, 16]	0.090, 0.096 [11]
$m_{v\perp}^*/m_{v\parallel}^*$	0.53 [15], 0.7/5, 0.8 [17], 1.02 [18]	0.45/1, 0.41 [16]	0.72 0.12
$\epsilon(0)_\perp/\epsilon(0)_\parallel$	8.45/9.12	9.29/10.16	10.2
$\epsilon(\infty)_\perp/\epsilon(\infty)_\parallel$	5.32	6.3/6.2	7.1
E_p [eV]			20.7 [11]
Δ_c [eV]	2.21 [19], 2.3 ~ 2.5 [20, 21]		
Δ_v [eV]	2.21 [19]		
χ [eV]	4.79 [22], 4.22 [23]	4.95 [22]	4.28 [22]
Parameter	Cubic ZnS	Wurtzite ZnS	Wurtzite ZnO
E_g [eV]	3.68	3.78	3.2
Δ_{so} [eV]		0.086	
m_c^*	0.184, 0.28 [11], 0.42 [24]	0.28	0.24, 0.19 [11]
$m_{v\perp}^*/m_{v\parallel}^*$	1.76, 0.61 [24] 0.23	0.49/1.4	0.59, 0.45 [17]
$\epsilon(0)_\perp/\epsilon(0)_\parallel$	8.0	9.6	7.8/8.75
$\epsilon(\infty)_\perp/\epsilon(\infty)_\parallel$	5.1	5.13	3.70/3.75
E_p [eV]	20.4 [11]		
Δ_c [eV]			
Δ_v [eV]			
χ [eV]	3.9 [22]		4.57 [22]

Advanced theories to account for the fine structures in the energy bands of II–VI semiconductor nanocrystals have been developed, e.g., the effective bond-orbital model for valence-band holes [26, 27], the sp^3s^* semi-empirical tight-binding theory [16, 24, 28], and sp^3d^5 nearest-neighbor tight-binding model [15, 29]. Advanced theories improve very much the energy positions of peaks in the optical spectrum [30], it is expected that the general principal features of the optical spectrum are determined by the geometric structures of the quantum dot. It has been shown numerically that the ground level energies obtained by the effective mass approximation of Table 3.3 are very close to the results from the sp^3d^5 nearest-neighbor tight-binding model (Fig. 3 in [15]).

Because of the rotational invariance, we pick wave functions of the Hamiltonian Eq. (3.166) in the form

$$\psi_{n\ell m}(r, \theta, \phi) = R_{n\ell}(r)Y_{\ell m}(\theta, \phi) \quad (3.168)$$

from which comes the radial Schrödinger equation

$$\frac{d^2 R_{n\ell}(r)}{dr^2} + \frac{2}{r} \frac{dR_{n\ell}(r)}{dr} - \frac{\ell(\ell+1)}{r^2} R_{n\ell}(r) + \frac{2m_c^*[E - V_c(r)]}{\hbar^2} R_{n\ell}(r) = 0 \quad (3.169)$$

where $Y_{\ell, -m}(\theta, \phi) = (-1)^m Y_{\ell m}^*(\theta, \phi)$ are the angular momentum eigen functions.

We discuss four cases.

Free Electron For $V_c(r) = 0$ and $E > 0$, we denote $k^2 = 2m_c^*E/\hbar^2$, the regular solution of Eq. (3.169) is $j_\ell(\rho)$, known as spherical Bessel functions

$$j_\ell(\rho) = (-\rho)^\ell \left(\frac{1}{\rho} \frac{d}{d\rho} \right)^\ell \left(\frac{\sin \rho}{\rho} \right) \quad (3.170)$$

where $\rho = kr$. And the irregular solution is called the spherical Neumann function $n_\ell(\rho)$

$$n_\ell(\rho) = -(-\rho)^\ell \left(\frac{1}{\rho} \frac{d}{d\rho} \right)^\ell \left(\frac{\cos \rho}{\rho} \right) \quad (3.171)$$

The first few values of these functions are

$$\begin{aligned} j_0(\rho) &= \frac{\sin \rho}{\rho} \\ j_1(\rho) &= \frac{\sin \rho}{\rho^2} - \frac{\cos \rho}{\rho} \\ j_2(\rho) &= \left(\frac{3}{\rho^3} - \frac{1}{\rho} \right) \sin \rho - \frac{3}{\rho^2} \cos \rho \\ n_0(\rho) &= -\frac{\cos \rho}{\rho} \\ n_1(\rho) &= -\frac{\cos \rho}{\rho^2} - \frac{\sin \rho}{\rho} \\ n_2(\rho) &= -\left(\frac{3}{\rho^3} - \frac{1}{\rho} \right) \cos \rho - \frac{3}{\rho^2} \sin \rho \end{aligned} \quad (3.172)$$

Infinite Potential Well For infinite potential well of

$$V_c(\mathbf{r}) = \begin{cases} 0 & |\mathbf{r}| < a \\ \infty & \text{otherwise} \end{cases} \quad (3.173)$$

Table 3.4 Bessel function roots of ka for a few values of ℓ

$\ell = 0$	1	2	3	4	5
3.1416 ($= \pi$)	4.4934	5.7635	6.9879	8.1826	9.3558
6.2832 ($= 2\pi$)	7.7252	9.0950	10.4171	11.7049	12.9665
9.4248 ($= 3\pi$)	10.9042	12.3229	13.6980	15.0397	16.3547
12.5664 ($= 4\pi$)	14.0661	15.5146	16.9236	18.3013	19.6531
15.7080 ($= 5\pi$)	17.2208	18.6890	20.1218	21.5255	22.9046

the solution that is regular at $r = 0$ is $R_\ell(r) = j_\ell(kr)$ with the eigen values determined by the condition that the solution vanishes at $r = a$, i.e., $j_\ell(ka) = 0$. The roots of ka for a few values of ℓ are listed in Table 3.4.

Square Potential Well In reality, the potential barrier for the electron in a quantum dot is always finite. Refer to Fig. 3.6, for a CdSe quantum dot (without the CdS shell) so that its potential energy is in the form of a square potential well

$$V_c(\mathbf{r}) = \begin{cases} -\chi & |\mathbf{r}| < a \\ 0 & \text{otherwise} \end{cases} \quad (3.174)$$

where a is the radius of the quantum dot. For continuum solutions $E > 0$, the solution for $r > a$ will be a combination of the regular and irregular solutions

$$R_\ell(r) = B_\ell j_\ell(kr) + C_\ell n_\ell(kr) \quad (3.175)$$

while the solution for $r \leq a$ must be the regular one of

$$R_\ell(r) = A_\ell j_\ell(qr) \quad (3.176)$$

where

$$q^2 = \frac{2m_c^*(E + \chi)}{\hbar^2}, \quad k^2 = \frac{2m_c^*E}{\hbar^2}$$

The two solutions and their derivatives must match at $r = a$. This leads to

$$q \frac{1}{j_\ell(\rho)} \frac{dj_\ell(\rho)}{d\rho} \Big|_{\rho=qa} = k \frac{B_\ell \frac{dj_\ell(\rho)}{d\rho} + C_\ell \frac{dn_\ell(\rho)}{d\rho}}{B_\ell j_\ell(\rho) + C_\ell n_\ell(\rho)} \Big|_{\rho=ka} \quad (3.177)$$

by which the ratio of B_ℓ/C_ℓ is determined. The ratio of A_ℓ/C_ℓ is

$$\frac{A_\ell}{C_\ell} = \frac{\frac{B_\ell}{C_\ell} j_\ell(ka) + n_\ell(ka)}{j_\ell(qa)} \quad (3.178)$$

and C_ℓ is obtained by the wave function normalization

$$|C_\ell|^2 \left\{ \left| \frac{A_\ell}{C_\ell} \right|^2 \int_0^a j_\ell^2(kr) r^2 dr + \int_a^\infty \left| \frac{B_\ell}{C_\ell} j_\ell(kr) + n_\ell(kr) \right|^2 r^2 dr \right\} = 1 \quad (3.179)$$

For bound states $-\chi \leq E \leq 0$, we write $\alpha^2 = -2m_c^* E/\hbar^2$. The solution for $r \leq a$ must be regular at the origin

$$R_\ell(r) = A_\ell j_\ell(qr) \quad (3.180)$$

and the solution for $r > a$ must vanish as $r \rightarrow \infty$,

$$R_\ell(r) = B_\ell h_\ell(i\alpha r) \quad (3.181)$$

where $h_\ell(\rho)$ is the spherical Hankel function

$$h_\ell(\rho) = j_\ell(\rho) + in_\ell(\rho) \quad (3.182)$$

The first three spherical Hankel functions are

$$\begin{aligned} h_0(\rho) &= -\frac{ie^{i\rho}}{\rho} \\ h_1(\rho) &= -\frac{e^{i\rho}}{\rho} \left(1 + \frac{i}{\rho}\right) \\ h_2(\rho) &= \frac{ie^{i\rho}}{\rho} \left(1 + \frac{3i}{\rho} - \frac{3}{\rho^2}\right) \end{aligned} \quad (3.183)$$

The solutions of bound states are given by the boundary conditions

$$q \frac{1}{j_\ell(\rho)} \frac{dj_\ell(\rho)}{d\rho} \Big|_{\rho=qa} = i\alpha \frac{1}{h_\ell(\rho)} \frac{dh_\ell(\rho)}{d\rho} \Big|_{\rho=i\alpha a} \quad (3.184)$$

A_ℓ/B_ℓ and B_ℓ are

$$\begin{aligned} \frac{A_\ell}{B_\ell} &= \frac{h_\ell(i\alpha a)}{j_\ell(qa)} \\ |B_\ell|^2 &\left[\left| \frac{A_\ell}{B_\ell} \right|^2 \int_0^a j_\ell^2(qr) r^2 dr + \int_a^\infty |h_\ell(i\alpha r)|^2 r^2 dr \right] = 1 \end{aligned} \quad (3.185)$$

The spherical Bessel functions have the following recurrence relation:

$$j_{\ell+1}(\rho) = \frac{\ell}{\rho} j_\ell(\rho) - \frac{dj_\ell(\rho)}{d\rho} = \frac{2\ell+1}{\rho} j_\ell(\rho) - j_{\ell-1}(\rho) \quad (3.186)$$

Alternatively, we can expression the relationship as

$$\begin{bmatrix} j_{\ell+1}(\rho) \\ j'_{\ell+1}(\rho) \end{bmatrix} = \begin{bmatrix} \ell/\rho & -1 \\ 1 - \ell(\ell+2)/\rho^2 & (\ell+2)/\rho \end{bmatrix} \begin{bmatrix} j_\ell(\rho) \\ j'_\ell(\rho) \end{bmatrix} \quad (3.187)$$

The same recurrence relation holds true for $n_\ell(\rho)$ and $h_\ell(\rho)$. Followings are the recurrence relation in different forms:

$$\begin{aligned}
 j_{\ell-1}(\rho) + j_{\ell+1}(\rho) &= \frac{2\ell+1}{\rho} j_\ell(\rho) \\
 \ell j_{\ell-1}(\rho) - (\ell+1)j_{\ell+1}(\rho) &= (2\ell+1) \frac{dj_\ell(\rho)}{d\rho} \\
 \frac{d}{d\rho} [\rho^{\ell+1} j_\ell(\rho)] &= x^{\ell+1} f_{\ell-1}(\rho) \\
 \frac{d}{d\rho} [\rho^{-\ell} j_\ell(\rho)] &= x^{-\ell} f_{\ell+1}(\rho)
 \end{aligned} \tag{3.188}$$

Arbitrary Potential Profile For more complicated quantum dots such as the CdSe-CdS core-shell quantum dot in Fig. 3.6 where the confinement potential varies with r , it is convenient to introduce the function $u_{n\ell}(r) = r R_{n\ell}(r)$ so that

$$\frac{d^2 u_{n\ell}(r)}{dr^2} + \frac{2m_c^*}{\hbar^2} \left[E - V_c(r) - \frac{\ell(\ell+1)\hbar^2}{2m_c^* r^2} \right] u_{n\ell}(r) = 0 \tag{3.189}$$

from Eq. (3.169) which looks very much like a one-dimensional Schrödinger equation (see Sect. 6.2) except that the definition of $u_\ell(r)$ and the finiteness of the wave function at the origin ($r=0$) require that $u_\ell(0) = 0$.

As mentioned early, the matrix element of optical transition between a valence-band state

$$\psi_i u_v = R_{\ell_i}(r) Y_{\ell_i m_i}(\theta, \phi) u_v$$

and a conduction-band state

$$\psi_j u_c = R_{\ell_j}(r) Y_{\ell_j m_j}(\theta, \phi) u_c$$

denoted as interband transition, is

$$\begin{aligned}
 \langle \psi_j u_c | \mathbf{A} \cdot \mathbf{p} | \psi_i u_v \rangle &= \mathbf{A} \cdot \mathbf{p}_{cv} \langle \psi_j | \psi_i \rangle \\
 &= \mathbf{A} \cdot \mathbf{p}_{cv} \int_0^\infty r^2 dr \int_0^\pi \sin \theta d\theta \int_0^{2\pi} d\phi R_{\ell_j}^*(r) Y_{\ell_j m_j}^*(\theta, \phi) R_{\ell_i}(r) Y_{\ell_i m_i}(\theta, \phi) \\
 &= \mathbf{A} \cdot \mathbf{p}_{cv} \delta_{\ell_j \ell_i} \delta_{m_j m_i} \int_0^\infty r^2 dr R_{\ell_j}^*(r) R_{\ell_i}(r)
 \end{aligned} \tag{3.190}$$

where \mathbf{A} is the vector potential of the electromagnetic field.

Between two states within the same band, either the conduction band or valence band, referred to as intraband transition, the optical matrix element is

$$\langle \psi_j u_c | \mathbf{A} \cdot \mathbf{p} | \psi_i u_c \rangle = \langle \psi_j | \mathbf{A} \cdot \mathbf{p} | \psi_i \rangle = m_0 \mathbf{A} \cdot \left\langle \psi_j \left| \frac{d\mathbf{r}}{dr} \right| \psi_i \right\rangle$$

$$\begin{aligned}
&= \frac{im_0}{\hbar} \mathbf{A} \cdot \langle \psi_j | [H, \mathbf{r}] | \psi_i \rangle \\
&= \frac{im_0(E_j - E_i)}{\hbar} \langle \psi_j | \mathbf{A} \cdot \mathbf{r} | \psi_i \rangle
\end{aligned} \tag{3.191}$$

and

$$\begin{aligned}
&\langle \psi_j | \mathbf{A} \cdot \mathbf{r} | \psi_i \rangle \\
&= \int_0^\infty r^2 dr \int_0^\pi \sin \theta d\theta \int_0^{2\pi} d\phi R_{\ell_j}^*(r) Y_{\ell_j m_j}^*(\theta, \phi) \mathbf{A} \cdot \mathbf{r} R_{\ell_i}(r) Y_{\ell_i m_i}(\theta, \phi)
\end{aligned} \tag{3.192}$$

Writing $\mathbf{r} = r\mathbf{r}_0$, where \mathbf{r}_0 is the unit vector, $\mathbf{A} = A_x \mathbf{x}_0 + A_y \mathbf{y}_0 + A_z \mathbf{z}_0$, \mathbf{x}_0 , \mathbf{y}_0 and \mathbf{z}_0 are unit vectors along x -, y - and z directions, respectively,

$$\begin{aligned}
\mathbf{A} \cdot \mathbf{r}_0 &= A_x \sin \theta \cos \phi + A_y \sin \theta \sin \phi + A_z \cos \theta \\
&= \sqrt{\frac{4\pi}{3}} \left(\frac{-A_x + iA_y}{\sqrt{2}} Y_{11} + A_z Y_{10} + \frac{A_x + iA_y}{\sqrt{2}} Y_{1\bar{1}} \right)
\end{aligned} \tag{3.193}$$

so that the angular integral in Eq. (3.192) now becomes

$$\int_0^\pi \sin \theta d\theta \int_0^{2\pi} d\phi Y_{\ell_j m_j}^*(\theta, \phi) Y_{1,m}(\theta, \phi) Y_{\ell_i m_i}(\theta, \phi) \tag{3.194}$$

for $m = -1, 0, 1$. The azimuthal integration yields

$$\int_0^{2\pi} d\phi e^{-im_j \phi} e^{im \phi} e^{im_i \phi} = 2\pi \delta_{m, m_j - m_i} \tag{3.195}$$

If we define the z axis to lie along the photon momentum direction \mathbf{s} so that $A_z = 0$ and thus $m = \pm 1$, the above equation indicates quantum selections of $m_j - m_i = \pm 1$.

By using the addition theorem for spherical harmonics

$$Y_{\ell_1 m_1}(\theta, \phi) Y_{\ell_2 m_2}(\theta, \phi) = \sum_{\ell=|\ell_1-\ell_2|}^{\ell_1+\ell_2} C(\ell, m_1 + m_2 | \ell_1 m_1; \ell_2 m_2) Y_{\ell, m_1+m_2}(\theta, \phi) \tag{3.196}$$

so that the angular integration in the optical matrix element vanishes unless $\ell_j = \ell_i + 1, \ell_i, |\ell_i - 1|$, which is the general form of the electric dipole radiation selection rule $\Delta \ell = 1, 0, -1$. In the above equation, $C(\ell m | \ell_1 m_1; \ell_2 m_2)$ are the Clebsch-Gordan (CG) coefficients which are nonzero only when $m = m_1 + m_2$. The expression below for the CG-coefficient is the most symmetric one of the various existing forms [31, 32]

Table 3.5 Clebsch-Gordan (CG) coefficients $C(\ell m|\ell_1 m_1; 1, m_2)$ when $m = m_1 + m_2$

ℓ	$m_2 = 1$	$m_2 = 0$	$m_2 = -1$
$\ell_1 + 1$	$\sqrt{\frac{(\ell_1+m)(\ell_1+m+1)}{(2\ell_1+1)(2\ell_1+2)}}$	$\sqrt{\frac{(\ell_1-m+1)(\ell_1+m+1)}{(2\ell_1+1)(\ell_1+1)}}$	$\sqrt{\frac{(\ell_1-m)(\ell_1-m+1)}{(2\ell_1+1)(2\ell_1+2)}}$
ℓ_1	$-\sqrt{\frac{(\ell_1+m)(\ell_1-m+1)}{2\ell_1(\ell_1+1)}}$	$\sqrt{\frac{m^2}{\ell_1(\ell_1+1)}}$	$\sqrt{\frac{(\ell_1-m)(\ell_1+m+1)}{2\ell_1(\ell_1+1)}}$
$\ell_1 - 1$	$\sqrt{\frac{(\ell_1-m)(\ell_1-m+1)}{2\ell_1(2\ell_1+1)}}$	$-\sqrt{\frac{(\ell_1-m)(\ell_1+m)}{\ell_1(2\ell_1+1)}}$	$\sqrt{\frac{(\ell_1+m)(\ell_1+m+1)}{2\ell_1(2\ell_1+1)}}$
$\ell_1 + 1$	$\sqrt{\frac{(\ell_1+m_1+1)(\ell_1+m_1+2)}{(2\ell_1+1)(2\ell_1+2)}}$	$\sqrt{\frac{(\ell_1-m_1+1)(\ell_1+m_1+1)}{(2\ell_1+1)(\ell_1+1)}}$	$\sqrt{\frac{(\ell_1-m_1+1)(\ell_1-m_1+2)}{(2\ell_1+1)(2\ell_1+2)}}$
ℓ_1	$-\sqrt{\frac{(\ell_1+m_1+1)(\ell_1-m_1)}{2\ell_1(\ell_1+1)}}$	$\sqrt{\frac{m_1^2}{\ell_1(\ell_1+1)}}$	$\sqrt{\frac{(\ell_1-m_1+1)(\ell_1+m_1)}{2\ell_1(\ell_1+1)}}$
$\ell_1 - 1$	$\sqrt{\frac{(\ell_1-m_1-1)(\ell_1-m_1)}{2\ell_1(2\ell_1+1)}}$	$-\sqrt{\frac{(\ell_1-m_1)(\ell_1+m_1)}{\ell_1(2\ell_1+1)}}$	$\sqrt{\frac{(\ell_1+m_1-1)(\ell_1+m_1)}{2\ell_1(2\ell_1+1)}}$

Table 3.6 Clebsch-Gordan (CG) coefficients $C(\ell, m + m_2|1m; \ell_2 m_2)$

	$m = 1$	$m = 0$	$m = -1$
$\ell = \ell_2 + 1$	$\sqrt{\frac{(\ell_2+m_2+1)(\ell_2+m_2+2)}{(2\ell_2+1)(2\ell_2+2)}}$	$\sqrt{\frac{(\ell_2-m_2+1)(\ell_2+m_2+1)}{(2\ell_2+1)(\ell_2+1)}}$	$\sqrt{\frac{(\ell_2-m_2+1)(\ell_2-m_2+2)}{(2\ell_2+1)(2\ell_2+2)}}$
$\ell = \ell_2$	$\sqrt{\frac{(\ell_2+m_2+1)(\ell_2-m_2)}{2\ell_2(\ell_2+1)}}$	$-\frac{m_2}{\sqrt{\ell_2(\ell_2+1)}}$	$-\sqrt{\frac{(\ell_2-m_2+1)(\ell_2+m_2)}{2\ell_2(\ell_2+1)}}$
$\ell = \ell_2 - 1$	$\sqrt{\frac{(\ell_2-m_2-1)(\ell_2-m_2)}{2\ell_2(2\ell_2+1)}}$	$-\sqrt{\frac{(\ell_2-m_2)(\ell_2+m_2)}{\ell_2(2\ell_2+1)}}$	$\sqrt{\frac{(\ell_2+m_2-1)(\ell_2+m_2)}{2\ell_2(2\ell_2+1)}}$

$$C(\ell m|\ell_1 m_1; \ell_2 m_2)$$

$$\begin{aligned}
&= \delta_{m, m_1+m_2} \sqrt{2\ell+1} \sqrt{\frac{(\ell_1+\ell_2-\ell)!(\ell_1-\ell_2+\ell)!(-\ell_1+\ell_2+\ell)!}{(\ell_1+\ell_2+\ell+1)!}} \\
&\times \sum_k \frac{(-1)^k \sqrt{(\ell_1+m_1)!(\ell_1-m_1)!(\ell_2+m_2)!(\ell_2-m_2)!(\ell+m)!(\ell-m)!}}{k!(\ell_1+\ell_2-\ell-k)!(\ell_1-m_1-k)!(\ell_2+m_2-k)!(\ell-\ell_2+m_1+k)!(\ell-\ell_1-m_2+k)!}
\end{aligned} \tag{3.197}$$

and the sum runs over all values of integer k which do not lead to negative factorials.

Table 3.5 below lists $C(\ell m|\ell_1 m_1; 1, m_2)$, where $m = m_1 + m_2$.

Table 3.6 lists $C(\ell m|\ell_1 m_1; 1m_2)$, where $m = m_1 + m_2$, obtained from the analytical expression of Refs. [31, 32].

Cylindrical Quantum Wire Finally we brief introduce the cylindrical quantum wire positioned along the x axis

$$\left[-\frac{\hbar^2}{2m_c^*} \left(\frac{d^2}{dy^2} + \frac{d^2}{dz^2} \right) + V_c(y, z) \right] \psi(y, z) = E \psi(y, z) \tag{3.198}$$

where

$$V_c(y, z) = \begin{cases} -\Delta_c & y^2 + z^2 \leq a^2 \\ 0 & y^2 + z^2 > a^2 \end{cases} \quad (3.199)$$

where we assume $\Delta_c \geq 0$. In the cylindrical coordinate of $(y, z) = (\rho, \phi)$ and by writing $\psi(y, z) = e^{im\phi} R_m(\rho)$,

$$\frac{d^2 R_m(\rho)}{d\rho^2} + \frac{1}{\rho} \frac{dR_m(\rho)}{d\rho} + \left\{ \frac{2m_c^*[E - V_c(\rho)]}{\hbar^2} - \frac{m^2}{\rho^2} \right\} R_m(\rho) = 0 \quad (3.200)$$

where m is an integer (negative or positive so the state of $|m|$ is two-fold degenerate).

For free electrons where $\Delta_c = 0$ and $E \geq 0$, Eq. (3.200) is the Bessel equation

$$\rho^2 \frac{d^2 R_m(\rho)}{d\rho^2} + \rho \frac{dR_m(\rho)}{d\rho} + (k^2 \rho^2 - m^2) R_m(\rho) = 0 \quad (3.201)$$

where $k^2 = 2m_c^* E / \hbar^2$.

For finite Δ_c and for energy states confined in the quantum wire, i.e., $-\Delta_c \leq E < 0$, the wave function outside of the quantum wire ($\rho > a$) is described by the Helmholtz equation

$$\rho^2 \frac{d^2 R_m(\rho)}{d\rho^2} + \rho \frac{dR_m(\rho)}{d\rho} - (\alpha^2 \rho^2 + m^2) R_m(\rho) = 0 \quad (3.202)$$

where $\alpha^2 = -2m_c^* E / \hbar^2$. Inside the quantum wire ($\rho \leq a$), we still have the Bessel equation

$$\rho^2 \frac{d^2 R_m(\rho)}{d\rho^2} + \rho \frac{dR_m(\rho)}{d\rho} + (q^2 \rho^2 - m^2) R_m(\rho) = 0 \quad (3.203)$$

where $q^2 = 2m_c^*(E + \Delta_c) / \hbar^2$.

The Bessel function

$$J_n(x) = \sum_{s=0}^{\infty} \frac{(-1)^s}{s!(n+s)!} \left(\frac{x}{2}\right)^{n+2s} = \frac{x^2}{2^n n!} - \frac{x^{n+2}}{2^{n+2}(n+1)!} + \dots \quad (3.204)$$

which also holds for $n < 0$. And $J_{-n}(x) = (-1)^n J_n(x)$. Recurrence relations are

$$J_{n-1}(x) + J_{n+1}(x) = \frac{2n}{x} J_n(x) \quad (3.205)$$

$$J_{n-1}(x) - J_{n+1}(x) = 2J'_n(x)$$

Neumann functions, which are Bessel functions of the second kind,

$$N_\nu(x) = \frac{\cos(\nu\pi) J_\nu(x) - J_{-\nu}(x)}{\sin(\nu\pi)} \quad (3.206)$$

is also a solution of Bessel's equation when $\nu = n$, an integer. The most general solution for the Bessel's equation can therefore be written as

$$A_n J_n(x) + B_n N_n(x) \quad (3.207)$$

$J_n(x)$ and $N_n(x)$ are independent. $N_n(x)$ diverges at least logarithmically ($\ln x$) so that any boundary condition that requires the solution to be finite at the origin automatically excludes $N_n(x)$. Conversely, in the absence of such a requirement $N_n(x)$ must be considered. Neumann functions satisfy the recurrence relations

$$\begin{aligned} N_{n-1}(x) + N_{n+1}(x) &= \frac{2n}{x} N_n(x) \\ N_{n-1}(x) - N_{n+1}(x) &= 2N'_n(x) \end{aligned} \quad (3.208)$$

The solutions of the Helmholtz equation, Eq. (3.202), are Bessel functions of imaginary argument, often written as the modified Bessel functions

$$I_n(x) = i^{-n} J_n(ix) \quad (3.209)$$

In terms of infinite series,

$$I_\nu = \sum_{s=0}^{\infty} \frac{1}{s!(s+\nu)!} \left(\frac{x}{2}\right)^{2s+\nu} \quad (3.210)$$

$I_n(x) = I_{-n}(x)$. The recurrence relations for $I_\nu(x)$ are

$$\begin{aligned} I_{\nu-1}(x) - I_{\nu+1}(x) &= \frac{2\nu}{x} I_\nu(x) \\ I_{\nu-1}(x) + I_{\nu+1}(x) &= 2I'_\nu(x) \end{aligned} \quad (3.211)$$

The second independent solution of Eq. (3.202) when ν is an integer is

$$K_\nu(x) = \frac{\pi}{2} \frac{I_{-\nu}(x) - I_\nu(x)}{\sin(\nu\pi)} \quad (3.212)$$

$K_\nu(x) = K_{-\nu}(x)$, which satisfies the recurrence relations

$$\begin{aligned} K_{\nu-1}(x) - K_{\nu+1}(x) &= -\frac{2\nu}{x} K_\nu(x) \\ K_{\nu-1}(x) + K_{\nu+1}(x) &= -2K'_\nu(x) \end{aligned} \quad (3.213)$$

Some values of the root of the equation $J_n(x) = 0$ are listed in Table 3.7.

Table 3.7 Roots of the equation $J_n(x) = 0$

	$n = 0$	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$
1	2.405	3.832	5.135	6.379	7.586	8.780
2	5.520	7.016	8.417	9.760	11.064	12.339
3	8.654	10.173	11.620	13.017	14.373	15.700
4	11.792	13.323	14.796	16.224	17.616	18.982
5	14.931	16.470	17.960	19.410	20.827	22.220

3.7 Exciton Polariton

For an optical absorption process, induced by an external electromagnetic field $\mathbf{E}(\mathbf{r}, t)$, in which an allowed electric dipole transition creates an exciton, $\Psi_{n\mathbf{K}}$, from a filled valence band, the wave function of the initial state is simply unity, which is denoted by Ψ_0 in the formalism of second quantization

$$|\Psi_{\mathbf{k}_e, \mathbf{k}_h}(\mathbf{r}_e, \mathbf{r}_h)\rangle \equiv \psi_{\mathbf{k}_e}(\mathbf{r}_e)\psi_{\mathbf{k}_h}(\mathbf{r}_h) = \alpha_{\mathbf{k}_e}^+ \beta_{\mathbf{k}_h}^+ |\Psi_0(\mathbf{r}_e, \mathbf{r}_h)\rangle \quad (3.214)$$

in which an electron has been raised to the conduction band at \mathbf{k}_e (creation of an electron in the conduction band, $\alpha_{\mathbf{k}_e}^+$), leaving a hole in the valence band at \mathbf{k}_h (creation of a hole in the valence band, $\beta_{\mathbf{k}_h}^+$). The exciton state can now be written as

$$|\Psi_{n\mathbf{K}}(\mathbf{r}_e, \mathbf{r}_h)\rangle = \sum_{\mathbf{k}_e, \mathbf{k}_h} |\Psi_{\mathbf{k}_e, \mathbf{k}_h}\rangle \langle \Psi_{\mathbf{k}_e, \mathbf{k}_h} | \Psi_{n\mathbf{K}} \rangle = \sum_{\mathbf{k}_e, \mathbf{k}_h} A_{n\mathbf{K}}(\mathbf{k}_e, \mathbf{k}_h) \alpha_{\mathbf{k}_e}^+ \beta_{\mathbf{k}_h}^+ |\Psi_0\rangle \quad (3.215)$$

where A 's are the same coefficients as in Eq. (3.118). In the following we denote $\langle \Psi_0 | H_0 | \Psi_0 \rangle = 0$ as the energy of the ground state (the filled valence band). The first-order perturbation Hamiltonian, V' , induced by an external electric field $\mathbf{E}(\mathbf{r}, t)$, is

$$V' = \int \mathbf{d}(\mathbf{r}) \cdot \mathbf{E}(\mathbf{r}, t) d\mathbf{r} \quad (3.216)$$

where $\mathbf{d}(\mathbf{r})$ is the dipole-moment operator of the exciton

$$\mathbf{d}(\mathbf{r}) = -e\mathbf{r}_e \delta(\mathbf{r} - \mathbf{r}_e) + e\mathbf{r}_h \delta(\mathbf{r} - \mathbf{r}_h) \quad (3.217)$$

By assuming weak quantum dot excitation so that the ground-state occupation of the quantum dot is only lightly perturbed by the external radiation, we express the time-dependent wave function

$$|\mathbf{r}_e, \mathbf{r}_h, t\rangle = c_0(t) |\Psi_0(\mathbf{r}_e, \mathbf{r}_h)\rangle + \sum_{n\mathbf{K}} c_{n\mathbf{K}}(t) |\Psi_{n\mathbf{K}}(\mathbf{r}_e, \mathbf{r}_h, t)\rangle \quad (3.218)$$

where

$$\Psi_{n\mathbf{K}}(\mathbf{r}_e, \mathbf{r}_h, t) = \Psi_{n\mathbf{K}}(\mathbf{r}_e, \mathbf{r}_h) e^{-i\omega_{n\mathbf{K}} t}$$

By inserting Eq. (3.218) into the time-dependent Schrödinger equation of

$$i\hbar \frac{\partial |\mathbf{r}_e, \mathbf{r}_h, t\rangle}{\partial t} = (H_0 + V') |\mathbf{r}_e, \mathbf{r}_h, t\rangle \quad (3.219)$$

we obtain directly

$$i\hbar \frac{dc_{n\mathbf{K}}(t)}{dt} = c_0(t) \langle \Psi_{n\mathbf{K}}(\mathbf{r}_e, \mathbf{r}_h) | \int \mathbf{d}(\mathbf{r}) \cdot \mathbf{E}(\mathbf{r}, t) d\mathbf{r} | \Psi_0(\mathbf{r}_e, \mathbf{r}_h) \rangle e^{i\omega_{n\mathbf{K}}t} \quad (3.220)$$

in the first-order approximation. In obtaining the above equation we have utilized the following relationship due to the symmetry consideration

$$\langle \Psi_{n\mathbf{K}}(\mathbf{r}_e, \mathbf{r}_h) | \int \mathbf{d}(\mathbf{r}) \cdot \mathbf{E}(\mathbf{r}, t) d\mathbf{r} | \Psi_{n\mathbf{K}}(\mathbf{r}_e, \mathbf{r}_h) \rangle = 0 \quad (3.221)$$

We express the optical radiation as

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}(\mathbf{r}, \omega) e^{-i\omega t} + \text{c.c.} \quad (3.222)$$

“c.c.” in the above equation denotes “complex conjugate”. Notice that $\mathbf{E}(\mathbf{r}, t)$ is a physical quantity so that it is real. The first-order perturbation becomes

$$i\hbar \frac{dc_{n\mathbf{K}}(t)}{dt} = c_0(t) \langle \Psi_{n\mathbf{K}}(\mathbf{r}_e, \mathbf{r}_h) | \int \mathbf{d}(\mathbf{r}) \cdot \mathbf{E}(\mathbf{r}, \omega) d\mathbf{r} | \Psi_0(\mathbf{r}_e, \mathbf{r}_h) \rangle e^{i(\omega_{n\mathbf{K}} - \omega)t} \quad (3.223)$$

where we have neglected the term corresponding to the less-likely photon emission during the creation of the exciton.

Assuming weak excitation so that the occupation of the initial state is only perturbed, i.e., $c_0(t) = 1$, a straightforward solution to the above equation is

$$c_{n\mathbf{K}}(t) = \frac{e^{i(\omega_{n\mathbf{K}} - \omega)t}}{\hbar(\omega_{n\mathbf{K}} - \omega)} \langle \Psi_{n\mathbf{K}}(\mathbf{r}_e, \mathbf{r}_h) | \int \mathbf{d}(\mathbf{r}) \cdot \mathbf{E}(\mathbf{r}, \omega) d\mathbf{r} | \Psi_0(\mathbf{r}_e, \mathbf{r}_h) \rangle \quad (3.224)$$

This however has neglected the decaying of the occupation of the initial state due to the excitation of the excited states. By including the damping rate γ of the excitonic state, we have thus arrived at the steady-state solution

$$c_{n\mathbf{K}}(t) = \frac{e^{i(\omega_{n\mathbf{K}} - \omega)t}}{\hbar(\omega_{n\mathbf{K}} - \omega + i\gamma)} \langle \Psi_{n\mathbf{K}}(\mathbf{r}_e, \mathbf{r}_h) | \int \mathbf{d}(\mathbf{r}) \cdot \mathbf{E}(\mathbf{r}, \omega) d\mathbf{r} | \Psi_0(\mathbf{r}_e, \mathbf{r}_h) \rangle \quad (3.225)$$

The excitation of the excited exciton state, i.e., the appearance of the electron and hole, results in an electric dipole, and the dielectric polarization is given by

$$\mathbf{P}_{n\mathbf{K}}(\mathbf{r}, t) = \langle \mathbf{r}_e, \mathbf{r}_h, t | \mathbf{d}(\mathbf{r}) | \mathbf{r}_e, \mathbf{r}_h, t \rangle = \langle \Psi_0 | \mathbf{d}(\mathbf{r}) | \Psi_{n\mathbf{K}}(\mathbf{r}_e, \mathbf{r}_h, t) \rangle c_{n\mathbf{K}}(t) + \text{c.c.} \quad (3.226)$$

We first calculate the matrix element of the electron momentum \mathbf{p}_e

$$\begin{aligned}
& \langle \Psi_{nK}(\mathbf{r}_e, \mathbf{r}_h) | \mathbf{p}_e | \Psi_0 \rangle \\
&= m_0 \left\langle \Psi_{nK}(\mathbf{r}_e, \mathbf{r}_h) \left| \frac{\partial \mathbf{r}_e}{\partial t} \right| \Psi_0 \right\rangle \\
&= \frac{im_0}{\hbar} \langle \Psi_{nK}(\mathbf{r}_e, \mathbf{r}_h) | [H_0, \mathbf{r}_e] | \Psi_0 \rangle \\
&= im_0(\omega_{nK} - \omega_0) \langle \Psi_{nK}(\mathbf{r}_e, \mathbf{r}_h) | \mathbf{r}_e | \Psi_0 \rangle
\end{aligned} \tag{3.227}$$

Here

$$\begin{aligned}
[H_0, \mathbf{r}_e] &= H_0 \mathbf{r}_e - \mathbf{r}_e H_0 \\
H_0 \Psi_{nK}(\mathbf{r}_e, \mathbf{r}_h) &= \hbar \omega_{nK} \Psi_{nK}(\mathbf{r}_e, \mathbf{r}_h) \\
H_0 \Psi_0 &= \hbar \omega_0 \Psi_0 \\
\hbar \omega_0 &= 0
\end{aligned}$$

so that

$$\langle \Psi_{nK}(\mathbf{r}_e, \mathbf{r}_h, t) | \mathbf{p}_e | \Psi_0(\mathbf{r}_e, \mathbf{r}_h) \rangle = i\omega_{nK} m_0 \langle \Psi_{nK}(\mathbf{r}_e, \mathbf{r}_h, t) | \mathbf{r}_e | \Psi_0(\mathbf{r}_e, \mathbf{r}_h) \rangle$$

where m_0 is the free electron mass. The dipole moment thus becomes

$$\begin{aligned}
& \langle \Psi_{nK}(\mathbf{r}_e, \mathbf{r}_h, t) | \mathbf{r}_e \cdot \mathbf{E}(\mathbf{r}_e) | \Psi_0(\mathbf{r}_e, \mathbf{r}_h) \rangle \\
&= \frac{-i}{\omega_{nK} m_0} \langle \Psi_{nK}(\mathbf{r}_e, \mathbf{r}_h, t) | \mathbf{p}_e \cdot \mathbf{E}(\mathbf{r}_e) | \Psi_0(\mathbf{r}_e, \mathbf{r}_h) \rangle
\end{aligned} \tag{3.228}$$

By expressing the momentum operator as

$$\mathbf{p}_e = \sum_{\mathbf{k}_e, \mathbf{k}_h} \alpha_{\mathbf{k}_e}^+ \beta_{\mathbf{k}_h}^+ \langle \psi_{\mathbf{k}_e}(\mathbf{r}_e) | \mathbf{p}_e | \psi_{\mathbf{k}_h}(\mathbf{r}_h) \rangle \delta_{\mathbf{r}_e, \mathbf{r}_h} \tag{3.229}$$

and by Eq. (3.214), the matrix element of the momentum is

$$\begin{aligned}
& \langle \Psi_{nK}(\mathbf{r}_e, \mathbf{r}_h) | \mathbf{p}_e \cdot \mathbf{E}(\mathbf{r}_e) | \Psi_0(\mathbf{r}_e, \mathbf{r}_h) \rangle \\
&= \sum_{\mathbf{k}_e, \mathbf{k}_h} A_{nK}^*(\mathbf{k}_e, \mathbf{k}_h) \langle \psi_{\mathbf{k}_e}(\mathbf{r}_e) | \mathbf{p}_e | \psi_{\mathbf{k}_h}(\mathbf{r}_h) \rangle \cdot \mathbf{E}(\mathbf{r}) \delta_{\mathbf{r}_e, \mathbf{r}_h}
\end{aligned} \tag{3.230}$$

where

$$\langle \Psi_0 | \sum_{\mathbf{k}'_e, \mathbf{k}'_h} \alpha_{\mathbf{k}'_e} \beta_{\mathbf{k}'_h} \sum_{\mathbf{k}_e, \mathbf{k}_h} \alpha_{\mathbf{k}_e}^+ \beta_{\mathbf{k}_h}^+ | \Psi_0 \rangle = \sum_{\mathbf{k}_e, \mathbf{k}_h} \tag{3.231}$$

For the spatial variation of the excitation field is small, e.g., $\mathbf{E}(\mathbf{r}) = \mathbf{E}_q e^{i\mathbf{q} \cdot \mathbf{r}}$ with small \mathbf{q} , it is then easy to show that the matrix element becomes [33]

$$\langle \Psi_{nK}(\mathbf{r}_e, \mathbf{r}_h) | \int \mathbf{p}_e \cdot \mathbf{E}(\mathbf{r}_e) d\mathbf{r} | \Psi_0(\mathbf{r}_e, \mathbf{r}_h) \rangle = \int \psi_{nK}^*(\mathbf{r}, \mathbf{r}) \mathbf{p}_{cv} \cdot \mathbf{E}(\mathbf{r}) d\mathbf{r} \tag{3.232}$$

when $\mathbf{p}_{cv} = \langle u_c(\mathbf{r}) | \mathbf{p} | u_v(\mathbf{r}) \rangle$, referred as the dipole momentum between conduction and valence bands, is approximated as constant over the range of \mathbf{k} involved. Here $\psi_{n\mathbf{K}}(\mathbf{r}_e, \mathbf{r}_h)$ is the exciton envelope function determined by Eq. (3.138).

It is easy to calculate $\langle \Psi_0 | \mathbf{d}(\mathbf{r}) | \Psi_{n\mathbf{K}}(\mathbf{r}_e, \mathbf{r}_h) \rangle$ similarly and the result is

$$\langle \Psi_0 | \mathbf{d}(\mathbf{r}) | \Psi_{n\mathbf{K}}(\mathbf{r}_e, \mathbf{r}_h) \rangle = \frac{-e}{im_0\omega_{n\mathbf{K}}} \psi_{n\mathbf{K}}(\mathbf{r}, \mathbf{r}) \mathbf{p}_{cv} \quad (3.233)$$

Inserting Eqs. (3.227) and (3.233) into Eq. (3.226), we get

$$\mathbf{P}_{n\mathbf{K}}(\mathbf{r}, t) = \frac{e^2 \mathbf{p}_{cv}}{\hbar(\omega_{n\mathbf{K}} - \omega) m_0^2 \omega_{n\mathbf{K}}^2} \psi_{n\mathbf{K}}(\mathbf{r}, \mathbf{r}) \int \psi_{n\mathbf{K}}^*(\mathbf{r}', \mathbf{r}') \mathbf{p}_{cv} \cdot \mathbf{E}(\mathbf{r}', t) d\mathbf{r}' \quad (3.234)$$

Coupling the nonradiative damping rate $\Gamma (= \hbar\gamma)$ of the exciton state (see Chap. 2 and more discussions late in this chapter), it can be shown that

$$\mathbf{P}_{n\mathbf{K}}(\mathbf{r}, t) = \frac{e^2 \mathbf{p}_{cv}}{\hbar(\omega_{n\mathbf{K}} - \omega - i\gamma) m_0^2 \omega_{n\mathbf{K}}^2} \psi_{n\mathbf{K}}(\mathbf{r}, \mathbf{r}) \int \psi_{n\mathbf{K}}^*(\mathbf{r}', \mathbf{r}') \mathbf{p}_{cv} \cdot \mathbf{E}(\mathbf{r}', t) d\mathbf{r}' \quad (3.235)$$

We shall discuss the damping more closely shortly.

Define

$$T_{n\mathbf{K}}(\omega) = \frac{\epsilon_\infty \omega_{\text{LT}} \pi a_{\text{Br}}^3}{\omega_{n\mathbf{K}} - \omega - i\gamma} \quad (3.236)$$

where

$$\epsilon_\infty \omega_{\text{LT}} a_{\text{Br}}^3 = \frac{4e^2 p_{cv}^2}{\hbar \omega_{n\mathbf{K}}^2 m_0^2}$$

$p_{cv} = |\mathbf{p}_{cv}|$, ω_{LT} and a_{Br} are the exciton longitudinal-transverse splitting and Bohr radius in the corresponding bulk semiconductor, the exciton contribution to the dielectric polarization is

$$\mathbf{P}_{n\mathbf{K}}(\mathbf{r}, t) = T_{n\mathbf{K}}(\omega) \psi_{n\mathbf{K}}(\mathbf{r}, \mathbf{r}) \int \psi_{n\mathbf{K}}^*(\mathbf{r}', \mathbf{r}') \mathbf{E}(\mathbf{r}', t) d\mathbf{r}' \quad (3.237)$$

where we have assumed $\mathbf{p}_{cv}(\mathbf{p}_{cv} \cdot \mathbf{E}) = p_{cv}^2 \mathbf{E}$.

We now discuss briefly the significance of such a dielectric polarization in a spherical quantum dot with radius R . Let us focus on the first excited exciton state. Repeat Eq. (3.138) and also add the quantum confinement potentials for the electron and hole

$$\left[\frac{\hbar^2 k_e^2}{2m_c^*} + \frac{\hbar^2 k_h^2}{2m_v^*} - \frac{e^2}{4\pi\epsilon_\infty |\mathbf{r}_e - \mathbf{r}_h|} + V_{\text{QD}}(\mathbf{r}_e) + V_{\text{QD}}(\mathbf{r}_h) \right] \psi(\mathbf{r}_e, \mathbf{r}_h) = (E - E_g) \psi(\mathbf{r}_e, \mathbf{r}_h) \quad (3.238)$$

For the first excited exciton state, the electron and hole are confined largely in the same spatial region, i.e., $|\mathbf{r}_e - \mathbf{a}| < R$ and $|\mathbf{r}_h - \mathbf{a}| < R$, where \mathbf{a} is the center

of the quantum dot, see Fig. 3.6. We hence approximate $V_{\text{QD}}(\mathbf{r}_e) \approx V_{\text{QD}}(\mathbf{r}_h) = V_{\text{QD}}(\mathbf{r})$. Follow derivations of Eq. (3.139), $\psi_{n\mathbf{K}}(\mathbf{r}_e, \mathbf{r}_h) = \psi_{\mathbf{K}}(\mathbf{r})\psi_n(\mathbf{r}_e - \mathbf{r}_h)$, and the equations for $\psi_{\mathbf{K}}(\mathbf{r})$ and $\psi_n(\mathbf{r}_e - \mathbf{r}_h)$ are

$$\left[-\frac{\hbar^2 \nabla_{\mathbf{r}}^2}{2m^*} + V_{\text{QD}}(\mathbf{r}) \right] \psi_{\mathbf{K}}(\mathbf{r}) = E_{\mathbf{K}} \psi_{\mathbf{K}}(\mathbf{r}) \quad (3.239)$$

$$\left(-\frac{\hbar^2 \nabla_{\mathbf{r}_e - \mathbf{r}_h}^2}{2\mu} - \frac{e^2}{4\pi\epsilon_{\infty}|\mathbf{r}_e - \mathbf{r}_h|} \right) \psi_n(\mathbf{r}_e - \mathbf{r}_h) = E_n \psi_n(\mathbf{r}_e - \mathbf{r}_h) \quad (3.240)$$

respectively.

By Eqs. (3.172) and Table 3.4 in the previous section, the solution of Eq. (3.239) is in the form of

$$\psi_{\mathbf{K}}(\mathbf{r}) = \frac{1}{|\mathbf{r} - \mathbf{a}|\sqrt{2\pi R}} \sin\left(\frac{\pi|\mathbf{r} - \mathbf{a}|}{R}\right) \quad (3.241)$$

The relative motion of the electron-hole pair of Eq. (3.240) was studied before, which is in the form of

$$\psi_n(\mathbf{r}_e - \mathbf{r}_h) = \frac{1}{\sqrt{\pi a_{\text{Br}}^3}} e^{-\frac{|\mathbf{r}_e - \mathbf{r}_h|}{a_{\text{Br}}}} \quad (3.242)$$

see Eq. (3.142). Note that the two wave functions in the above equations are normalized. Thus, the first-excited exciton state in a spherical quantum dot with radius R is described as

$$\psi_{n\mathbf{K}}(\mathbf{r}_e, \mathbf{r}_h) = \frac{1}{|\mathbf{r} - \mathbf{a}|\sqrt{2\pi R}} \sin\left(\frac{\pi|\mathbf{r} - \mathbf{a}|}{R}\right) \frac{1}{\sqrt{\pi a_{\text{Br}}^3}} e^{-\frac{|\mathbf{r}_e - \mathbf{r}_h|}{a_{\text{Br}}}} \quad (3.243)$$

which is confined and real. Note that the above expression is valid when $R \approx a_{\text{Br}}$ so that one may neglect the internal motion of the electron-hole pair in the quantum dot. And

$$\mathbf{r} = \frac{m_e \mathbf{r}_e + m_h \mathbf{r}_h}{m_e + m_h} \quad (3.244)$$

is the exciton center of mass.

For $\mathbf{r}_e = \mathbf{r}_h = \mathbf{r}$,

$$\psi_{n\mathbf{K}}(\mathbf{r}, \mathbf{r}) = \frac{1}{|\mathbf{r} - \mathbf{a}|\sqrt{2\pi R}} \sin\left(\frac{\pi|\mathbf{r} - \mathbf{a}|}{R}\right) \frac{1}{\sqrt{\pi a_{\text{Br}}^3}} \quad (3.245)$$

Assuming small spatial variation of the excitation field within the quantum dot, i.e., $\mathbf{E}(\mathbf{r}) \approx \mathbf{E}(\mathbf{a})$,

$$\int \psi_{n\mathbf{K}}^*(\mathbf{r}', \mathbf{r}') \mathbf{E}(\mathbf{r}') d\mathbf{r}' = \mathbf{E}(\mathbf{a}) \int \psi_{n\mathbf{K}}^*(\mathbf{r}', \mathbf{r}') d\mathbf{r}' = \frac{4R^2}{\pi \sqrt{2R a_{\text{Br}}^3}} \mathbf{E}(\mathbf{a}) \quad (3.246)$$

In other words, it is assumed that the electric field inside the quantum dot is uniform since the wavelength of interest is about 1 μm (visible optical range) whereas the quantum dot diameter is about 20 nm or less.

Let $\alpha_{n\mathbf{K}} = \pi|\mathbf{r} - \mathbf{a}|/R$ and by Eqs. (3.236) and (3.246)

$$\mathbf{P}_{n\mathbf{K}}(\mathbf{r}) = \frac{2\epsilon_{\infty}\omega_{\text{LT}}}{\omega_{n\mathbf{K}} - \omega - i\gamma} \frac{\sin \alpha_{n\mathbf{K}}}{\alpha_{n\mathbf{K}}} \mathbf{E}(\mathbf{a})$$

By approximating $\mathbf{E}(\mathbf{a}) \approx \mathbf{E}(\mathbf{r})$, we obtain the dielectric polarization of the exciton due to its optical generation by the external optical field $\mathbf{E}(\mathbf{r})$

$$\mathbf{P}_{n\mathbf{K}}(\mathbf{r}) = \frac{2\epsilon_{\infty}\omega_{\text{LT}}}{\omega_{n\mathbf{K}} - \omega - i\gamma} \frac{\sin \alpha_{n\mathbf{K}}}{\alpha_{n\mathbf{K}}} \mathbf{E}(\mathbf{r}) \quad (3.247)$$

Due to such an exciton polarization, the nonlocal material equation relating the electric field $\mathbf{E}(\mathbf{r})$ and the displacement vector $\mathbf{D}(\mathbf{r})$ is

$$\mathbf{D}(\mathbf{r}) = \epsilon_{\infty}\mathbf{E}(\mathbf{r}) + \mathbf{P}(\mathbf{r}) = \epsilon_{\text{QD}}(\mathbf{r})\mathbf{E}(\mathbf{r}) \quad (3.248)$$

where an effective permittivity can be defined for the quantum-dot exciton polariton

$$\epsilon_{\text{QD}}(\mathbf{r}) = \epsilon_{\infty} \left(1 + \frac{2\omega_{\text{LT}}}{\omega_{n\mathbf{K}} - \omega + i\gamma} \frac{\sin \alpha_{n\mathbf{K}}}{\alpha_{n\mathbf{K}}} \right) \quad (3.249)$$

By averaging $\epsilon_{\text{QD}}(\mathbf{r}, \omega)$ over the quantum dot, we obtain the expression for the effective permittivity of the quantum dot due to the ground-state exciton excited by an external electromagnetic field

$$\epsilon_{\text{QD}} = \frac{1}{\Omega} \int \epsilon_{\text{QD}}(\mathbf{r}) d\mathbf{r} = \epsilon_{\infty} \left[1 + \frac{6\omega_{\text{LT}}}{\pi^2(\omega_{n\mathbf{K}} - \omega + i\gamma)} \right] \quad (3.250)$$

where $\Omega = 4\pi R^3/3$ is the volume of the quantum dot with a radius of R .

We have thus far assumed that the quantum dot stays largely at its ground state and the probability of optical excitation is very small. Let us exam this assumption closely in this section. Due to the external radiation, the time-dependent wave function

$$|\mathbf{r}_e, \mathbf{r}_h, t\rangle = c_0(t)|\Psi_0(\mathbf{r}_e, \mathbf{r}_h)\rangle + \sum_n c_n(t)|\Psi_n(\mathbf{r}_e, \mathbf{r}_h, t)\rangle \quad (3.251)$$

from which

$$\begin{aligned} i\hbar \frac{dc_n(t)}{dt} &= c_0(t) \langle \Psi_n(\mathbf{r}_e, \mathbf{r}_h) | V'(t) | \Psi_0(\mathbf{r}_e, \mathbf{r}_h) \rangle e^{i\omega_n t} \\ i\hbar \frac{dc_0(t)}{dt} &= \sum_{n(\neq 0)} c_n(t) \langle \Psi_0(\mathbf{r}_e, \mathbf{r}_h) | V'(t) | \Psi_n(\mathbf{r}_e, \mathbf{r}_h) \rangle e^{-i\omega_n t} \end{aligned} \quad (3.252)$$

where $V'(t)$ is the perturbation due to the optical field. Here we have ignored the exciton wave vector \mathbf{K} since it is zero for exciton states confined in a quantum dot. Note that optical couplings among exciton states are zero because of the selection rules. Furthermore, only a limited number of exciton states have nonzero optical couplings with the ground state.

Let $V'(t) = V'e^{i\omega t}$. We consider a two-state system

$$\begin{aligned} i\hbar \frac{dc_0(t)}{dt} &= c_1(t)e^{-i(\omega_1-\omega)t} \langle 0|V'|1\rangle \equiv \hbar\gamma c_1(t)e^{-i(\omega_1-\omega)t} \\ i\hbar \frac{dc_1(t)}{dt} &= \hbar\gamma c_0(t)e^{i(\omega_1-\omega)t} \end{aligned} \quad (3.253)$$

where $\langle 0|V'|1\rangle = \hbar\gamma/2$ simplifies the optical transition matrix, $\hbar\omega_1$ is the eigenvalue of $|1\rangle$ and the eigenvalue of $|0\rangle$ is set to be zero. In terms of the detuning parameter $\delta = \omega - \omega_1$, the above equations read

$$\begin{aligned} i\hbar \frac{dc_0(t)}{dt} &= \frac{\hbar\gamma}{2} c_1(t)e^{i\delta t} \\ i\hbar \frac{dc_1(t)}{dt} &= \frac{\hbar\gamma}{2} c_0(t)e^{-i\delta t} \end{aligned} \quad (3.254)$$

Differentiating Eqs. (3.254) with respect to time yields

$$\frac{d^2c_0}{dt^2} = \frac{\gamma\delta}{2} c_1 e^{i\delta t} - \frac{i\gamma}{2} \frac{dc_1}{dt} e^{i\delta t} = i\delta \frac{dc_0}{dt} - \frac{\gamma^2}{4} c_0 \quad (3.255)$$

i.e.,

$$\frac{d^2c_0}{dt^2} = i\delta \frac{dc_0}{dt} - \frac{\gamma^2}{4} c_0 \quad (3.256)$$

A trial solution of the form

$$c_0 = e^{-i\chi t}$$

leads to

$$\chi^2 + \delta\chi - \gamma^2 = 0, \quad \chi = \chi_{\pm} = -\frac{\delta}{2} \pm \frac{1}{2}\sqrt{\delta^2 + \gamma^2} \quad (3.257)$$

Thus a general solution is

$$c_0 = e^{i\delta t/2} \left[A \cos\left(\frac{t}{2}\sqrt{\delta^2 + \gamma^2}\right) + B \sin\left(\frac{t}{2}\sqrt{\delta^2 + \gamma^2}\right) \right] \quad (3.258)$$

By the first equation of Eqs. (3.254),

$$\begin{aligned}
c_1 &= \frac{2i}{\gamma} e^{-i\delta t} \frac{dc_0}{dt} \\
&= -e^{-i\delta t/2} \left\{ \frac{\delta}{\gamma} \left[A \cos \left(\frac{t}{2} \sqrt{\delta^2 + \gamma^2} \right) + B \sin \left(\frac{t}{2} \sqrt{\delta^2 + \gamma^2} \right) \right] \right. \\
&\quad \left. + i \frac{\sqrt{\delta^2 + \gamma^2}}{\gamma} \left[A \sin \left(\frac{t}{2} \sqrt{\delta^2 + \gamma^2} \right) - B \cos \left(\frac{t}{2} \sqrt{\delta^2 + \gamma^2} \right) \right] \right\}
\end{aligned} \tag{3.259}$$

If the system is in the $|0\rangle$ state at time $t = 0$, i.e., $c_0(0) = 1$ and $c_1(0) = 0$, $A = 1$ and $B = -i\delta/\sqrt{\delta^2 + \gamma^2}$, so that at a later time

$$\begin{aligned}
c_0(t) &= e^{i\delta t/2} \left[\cos \left(\frac{t}{2} \sqrt{\delta^2 + \gamma^2} \right) - \frac{i\delta}{\sqrt{\delta^2 + \gamma^2}} \sin \left(\frac{t}{2} \sqrt{\delta^2 + \gamma^2} \right) \right] \\
c_1(t) &= e^{-i\delta t/2} \sin \left(\frac{t}{2} \sqrt{\delta^2 + \gamma^2} \right) \frac{2i\gamma}{\sqrt{\delta^2 + \gamma^2}}
\end{aligned} \tag{3.260}$$

We can then discuss the probability of an electron transiting from its initial state $|0\rangle$ to $|1\rangle$ under the influence of $V'(t)$. Expressing the total wave function as $c_0(t)|0\rangle + c_1(t)|1\rangle$, the probability of electron occupying state $|1\rangle$ at time t is simply

$$|c_1|^2 = e^{-i\delta t} \left[1 - \cos \left(t \sqrt{\delta^2 + \gamma^2} \right) \right] \frac{\gamma^2}{\delta^2 + \gamma^2} \tag{3.261}$$

We now calculate the matrix element of momentum \mathbf{p} (can be any physical parameter)

$$\langle (1|c_1^*(t) + \langle 0|c_0^*(t)) \mathbf{p} (c_0(t)|0\rangle + c_1(t)|1\rangle) = c_0(t)c_1^*(t) \langle 1|\mathbf{p}|0\rangle + c_0^*(t)c_1(t) \langle 0|\mathbf{p}|1\rangle$$

Note that $\langle 0|\mathbf{p}|0\rangle = \langle 1|\mathbf{p}|1\rangle = 0$. The above equation shows that the physical variable is to be evaluated by $c_0(t)c_1^*(t)$, which is

$$c_0(t)c_1^*(t) = e^{i\delta t} \left[\left(\sqrt{\delta^2 + \gamma^2} - i\delta \right) \cos \left(t \sqrt{\delta^2 + \gamma^2} \right) - i\delta \right] \frac{-i\gamma}{\delta^2 + \gamma^2} \tag{3.262}$$

by Eqs. (3.260). Including the term $e^{i\omega_1 t}$ in $\langle 1|\mathbf{p}|0\rangle$ (remember that the total wave function of an eigen solution contains such a term) and making a time average

$$\langle c_0(t)c_1^*(t) \langle 1|\mathbf{p}|0\rangle \rangle_t = |\langle 1|\mathbf{p}|0\rangle| \frac{(\omega_1 - \omega)\gamma}{(\omega_1 - \omega)^2 + \gamma^2} \tag{3.263}$$

$\langle \rangle_t$ denotes the time average. Note that $\hbar\gamma/2 = \langle 0|V|1\rangle$ is the transition matrix element. It is easy to see that the above equation is equivalent to the real part of the

dielectric polarization in Eq. (3.235): Repeating Eq. (3.235) here

$$\mathbf{P}_{n\mathbf{K}}(\mathbf{r}, t) = \frac{e^2 \mathbf{p}_{cv}}{\hbar(\omega_{n\mathbf{K}} - \omega - i\gamma)m_0^2 \omega_{n\mathbf{K}}^2} \psi_{n\mathbf{K}}(\mathbf{r}, \mathbf{r}) \int \psi_{n\mathbf{K}}^*(\mathbf{r}', \mathbf{r}') \mathbf{p}_{cv} \cdot \mathbf{E}(\mathbf{r}', t) d\mathbf{r}'$$

where apart from physical constants

$$\frac{1}{\omega_{n\mathbf{K}} - \omega - i\gamma} = \frac{\omega_{n\mathbf{K}} - \omega}{(\omega_{n\mathbf{K}} - \omega)^2 + \gamma^2} + i \frac{\gamma}{(\omega_{n\mathbf{K}} - \omega)^2 + \gamma^2}$$

Furthermore,

$$\gamma \propto \langle 0|V'|1 \rangle \propto \int \psi_{n\mathbf{K}}^*(\mathbf{r}', \mathbf{r}') \mathbf{p}_{cv} \cdot \mathbf{E}(\mathbf{r}', t) d\mathbf{r}'$$

Here we see the damping rate of the initial state due to the excitation of the excited states. A more advanced treatment of multiple states (we considered only two states in the above derivations) can be done by the generalized Golden rule, which will result in a damping rate due to the interactions among all the states, see Chap. 2.

We now calculate the energy dissipation. First we consider Poynting's theorem describing conservation of energy. Repeat Eq. (3.26)

$$\oint \mathbf{E} \times \mathbf{H} \cdot d\mathbf{S} = - \int \left[\mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} + \mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t} \right] d\Omega \quad (3.264)$$

where S denotes the surface and Ω the volume. Denote $\epsilon = \epsilon' + i\epsilon''$ and $\mu = \mu' + i\mu''$. For monochromatic light with a frequency ω , an amplitude E_0 for its \mathbf{E} field and H_0 for \mathbf{H} field, the time average of Poynting's theorem in Eq. (3.26) gives

$$\oint \langle \mathbf{E} \times \mathbf{H} \rangle_t \cdot d\mathbf{S} = - \int \left[\omega \epsilon'' \frac{E_0^2}{2} + \omega \mu'' \frac{H_0^2}{2} \right] d\Omega \quad (3.265)$$

Loss requirements are satisfied through the signs of ϵ'' and μ'' . Assuming $\mu'' = 0$ and a small volume $\Delta\Omega$, the power dissipated is thus

$$- \oint \langle \mathbf{E} \times \mathbf{H} \rangle_t \cdot d\mathbf{S} = \frac{E_0^2}{2} \omega \epsilon'' \Delta\Omega \quad (3.266)$$

On the other hand, consider a thin slab of material with thickness Δz and normal surface area S , and a plane wave traveling in the z direction with complex wave number $k = k' + ik''$. The time average Poynting vector, assuming a single traveling wave, is

$$\mathbf{S} = z_0 \frac{E_0^2}{2} \Re \left\{ \frac{1}{\eta^*} e^{-2k''z} \right\} \quad (3.267)$$

where $\eta = \sqrt{\mu/\epsilon}$ is the complex wave impedance and z_0 is the unit vector of the z direction. $\Re A$ denotes the real part of a complex variable A . The closed surface integral of the Poynting vector, i.e., the power dissipated in $\Delta\Omega = S\Delta z$, is thus

$$P_d = \frac{SE_0^2}{2} \Re \left\{ \frac{1}{\eta^*} (1 - e^{-2k'' \Delta z}) \right\} \quad (3.268)$$

Writing $\eta = \sqrt{\mu_0/\epsilon}$ and assuming Δz very small so that $e^{-2k'' \Delta z} \ll 1$,

$$P_d = \frac{SE_0^2}{2\sqrt{\mu_0}} \Re \{ 2k'' \Delta z (\sqrt{\epsilon})^* \} \quad (3.269)$$

Since $k = \omega\sqrt{\epsilon}/c$, where c is the speed of light,

$$\Re \{ k'' (\sqrt{\epsilon})^* \} = \frac{\omega \epsilon''}{2c} \quad (3.270)$$

so that

$$P_d = \frac{SE_0^2}{2} \omega \epsilon'' \Delta z = \frac{E_0^2}{2} \omega \epsilon'' \Delta \Omega \quad (3.271)$$

as Eq. (3.29). Therefore, the energy and Poynting vector descriptions for dissipation are in agreement, as they must be from Eq. (3.26).

We now calculate the optical loss of the external electromagnetic field due to the quantum-dot exciton excitation by writing the continuity equation for the photon density

$$\frac{dn_{\hbar\omega}(z)}{dt} = \frac{\partial n_{\hbar\omega}(z)}{\partial t} - \frac{\partial [cn_{\hbar\omega}(z)]}{\partial z} \quad (3.272)$$

Here we consider a beam of photons traveling along the z axis. The first term on the right side of the above equation represents the absorption rate of photons and the second term the photons leaving due to the photon flux. c is the speed of light in the medium. We have used the above equation in Sect. 3.3 when discussing the optical spectrum of the materials. At steady state we have, in general

$$n_{\hbar\omega}(z) = n_{\hbar\omega}(0) e^{-\epsilon''(\omega)z} \quad (3.273)$$

which defines the optical loss $\epsilon''(\omega)$ as in Eqs. (3.29) and (3.271). Since

$$\frac{\partial n_{\hbar\omega}(z)}{\partial t} = p_{n\mathbf{K} \leftarrow 0} (1 - f_{n\mathbf{K}}) f_0 - p_{0 \leftarrow n\mathbf{K}} (1 - f_0) f_{n\mathbf{K}} \quad (3.274)$$

where f_0 and $f_{n\mathbf{K}}$ are the occupations of state Ψ_0 and $\Psi_{n\mathbf{K}}$, respectively. $f_0 = |c_0|^2$ and $f_{n\mathbf{K}} = |c_{n\mathbf{K}}(\infty)|^2$. $p_{n\mathbf{K} \leftarrow 0}$ is the transition probability per unit time

$$\begin{aligned} p_{n\mathbf{K} \leftarrow 0} &= |c_{n\mathbf{K}}(\infty)|^2 \gamma \\ &= \frac{e^2 \gamma}{\hbar^2 [(\omega_{n\mathbf{K}} - \omega)^2 + \gamma^2] \omega_{n\mathbf{K}}^2 m_0^2} \left[\int \psi_{n\mathbf{K}}(\mathbf{r}', \mathbf{r}') \mathbf{p}_{cv} \cdot \mathbf{E}(\mathbf{r}') d\mathbf{r}' \right]^2 \end{aligned} \quad (3.275)$$

At steady state,

$$\frac{dn_{\hbar\omega}(z)}{dt} = 0$$

so that

$$\epsilon'' = \frac{e^2\gamma(f_0 - f_{n\mathbf{K}})}{\hbar^2[(\omega_{n\mathbf{K}} - \omega)^2 + \gamma^2]cn_{\hbar\omega}\omega_{n\mathbf{K}}^2m_0^2} \left[\int \psi_{n\mathbf{K}}(\mathbf{r}', \mathbf{r}') \mathbf{p}_{cv} \cdot \mathbf{E}(\mathbf{r}') d\mathbf{r}' \right]^2 \quad (3.276)$$

We assume again that the electric field inside the quantum dot is uniform and note that the energy density of the electromagnetic field equals $\epsilon_\infty E^2 = \hbar\omega_{\hbar\omega}/\Omega$, where $\Omega = 4\pi R^3/3$ is the volume of the quantum dot,

$$\epsilon'' = \frac{e^2 p_{cv}^2 \gamma \omega (f_0 - f_{n\mathbf{K}})}{\Omega \hbar [(\omega_{n\mathbf{K}} - \omega)^2 + \gamma^2] c \epsilon_\infty \omega_{n\mathbf{K}}^2 m_0^2} \left[\int \psi_{n\mathbf{K}}(\mathbf{r}', \mathbf{r}') d\mathbf{r}' \right]^2 \quad (3.277)$$

By inserting Eq. (3.245) into the above equation, we finally have

$$\epsilon'' = \frac{6\gamma\omega\omega_{\text{LT}}(f_0 - f_{n\mathbf{K}})}{\pi^2 c [(\omega_{n\mathbf{K}} - \omega)^2 + \gamma^2]} = \frac{6\gamma\omega\omega_{\text{LT}}}{\pi^2 c [(\omega_{n\mathbf{K}} - \omega)^2 + \gamma^2]} \quad (3.278)$$

so that by Eq. (3.33)

$$\epsilon''(\omega) = \frac{6\gamma\omega_{\text{LT}}}{\pi^2 [(\omega_{n\mathbf{K}} - \omega)^2 + \gamma^2]} \quad (3.279)$$

per quantum dot. The last equality in Eq. (3.278) is obtained since $f_{n\mathbf{K}} \ll 1$ and $f_0 \approx 1$. Note that Eq. (3.279) is exactly the imaginary part of the complex effective permittivity of Eq. (3.250).

Without any extra external modifications expect the incident electromagnetic field of $\mathbf{E}(\mathbf{r}, t)$, $f_{n\mathbf{K}}(t) = |c_{n\mathbf{K}}(t)|^2$, which is small (the first-order perturbation condition). High pumping rate can be achieved by increasing the optical power which, however, is not the case for common optoelectronic applications. The occupation of the excited state is normally very smaller than the ground state so that $f_0 > f_{n\mathbf{K}}$ and optical absorption dominates. This means that the optical field $\mathbf{E}(\mathbf{r}, t)$ is absorbed during the exciton excitation following its propagation through the quantum dot. The situation is normally referred to as optical loss, for examples, in photodetectors. Optical gain is achieved in light emitting and laser devices by population inversion of $f_0 < f_{n\mathbf{K}}$.

When we consider the extra pumping process that the exciton population $f_{n\mathbf{K}}$ is finite so the time-dependent wave function becomes

$$|\mathbf{r}_e, \mathbf{r}_h, t\rangle = [c_0^{(0)} + c_0(t)] |\Psi_0(\mathbf{r}_e, \mathbf{r}_h)\rangle + \sum_{n\mathbf{K}} [c_{n\mathbf{K}}^{(0)} + c_{n\mathbf{K}}(t)] |\Psi_{n\mathbf{K}}(\mathbf{r}_e, \mathbf{r}_h, t)\rangle \quad (3.280)$$

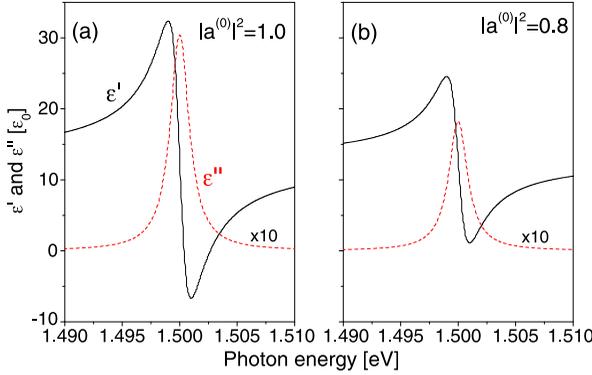


Fig. 3.7 Dielectric coefficients of a quantum dot before and after it becomes excited. $\epsilon_\infty = 12.8$, $\hbar\omega_{nK} = 1.5$ eV, $\hbar\omega_{LT} = 5$ meV, and $\hbar\gamma = 1$ meV for typical II–VI semiconductor quantum dots, e.g., colloidal PbSe/ZnSe quantum dots (Reprinted with permission from Y. Fu, L. Thylen, and H. Ågren, A lossless negative dielectric constant from quantum dot exciton polaritons, *Nano Lett.*, vol. 8, pp. 1551–1555, 2008. Copyright 2008, American Chemical Society)

due to the external electromagnetic field $\mathbf{E}(\mathbf{r}, t)$. By following the same mathematical procedures as what we did before, we can obtain

$$\epsilon_{\text{QD}}(\omega) = \epsilon_\infty \left[1 + \sum_{nK} \frac{6(2|c_0^{(0)}|^2 - 1)\omega_{LT}}{\pi^2(\omega_{nK} - \omega + i\gamma)} \right] \quad (3.281)$$

Here we observe that the quantum dot can become lossless if $|c_0^{(0)}|^2 = 1/2$. At the same time, the polarization changes sign. Figure 3.7 shows the dielectric coefficients before and after the quantum dot becomes populated.

Equation (3.281) and Fig. 3.7 show that resonant exciton states in quantum dots provide a source for both the high and negative dielectric constant.

3.8 Multiphoton Process

One important aspect concerning the optical properties of quantum dots is the multiphoton process, which has been widely applied in recent years in biological and medical imaging after the pioneer work of Goepfert-Mayer [2–5]. Early theoretical works normally try to solve perturbative and time-independent Schrödinger equation, see the end of Sect. 3.2, while today’s experimental setups of ultrafast (femtosecond = fs = 10^{-15} s, even attosecond = as = 10^{-18} s) and ultraintense (in many experiments the optical power of the laser pulse peak can reach GW/cm^2) are to be understood only by trying to solve the nonperturbative time-dependent Schrödinger equation.

We now study the dynamical optical transitions in a spherical quantum dot. By Sect. 3.2 and Eq. (3.54), the total Hamiltonian of the electron and the electromagnetic field is

$$H(\mathbf{r}, t) = H_0(\mathbf{r}) + \frac{e}{m_0} \mathbf{A} \cdot \mathbf{p} + \sum_s \hbar\omega_s \left(b_s^\dagger b_s + \frac{1}{2} \right) \quad (3.282)$$

where H_0 is the Hamiltonian of the electron in the quantum dot. b_s and b_s^\dagger are annihilation and creation operators of the photon of state s which carries a momentum $\hbar s$ and an energy $\hbar\omega_s$. As before, we denote

$$H_0(\mathbf{r})\Psi_k(\mathbf{r}) = E_k\Psi_k(\mathbf{r}) \quad (3.283)$$

to characterize the eigen state of the electron in the quantum dot. After the switch on of an external electromagnetic field in Eq. (3.282), the total wave function of the electron-photon system becomes

$$\Psi(\mathbf{r}, t) = \sum_k C_k(t)\Psi_k(\mathbf{r})|N_k\rangle \exp[-i(E_k + N_k\hbar\omega_s)t/\hbar] \quad (3.284)$$

where $|N_k\rangle$ describes the photon field with N_k photons associated with electron state k . Here we focus on monochromatic light with photon energy $\hbar\omega_s$ and photon momentum $\hbar s$. Let us now solve the time-dependent Schrödinger equation

$$i\hbar \frac{\partial \Psi(\mathbf{r}, t)}{\partial t} = H(\mathbf{r}, t)\Psi(\mathbf{r}, t) \quad (3.285)$$

Following the generalized Golden rule presented in Chap. 2 we now consider the optical transition induced by the external excitation radiation. We insert the total wave function expression Eq. (3.284) into the above Schrödinger equation, multiple both sides of the equation by $\langle \Psi_q N_q | \exp[i(E_q + N_q\hbar\omega_s)t/\hbar]$ then integrate over the electron and photon coordinations

$$\begin{aligned} i\hbar \frac{dC_q(t)}{dt} &= \frac{e}{m_0} \sqrt{\frac{\hbar}{2\omega_s \epsilon \Omega}} \sum_k \langle \Psi_q | \mathbf{a} \cdot \mathbf{p} | \Psi_k \rangle C_k(t) \\ &\times \left\{ \sqrt{N_k} \exp\left[\frac{i(E_q - E_k - \hbar\omega_s)t}{\hbar} \right] \delta_{N_q, N_k-1} \right. \\ &\left. + \sqrt{N_k + 1} \exp\left[\frac{i(E_q - E_k + \hbar\omega_s)t}{\hbar} \right] \delta_{N_q, N_k+1} \right\} \quad (3.286) \end{aligned}$$

Here we have set $e^{i\mathbf{s} \cdot \mathbf{r}} = 1$ in Eq. (3.38) because of the small wave number of the photon compared with the electron momentum.

The above equation shows again, see Eq. (3.69), that an electron originally occupying electron state k will transit to electron state q ($k \neq q$) after absorbing (emitting) one photon $\hbar\omega_s$, thus the number of photons in the photon field is reduced

(increased) by 1. Equation (3.286) itself describes absorption and stimulated emission processes. For spontaneous radiative transition when the electron transits from state \mathbf{k} to \mathbf{q} , we can use a similar equation with a photon resonance energy that equals $E_{\mathbf{k}} - E_{\mathbf{q}}$. Since the spontaneously emitted photon does not accumulate in the quantum dot, the number of such photon inside the quantum dot is zero before the transition. It becomes one directly after the transition. $\langle \mathbf{q} | \mathbf{a} \cdot \mathbf{p} | \mathbf{k} \rangle$ in Eq. (3.286) was already calculated numerically earlier in the previous section.

As we have discussed in Sect. 2.2, Eq. (3.286) is perfect within the time interval $[t, t + dt]$ such that the eventual change in $C_{\mathbf{k}}(t)$ is not significant so that $C_{\mathbf{k}}(t)$ can be approximated as constant on the right side of the equation.

We now use Eq. (3.286) to design the following Monte Carlo scheme to study the temporal development of electron distribution in the quantum-dot states under the influence of an photon field. Before switching on the excitation photon field, all valence-band (VB) states are occupied and the conduction-band (CB) ones are unoccupied. After the switch on of the photon field, electrons originally occupying VB states will be photoexcited to CB states, leaving holes in the VB band. During the time, the occupation of each electron state is either 1 or 0, while Eq. (3.286) shows the transition probability.

For nonradiative energy relaxations, we focus on the predominant electron-phonon scattering in the following form

$$i \frac{dC_{\mathbf{q}}(t)}{dt} = \sqrt{\frac{\pi(D_t K)^2}{\rho \Omega \omega_p}} \sum_{\mathbf{k}} C_{\mathbf{k}}(t) \left\{ \sqrt{M_{\mathbf{k}}} \exp \left[\frac{i(E_{\mathbf{q}} - E_{\mathbf{k}} - \hbar \omega_p)t}{\hbar} \right] \delta_{M_{\mathbf{q}}, M_{\mathbf{k}}-1} \right. \\ \left. + \sqrt{M_{\mathbf{k}} + 1} \exp \left[\frac{i(E_{\mathbf{q}} - E_{\mathbf{k}} + \hbar \omega_p)t}{\hbar} \right] \delta_{M_{\mathbf{q}}, M_{\mathbf{k}}+1} \right\} \quad (3.287)$$

where $M_{\mathbf{k}}$ is the number of phonons associated with electron state \mathbf{k} . $\hbar \omega_p$ is the phonon energy. Note that the above equation is consistent with the common expression about the scattering rate for an electron to relax from state \mathbf{k} to \mathbf{q} by emitting a phonon $\hbar \omega_p$ [34–36]

$$P_{\mathbf{kq}} = \frac{\pi(D_t K)^2}{\rho \Omega \omega_p} (M_{\mathbf{q}} + 1) \frac{\Gamma}{(E_{\mathbf{k}} - E_{\mathbf{q}} + \hbar \omega_p)^2 + \Gamma^2} \quad (3.288)$$

where $D_t K$ is the optical deformation potential and $\rho = 5.8 \text{ g/cm}^3$ is the mass density of the CdSe material. $\hbar \omega_p = 37 \text{ meV}$ is the optical phonon energy of CdS at which the phonon density of states is maximal [37, 38]. Γ , the relaxation energy, is obtained similarly by Eq. (2.85), which is about 2.04 meV by adopting $D_t K = 10^8 \text{ eV/cm}$ for most commonly used semiconductors [35] which gives a scattering rate of $3.09 \times 10^{12} \text{ s}^{-1}$ at phonon emission resonance by including the phonon density of states. Such a rate is typical in common semiconductor bulk materials [39] as well as in quantum-dot-based light-emitting devices [40, 41].

In the Monte Carlo scheme, the time interval dt is chosen in such a way that maximally only one transition event can occur. Thus, the time duration of each event

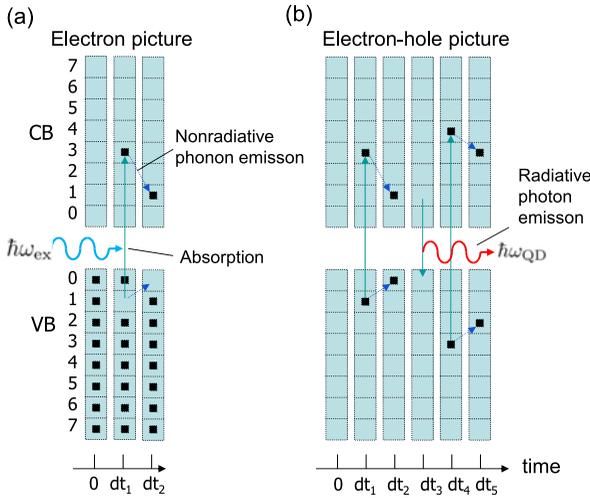


Fig. 3.8 Schematic diagram of transition process. (a) Electron picture. *Black dots* are electrons. (b) Electron-hole picture. *Black dots* in the conduction band (CB) are electrons. They are holes in the valence band (VB). $\hbar\omega_{ex}$ denotes the excitation photon field, $\hbar\omega_{QD}$ denotes the quantum-dot emission field (Reprinted with permission from T.-T. Han, Y. Fu, and H. Ågren, Dynamic photon emission from multiphoton-excited semiconductor quantum dot, J. Appl. Phys., vol. 103, p. 93703(6), 2008. Copyright 2008, American Institute of Physics)

varies, which is denoted by adding a subscript to dt_i . For the whole system containing n electrons and N electron states, there are $n \times (N - n)$ probable final states for a one-electron-transition process. After calculating the probability for every possible transition within dt , a random number is generated in the Monte Carlo scheme to determine the occurrence of transition events.

We switch on the excitation photon field at $t = 0$. After $t = dt_1$, one electron originally occupying one VB state, may be excited to a CB state, via interband absorption, and the photon field is reduced by one photon. After $t = dt_1 + dt_2$, the CB electron may recombine with the VB hole; the CB electron may absorb a new photon then transit to another CB state via intraband transition; It may also emit a phonon . . . Schematic temporal development of the sublevel occupation in the quantum dot is shown in Fig. 3.8. Note that all the sublevels are identified by an index number according to their relative energies. For example, the ground sublevel is tagged as number “0”, and the first-excited sublevel as “1”. The occupied sublevels in the CB and the unoccupied sublevels in the VB (i.e., holes) are marked by black dots. As schematically shown in Fig. 3.8, an electron originally staying at VB state 1 at time 0 is excited to CB state 3 at time interval 1 by absorbing one photon. Transitions in time interval 2 are nonradiative phonon scattering processes from CB3 to CB1 and from VB1 to VB0.

We now consider a single CdS quantum dot with a radius of 3.7 nm in vacuum, where the quantum-dot bandgap (energy separation between the ground states in the CB and VB) is 2.655 eV. Figure 3.9 shows the dynamic occupations of the CB

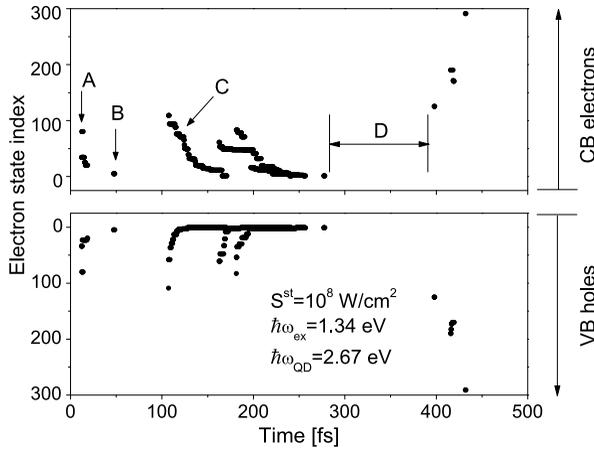


Fig. 3.9 Temporal development of occupations of CB and VB states in a CdS quantum dot with a radius of 3.7 nm (its energy bandgap is $\hbar\omega_{\text{QD}} = 2.67$ eV). The excitation power $S^{\text{st}} = 10^8$ W/cm² and excitation photon energy $\hbar\omega_{\text{ex}} = 1.34$ eV, corresponds to a two-photon excitation (Reprinted with permission from T.-T. Han, Y. Fu, and H. Ågren, Dynamic photon emission from multiphoton-excited semiconductor quantum dot, *J. Appl. Phys.*, vol. 103, p. 93703(6), 2008. Copyright 2008, American Institute of Physics)

and VB states using the electron-hole picture after a continuous-wave optical excitation is switched on at $t = 0$. This Monte Carlo scheme includes all the probability calculations of possible transitions. And a random number is generated to pick up one transition when it is smaller than the sum of the probabilities of all possible transitions. Otherwise the system remains unmodified until the next time interval. “A” corresponds to an optical excitation which occurs just a few fs after the turn-on of the excitation photon field. “B” marks a radiative recombination process. For the excited electrons and holes, stepwise tracks are observed in Fig. 3.9 (such as a series of points marked with “C”) due to the electron-phonon nonradiative relaxation processes from high-energy- to low-energy-states. And from time to time, the quantum dot can stay idle (“D”).

Figure 3.10 shows the occupation of the ground CB state, i.e., $|C_{c0}\rangle^2$ in Eq. (3.286), after the continuous-wave optical excitation is switched on at $t = 0$. We observe that the excitation spectral range is very broad. Single-photon excitation (i.e., photon energy larger than 2.655 eV) is very fast and strong, whereas for the photon energy from about 1.3 eV to 2.6 eV, i.e., two-photon excitation between the two arrows along the photon energy axis in Fig. 3.10, the calculated excited exciton rate is also very significant, even when the photon energy is below 1.3 eV (more-than-two-photon excitation).

Table 3.8 lists the details about three types of core-multishell CdSe quantum dots, which are denoted as QD556, QD600, and QD622, respectively, that were chemically synthesized. Before optical characterization, the quantum dot samples were diluted by toluene to concentrations of 2.0×10^{-6} , 1.3×10^{-6} , and 1.6×10^{-6} mol/L. Fluorescence spectroscopy measurements were performed using

Fig. 3.10 Multiphoton excitation of quantum-dot exciton in an ultrashort period excited by a CW light source switched on at $t = 0$. S^{st} denotes the optical power of the excitation light source (Reprinted with permission from Y. Fu, T.-T. Han, Y. Luo, and H. Ågren, Multi-photon excitation of quantum dots by ultra-short and ultra-intense laser pulse, *Appl. Phys. Lett.*, vol. 88, p. 221114(3), 2006. Copyright 2006, American Institute of Physics)

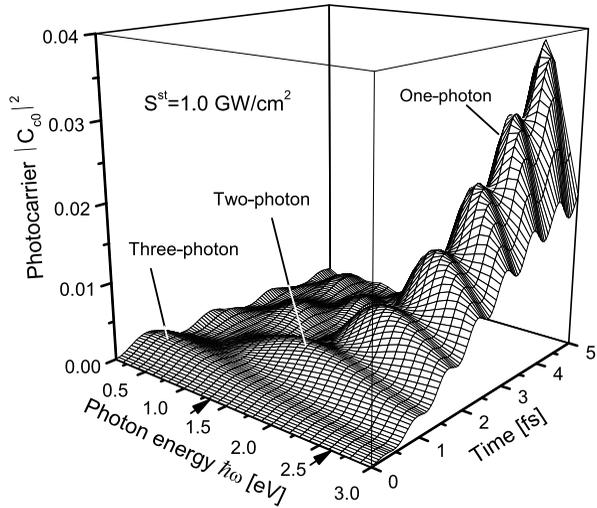


Table 3.8 Structure and optical properties of core-multishell CdSe quantum dots

	QD556	QD600	QD622
core CdSe diameter [nm]	3.6	4.5	5.4
shell CdS [ML]	1	2	3
shell Cd _{0.5} Zn _{0.5} S [ML]	1	1	1
shell ZnS [ML]	1	1	1
fluorescence peak wavelength [nm]	556	600	622
full width at half maximum [nm/meV]	32/121	29/104	24/77
quantum dot exciton energy $\hbar\omega_{QD}$ [eV]	2.23	2.07	1.99

the fluorescence spectrometer (Cary Eclipse, VARIAN) at excitation wavelengths of 350, 400, 450 nm (one-photon excitation), respectively, which shows that the one-photon fluorescence spectra of these quantum dots are independent of the excitation wavelength. Two-photon-induced fluorescence experiments were performed using a femtosecond laser with a pulse width of 130 fs, peak wavelength of 800 nm, full width at half maximum (FWHM) is 7 nm, and average power output of about 630 mW at 800 nm at a pulse frequency of 76 MHz. The laser pulse beam was focused into the center of a 1.0 cm path length square sample cell using a 10 cm focal length lens, and the laser beam had a diameter of approximately 30 μm on the sample. The two-photon excitation fluorescence of the samples was led to a monochromator. The results of the three quantum dot assemblies are listed in Table 3.8 and presented in Fig. 3.11.

Except the differences in the peak wavelengths, fluorescence spectra of the three types of quantum dots are rather similar both qualitatively and quantitatively with a full width at half maximum of about 30 nm. We observe the decrease of the full width at half maximum when the quantum dot is coated, which is believed to be the

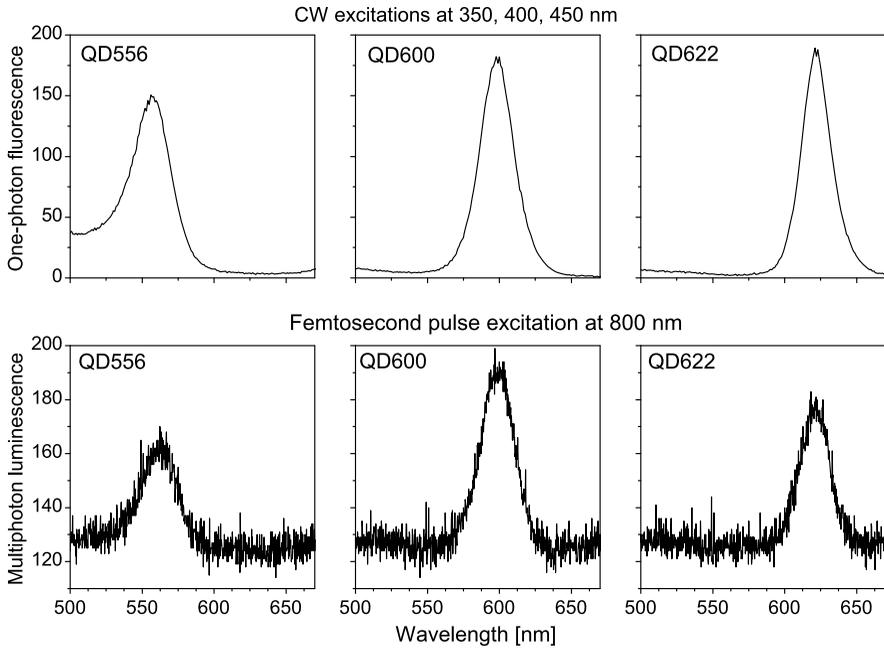


Fig. 3.11 One-photon and multiphoton fluorescence spectra of the three quantum dot assemblies (Reprinted with permission from Y. Fu, T.-T. Han, H. Ågren, L. Lin, P. Chen, Y. Liu, G.-Q. Tang, J. Wu, Y. Yue, and N. Dai, Design of semiconductor CdSe-core ZnS/CdS-multishell quantum dots for multiphoton applications, *Appl. Phys. Lett.*, vol. 90, p. 173102(3), 2007. Copyright 2007, American Institute of Physics)

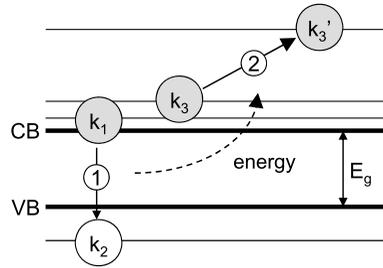
result of reduced surface states that deteriorate the optical properties of low-energy states confined in the CdSe core. Most importantly, after removing the base level count of about 125, the multiphoton fluorescence spectra are strikingly identical to the one-photon spectra, both the shapes including the full widths at half maximum and the relative intensities.

3.9 Auger Recombination and Impact Ionization

The Auger recombination is a process in which an electron and a hole recombine and the excess energy is transferred to either another electron or a hole, see Fig. 3.12. We first examine the case in semiconductor bulk materials, where we denote that state k_1 is occupied by electron, k_2 empty (hole), and k_3 is occupied (electron) (or empty when the third partner is a hole). The probability of such a configuration is

$$W(k_1, k_2, k_3) = f(k_1)[1 - f(k_2)]f(k_3) \quad (3.289)$$

Fig. 3.12 Schematic depiction of Auger recombination process of an electron k_1 and a hole k_2 recombine (process 1) and the excess energy is transferred (process 2) to another electron k_3 which then transits to state k'_3



where

$$\begin{aligned}
 f(k_1) &= \frac{n}{N_c} \exp\left(\frac{-E_{ck_1}}{k_B T}\right) \\
 1 - f(k_2) &= \frac{p}{N_v} \exp\left(\frac{-E_{vk_2}}{k_B T}\right) \\
 f(k_3) &= \frac{n}{N_c} \exp\left(\frac{-E_{ck_3}}{k_B T}\right)
 \end{aligned}
 \tag{3.290}$$

for nondegenerate semiconductors. Here n and p are electron and hole concentrations and N_c and N_v are CB and VB intrinsic densities of states, see Eq. (4.5).

We further assume that the final state k'_3 for the third electron is always available (i.e., empty) since it is a high-energy electron state in the conduction band. The total probability is thus

$$W(k_1, k_2, k_3) \approx \frac{n}{N_c} \exp\left(-\frac{E_g + E_{ck_1} + E_{vk_2} + E_{ck_3}}{k_B T}\right)
 \tag{3.291}$$

We now try to minimize the energy

$$\frac{\hbar^2 k_1^2}{2m_c^*} + \frac{\hbar^2 k_2^2}{2m_v^*} + \frac{\hbar^2 k_3^2}{2m_c^*}
 \tag{3.292}$$

(in the parabolic band approximation) for optimal probability factor, which is consistent with energy and momentum conservations. It is found that the threshold is reached when the high energy state k'_3 lines opposite to the other three states. And the conditions of

$$k_1 = k_3 = \mu k_2
 \tag{3.293}$$

give us a maximum probability factor

$$W(k_1, k_2, k_3) = \frac{n}{N_c} \exp\left[-\frac{(1 + 2\mu)E_g}{(1 + \mu)k_B T}\right]
 \tag{3.294}$$

where $\mu = m_c^*/m_v^*$.

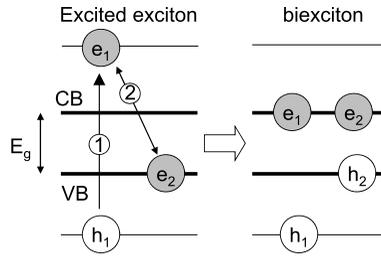


Fig. 3.13 Schematic depiction of impact ionization of a high-energy electron-hole pair. (1) Electron-hole pair (e_1 and h_1) is photogenerated. (2) e_2 gets excited from a valence-band state to a conduction-band state via Coulombic interaction with e_1 , leaving hole h_2 behind. The reverse process is referred to be Auger recombination, see Fig. 3.12

The final Auger rate is obtained by evaluating the multiple integral of the matrix element multiplied by the probability factor over all the available states. The interaction is Coulombic among the charged particles.

In general, the Auger rate increases exponentially as the bandgap is decreased. They also increase exponentially as the temperature increases. These are the direct results of the energy and momentum conservations and the carrier statistics. Auger processes are more or less unimportant in semiconductors with bandgaps larger than approximately 1.5 eV (e.g., GaAs, AlGaAs, InP). However, they become quite important in narrow bandgap materials such as InGaAs and HgCdTe ($E_g < 0.5$ eV), and are thus a serious hindrance for the development of long wavelength lasers. In general, Auger processes can be mediated by defects as well. Moreover, deep levels in the bandgap can be involved in the Auger processes.

We now consider the case of semiconductor nanoparticles. Nozik proposed that impact ionization (i.e., the reverse process of the Auger effect) might be enhanced in semiconductor quantum dots to increase the efficiency of solar cells up to about 66 % [42]. This was later verified experimentally by Schaller and Klimov [43]. Impact ionization was reported to produce multiple excitons per photon in one quantum dot, which is referred to multiple exciton generation (MEG), that results in a very high quantum yield (up to 300 % when the photon energy reaches 4 times the quantum-dot bandgap) in quantum-dot solar cells [44]. Multiple carrier extraction (> 100 %) was observed at photon energies greater than 2.8 times the PbSe quantum-dot bandgap with about 210 % measured at 4.4 times the bandgap [45]. In a recent communication in Nano Letters, Trinh et al. [46] showed compelling support for carrier multiplication in PbSe quantum dots, while at the same time the obtained MEG efficiency was very close to the one in bulk material.

Figure 3.13 shows a highly photoexcited electron and hole pair that evolves into a multiple-exciton state through the impact ionization. The carrier-carrier interaction is expressed by the Coulombic potential

$$V = \frac{e^2}{4\pi\epsilon|\mathbf{r}_1 - \mathbf{r}_2|} \quad (3.295)$$

to account for the two-body interaction of two electrons from an initial state $\phi_j(\mathbf{r}_1)\psi_i(\mathbf{r}_2) \exp[-i(E_j + E_i)t/\hbar]$ to a final state

$$\frac{1}{\sqrt{2}} \left| \begin{array}{cc} \psi_n(\mathbf{r}_1)\psi_m(\mathbf{r}_1) \\ \psi_n(\mathbf{r}_2)\psi_m(\mathbf{r}_2) \end{array} \right| \exp[-i(E_n + E_m)t/\hbar] \quad (3.296)$$

Note that the two electrons in the final state are not distinguishable so that we use the Slater determinant.

Let us discuss the scattering matrix element of the Auger-type process

$$\int \int \Psi_{ck_1}^*(\mathbf{r}_1)\Phi_{vk_2}^*(\mathbf{r}_2) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \Psi_{ck'_1}(\mathbf{r}_1)\Psi_{ck'_2}(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \quad (3.297)$$

where an incident high-energy electron in the conduction band with a wave vector \mathbf{k}_1 collides with a second electron that occupies a VB state \mathbf{k}_2 , resulting in two CB electrons \mathbf{k}'_1 and \mathbf{k}'_2 . See, for example, the textbook of Landau and Lifshitz [47]. Here $\Psi_{ck}(\mathbf{r})$ [$\Phi_{vk}(\mathbf{r})$] denotes the total wave function of CB electron (VB hole) state \mathbf{k} . For semiconductor systems and within the effective mass approximation, $\Psi_{ck}(\mathbf{r}) = \psi_{ck}(\mathbf{r})u_{ck}(\mathbf{r})$ and $\Phi_{vk}(\mathbf{r}) = \phi_{vk}(\mathbf{r})u_{vk}(\mathbf{r})$, where ψ and ϕ are envelope functions of CB electron and VB hole, respectively, and u 's are periodic Bloch functions, the above expression becomes

$$\int \int \psi_{ck_1}^*(\mathbf{r}_1)u_{ck_1}^*(\mathbf{r}_1)\phi_{vk_2}^*(\mathbf{r}_2)u_{vk_2}^*(\mathbf{r}_2) \times \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \psi_{ck'_1}(\mathbf{r}_1)u_{ck'_1}(\mathbf{r}_1)\psi_{ck'_2}(\mathbf{r}_2)u_{ck'_2}(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \quad (3.298)$$

See, for example, Eq. (6.128) in Ref. [35]. Following overlap integrals are thus involved

$$I_1 = \int_{\text{cell}} u_{ck_1}^*(\mathbf{r}_1)u_{ck'_1}(\mathbf{r}_1) d\mathbf{r}_1 \quad (3.299)$$

$$I_2 = \int_{\text{cell}} u_{vk_2}^*(\mathbf{r}_2)u_{ck'_2}(\mathbf{r}_2) d\mathbf{r}_2$$

in the evaluation of the Auger-type scattering processes. The first overlap integral can be approximated to unity. Because of the orthogonality of u functions for different bands but the same \mathbf{k} , the second overlap integral I_2 is, in crudest approximation, zero. This however is not correct since the periodic Bloch functions are functions of wave vectors, see Eqs. (3.128), (3.129). For small \mathbf{k} we can write

$$u_{ck}(\mathbf{r}) = u_c(\mathbf{r}) + \mathbf{k}\nabla_{\mathbf{k}}u_c(\mathbf{r}) + \dots \quad (3.300)$$

$$u_{vk}(\mathbf{r}) = u_v(\mathbf{r}) + \mathbf{k}\nabla_{\mathbf{k}}u_v(\mathbf{r}) + \dots$$

where $u_c(\mathbf{r})$ and $u_v(\mathbf{r})$ are periodic Bloch functions at the CB and VB edges, respectively. $\int_{\text{cell}} u_v^*(\mathbf{r}_2)u_c(\mathbf{r}_2) d\mathbf{r}_2 = 0$. Substituting these expressions into the over-

lap integrals we obtain the squared overlap integral in terms of the heavy-hole mass m_v [48, 49]

$$|I_2|^2 = \frac{\hbar^2}{2E_g} \left(\frac{1}{m_0} + \frac{1}{m_v} \right) |\mathbf{k}_2 - \mathbf{k}'_2|^2 \quad (3.301)$$

where E_g is the energy bandgap of the bulk material. By using the inverse Bohr radius of shallow impurities as a measure about the \mathbf{k} values, Landsberg and Adams obtained that $|I_2| = 0.223$ for shallow-impurity-assisted Auger-type processes in bulk CdS and $|I_2| = 0.265$ in GaAs [49]. Note that the inverse Bohr radius of shallow impurities in bulk semiconductors is small. For quantum dots under investigation, the effective Bohr radius of the electron and hole distribution is largely determined by the quantum dot size, which is about 5 nm, i.e., very small compared with the Bohr radius of shallow impurities in bulk semiconductors (about 100 nm in CdS and GaAs [49]). This results in a large value of $|I_2|$ that may exceed unity depending on the quantum confinement of the carrier inside the quantum dots. Under this specific circumstance, we approximate $|I_2| = 1$. In other words, the electrons and holes in quantum dots, described by effective masses with the presence of the quantum-dot confinement potentials, interact with each other via the Coulombic force of Eq. (3.295).

By the scattering theory and the generalized Golden rule presented in Chap. 2 and denoting $\langle \phi_j \psi_i | \hat{T}(t) | \phi_j \psi_i \rangle \approx e^{-w_{ji}t/2}$ as the temporal development $\hat{T}(t)$ of state ji , it is easy to obtain

$$w_{ji} = \frac{2\pi}{\hbar} \sum_{nm \neq ji} |\langle \phi_j \psi_i | V | \psi_n \psi_m \rangle|^2 \delta(E_j + E_i - E_n - E_m) \quad (3.302)$$

$\Gamma_{ji} = \hbar w_{ji}/2$ is the relaxation energy and $1/w_{ji}$ the decaying time. In numerical calculations the δ functions in the above equation are replaced by $\Gamma_{ji}/[\Gamma_{ji}^2 + (E_j + E_i - E_n - E_m)^2]$.

To estimate the Coulombic potential of Eq. (3.295), we make the following analysis. Consider the best configuration for Eq. (3.295) that the four wave functions involved in the impact ionization are all uniformly distributed within the quantum dot volume. The corresponding Coulombic potential is

$$\frac{e}{8\pi\epsilon} \begin{cases} \frac{2}{r}, & r > a \\ \frac{1}{a} \left(3 - \frac{r^2}{a^2} \right), & r \leq a \end{cases} \quad (3.303)$$

where a is the radius of the quantum dot. When the hole is assumed to distribute also uniformly in the same quantum-dot volume, the Coulombic interaction energy between this electron-hole pair is $3e^2/10\pi\epsilon a$, which is about 28.8 meV when $a = 5$ nm and $\epsilon = 12$ (common semiconductors) when we neglect the energy conservation requirement, i.e., we neglect the δ function in Eq. (3.302). This is very large compared with the value in bulk because of the quantum confinements (electrons and holes are now confined within a spherical volume with a radius of a). Such an interaction energy is much large as compared with the light-matter interaction which

is normally less than 0.1 meV in similar quantum dots. Large impact ionization and also Auger recombination are thus expected in colloidal quantum dots.

To calculate the impact ionization interaction from an initial state of

$$R_{\ell_1}(r_1)Y_{\ell_1 m_1}(\theta_1, \phi_1)R_{\ell_2}(r_2)Y_{\ell_2 m_2}(\theta_2, \phi_2)$$

to a final state

$$R_{\ell_3}(r_1)Y_{\ell_3 m}(\theta_2, \phi_2)R_{\ell_4}(r_2)Y_{\ell_4 m_4}(\theta_2, \phi_2)$$

(the first term one in Eq. (3.296)), we notice

$$\frac{1}{4\pi|\mathbf{r}_1 - \mathbf{r}_2|} = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \frac{1}{2\ell+1} \frac{r_{<}^{\ell}}{r_{>}^{\ell+1}} Y_{\ell m}^*(\theta_2, \phi_2) Y_{\ell m}(\theta_1, \phi_1) \quad (3.304)$$

where $r_{<} = \min\{|\mathbf{r}_1|, |\mathbf{r}_2|\}$ and $r_{>} = \max\{|\mathbf{r}_1|, |\mathbf{r}_2|\}$ so that the impact ionization energy consists of

$$\begin{aligned} & \frac{r_{<}^{\ell}}{r_{>}^{\ell+1}} R_{\ell_1}(r_1)R_{\ell_2}(r_2)R_{\ell_3}(r_1)R_{\ell_4}(r_2) \\ & \times Y_{\ell_1 m_1}^*(\theta_1, \phi_1)Y_{\ell_2 m_2}^*(\theta_2, \phi_2)Y_{\ell m}^*(\theta_2, \phi_2) \\ & \times Y_{\ell m}(\theta_1, \phi_1)Y_{\ell_3 m_3}(\theta_1, \phi_1)Y_{\ell_4 m_4}(\theta_2, \phi_2) \end{aligned} \quad (3.305)$$

for which we utilize the addition theorem for spherical harmonics, see Sect. 3.6.

A little detailed analysis shows a few selection rules about ℓ and m such as $m_4 - m_2 = m_1 - m_3$. However, these selection rules can be easily fulfilled in nanoscale quantum dots because of the large number of available states confined in the quantum dots. The most important qualitative selection rules are about the radial functions that

$$\begin{aligned} & \int \int \frac{r_{<}^{\ell}}{r_{>}^{\ell+1}} R_{\ell_1}(r_1)R_{\ell_2}(r_2)R_{\ell_3}(r_1)R_{\ell_4}(r_2) \\ & = \int_0^a R_{\ell_1}(r_1)R_{\ell_3}(r_1) \left[\int_0^{r_1} \frac{r_2^{\ell}}{r_1^{\ell+1}} R_{\ell_2}(r_2)R_{\ell_4}(r_2)r_2^2 dr_2 \right. \\ & \quad \left. + \int_{r_1}^a \frac{r_1^{\ell}}{r_2^{\ell+1}} R_{\ell_2}(r_2)R_{\ell_4}(r_2)r_2^2 dr_2 \right] r_1^2 dr_1 \end{aligned} \quad (3.306)$$

which shows that a direct spatial overlapping between $R_{\ell_2}(r_2)$ (VB hole state) and $R_{\ell_4}(r_4)$ (CB electron state) will result in a large impact ionization. Impact ionization in type-II quantum dots is thus negligible since the electron-hole wave function overlapping is small. Furthermore, to ensure energy conservation, $R_{\ell_1}(r)$ is normally a high-energy CB state, while both $R_{\ell_3}(r)$ and $R_{\ell_4}(r)$ are low-energy CB states. Impact ionization in many core-shell-structured quantum dots will be small, since the

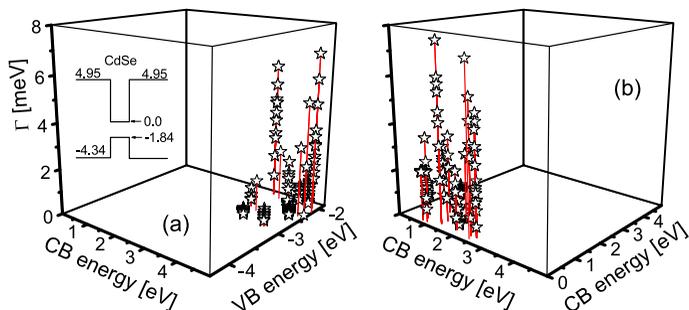


Fig. 3.14 (a) Impact ionization energy between initial CB state(s) and initial VB state(s). *Inset* shows the band structure of the CdSe quantum dot in vacuum. (b) Auger recombination. The quantum-dot radius $a = 4$ nm (Reprinted with permission from Y. Fu, Y.-H. Zhou, H. Su, F.Y.C. Boey, and H. Ågren. Impact ionization and Auger recombination rates in semiconductor quantum dots, *J. Phys. Chem. C*, vol. 114, pp. 3743–3747, 2010. Copyright 2010, American Chemical Society)

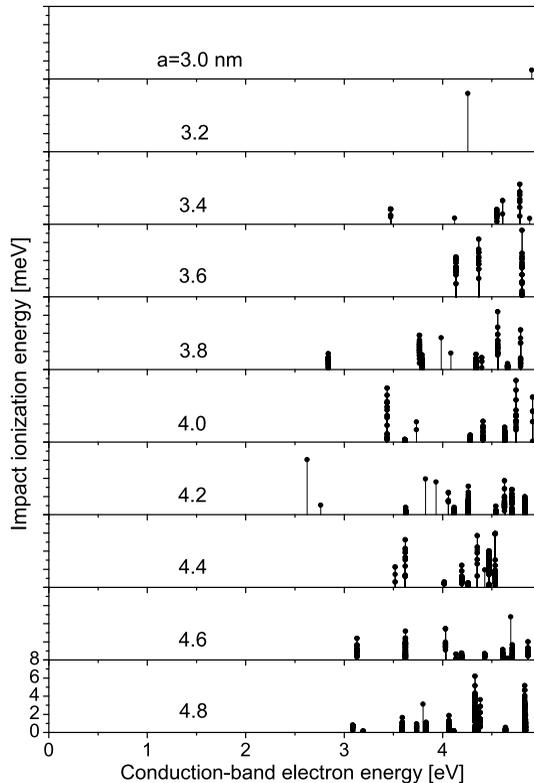
high-energy wave functions are much more extended (extended into shells) than the ground-state wave function (deeply confined in the core region).

We pick up one of the CB state and one VB state to calculate the total impact ionization energy to all possible combinations of two final CB states, which is shown in Fig. 3.14(a).

Figure 3.14(a) agrees what can be expected intuitively that the initial CB state has to be high while the VB state low in order to fulfill the energy conservation requirement. Furthermore, as mentioned before, high-energy states can be highly degenerate because of large ℓ , so Γ can be also high. For the case of Fig. 3.14 where $a = 4$ nm, we observe a maximal Γ of about 7 meV. We further observe tens of Γ over 4 meV while majorities are about 1 meV. These are very high as compared with the light-matter interaction in a similar CdSe quantum dot; thus, we can be very optimistic about MEG processes in colloidal quantum dots, as has been very much reported as well as anticipated for significant solar cell applications. This is more clearly reflected in Fig. 3.15 about the relationship between the impact ionization and initial CB state(s). For the CdSe quantum dots under investigation, MEG effect can be expected when the excitation radiation energy exceeds a threshold of about $2E_g$ ($E_g = 1.84$ eV for CdSe) when the quantum-dot radius $a = 4.8$ nm, while the threshold energy will be $3E_g$ when $a = 4$ nm, and $4E_g$ for $a = 3$ nm.

Auger recombination rates (reverse processes of the impact ionizations) shown in Fig. 3.14(b) can be similarly analyzed. Note that the energy range of the initial CB states in Auger recombination processes is much wider than the impact ionization, especially for small quantum dots. This is due to the large density of states at high energy so that there are more available final CB states for two initial CB states to interact via the Auger recombination process.

Fig. 3.15 Impact ionization energy vs. initial CB states as functions of the quantum dot radius a (Reprinted with permission from Y. Fu, Y.-H. Zhou, H. Su, F.Y.C. Boey, and H. Ågren, Impact ionization and Auger recombination rates in semiconductor quantum dots, *J. Phys. Chem. C*, vol. 114, pp. 3743–3747, 2010. Copyright 2010, American Chemical Society)



References

1. Sze SM (1981) *Physics of semiconductor devices*, 2nd edn. Wiley, New York, p 32
2. Goepfert Mayer M (1931) Elementary processes with two quantum jumps. *Ann Phys (Leipz)* 9:273–294
3. Lami J-F, Gilliot P, Hirlimann C (1996) Observation of interband two-photon absorption saturation in CdS. *Phys Rev Lett* 77:1632–1635
4. Helmchen F, Svododa K, Denk W, Kleinfeld D, Tank DW (1996) In vivo dendritic calcium dynamics in deep-layer cortical pyramidal neurons. *Nat Neurosci* 2:989–996
5. Yokoyama H, Guo H, Yoda T, Takashima K, Sato K-I, Taniguchi H, Ito H (2006) Two-photon bioimaging with picosecond optical pulses from a semiconductor laser. *Opt Express* 14:3467–3471
6. Wherrett BS (1984) Scaling rules for multiphoton interband absorption in semiconductors. *J Opt Soc Am B* 1:67–72
7. Schmidt ME, Blanton SA, Hines MA, Guyot-Sionnest P (1996) Size-dependent two-photon excitation spectroscopy of CdSe nanocrystals. *Phys Rev B* 53:12629–12632
8. Haken H (1963) Theory of exciton II. In: Kuper CG, Whitfield GD (eds) *Polarons and excitons*. Plenum, New York, p 295
9. Dimmock JO (1967) Introduction to the theory of exciton states in semiconductors. In: Willardson RK, Beer AC (eds) *Semiconductors and semimetals*, vol 3. Academic Press, New York, pp 259–319, Chap. 7
10. Haken H (1983) *Quantum field theory of solids*. North-Holland, Amsterdam, p 151

11. Lawaetz P (1971) Valence-band parameters in cubic semiconductors. *Phys Rev B* 4:3460–3467
12. Madelung O (ed) (1991) *Semiconductors group IV elements and III–V compounds*. Springer, Berlin
13. Miller DAB, Chemla DS, Eilenberg DJ, Smith PW, Gossard AC, Tsang WT (1982) Large room-temperature optical nonlinearity in GaAs/Ga_{1-x}Al_xAs multiple quantum well structures. *Appl Phys Lett* 41:679–681
14. Sun HD, Makino T, Segawa Y, Kawasaki M, Ohtomo A, Tamura K, Koinuma H (2002) Enhancement of exciton binding energies in ZnO/ZnMgO multiquantum wells. *J Appl Phys* 91:1993–1997
15. Sapra S, Sarma DD (2004) Evolution of the electronic structure with size in II–VI semiconductor nanocrystals. *Phys Rev B* 69:125304
16. Lippens PE, Lannoo M (1990) Comparison between calculated and experimental values of the lowest excited electronic state of small CdSe crystallites. *Phys Rev B* 41:6079–6081
17. Brus LE (1984) Electron-electron and electron-hole interactions in small semiconductor crystallites: the size dependence of the lowest excited electronic state. *J Chem Phys* 80:4403–4409
18. Nair SV, Sinha S, Rustagi KC (1987) Quantum size effects in spherical semiconductor microcrystals. *Phys Rev B* 35:4098–4101
19. Kayanuma Y, Momiji H (1990) Incomplete confinement of electrons and holes in microcrystals. *Phys Rev B* 41:10261–10263
20. Grabovskis VYa, Dzenis YaYa, Ekimov AI, Kudryavtsev IA, Tolstoi MN, Rogulis UT (1989) Photoionization of semiconducting microcrystals in glass [luminescence studies]. *Fiz Tverd Tela* 31:272–275
21. Grabovskis VYa, Dzenis YaYa, Ekimov AI, Kudryavtsev IA, Tolstoi MN, Rogulis UT (1989) *Sov Phys, Solid State* 31:149–151
22. Swank RK (1967) Surface properties of II–VI compounds. *Phys Rev* 153:844–849
23. Bujatti M (1968) CdS-metal barriers from photovoltage measurements. *Brit J Appl Phys (J Phys D)*, Ser 2 1:581–584
24. Lippens PE, Lannoo M (1989) Calculation of the bandgap for small CdS and ZnS crystallites. *Phys Rev B* 39:10935–10942
25. Madelung O (ed) (1992) *Data in science and technology: semiconductors other than group IV elements and III–V compounds*. Springer, Boston
26. Einevoll GT (1992) Confinement of excitons in quantum dots. *Phys Rev B* 45:3410–3417
27. Nair SV, Ramaniah LM, Rustagi KC (1992) Electron states in a quantum dot in an effective-bond-orbital model. *Phys Rev B* 45:5969–5979
28. Vogl P, Hjalmarson HP, Dow JD (1983) A semi-empirical tight-binding theory of the electronic structure of semiconductor. *J Phys Chem Solids* 44:365–378
29. Sapra S, Shanthi N, Sarma DD (2002) Realistic tight-binding model for the electronic structure of II–VI semiconductors. *Phys Rev B* 66:205202
30. Jiang J, Gao B, Han T-T, Fu Y (2009) Ab initio study of energy band structures of GaAs nanoclusters. *Appl Phys Lett* 94, 092110
31. van der Waerden BL (1932) *Die Gruppentheoretische Methode in der Quantenmechanik*. Springer, Berlin
32. Racah G (1942) Theory of complex spectra. II. *Phys Rev* 62:438–462
33. Fu Y, Willander M, Ivchenko EL (2000) Photonic dispersions of semiconductor-quantum-dot-array-based photonic crystals in primitive and face-centered cubic lattices. *Superlattices Microstruct* 27:255–264
34. Jacobini C, Reggiani L (1983) The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials. *Rev Mod Phys* 55:645–705
35. Ridley BK (1988) *Quantum processes in semiconductors*, 2nd edn. Clarendon, Oxford
36. Han T-T, Fu Y, Ågren H (2008) Dynamic photon emission from multiphoton-excited semiconductor quantum dot. *J Appl Phys* 103:93703(6)
37. Reynolds DC, Litton CW, Collins TC (1971) Bound-phonon quasiparticle in CdS. *Phys Rev B* 4:1868–1872

38. Pan AL, Liu RB, Zou BS (2006) Phonon-assisted stimulated emission from single CdS nanoribbons at room temperature. *Appl Phys Lett* 88:173102(3)
39. Rustagi KC, Weber W (1976) Adiabatic bond charge model for the phonons in A^3B^5 semiconductors. *Solid State Commun* 18:673–675
40. Sugawara M, Mukai K, Shoji H (1997) Effect of phonon bottleneck on quantum-dot laser performance. *Appl Phys Lett* 71:2791
41. Murdin BN, Hollingworth AR, Kamal-Saadi M, Kotitschke RT, Ciesla CM, Pidgeon CR, Findlay PC, Pellemans HPM, Langerak CJGM, Rowe AC, Stradling RA, Gornik E (1999) Suppression of LO phonon scattering in Landau quantized quantum dots. *Phys Rev B* 59:R7817–R7820
42. Nozik AJ (2002) Quantum dot solar cells. *Physica E* 14:115–120
43. Schaller RD, Klimov VI (2004) High efficiency carrier multiplication in PbSe nanocrystals: implications for solar energy conversion. *Phys Rev Lett* 92:186601
44. Hanna M, Ellingson RJ, Beard M, Yu P, Nozik AJ (2004) Quantum dot solar cells: high efficiency through impact ionization. In: DOE solar energy technologies program review meeting, October 25–28, 2004, Denver, USA
45. Kim SJ, Kim WJ, Sahoo Y, Cartwright AN, Prasad PN (2008) Multiple exciton generation and electrical extraction from a PbSe quantum dot photoconductor. *Appl Phys Lett* 92:31107(3)
46. Trinh MT, Houtepen AJ, Schins JM, Hanrath T, Piris J, Knulst W, Goossens APLM, Siebbeles LDA (2008) In spite of recent doubts carrier multiplication does occur in PbSe nanocrystals. *Nano Lett* 8:1713–1718
47. Landau LD, Lifshitz EM (1962) *Quantum mechanics*, 3rd edn. Pergamon, Oxford, p 278
48. Ridley BK (1988) *Quantum processes in semiconductors*. Clarendon, Oxford, pp 269–278
49. Landsberg PT, Adams MJ (1973) Theory of donor-acceptor radiative and Auger recombination in simple semiconductors. *Proc R Soc Lond A* 334:523–539

Chapter 4

Electronic Quantum Devices

Abstract Starting with the introduction of electron transport in conventional $p - n$ junction and field-effect transistor, we first discuss semiclassical versus quantum mechanical considerations about carrier transport in solids after which we focus on the tunneling of an electron wave through a potential barrier in resonant tunneling diode and heterostructure barrier varactor. Quantum mechanical engineering of nano-scale transistors, including high-electron-mobility transistor and single-electron transistor, are then presented for higher carrier mobility and better current-voltage control.

4.1 $p - n$ Junction and Field-Effect Transistor

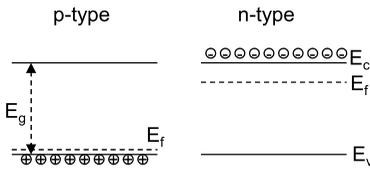
As we learned from Chap. 1, an intrinsic semiconductor bulk material has a completely filled valence band and a completely empty conduction band separated by a bandgap E_g . It is not conducting under a moderate external bias which is less than a few volts, see Sect. 2.4. When shining the semiconductor material with a light with photon energy $\hbar\omega$ larger than E_g , electrons, initially occupying the valence band, will transit to the conduction band after absorbing the photon. Both the conduction band and the valence band are now partially filled and the semiconductor becomes electric conducting. Furthermore, the semiconductor can become electric conducting when we incorporate dopants into the semiconductor, either n -type or p -type dopants which have one or more extra or less valence electrons than the atoms in the host semiconductor. The extra valence electrons will occupy the conduction band, while the missing valence electrons will make the valence band partially unoccupied.

By the drift and diffusion model of Sect. 2.7, we can calculate the current density j of a uniform, bulk, and n -type doped semiconductor

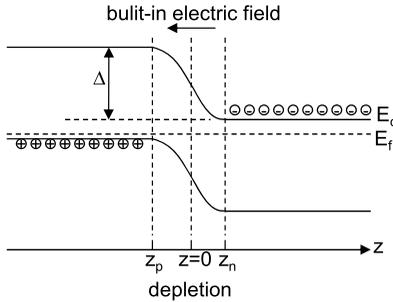
$$j = -en\mu E \tag{4.1}$$

where n is the concentration of free electrons, μ is the carrier mobility and E is the electric field due to the external bias. In normal situations, free electrons in a semiconductor materials come from dopings such as in a Si-doped GaAs material where each Si atom contribute one free electron to the conduction band in GaAs so

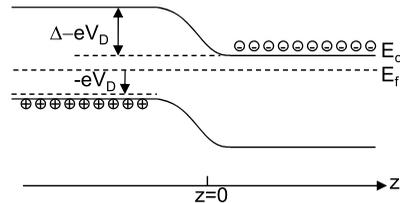
(a) p-type and n-type semiconductors



(b) pn junction at equilibrium



(c) forward bias V_D



(d) I-V characteristics

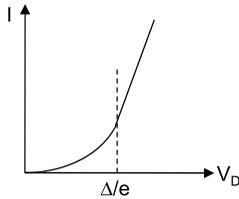


Fig. 4.1 Energy band structures of (a) *p*-type and *n*-type doped semiconductor, (b) *p* – *n* junction at equilibrium, (c) forward bias V_D , (d) Current-voltage $I - V$ characteristics

that the concentration of the free electrons equals to the Si doping concentration. A similar equation can be written down for a *p*-type doped semiconductor.

The story of electric conduction becomes more much interesting when we put one half-infinitely large *n*-type semiconductor and one half-infinitely large *p*-type semiconductor together to form a *p* – *n* junction. Figure 4.1 shows the energy band structures of the *p*-type, *n*-type doped semiconductor, *p* – *n* junction at equilibrium and with a forward bias V_D , as well as the general feature of the current-voltage $I - V$ characteristics.

The *p*-type and the *n*-type semiconductors are normally characterized by their doping concentrations, n_A and n_D , respectively, where subscript “A” denotes acceptor (to accept one electron from the host material, thus leaving one hole there) and “D” donors (to donate one electron into the host semiconductor). We adopt the

concept of Fermi level E_f to characterize quantitatively the semiconductor

$$n_D = \int \frac{1}{1 + e^{(E-E_f)/k_B T}} N_3(E) dE \quad (4.2)$$

where $N_3(E)$ is the three-dimensional density of states, see Eq. (1.194). The numerical calculation of E_f as a function of n_D at different temperature T is not trivial, see Sect. 6.1. In practical device applications, n_D is not so high so that E_f is far below E_c . In this case,

$$\frac{1}{1 + e^{(E-E_f)/k_B T}} \approx e^{-(E-E_f)/k_B T} \quad (4.3)$$

for $E > E_c$, thus

$$n_D = \int e^{-(E-E_f)/k_B T} N_3(E) dE = N_c e^{-(E_c-E_f)/k_B T} \quad (4.4)$$

where

$$N_c = \int e^{-(E-E_c)/k_B T} N_3(E) dE \quad (4.5)$$

is a physical constant about the semiconductor, normally known as the intrinsic electron concentration in the conduction band. Equation (4.4) indicates that E_f will be close to E_c when we increase n_D . A similar relationship can be obtained between E_f and n_A . The general energy band structures of the p -type and n -type semiconductors are shown in Fig. 4.1(a).

By joining a half-infinite p -type semiconductor ($z < 0$) and a half-infinite n -type semiconductor ($z \geq 0$) at $z = 0$, see Fig. 4.1(b), a depletion region will be formed where holes will be depleted in $(z_p, 0)$ and electrons in $(0, z_n)$, leaving ionized dopants that induce a built-in electric field, thereafter a potential barrier Δ . Electrons in the n -type region will diffuse to the p -type region through the depletion region because of the non-uniform electron distribution. However, such a diffusion is suppressed by the built-in electric field in the depletion region so that eventually at equilibrium, there is no net electron transport through the depletion region. Same situation occurs for the holes in the valence band. A uniform Fermi level is well defined which is flat across the whole $p - n$ junction.

Now we put the p side of the junction grounded and a bias V_D on the n side, see Fig. 4.1(c). The bias will apply largely across the depletion region because of Ohm's law, which will reduce the electric field there. Diffusion of the electrons from the n side to the p side will thus surpass the drift, resulting in a net electron transport.

Let us calculate the current induced by V_D , i.e., the $I - V$ characteristics of the $p - n$ junction. Instead of detailed analysis of the formation of the depletion region, the strength of the electric field in the depletion region, then the diffusion and drift of the electrons across the depletion region, we consider the following equivalent approach: electron transport from the n side to the p side is largely stopped by the potential barrier Δ . A forward bias V_D will decrease the barrier height to $\Delta - eV_D$

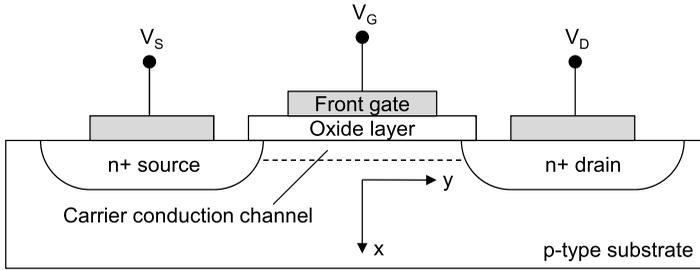


Fig. 4.2 An n -channel metal-oxide-semiconductor field-effect transistor

to assist the electron transport. By assuming quasi equilibrium state, the electron density in the p side of the junction is

$$n = N_c e^{-(\Delta - eV_D + E_c - E_f)/k_B T} \quad (4.6)$$

Since the current is proportional to the number of electrons that overcome the potential barrier, we have

$$j \propto N_c e^{-(\Delta - eV_D + E_c - E_f)/k_B T} \quad (4.7)$$

We further know that the net current is zero at zero bias. Thus

$$j = j_0 (e^{eV_D} - 1) \quad (4.8)$$

where

$$j_0 \propto N_c e^{-(\Delta + E_c - E_f)/k_B T} \quad (4.9)$$

When $eV_D > \Delta$, there is no potential barrier for the electrons to transport from the n side to the p side, a linear $I - V$ relationship will be expected. Figure 4.1(d) shows schematically the $I - V$ characteristics of the $p - n$ junction.

We have thus be able to obtain a variable current through a $p - n$ junction. We now study the base of the modern electronics, namely, the metal-oxide-semiconductor field-effect transistor, i.e., MOSFET. Figure 4.2 shows the schematic device structure of an n -channel MOSFET.

We first look at the metal-oxide-semiconductor structure along the x axis at the middle of the MOSFET, which is shown in Fig. 4.3(a). Referring to the vacuum energy level E_0 , the distance between E_0 and the Fermi level of a metal $E_{f,m}$ is known as the work function $e\phi_m$, i.e., the energy needed to excite an electron at the Fermi level into the vacuum level. Another physical parameter is $e\chi$, electron affinity, is the energy needed to excite an electron at the conduction bandedge into the vacuum level. Table 4.1 lists work functions $e\phi_m$ of important metal elements and electron affinities $e\chi$ of commonly used semiconductors. Note that the values of $e\phi_m$ and $e\chi$ are rather scattered in literatures and Table 4.1 serves only as a reference.

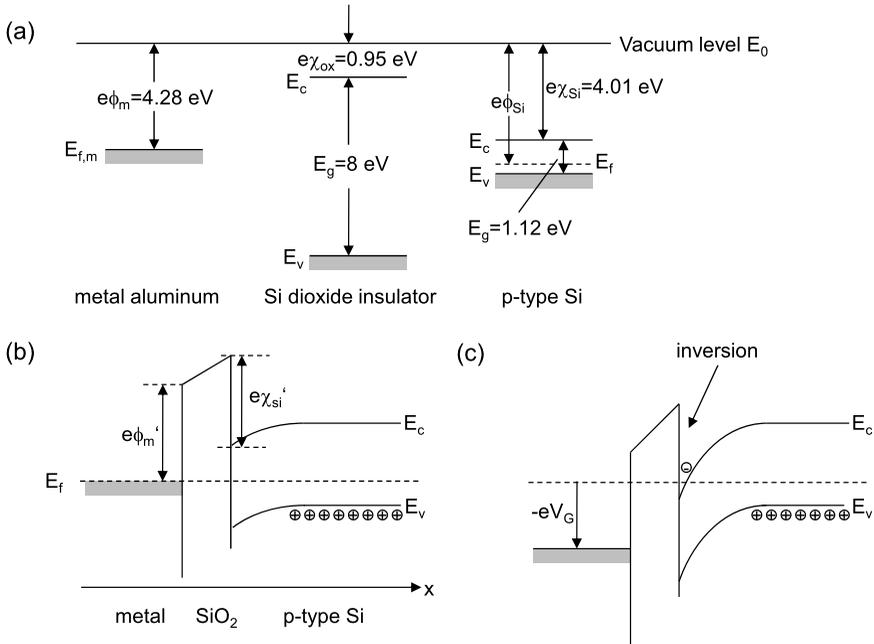


Fig. 4.3 (a) Energy band structures of the metal, oxide, and *p*-type Si semiconductor. (b) Energy band alignment of the metal-oxide-*p*-type Si structure. (c) Energy band structure of the under a gate bias V_{GT} where an inversion layer (*n*-type) is formed at the oxide-semiconductor interface

Table 4.1 Work functions $e\phi_m$ of important metal elements and electron affinities $e\chi$ of commonly used semiconductors [eV]

Element	$e\phi_m$	Semiconductor	$e\chi$
Ag	4.26	Ge	4.13
Al	4.28	Si	4.01
Au	5.1	GaAs	4.07
Cr	4.5	AlAs	3.5
Mo	4.6	SiO ₂	0.95
Ni	5.15		
Pd	5.12		
Pt	5.65		
Ti	4.33		
W	4.55		

When we join the metal, SiO₂ and *p*-type Si in Fig. 4.3(a) to form a metal-oxide-semiconductor structure in Fig. 4.3(b), various energy levels are to be aligned in order to reach equilibrium. Without any external bias, thermal equilibrium is

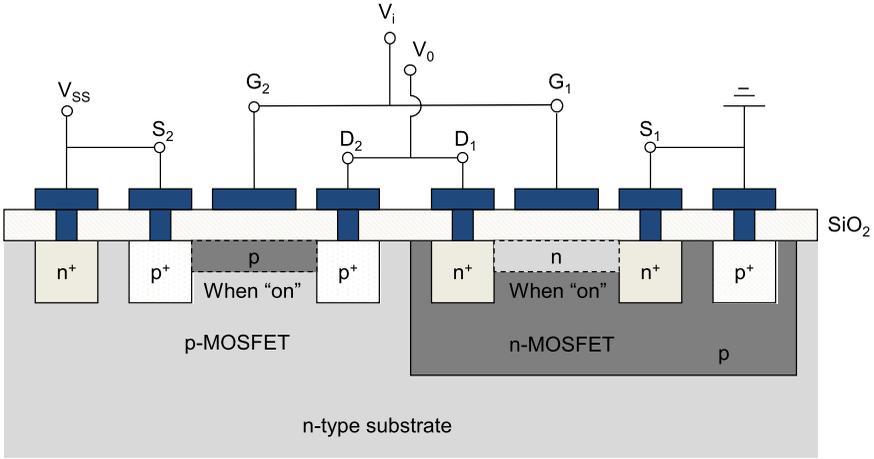


Fig. 4.4 Complementary MOSFET device

reached which is described by Fermi level E_f . Referring to Fig. 4.3(a), we observe that the MOS structure is characterized by the so-called modified work function $e\phi'_m = e\phi_m - 0.95$ eV between the metal and SiO₂ and modified electron affinity $e\chi'_{Si} = e\chi_{Si} - 0.95$ eV between Si and SiO₂, where 0.95 eV is the electron affinity of SiO₂.

When we put a gate bias V_G with respect to the p -type Si, the local Fermi level in metal is lowered by $-eV_G$. Because of the high free carrier concentrations in the metal and p -type Si, the gate bias applies largely in SiO₂ and the SiO₂-Si interface where the free carrier concentration is low. At high enough V_G , the conduction band edge of Si close to the SiO₂-Si interface will get close to the local Fermi level in Si so that electrons begin to accumulate there, forming the so-called inversion layer, see Fig. 4.3(c). Refer back to Fig. 4.2, we thus obtain a carrier conduction channel along the y axis between the n^+ -source and n^+ -drain, where originally there is no electrons at zero gate bias $V_G = 0$. A conduction current I_D will be induced when a drain bias V_D is applied (the source is normally grounded $V_S = 0$).

The revolution of the MOSFET device is that we are able to control the current between the source and the drain by the gate. The gate modifies the device through its electric field at the SiO₂-Si interface, thus the term “field-effect”, while the word “transistor” was formed by two words “transfer” and “resistor”, meaning that the resistance is now transferable. Putting a p^+ source and a p^+ drain in an n -type substrate we obtain the so-called p -channel MOSFET. Complimentary configuring the n -MOSFET and p -MOSFET we obtain the CMOS device, see Fig. 4.4, on which the current digital world is based. The whole device is fabricated on an n -type substrate. The left side is processed into a p -MOSFET, while the right side is first p -type doped into a p -type region, then processed into an n -type MOSFET. Wiring various electrodes as in Fig. 4.4 and setting V_{SS} high, the CMOS device becomes an inverter such that when the input V_i is high, the output V_o is low, while when V_i is low, $V_o =$ high.

When $V_i = \text{high}$, $G_1 = G_2 = \text{high}$, SiO₂-Si interface in the n -MOSFET becomes inverted (the SiO₂-Si interface in the p -MOSFET remains n -type), so that D_1 is conducting with S_1 . The situation of the n -MOSFET is called “on” and the p -MOSFET is “off”. Since S_1 is grounded thus low, D_1 and therefore $V_o = \text{low}$. When $V_i = \text{low}$, the p -MOSFET becomes “on” while the n -MOSFET is “off”, resulting in $V_o = S_2 = V_{SS} = \text{high}$.

Most of the electronics nanotechnology in the past 50 years is about scaling down the size of the CMOS device. One basic driving force of scaling down is to increase the density of CMOS devices in order to enhance the functions of the electronics, while the second most important driving force is to reduce the current in order to minimize the power consumption of the electronics. The process of scaling down has been continuous and the devices are already so small that the effective area of the devices where the electrons transfer signals is in the order of tens nm.

4.2 Semiclassical vs Quantum Considerations

To accurately analyze a semiconductor structure which is intended as a self-contained device under various operating conditions, a mathematical model has to be given. The equations which form this mathematical model are commonly called the basic semiconductor equations.

For the well-established semiclassical particle picture, the basic semiconductor equations consist of Poisson’s equation, the continuity equations for electrons and holes and the current relations for electrons and holes. For some applications it is desired to add to this set the heat flow equation [1].

- Poisson’s equation

$$\nabla \cdot (\epsilon \nabla \phi) = -\frac{e}{\epsilon_0} (n - p - N_A + N_D) \quad (4.10)$$

- Continuity equations

$$\begin{aligned} \nabla \cdot \mathbf{J}_n - e \frac{\partial n}{\partial t} &= eR \\ \nabla \cdot \mathbf{J}_p + e \frac{\partial p}{\partial t} &= -eR \end{aligned} \quad (4.11)$$

- Carrier transport equations

$$\begin{aligned} \mathbf{J}_n &= en\mu_n \mathbf{E} + eD_n \nabla n \\ \mathbf{J}_p &= ep\mu_p \mathbf{E} - eD_p \nabla p \end{aligned} \quad (4.12)$$

with the Einstein relation

$$D = \frac{k_B T}{e} \mu \quad (4.13)$$

Table 4.2 Physical parameters in the basic semiconductor device equations

n, p	carrier concentration
$\mathbf{J}_n, \mathbf{J}_p$	current density
ϕ, \mathbf{E}	electric potential and field
N_A, N_D	doping concentration
R	net generation rate/recombination rate
μ_n, μ_p	carrier mobility
D_n, D_p	diffusion coefficient
ϵ	dielectric constant
H	thermal generation
c_p	specific heat capacity
ρ	mass density
k	thermal conductivity

and

$$\mathbf{E} = -\nabla\phi \quad (4.14)$$

is the electric field.

- Heat flow equation

$$c_p\rho\frac{\partial T}{\partial t} - H = \nabla \cdot (k\nabla T) \quad (4.15)$$

The physical parameters used in the basic semiconductor equations are listed in Table 4.2.

The above basic equations are based on the concept of a probability distribution function $f(\mathbf{r}, \mathbf{p}, t)$ well defined over the phase space (both the \mathbf{r} - and \mathbf{k} -space) of the carriers. It is however limited when discussing wave transport in the submicron region where a phase-space distribution function becomes inadequate:

1. Due to the uncertainty principle

$$\Delta\mathbf{r} \cdot \Delta\mathbf{p} \geq \hbar$$

$f(\mathbf{r}, \mathbf{p}, t)$ does not have a precise meaning as a probability function in both the position (\mathbf{r}) and momentum (\mathbf{p}) spaces.

2. Semiclassical picture treats carriers as essentially free particles which are only scattered at specific spatial locations by phonons, impurities, imperfections etc. but are otherwise free between collisions. It means that the scattering takes place locally in both the space and the time domains. Quantum mechanically however, scattering potentials are extended in the space and scattering processes take a finite amount of time to complete.

Table 4.3 Issues relevant to quantum wave transport in semiconductors

Strong electric and magnetic fields
Periodic crystal potential; energy band structure effects
Carrier-carrier scattering
Pauli exclusion principle
Inhomogeneous electric fields
Energy dissipation, charge extraction and injection
Band-engineered nanostructures (potential barrier and quantum well)
Dynamic screening
Tunneling
Transient effects
Scattering from defects, phonons
Temperature dependence
Many-body effects

3. External force F should be essentially constant over the width of the electron wave packet so that the acceleration theorem

$$\hbar \dot{k} = F$$

become justified.

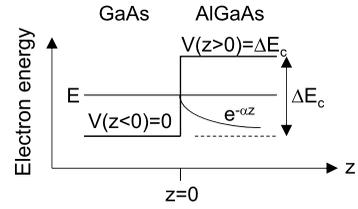
4. Correlations among successive scattering events become important when scattering centers are densely populated.
5. Carrier-carrier interaction becomes significant for degenerate electron systems.
6. It is really a Thomas-Fermi approximation when interpreting $f(\mathbf{r}, \mathbf{p}, t)$ as a probability density in position and momentum spaces. It is however grossly inaccurate when describing critical doping density of semiconductor-metal (Mott) transition in multiple-valley semiconductors. Furthermore, low-temperature conductivity of degenerately doped semiconductors, with highly anisotropic energy band structures (e.g., Si and Ge), can be understood only through the use of dielectric screening, as the use of Thomas-Fermi screening significantly underestimates the screening effect.

Hence, quantum mechanical considerations must be involved when discussing electron systems whose geometric sizes are comparable to the wavelengths of electrons. The most important issues relevant to quantum transport are listed in Table 4.3.

4.3 Resonant Tunneling Diode

Consider a hetero-interface structure along the z axis such as GaAs when $z \leq 0$ and AlGaAs when $z > 0$, see Fig. 4.5. When we take the conduction bandedge of GaAs

Fig. 4.5 GaAs/AlGaAs hetero interface



as potential reference $V(z \leq 0) = 0$, the conduction bandedge of the AlGaAs side is characterized as a potential barrier with a potential energy $V(z > 0) = \Delta E_c$, which is the conduction band offset between GaAs and AlGaAs. If the total energy, E , of an electron is less than ΔE_c , the Schrödinger equation of the electron in AlGaAs

$$\left(-\frac{\hbar^2}{2m_c^*} \frac{d^2}{dz^2} + \Delta E_c \right) \psi(z) = E \psi(z) \quad (4.16)$$

becomes

$$\frac{\hbar^2}{2m_c^*} \frac{d^2 \psi(z)}{dz^2} = (\Delta E_c - E) \psi(z) \quad (4.17)$$

Here m_c^* is the electron effective mass and ΔE_c is constant for $z > 0$. The electron wave function $\psi(z)$ varies as $e^{-\alpha z}$ with

$$\alpha = \sqrt{\frac{2m_c^*(\Delta E_c - E)}{\hbar^2}}$$

(The general solution is $\psi(z) = Ae^{-\alpha z} + Be^{\alpha z}$ mathematically, while we have to let $B = 0$ so that $\psi(z)$ remains finite in the whole region of $z > 0$.) The momentum in this case is clearly undefined, i.e., it is imaginary for $z > 0$, and there is a finite probability of the electron wave existing on the ‘wrong’ side ($z > 0$) of the barrier, this probability decreasing with the penetration into the potential barrier. Such a wave is termed *evanescent*.

Evanescent waves are the key to the quantum mechanical tunneling effect. With very thin potential barriers there is a finite probability of an electron appearing on the far side of the barrier, although on a classical argument the potential barrier is too high for any penetration to occur. This is the basis for the tunnel process in the tunnel diode. Electrons can tunnel through thin insulating films between metals and can also tunnel across the bandgap at high enough electric field. This latter process is the Zener effect.

The concept of tunneling diodes goes back several decades, and was first implemented in very heavily doped $p - n$ junction diodes [2]. In this case the tunneling is through the forbidden gap region, and involves electrons making transitions from the conduction band to the valence band, and vice versa. In the present context, however, we can separate two GaAs regions by a thin barrier region of AlGaAs, and the tunnel barrier is formed by the conduction (and valence) band discontinuity at the AlGaAs/GaAs heterointerfaces.

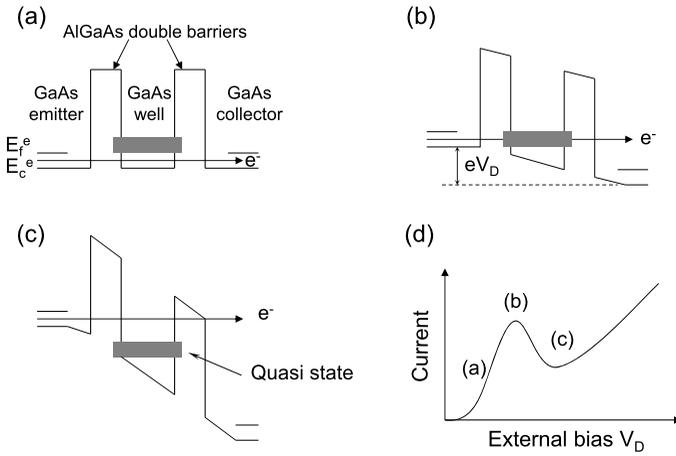


Fig. 4.6 Structure and carrier transport of an AlGaAs/GaAs double barrier resonant tunneling diode. (a) There is no net current at zero external bias. (b) Properly biased so that electrons from the emitter will tunnel through the device via the quasi state in the central GaAs well to empty states in the collector. (c) Off resonance and the carrier transport is low. (d) $I - V$ characteristics shows a resonant tunneling current peak

The tunneling of electrons through heterojunction barriers is the basic feature of many new quantum electronic devices [3–7]. Most of these devices can be categorized into two broad areas. The first ones are three-terminal devices and use the barriers associated with heterojunctions as a means of injecting fast and hot carriers into a narrow base layer. Ideally the scattering rate in the base should be small so that the electrons can traverse a collector barrier (also a heterojunction) with a reasonably high efficiency. An example of such a device is the tunneling hot electron transistor [5].

The second type of devices are based on “resonant” tunneling, involving either a superlattice structure [4] or double barrier quantum well [5]. These are usually two-terminal devices, though several three-terminal transistor variants have been proposed [6, 7]. Figure 4.6(a) shows a typical n -type AlGaAs/GaAs double-barrier resonant-tunneling diode. The GaAs emitter and collector are n -type doped which are characterized by Fermi level E_f^e in the emitter and E_c^e in the collector. At zero bias, the two Fermi levels are aligned so that there is no net current. The central GaAs well is about 10 nm thick sandwiched between two AlGaAs barriers. Because of the finite thickness and the surrounding barriers, electron states in the central GaAs well become quantized, see the shaded thick line which denotes the ground state. Note that the quantized states here are not infinitely narrow in energy because of the finite barrier thickness. They are normally referred to be quasi states. By applying an external bias, see Fig. 4.6(b), in such a condition that the quasi state is aligned between E_f^e and E_c^e , electrons in the emitter will tunnel to the quasi state in the central GaAs well through the evanescent state of the first AlGaAs barrier (provided that the AlGaAs barrier is not too thick), then further tunnel to the empty

states in the collector, forming the tunneling current. Further increasing V_D pushes the quasi state below E_c^e , see Fig. 4.6(b), so that there will be no quasi state available in the central GaAs well to receive the electron from the emitter, carrier transport through the device will be low. Increasing V_D furthermore will push the second AlGaAs barrier below E_c^e so that the current increases. A current peak exists as a function of V_D , which leads to a negative differential conductivity between (b) and (c), see Fig. 4.6(d). Such a negative differential conductivity has considerable potential as millimeter-wave devices as well as functional devices for digital and optical applications.

Four important experimental results should be mentioned:

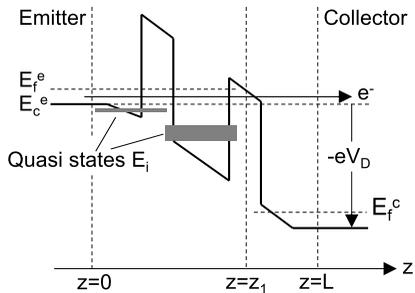
1. Photoluminescence and photoluminescence excitation spectra indicate the existence of quasi states in the central quantum well of the resonant tunneling diode [8]. This is a support of Luryi's opinion [9] that the resonance observed in the resonant tunneling diode can be explained as the tunneling of electrons from the three dimensional states in the emitter to the two dimensional quasi states in the well.
2. The dc $I - V$ characteristic measurements show not only the principal resonant peak as predicted by the theory, but also an oscillation, a bistability and a shoulder peak in the negative differential resistance region [10, 11]. These results are interpreted by Sollner as results of dynamic properties in the measurement circuits [12], while other explanations remain.
3. An experiment shows a peak in the capacitance-voltage characteristic at the resonant tunneling diode's resonance state [13].
4. The ac measurement results indicate that the charge transit time through the resonant tunneling diode is in the order of $t = 60$ fs and the resonance strength decreases when the ac frequency increases [10, 14].

4.3.1 $I - V$ Characteristics at Steady State

Refer to Fig. 4.7, we assume that the tunneling diode is positioned along the z direction. The left side $z < 0$ is denoted as the carrier emitter and the right side $z \geq L$ as the carrier collector. Let N_D be the Si concentration in the GaAs regions on the left side of the left AlGaAs barrier and on the right side of the right AlGaAs barrier (thus n -type doping), while the two AlGaAs barriers and the central GaAs well are undoped. The high doping levels in the electrodes are designed to provide free electrons, while the low-doping level in the central part is aimed to minimize various scattering processes that perturb the electron transport from the emitter to the collector.

Because of the barriers, the tunneling current via evanescent waves is normally rather low so that we assume that the electrons in the regions of the emitter and the collector that are far away from the AlGaAs barriers and GaAs well, i.e., $z < 0$ and $z > L$, see Fig. 4.7, are at equilibrium with their donors. Note that we still do not know exactly where $z = 0$ and $z = L$ are, which are to be determined.

Fig. 4.7 Energy band diagram of double-barrier resonant-tunneling diode under bias V_D



The emitter and collector are highly doped so that their resistances are small, while the AlGaAs barriers and the central GaAs well are not doped so are highly resistive. Grounding the emitter and biasing the collector with V_D , the bias will be largely applied in the resistive region, i.e., $0 \leq z \leq L$, see Fig. 4.7, between the emitter and the collector, because of the Ohm's law. Thus, $\phi = 0$ for $z < 0$ and $\phi = V_D$ for $z > L$.

For the bulk GaAs of $z < 0$, the energy of one electron at state \mathbf{k} is

$$E(\mathbf{k}) = E_c^e + \frac{\hbar^2 k^2}{2m_c^*} \quad (4.18)$$

where E_c^e is the conduction bandedge of the emitter (superscript e denotes "emitter").

In general, the total energy of the electron in a bulk material is

$$E = E(\mathbf{k}) - e\phi$$

after including the Coulombic potential energy ϕ . The electron concentration is given by

$$n = \int \frac{1}{1 + e^{[E(\mathbf{k}) - e\phi - E_f]/k_B T}} \frac{2d\mathbf{k}}{(2\pi)^3} = \frac{1}{2\pi} \left(\frac{m_c^* k_B T}{\hbar^2} \right)^{3/2} F_{1/2} \left(\frac{e\phi + E_f}{k_B T} \right) \quad (4.19)$$

where E_f is the Fermi energy and $F_{1/2}$ is the Fermi-Dirac integral of order 1/2

$$F_{1/2}(\eta) = \int_0^\infty \frac{x^{1/2} dx}{1 + e^{x-\eta}} \quad (4.20)$$

E_f can be determined by letting $n = N_D$ at equilibrium.

By Eq. (4.19), we can determine the Fermi level and conduction bandedge E_f^e and E_c^e in the emitter, and $E_f^c - eV_D$ and $E_c^c - eV_D$ in the collector (superscript c denotes "collector"), see Fig. 4.7.

We now determine the Coulombic potential $\phi(z)$ in the region of $0 \leq z \leq L$ by solving the Poisson equation

$$\frac{d}{dz} \left[\epsilon(z) \frac{d\phi(z)}{dz} \right] = -e[n(z) - N_D(z)] \quad (4.21)$$

In the above equation, $\epsilon(z)$ is known when we know the geometric structure of the diode. $N_D(z)$ is also known. And the boundary conditions for $\phi(z)$ are: $\phi(z)|_{z < 0} = 0$ and $\phi(z)|_{z > L} = V_D$. $n(z)$ is the electron distribution which is unknown for the moment.

Including the Coulombic potential, the Schrödinger equation in the effective-mass approximation has now the form

$$\left[-\frac{\hbar^2 \nabla^2}{2m_c^*} + E_c(z) - e\phi(z) \right] \psi(\mathbf{r}) = E\psi(\mathbf{r}) \quad (4.22)$$

Here $E_c(z)$ describes the spatial variation of the conduction bandedge of the diode, which is zero when we set the GaAs conduction bandedge as the potential reference, and ΔE_c the conduction band offset between AlGaAs and GaAs, see Fig. 4.5.

Since the resonant tunneling diode is one-dimensional along the z direction while extended in the xy plane,

$$\begin{aligned} \psi(\mathbf{r}) &= e^{i(k_x x + k_y y)} \psi(z) \\ E &= \frac{\hbar^2(k_x^2 + k_y^2)}{2m_c^*} + E_z + E_c \\ \left[-\frac{\hbar^2}{2m_c^*} \frac{\partial^2}{\partial z^2} + E_c(z) - e\phi(z) \right] \psi(z) &= E_z \psi(z) \end{aligned} \quad (4.23)$$

Refer to Fig. 4.7, there are electron states in the emitter ($z < 0$) above E_c , which are occupied by electrons according to the Fermi distribution. These electrons are three dimensional

$$\psi_{k_x k_y k_z}^e(x, y, z) = e^{i(k_x x + k_y y + k_z z)}, \quad E^e(k_x, k_y, k_z) = E_c + \frac{\hbar^2(k_x^2 + k_y^2 + k_z^2)}{2m^*} \quad (4.24)$$

which are two-fold degenerate in terms of k_z . The one with $k_z > 0$ will travel from the emitter to the central region. It will be partially reflected by the double barriers and partially tunnel through the central region. Such an electron transport is quantitatively determined by the Schrödinger equation Eq. (4.23). Since the potential energy of the electrons is x and y independent, k_x and k_y will be conserved during the transport and we obtain the following simple boundary conditions for the last equation of Eqs. (4.23)

$$\psi_{k_z}^e(z) = \begin{cases} e^{ik_z z} + r_{k_z} e^{-ik_z z}, & z < 0 \\ t_{k_z} e^{iq_z z}, & z > L \end{cases} \quad (4.25)$$

where r_{k_z} and t_{k_z} are the amplitudes of reflected and transmitted waves. For coherent tunneling, the total energy of the electron is conserved so that

$$\frac{\hbar^2 k_z^2}{2m^*} = \frac{\hbar^2 q_z^2}{2m^*} - eV_D$$

Electron state $\psi_{k_z}^c(z)$ in the collector is determined in the similar way.

The spatial distribution of all these electrons is

$$n'(x, y, z) = \int \left\{ \frac{|\psi_{k_x k_y k_z}^e(x, y, z)|^2}{1 + e^{[E^e(k) - E_f^e]/k_B T}} + \frac{|\psi_{k_x k_y k_z}^c(x, y, z)|^2}{1 + e^{[E^c(k) - E_f^c + eV_D]/k_B T}} \right\} \frac{2d\mathbf{k}}{(2\pi)^3} \quad (4.26)$$

Since $|\psi_{k_x k_y k_z}^e(x, y, z)|^2$ and $|\psi_{k_x k_y k_z}^c(x, y, z)|^2$ are k_x and k_y independent, see the first equation of Eqs. (4.23), the integrations over k_x and k_y in the above equation can be carried out analytically and we obtain $n'(x, y, z) = n'(z)$, i.e., the electron distribution is x and y independent, and

$$n'(z) = \frac{m_c^* k_B T}{2\pi \hbar^2} \int \left\{ |\psi_{k_z}^e(z)|^2 \ln [1 + e^{[E_f^e - E_c^e - E^e(k_z)]/k_B T}] \right. \\ \left. + |\psi_{k_z}^c(z)|^2 \ln [1 + e^{[E_f^c - E_c^c - E^c(k_z)]/k_B T}] \right\} \frac{2dk_z}{2\pi} \quad (4.27)$$

since

$$\int \frac{1}{1 + \eta e^{\alpha(k_x^2 + k_y^2)}} \frac{dk_x dk_y}{(2\pi)^2} = \frac{1}{4\pi\alpha} \ln(1 + \eta^{-1}) \quad (4.28)$$

Here

$$\frac{m_c^* k_B T}{2\pi \hbar^2} \ln [1 + e^{[E_f^e - E_c^e - E^e(k_z)]/k_B T}]$$

is normally referred to as the sheet density of state $E^e(k_z)$.

We notice that there are a few quasi states which are confined in the triangle quantum well before the left barrier and in the quantum well between the two barriers, i.e., E_i in Fig. 4.7. These states are definitely confined in the emitter since their energies are below E_c^e so that their wave functions are decaying, see Fig. 4.5. On the collector side, they are open if we assume that their wave functions can extend coherently all over the space. However, as we mentioned before, the barriers are normally thick so that the transmission through the tunneling structure is low. If electrons from the emitter happen to occupy these states, they will stay there long time before tunneling through the barriers to reach the collector. We can thus assume that electrons occupying E_i are at equilibrium with the electrons in the emitter. Their contribution to the total electron distribution is

$$n_s(z) = \frac{m^* k_B T}{2\pi \hbar^2} \sum_i |\psi_i(z)|^2 \ln [1 + e^{(E_f^e - E_i)/k_B T}] \quad (4.29)$$

Table 4.4 Tunneling diode structures. Layer 4, 5, 6, 7, and 8 are unintentionally doped

	Material	Thickness [nm]	Doping [10^{18} cm^{-3}]
1	GaAs	100	2.0
2	GaAs	80	0.2
3	GaAs	50	0.02
4	GaAs spacer	20	
5	AlAs barrier	3	
6	GaAs quantum well	8	
7	AlAs barrier	3	
8	GaAs spacer	20	
9	GaAs	50	0.02
10	GaAs	80	0.2
11	GaAs	300	2.0
12	AlAs etch-stop layer	15	
13	GaAs buffer	400	
14	GaAs substrate		

The derivation of the above equation is very similar to Eq. (4.27) when we realize that $2k_z/2\pi$ is actually the density of states along the z direction for extended states k_z .

In a brief summary, the theoretical consideration consists of: We start with a guess of the potential distribution $\phi(z)$ under V_D such as V_D is linearly applied between $z = 0$ and $z = L$. The Schrödinger equation of Eqs. (4.23) is then defined from which we find $\psi_{k_z}^e$, $\psi_{k_z}^c$ and ψ_i . After that, the total electron distribution $n(z) = n'(z) + n_s(z)$ are obtained from Eqs. (4.27), (4.29), which will be inserted into the Poisson equation Eq. (4.21) for an updated potential $\phi(z)$. The process then repeats until self consistency is reached.

The definition, or rather, the determination of $z = 0$ and $z = L$ is quite arbitrary at the very beginning of the self-consistent calculation. It is only certain that $z = 0$ and $z = L$ should be deep enough in the emitter and collector. The basic method is thus only trial by error. Try a narrow central region (between $z = 0$ and $z = L$) then expand the region and compare the two results. A proper set of $z = 0$ and $z = L$ is reached when the two results converge.

We now apply the above equations to study a n -type AlGaAs/GaAs double barrier resonant tunneling diode schematically shown in Fig. 4.6. The double barrier layer structures were grown by molecular beam epitaxy (MBE) on semi-insulating GaAs substrates and nominal material layer structure is listed in Table 4.4 [15]. Typical $I - V$ characteristics at 77 and 300 K is shown in Fig. 4.8, which clearly demonstrates the resonant tunneling effect.

Theoretical expectation of the $I - V$ characteristics is shown in Fig. 4.9, where one tunneling current peak is observed at an external bias of 0.28 V, corresponding to the zero-bias resonant state at 0.12 eV in the tunneling probability $|t_{k_z}|^2$ spectrum as a function of the electron kinetic energy $\hbar^2 k_z^2 / 2m_c^*$ shown in the inset.

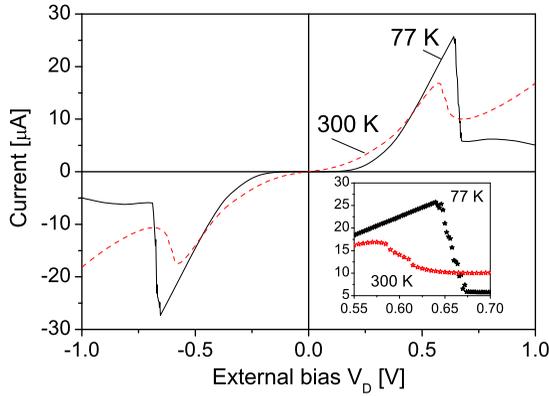


Fig. 4.8 $I - V$ characteristics of a GaAs/AlAs tunneling diode at 77 K (solid line) and 300 K (dashed line). Inset shows details of the resonant tunneling current (Reprinted with permission from Y. Hou, W.-P. Wang, N. Li, W. Lu, and Y. Fu, Effects of series and parallel resistances on the current-voltage characteristics of small-area air-bridge resonant tunneling diode, J. Appl. Phys., vol. 104, p. 074508(5), 2008. Copyright 2008, American Institute of Physics)

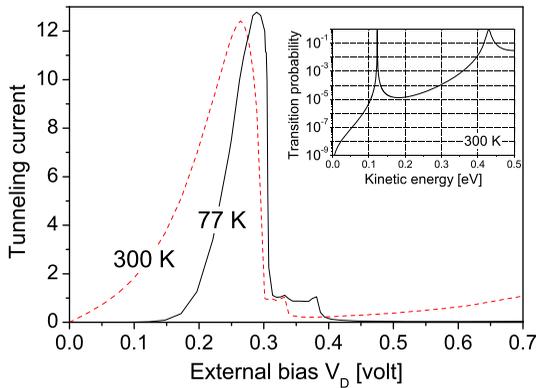


Fig. 4.9 Theoretical $I - V$ characteristics at 77 K (solid line) and 300 K (dashed line). Inset shows the tunneling probability $|t_{k_z}|^2$ as a function of the electron kinetic energy $\hbar^2 k_z^2 / 2m_c^*$, see Eq. (4.25) (Reprinted with permission from Y. Hou, W.-P. Wang, N. Li, W. Lu, and Y. Fu, Effects of series and parallel resistances on the current-voltage characteristics of small-area air-bridge resonant tunneling diode, J. Appl. Phys., vol. 104, p. 074508(5), 2008. Copyright 2008, American Institute of Physics)

We first have a close look at the tunneling probability spectrum which shows two tunneling peaks. By referring to the transport mechanism of Fig. 4.6 it can be realized that these two tunneling peaks correspond to two quasi states in the central GaAs well, one at 0.12 eV and the other at 0.42 eV. The tunneling peaks are very narrow, especially the peak at 0.12 eV. In other words, only those electrons in the emitter with a certain k_z which matches with the quasi state in the central GaAs

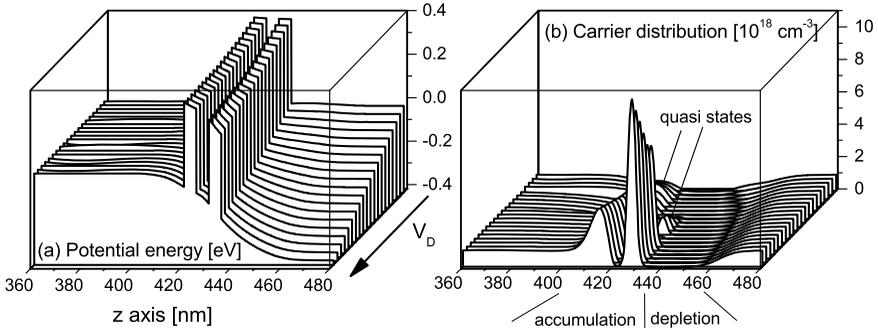


Fig. 4.10 (a) Potential energy profile and (b) carrier distribution as functions of external bias V_D at 77 K. The arrow indicates the linear increase of the external bias V_D from 0.0 to 0.44 V (Reprinted with permission from Y. Hou, W.-P. Wang, N. Li, W. Lu, and Y. Fu, Effects of series and parallel resistances on the current-voltage characteristics of small-area air-bridge resonant tunneling diode, *J. Appl. Phys.*, vol. 104, p. 074508(5), 2008. Copyright 2008, American Institute of Physics)

well can tunnel to the collector. All other electrons will be reflected. Thus, the total tunneling current is very weak. This agrees with what we have pre-assumed that the electrons in the emitter are at equilibrium state which can be described by Fermi level E_f^e . Figure 4.10 shows the numerical expectations of the potential energy and carrier distribution of the tunneling diode as functions of the external bias, which all agree well with what we expected by Fig. 4.6. Here we observe the lowering of the GaAs well by V_D in Fig. 4.10(a), the filling of the quasi states in the triangle quantum well on the left side of the first AlAs barrier and in the GaAs well, which is normally referred to as the accumulation. We also see the depletion on the right side of the second barrier in the collector. The carrier accumulation in the quasi states is reflected in the $I - V$ characteristics as a shoulder on the right side of the tunneling peak in Fig. 4.9.

Discrepancies between experimental Fig. 4.8 and theoretical expectations in Fig. 4.9 exist. Most prominent difference is the voltage of the current peak, which is 0.63 V while theoretically it is only 0.28 V, at 77 K. Another significant difference is the high peak-to-valley ratio predicted by the theory as compared with the experimental one. The first difference can be explained by a series resistance, mostly in the circuit such as contacts, so that only a portion of the total experimentally applied bias is distributed to the active central region of $0 < z < L$ in Fig. 4.10. The second difference can be explained by a parallel resistance, such as edges of the tunneling diode since in reality the extension in the xy plane can never be infinite. Moreover, the theoretical model assumes coherent tunneling, while in reality, the tunneling electrons are under the influences of many energy relaxation scatterings including defects at GaAs/AlAs hetero interfaces, residual dopants in layers 4 to 8 (see Table 4.4).

The series resistance has a much profound effect. Figure 4.11 shows the $I - V$ characteristics of another tunneling diode, which has the same nominal device design structure as in Table 4.4 but fabricated in another process. As compared with

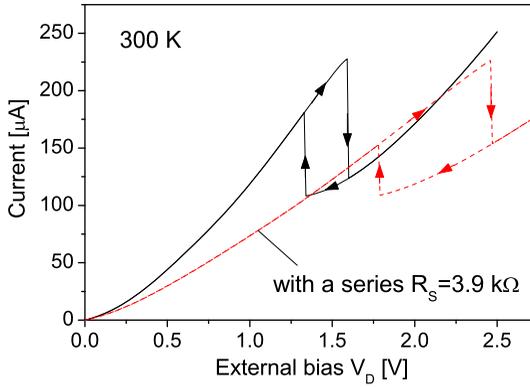


Fig. 4.11 $I - V$ characteristics of another tunneling diode at 300 K. *Arrows* indicate the changing directions of the external bias. *Dashed lines* are the $I - V$ characteristics when the tunneling diode was in series connection with an external resistance $R_S = 3.9 \text{ k}\Omega$ (Reprinted with permission from Y. Hou, W.-P. Wang, N. Li, W. Lu, and Y. Fu, Effects of series and parallel resistances on the current-voltage characteristics of small-area air-bridge resonant tunneling diode, *J. Appl. Phys.*, vol. 104, p. 074508(5), 2008. Copyright 2008, American Institute of Physics)

the tunneling diode of Fig. 4.8, the current through the latter diode is about one order of magnitude higher. Furthermore, the latter diode shows a clear bistability effect when we increase the external bias V_D from 0.0 to 3.0 V then decrease from 3.0 to 0.0 V.

By denoting $I(V) - V$ as the original $I - V$ characteristics of the tunneling diode without any external series resistor, we can expect theoretically the $I(V^*) - V^*$ curve when the tunneling diode is in series connection with a series resistor R_S

$$V^* = V + I(V) \times R_S \tag{4.30}$$

This describes exactly the relationship between the two sets of measurement data in Fig. 4.11. Here we can see that the external series resistance shifts the current-peak bias to a higher value, it also widens the bistability loop.

4.3.2 Response to a Time-Dependent Perturbation

We now study the effect of a small external ac bias V_{ac} on the carrier conduction through the double-barrier resonant-tunneling diode. Using the effective-mass approximation, the Schrödinger equation for a one-dimensional electron in a time-dependent ac electric field is given by:

$$i\hbar \frac{\partial \psi(z, t)}{\partial t} = \left[-\frac{\hbar^2}{2m_c^*} \frac{\partial^2}{\partial z^2} + E_c - e\phi(z) + V_{ac}e^{-i\omega t} + V_{ac}^*e^{i\omega t} \right] \psi(z, t) \tag{4.31}$$

where E_c is the conduction bandedge, $\phi(z)$ is the Coulombic potential originated from carrier distribution, dopants and a external dc bias V_D , see Eq. (4.21). V_{ac} is the amplitude of the small ac bias at frequency of ω . Here we write the ac perturbation as $V_{ac}e^{-i\omega t} + V_{ac}^*e^{i\omega t}$ to make sure that the ac perturbation is real.

As in the previous section, we assume that the energy band diagrams and thus the eigen wave functions in the emitter and collector are not affected by the ac perturbation since the doping levels there are high. Moreover, since V_{ac} is a perturbation, it is expected that the conduction band structure under the bias of V_D is not affected by V_{ac} . We use the dc conduction band structure of Fig. 4.10(a) to calculate the effect of the ac perturbation on the current conduction.

The electron state at E in the region between $z = 0$ and $z = L$ is modified by the ac perturbation. Between $z = 0$ and $z = L$, we denote the eigen solution of electron state E_i without the ac perturbation as

$$\left[-\frac{\hbar^2}{2m_c^*} \frac{\partial^2}{\partial z^2} + E_c - e\phi(z) \right] \psi_i^0(z) = E_i \psi_i^0(z) \quad (4.32)$$

the wave function of the electron state under the ac perturbation is written as

$$\psi(z, t) = \sum_i C_i(t) e^{-iE_i t/\hbar} \psi_i^0(z) \quad (4.33)$$

By the first-order perturbation it is easy to obtain

$$i\hbar \frac{dC_i(t)}{dt} = (V_{ac}e^{-i\omega t} + V_{ac}^*e^{i\omega t}) C_i(t) \quad (4.34)$$

which results in

$$C_i(t) = \exp\left(\frac{V_{ac}e^{-i\omega t}}{\hbar\omega} - \frac{V_{ac}^*e^{i\omega t}}{\hbar\omega}\right) \quad (4.35)$$

Here we see that the ac perturbation does not mix up eigen state E_i with other eigen states. It only modify its wave function to

$$\exp\left(\frac{V_{ac}e^{-i\omega t}}{\omega} - \frac{V_{ac}^*e^{i\omega t}}{\omega}\right) e^{-iE_i t/\hbar} \psi_i^0(z) \quad (4.36)$$

In order to simplify notations, we drop the subscript “ i ” on E_i and write ψ_i^0 as ψ_E in Eq. (4.32).

We consider the electron in the form of plane wave

$$e^{i(kz - Et/\hbar)}$$

injected into the active region from the emitter, where

$$E = \frac{\hbar^2 k^2}{2m_c^*}$$

is the electron kinetic energy. The electron wave is partially reflected from the active layer and partially transmitted through the active layer, similar to Eq. (4.25). The general form of the wave function, in the emitter $z < 0$, the active region $0 < z < L$, and collector $z > L$, is

$$\psi_e(z, t) = \begin{cases} e^{i(kz - Et/\hbar)} + \sum_n B_n e^{-i[k_n z + (E + n\hbar\omega)t/\hbar]} & z < 0 \\ \sum_n C_n \exp\left(\frac{V_{ac} e^{-i\omega t}}{\omega} - \frac{V_{ac}^* e^{i\omega t}}{\omega}\right) \psi_{E+n\hbar\omega}(z) e^{-i(E+n\hbar\omega)t/\hbar} & 0 \leq z \leq L \\ \sum_n A_n e^{i[q_n z - (E+n\hbar\omega)t/\hbar]} & z > L \end{cases} \quad (4.37)$$

where n is an integer. Here k_n and q_n are wave vectors in the emitter and collector, respectively, corresponding to the total electron energy $E + n\hbar\omega$

$$\frac{\hbar^2 k_n^2}{2m_c^*} = E + n\hbar\omega, \quad \frac{\hbar^2 q_n^2}{2m_c^*} = E + n\hbar\omega + eV_D \quad (4.38)$$

We may expect that the ac perturbation couples all possible electron states, whereas only states $E + n\hbar\omega$ are explicitly involved in Eq. (4.37). The reason is as follows. We expand the modified wave function of Eq. (4.36)

$$\begin{aligned} & \exp\left(\frac{V_{ac} e^{-i\omega t}}{\omega} - \frac{V_{ac}^* e^{i\omega t}}{\omega}\right) \psi_{E+n\hbar\omega}(z) e^{-i(E+n\hbar\omega)t/\hbar} \\ &= \sum_{jk} \frac{(V_{ac}/\hbar\omega)^k}{k!} \cdot \frac{(-V_{ac}^*/\hbar\omega)^j}{j!} \psi_{E+n\hbar\omega}(z) e^{-i[E+(n+k-j)\hbar\omega]t/\hbar} \end{aligned} \quad (4.39)$$

where j and k are integers, which requires n being an integer so that C_n is nonzero.

Equation (4.37) is solved numerically to obtain the A_n . The total wave function of electron state E in the collector region is

$$\Psi = u_c \sum_n A_n e^{i[q_n z - (E+n\hbar\omega)t/\hbar]}$$

where u_c is the periodic Bloch function of the conduction band. By inserting the above expression into the probability flow density Eq. (2.10), the transmitted conduction current of such electron state E originated from the emitter is

$$i(E) = e \sum_n \frac{\hbar q_n}{m_c^*} |A_n|^2 \quad (4.40)$$

where e is the electron charge.

Figure 4.12(a) shows the calculated tunneling probabilities as functions of incident electron energy when $V_D = 0$. Here the tunneling diode consists of a 5.6-nm thick GaAs well sandwiched between two 8.5 nm thick $\text{Al}_{0.4}\text{Ga}_{0.6}\text{As}$ barriers. The dashed line in Fig. 4.12(a) is obtained for the case without the ac perturbation, while the solid line is calculated when $V_{ac} = 2.5$ meV and $\hbar\omega = 5$ meV. Here we clearly observe that V_{ac} couples the electron states $E \pm n\hbar\omega$. Side-bands are expected due

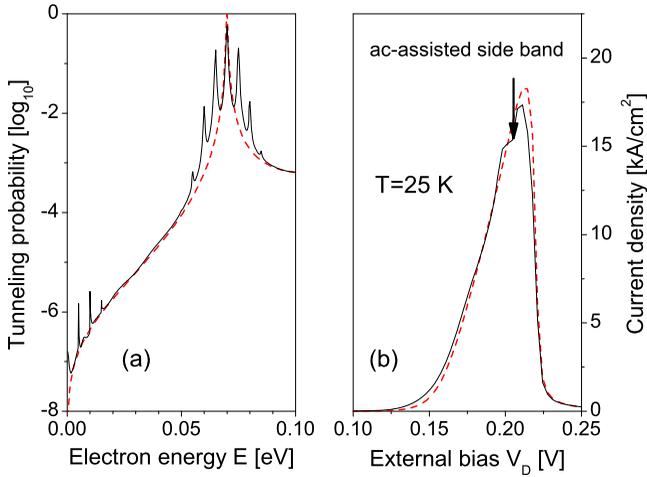


Fig. 4.12 (a) Calculated tunneling probabilities as functions of incident electron energy from the emitter. $V_D = 0$. (b) Calculated $I - V$ spectra at $T = 25$ K. *Dashed lines:* $V_{ac} = 0$; *solid lines:* $V_{ac} = 2.5$ meV and $\hbar\omega = 5$ meV (Reprinted with permission from Y. Fu and M. Willander, Response of a semiconductor tunneling structure to an ac perturbation, *J. Appl. Phys.*, vol. 72, pp. 3593–3597, 1992. Copyright 1992, American Institute of Physics)

to the ac-assisted tunneling, see Eq. (4.39). Also indicated in Fig. 4.12(a) is the reduced peak value of the tunneling probability.

The calculated $I - V$ spectra at $T = 25$ K are shown in Fig. 4.12(b). The ac-assisted tunneling side-bands in Fig. 4.12(a) are clearly reflected in Fig. 4.12(b). The current density at resonance is reduced while the resonance peak is broadened.

The current density at resonance is plotted in Fig. 4.13 as a function of V_{ac} and $\hbar\omega$. Since the perturbation approach is used here, V_{ac} should always be smaller than $\hbar\omega$ in the calculation so that the condition

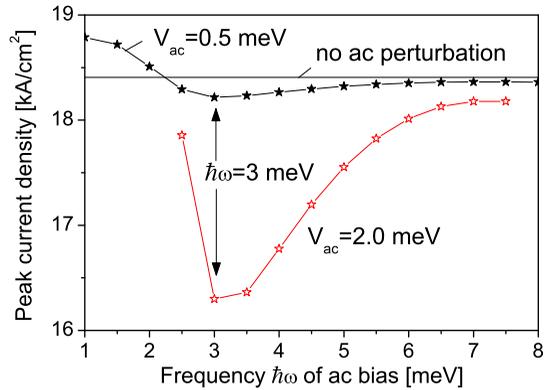
$$\frac{V_{ac}}{\hbar\omega} < 1$$

is always established, see Eq. (4.39)

Figure 4.13 indicates that the current density at resonance is larger than its dc value when V_{ac} and $\hbar\omega$ are quite small. This is easy to understand because of the ac-assisted tunneling side-bands in Fig. 4.12(a). When $\hbar\omega$ is large, the current decreases and is below its dc value. The current density reaches its lowest value in response to the ac perturbation when $\hbar\omega = 3$ meV. The larger the amplitude of the ac perturbation, the more is the current density lowered, but the frequency of the ac perturbation at which the current density reaches its lowest value is almost independent of the amplitude of the ac perturbation, i.e., always at $\hbar\omega = 3$ meV.

When $\hbar\omega$ is further increased, the current density at resonance gradually retrieves its dc value. By Fig. 4.13, when $\hbar\omega > 10$ meV, the current density at resonance resumes its dc value. This suggests that the response of the correlated wave functions

Fig. 4.13 Peak current density response as a function of the ac signal frequency $\hbar\omega$ (Reprinted with permission from Y. Fu and M. Willander, Response of a semiconductor tunneling structure to an ac perturbation, J. Appl. Phys., vol. 72, pp. 3593–3597, 1992. Copyright 1992, American Institute of Physics)



at $E + n\hbar\omega$ to the ac perturbation is time limited. When the frequency of the perturbation is low or comparable to the speed of the wave function response, the correlation of the wave functions can follow the temporal variation of the perturbation. When the ac perturbation frequency is too high, e.g., $\hbar\omega = 10$ meV for the present tunneling diode, the wave functions have difficulties to follow up with the temporal variation of the ac perturbation so that the current returns to its normal dc value.

Here we understand that the speed of the wave function response to the external perturbation is finite, as demonstrated experimentally [10, 14].

4.3.3 Phonon-Assisted Tunneling

Electron-phonon scattering is a major issue in semiconductor devices. Moreover, the phonon state becomes also confined in low-dimensional system so that the strength of the electron-phonon interaction is greatly affected, see Sect. 2.8.1. Here we shall study the effect of the phonon scattering on the tunneling current of a tunneling diode.

Using the effective-mass approximation and similar to the independent-boson model when describing the electron-phonon scattering, the Schrödinger equation for a one-dimensional electron state in the tunneling diode is given by:

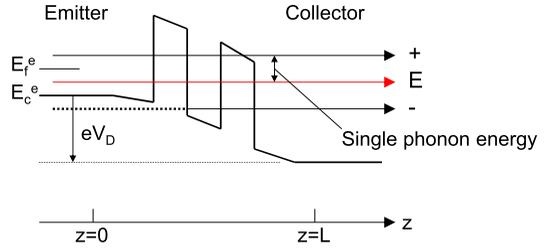
$$i\hbar \frac{\partial \psi(z, t)}{\partial t} = \left[-\frac{\hbar^2}{2m_c^*} \frac{\partial^2}{\partial z^2} + E_c - e\phi(z) + W e^{-i\omega t} + W^* e^{i\omega t} \right] \psi(z, t) \quad (4.41)$$

where

$$W = \sum_{\mathbf{q}} M(\mathbf{q}) a_{\mathbf{q}}$$

$M(\mathbf{q})$ is the electron-phonon scattering matrix and $a_{\mathbf{q}}$ is the phonon annihilation operator, $\hbar\omega$ is the phonon energy, and $\hbar\mathbf{q}$ is the phonon momentum.

Fig. 4.14 Schematic band diagram of a biased RTD and the three tunneling channels in a phonon field. Channel “-” is the evanescent tunneling channel which is below E_c^e



In the bulk semiconductor, the quantity W in Eq. (4.41) depends on the electron energy, see Sect. 2.8.1. The heterointerfaces in the tunneling diode modify substantially the phonon modes [16], so that W is further complicated and becomes dependent on the electron position. To avoid complications, here we assume that W is a constant, independent of the electron energy and position. Then Eq. (4.41) is solved in the similar way as we solved Eq. (4.31) in the last section.

Experimentally, only electrons whose energies are between the conduction band edge E_c^e and the Fermi level E_f^e in the emitter are available to transport from the emitter to the collector, see Fig. 4.6. The external voltage V_D is applied to lower the resonance state in the central quantum well to the position between E_c^e and E_f^e so that resonant tunneling processes can occur.

Note that the energy of optical phonons that are most important in electron-phonon scattering in commonly used semiconductors is about 30 meV, while for n -type GaAs doped at 10^{18} cm^{-3} , the Fermi level is about 100 meV above the conduction bandedge (see Fig. 6.1). Thus, even for highly doped semiconductors, there are a large portion of total electron states available in the emitter which does not have enough energy to emit a phonon in the emitter. On the other hand, since

$$E_c^e - E_c^c = eV_D$$

where E_c^c is the conduction bandedge of the collector, an electron can gain energy from the electric field V_D during its transport from the emitter to the collector. If

$$eV_D > \hbar\omega$$

the electron can emit a phonon when it gains enough energy from the electric field even though the phonon emission may not be possible in the emitter region.

In Fig. 4.14 we show three major tunneling channels (we neglect multiple phonon emission and absorption processes because of their low probabilities) where we take the Fermi level E_f^e into consideration, namely, the energy of the incident electron should be less than E_f^e at low temperature. Channel E is the principal one (assuming the conservations of the energy and momentum in the xy plane), the other two are one-phonon processes, absorption of a phonon (channel “+”, largely possible at high temperature) and emission of a phonon (channel “-”), respectively. The local Fermi level is $E_f^e = 18 \text{ meV}$ above the conduction bandedge when the doping level is $2 \times 10^{17} \text{ cm}^{-3}$ for a GaAs/AlGaAs tunneling diode. Since the phonon energy $\hbar\omega$ is 36.2 meV [17] in the GaAs material, the phonon-emission channel is always below E_c^e .

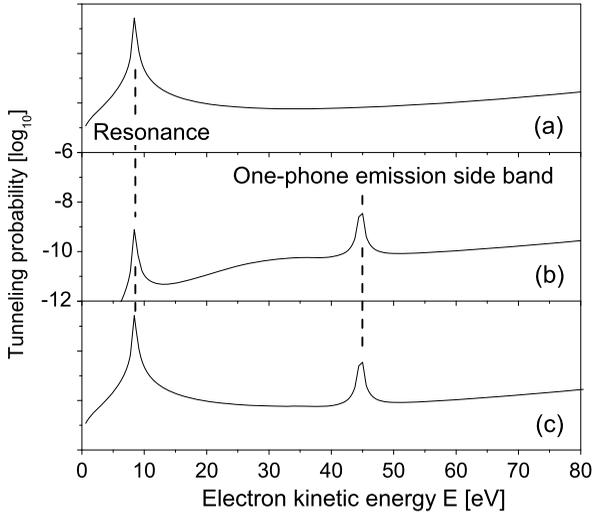


Fig. 4.15 Tunneling probability as function of electron kinetic energy. The sample is biased at $V_D = 0.12$ V at $T = 4.2$ K. **(a)** Without the phonon field; **(b)** With a fixed electron-phonon interaction g_{\max} ; **(c)** With a reduced electron-phonon interaction (Reprinted with permission from Y. Fu and M. Willander, *Evanescent channels in calculation of phonon-associated tunneling spectrum of a semiconductor tunneling structure*, *J. Appl. Phys.*, vol. 73, pp. 1848–1852, 1993. Copyright 1993, American Institute of Physics)

When a tunneling channel at E is below E_c^e , its corresponding wave function is bounded in the emitter ($z < 0$), namely it is an evanescent state having the form of $e^{\alpha z}$, where

$$\alpha = \sqrt{\frac{2m^*(E_c^e - E)}{\hbar^2}} \quad (4.42)$$

Let us now explicitly analyze a simplified situation by maintaining the perturbation expansion to the order of $|W|^2$. The wave function are expressed as

For $z < 0$,

$$(e^{ikz} + B_E e^{-ikz})e^{-iEt/\hbar} + B_+ e^{-i[k+z+(E+\hbar\omega)t/\hbar]} + B_- e^{-i[k-z+(E-\hbar\omega)t/\hbar]} \quad (4.43)$$

for $0 \leq z \leq L$, we refer to Eq. (4.39),

$$\left[\left(1 - \frac{|W|^2}{\hbar^2 \omega^2} \right) \psi_E + \frac{W}{\hbar \omega} \psi_- - \frac{W^*}{\hbar \omega} \psi_+ \right] e^{-iEt/\hbar} + \left(\frac{W}{\hbar \omega} \psi_E + \psi_+ \right) e^{-i(E+\hbar\omega)t/\hbar} + \left(-\frac{W^*}{\hbar \omega} \psi_E + \psi_- \right) e^{-i(E-\hbar\omega)t/\hbar} \quad (4.44)$$

for $z > L$,

$$A_E e^{i(qz - Et/\hbar)} + A_+ e^{i[q+z-(E+\hbar\omega)t/\hbar]} + A_- e^{i[q-z-(E-\hbar\omega)t/\hbar]} \quad (4.45)$$

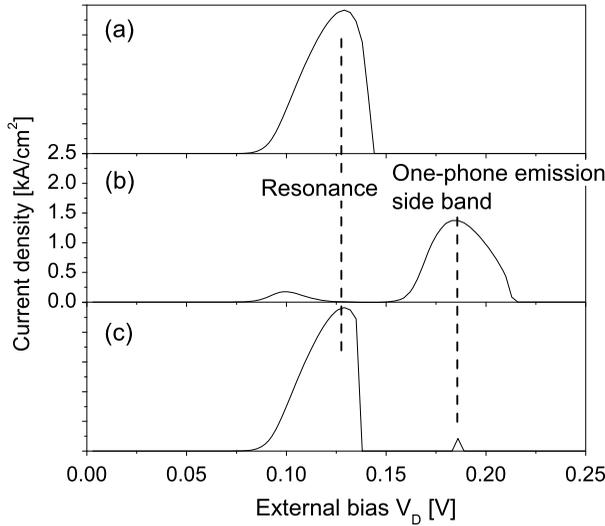


Fig. 4.16 Calculated tunneling current-voltage characteristics at $T = 4.2$ K. (a) Without the phonon field; (b) With a fixed electron-phonon interaction g_{\max} ; (c) With a reduced electron-phonon interaction (Reprinted with permission from Y. Fu and M. Willander, *Evanescent channels in calculation of phonon-associated tunneling spectrum of a semiconductor tunneling structure*, *J. Appl. Phys.*, vol. 73, pp. 1848–1852, 1993. Copyright 1993, American Institute of Physics)

where the subscript E , “+” and “-” indicate the tunneling channels of E , $E + \hbar\omega$ and $E - \hbar\omega$, respectively.

When $E - \hbar\omega < E_c^e$, k_- is imaginary and ψ_- as well as $B_-e^{-ik_-z}$ are evanescent, and therefore are expected to vanish in the emitter. Since the matching condition for channel “-” at $z = 0$ demands

$$-\frac{W^*}{\hbar\omega}\psi_E + \psi_- = B_-e^{-ik_-z}$$

which in its turn requires

$$\frac{W^*}{\hbar\omega}\psi_E$$

to be small. If the electron-phonon interaction W is finite, ψ_E must vanish. This means that B_E in Eq. (4.43) is close to 1. A large B_E implies a large reflection current, and the tunneling current is therefore small.

This phenomenon is clearly demonstrated in Fig. 4.15 which shows the tunneling probability as a function of incident electron energy. Without the phonon field, see Fig. 4.15(a), the tunneling probability at resonance is quite large compared with the off-resonance situations. When the phonon field is turned on, B_E increases so that the tunneling probability in the principal tunneling channel is drastically decreased, see Fig. 4.15(b). However, such result is unreasonable as compared with experiment

[17] and the theoretical work of Wingreen et al. [18], where it was shown that the tunneling resonance is still dominant even when there exists a phonon field.

Such a problem is caused by the assumption of a constant electron-phonon interaction, which is reasonable to treat the electron-phonon coupling between two electron states which are well above E_c^e . Let us take it as the upper limit g_{\max} for the interaction between an evanescent state which is below E_c^e and an unbounded state which is above E_c^e

$$g_{\max} = \sum_{\mathbf{q}} \left[\frac{|M(\mathbf{q})|}{\omega} \right]^2 \quad (4.46)$$

The specific value of g at certain incident electron energy is obtained as between 0 and g_{\max} in such a way that boundary conditions for evanescent state and wave function matching conditions are satisfied at the same time. The resulting tunneling probability is shown as Fig. 4.15(c). Figure 4.16 is the current-voltage characteristics with $g_{\max} = 0.03$. In both Figs. 4.15 and 4.16, spectrum (a) is obtained in the absence of the phonon field. A phonon field with constant g_{\max} results in spectrum (b). And spectrum (c) is obtained when $0 < g < g_{\max}$.

Let us now discuss the electron-phonon interaction when an evanescent state is involved. Generally speaking, the electron in a crystal is under the influence of the lattice potential, see Eq. (1.32). Phonon is a quasi particle of lattice vibration from its equilibrium position. Let \mathbf{Q}_n be the displacement of a lattice vector \mathbf{R}_n , the modification in the lattice potential, i.e., the electron-phonon interaction can be expressed as

$$H_{\text{ep}} = \sum_n \mathbf{Q}_n \cdot \nabla V(\mathbf{r} - \mathbf{R}_n) \quad (4.47)$$

where \mathbf{r} is the electron position, V is the lattice potential. Consider only one phonon mode

$$\mathbf{Q}_n = \frac{1}{2} \mathbf{A} e^{i(\mathbf{q} \cdot \mathbf{R}_n - \omega t)} + \text{c.c.} \quad (4.48)$$

where \mathbf{q} is the phonon wave vector. "c.c." denotes complex conjugate. The coupling between two bulk electron states

$$|\mathbf{k}_i\rangle = \frac{1}{\sqrt{N}} u_c e^{i\mathbf{k}_i \cdot \mathbf{r}}, \quad |\mathbf{k}_f\rangle = \frac{1}{\sqrt{N}} u_c e^{i\mathbf{k}_f \cdot \mathbf{r}}$$

via the electron-phonon interaction is

$$\langle \mathbf{k}_f | H_{\text{ep}} | \mathbf{k}_i \rangle = \frac{1}{2N} \sum_n e^{i\mathbf{q} \cdot \mathbf{R}_n} \int_{\text{cell}} e^{i(\mathbf{k}_i - \mathbf{k}_j) \cdot \mathbf{r}} |u_c|^2 \mathbf{A} \cdot \nabla V(\mathbf{r} - \mathbf{R}_n) d\mathbf{r} \quad (4.49)$$

with a term $e^{i(E_{k_i} - E_{k_j} - \hbar\omega)t/\hbar}$ indicating energy conservation. Here N is the number of unit cells in the lattice and u_c is the Bloch function of the conduction band.

Assume that the electrodes are doped at $2 \times 10^{17} \text{ cm}^{-3}$ so that the Fermi level is about 18 meV above the conduction bandedge. $e^{i(\mathbf{k}_i - \mathbf{k}_j) \cdot \mathbf{r}}$ can be replaced by $e^{i(\mathbf{k}_i - \mathbf{k}_j) \cdot \mathbf{R}_n}$ since $|\mathbf{k}_i|$ and $|\mathbf{k}_j|$ are very small. Moreover,

$$\int_{\text{cell}} |u_c|^2 \mathbf{A} \cdot \nabla V(\mathbf{r} - \mathbf{R}_n) d\mathbf{r} = A$$

is the same for every cell, we obtain

$$\langle \mathbf{k}_f | H_{\text{ep}} | \mathbf{k}_i \rangle = \frac{A}{2N} \sum_n e^{i(q_z + k_i - k_j) \cdot \mathbf{R}_n} \quad (4.50)$$

Because the electron energy and the wave vector in the xy plane are conserved during the tunneling process, the above expression becomes further simplified:

$$\langle \mathbf{k}_f | H_{\text{ep}} | \mathbf{k}_i \rangle = \frac{A}{2N} \sum_n e^{i(q_z + k_i - k_j)na} \quad (4.51)$$

where na is the position of lattice n in the z direction, and a is the lattice constant in the z direction. Because

$$\frac{1}{2N} \sum_n e^{i(q_z + k_i - k_j)na} = \delta(q_z + k_i - k_j) \quad (4.52)$$

we finally obtain the result of the electron-phonon interaction in bulk material:

$$\langle \mathbf{k}_f | H_{\text{ep}} | \mathbf{k}_i \rangle = A \delta(q_z + k_i - k_j) \quad (4.53)$$

which is expected as the momentum conservation during the electron-phonon interaction.

We now take a look at the case when an evanescence state is involved. Let superscript ‘+’ and ‘-’ represent the spatial region of $z > 0$ and $z < 0$, respectively. Channel E in Fig. 4.14 can be approximated by real wave vector k_i^- for $z < 0$ and k_i^+ for $z > 0$. Channel “-”, the evanescent state, is represented by $(k_j^- - i\beta)$ ($\beta > 0$) for $z < 0$ and k_j^+ for $z > 0$. Let

$$q_z + k_i^- - k_j^- \equiv k^-, \quad q_z + k_i^+ - k_j^+ \equiv k^+$$

the summation in Eq. (4.50) becomes

$$\begin{aligned} \langle \mathbf{k}_f | H_{\text{ep}} | \mathbf{k}_i \rangle &= \frac{A}{2N} \left(\sum_{n < 0} e^{iak^-n + a\beta n} + \sum_{n > 0} e^{iak^+n} \right) \\ &= \frac{A}{2} \delta(k^+) + \frac{A}{2N} \frac{\beta}{a(k^-)^2 + a\beta^2} \end{aligned} \quad (4.54)$$

By this simple analysis we see clearly that the coupling between two electron states via the electron-phonon interaction is reduced by half when an evanescent

state is involved (the second term in the above equation vanishes when N goes to ∞). Thus, the matching condition of small

$$\frac{W^*}{\hbar\omega} \psi_E$$

does not necessary demand a small ψ_E , it can indicate a weakened electron-phonon interaction.

4.4 Heterostructure Barrier Varactor

Solid-state devices have long been used for sensitive receivers in the millimeter- and sub-millimeter-wavelength ranges (0.03 ~ 3 THz). Typical applications at these frequencies can be found in advanced scientific purposes like radio astronomy but recently also in commercial applications such as intelligent cruise control. The use of solid-state THz sources offers significant advantages in weight, size, and reliability over their high frequency continuous-wave tube counterparts.

To provide power in the submillimeter wavelength range, a reverse biased Schottky diode is commonly used. When the Schottky diode is pumped with a sinusoidal signal, its non-linear capacitance generates harmonics and an external circuit extracts the desired harmonic. An alternative method to produce power is to use a direct generator such as Gunn or IMPATT oscillators, but decreasing efficiency and shrinking dimensions make the output power decrease rapidly with the frequency. Therefore, frequency multipliers are normally employed at this wavelength range.

In recent years new varactor structures with a symmetry $C - V$ characteristic have been proposed. A symmetric $C - V$ characteristic and an anti-symmetric $I - V$ characteristic will mainly create odd harmonics in a frequency multiplier. For the frequency tripler case, only the circuit impedances at the input frequency and the output frequency are of major importance. In contrast, a tripler circuit using the Schottky barrier varactor diode must also impedance-match the second harmonic (idler) properly to convert a reasonable amount of the pump power to the third harmonic.

A promising symmetric varactor device is the heterostructure barrier varactor diode (HBV) diode [19, 20], e.g., single-barrier varactor [21], double-barrier quantum-well varactor [22], barrier intrinsic-N structure, barrier-N-N⁺ structure [23, 24], multi-stack quantum-barrier varactor [25]. The possibility to epitaxially stack several barriers for a high power capability and a low capacitance value per unit area makes this device very promising for millimeter and submillimeter wave power generation.

The heterostructure barrier varactor consists of an undoped high-bandgap semiconductor (barrier) sandwiched between two moderately n -doped low-bandgap semiconductors, see Table 4.5. The barrier prevents electron transport through the structure. The first heterostructure barrier varactor was fabricated in the AlGaAs material system [19]. The disadvantage of this system is the low barrier height leading to an excessive conduction current even for Al-rich barriers. The reason is that

Table 4.5 Heterostructure barrier varactor sample structures

Thickness [nm]	InGaAs/InAlAs	AlGaAs/GaAs
300	InGaAs, $n = 10^{17}$	GaAs, $n = 10^{17}$
5	InGaAs	GaAs
L	InAlAs	AlGaAs
3	AlAs	AlAs
L	InAlAs	AlGaAs
5	InGaAs	GaAs
300	InGaAs, $n = 10^{17}$	GaAs, $n = 10^{17}$

$\text{Al}_x\text{Ga}_{1-x}\text{As}$ is an indirect bandgap semiconductor for $x > 0.41$, and for barriers with thickness of more than 4.0 nm a parasitic Γ -X transfer current dominates over the direct Γ - Γ current [26, 27].

To obtain a high multiplier conversion efficiency, the displacement current should dominate over the conduction current, i.e., a varactor type of multiplier. The conduction current can be reduced by placing a thin AlAs layer in the center of an $\text{Al}_{0.4}\text{Ga}_{0.6}\text{As}$ barrier, which increases the effective barrier height [28].

The differential capacitance is defined as

$$C = \frac{dQ}{dV_D} \quad (4.55)$$

where Q is the total charge accumulated in the emitter region, or the total charge depleted from the collector region. V_D is the external voltage. The $C - V$ characteristic is very important for the performance of a frequency multiplier. The capacitance modulation ratio

$$\frac{C_{\max}}{C_{\min}}$$

is an important figure of merit, where the maximum differential capacitance, C_{\max} , is given at zero external bias and the minimum capacitance, C_{\min} , at the maximum external bias during a pump cycle.

It has been shown that the multiplier efficiency increases monotonously for increasing C_{\max}/C_{\min} ratio. Moreover, the efficiency does not only depend on C_{\max} and C_{\min} , but also on the shape of the $C - V$ characteristic. The multiplier efficiency is improved if the shape of the $C - V$ characteristic is sharp near zero bias [29]. It is therefore important to characterize the capacitance well, especially close to zero bias.

Here we employ a similar self-consistent analysis of the Poisson and Schrödinger equations, that we developed in the previous section, to study the heterostructure barrier varactor. As discussed in the previous section, the emitter and the collector are approximated by two local Fermi levels, E_f^e and E_f^c . Here let us discuss a little bit more about this local quasi-equilibrium-state approach. The validity of local

Fermi level concept depends on the small conduction current which does not destroy the local equilibrium state. The conduction current is normally low when the external bias V_D is small. We now analyze the conduction current when the bias is increased.

An electron wave is injected from the emitter in the form of

$$e^{ik_e z}$$

where k_e is the wave vector in the emitter. It is partially reflected

$$r_e e^{-ik_e z}$$

and partially transmitted

$$t_e e^{iq_e z}$$

Assume that the central active layer is thin enough so that we can express the wave function as

$$\psi(z) = \begin{cases} e^{ik_e z} + r_e e^{-ik_e z} & z < 0 \\ t_e e^{iq_e z} & z \geq 0 \end{cases} \quad (4.56)$$

The continuations of the wave function and its first-order derivative at $z = 0$ give

$$1 + r_e = t_e, \quad k_e(1 - r_e) = q_e t_e \quad \Rightarrow \quad t_e = \frac{2}{1 + \frac{q_e}{k_e}} \quad (4.57)$$

where

$$\frac{\hbar^2 k_e^2}{2m_c^*} = \frac{\hbar^2 q_e^2}{2m_c^*} - eV_D$$

due to the conservation of the total energy. It is thus observed that the amplitude of the transmitted wave, t_e , decreases following the increase of the bias V_D . The carriers from the emitter are largely reflected and the transmission is small. The carriers from the collector are completely reflected when the sample is biased in such a way that E_f^c is below E_c^e .

The amplitude of the transmitted wave, t_e , is further reduced when the finite thickness of the barrier and the finite barrier height are taken into account. It has thus been shown that the local equilibrium assumption of E_f^e and E_f^c is valid because of the low carrier transmission. The two quasi Fermi levels are aligned within their corresponding regions and are separated from each other by eV_D .

The conduction bandedge and electron distribution as functions of the external bias V_D are plotted in Fig. 4.17 for an InGaAs/InAlAs varactor. Here the heterostructure barrier is composed of a 3-nm thick AlAs layer sandwiched by two 8-nm thick $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ layers. The central barrier is embedded in 10^{17} cm^{-3} doped $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$. And the device temperature is set at 300 K.

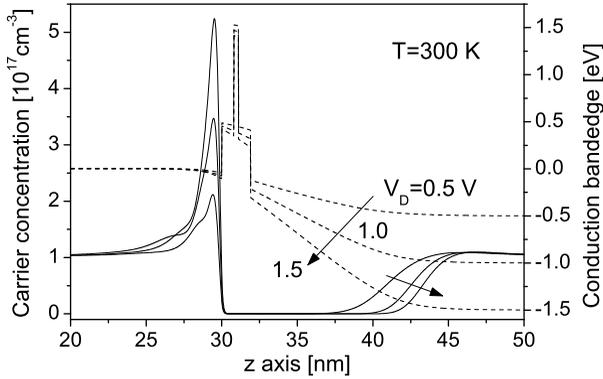


Fig. 4.17 Conduction bandgaps (*dashed lines*) and electron distributions (*solid lines*) in an InGaAs/InAlAs heterostructure barrier varactor at 300 K. *Arrows* indicate the increase of $V_D = 0.5, 1.0, 1.5$ V (Reprinted with permission from Y. Fu, L. Dillner, J. Stake, M. Willander, and E. L. Kollberg, Capacitance analysis of AlGaAs/GaAs and InAlAs/InGaAs heterostructure barrier varactors, *J. Appl. Phys.*, vol. 83, pp. 1457–1462, 1998. Copyright 1998, American Institute of Physics)

4.4.1 Conduction Current

The $I - V$ characteristic is mainly composed of thermionic electron emission over the heterostructure barrier as well as tunneling through the barrier. The dominating mechanism depends on the electron temperature, the barrier height, and the barrier thickness. To ensure varactor mode operation over a wide range of pump power in a multiplier circuit, it is important to minimize the conduction current.

The current contribution of state k_e originated from the emitter is expressed by Eq. (4.40). Including the Fermi distribution and the density of states, see Eq. (4.27), the total conduction current from the emitter to the collector is

$$I_{\text{emitter} \rightarrow \text{collector}} = \frac{-em_c^*k_B T}{2\pi\hbar^2} \int \frac{\hbar q_e}{m_c^*} |t_e|^2 \ln \left[1 + e^{(E_f^e - E_c^e - E_e)/k_B T} \right] \frac{2dk_e}{2\pi} \quad (4.58)$$

By adding the current from the collector to the emitter, the total conduction current along the z axis is calculated from the following equation

$$I = \frac{em_c^*k_B T}{2\pi\hbar^2} \int \frac{\hbar k_c}{m_c^*} |t_c|^2 \ln \left[1 + e^{(E_f^c - E_c^c - E_c)/k_B T} \right] \frac{2dq_c}{2\pi} - \frac{em_c^*k_B T}{2\pi\hbar^2} \int \frac{\hbar q_e}{m_c^*} |t_e|^2 \ln \left[1 + e^{(E_f^e - E_c^e - E_e)/k_B T} \right] \frac{2dk_e}{2\pi} \quad (4.59)$$

where t_e and t_c are amplitudes of transmitted waves when transmitting from the emitter to collector and from the collector to the emitter, respectively, with wave

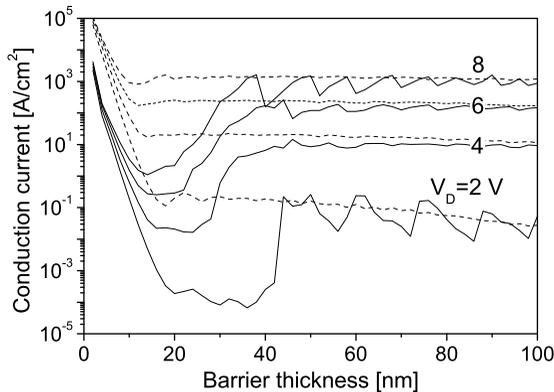


Fig. 4.18 The relations between conduction current and barrier thickness of an InGaAs/InAlAs varactor. *Dashed lines*: central barrier region does not have an AlAs layer; *Solid lines*: with a 3-nm AlAs layer. $V_D = 2, 4, 6$ and 8 V, respectively (Reprinted with permission from Y. Fu, L. Dillner, J. Stake, M. Willander, and E. L. Kollberg, Capacitance analysis of AlGaAs/GaAs and InAlAs/InGaAs heterostructure barrier varactors, *J. Appl. Phys.*, vol. 83, pp. 1457–1462, 1998. Copyright 1998, American Institute of Physics)

vector k_e and q_c . Relationships between various parameters are

$$E_e = \frac{\hbar^2 k_e^2}{2m_c^*}, \quad E_e + eV_D = \frac{\hbar^2 q_e^2}{2m_c^*}, \quad E_c = \frac{\hbar^2 q_c^2}{2m_c^*}, \quad E_c - eV_D = \frac{\hbar^2 k_c^2}{2m_c^*}$$

In order to minimize the conduction current, its relation with the barrier thickness is to be investigated. The results are presented in Fig. 4.18. The conduction current decreases generally following the increase of the barrier thickness, as expected for the situations without central AlAs layer.

With a 3-nm thick AlAs layer introduced to the barrier region, the situation becomes complicated. We can discuss the situation in three steps.

- The AlAs layer is dominant when the total barrier is thin enough that the conduction current decreases following the increase of the barrier thickness. It is easy to show that the barrier height does not affect the transmission when the barrier is very thin. We expect the same current-thickness relation for cases with and without the central AlAs layer.
- When the total barrier is very thick, the effect from the 3-nm AlAs layer becomes negligible, the current-barrier-thickness relation converges to the one when the AlAs layer is absent, as demonstrated in the figures.
- It is expected that the conduction current is reduced by introducing the AlAs layer because of the higher barrier height. A valley is then anticipated in the current-thickness-relation in the presence of the AlAs layer.

The above analysis is clearly demonstrated by Fig. 4.18.

Minimization of the conduction current with respect to the proper barrier structure design is thus possible. Figure 4.18 indicates an optimal InGaAs/InAlAs var-

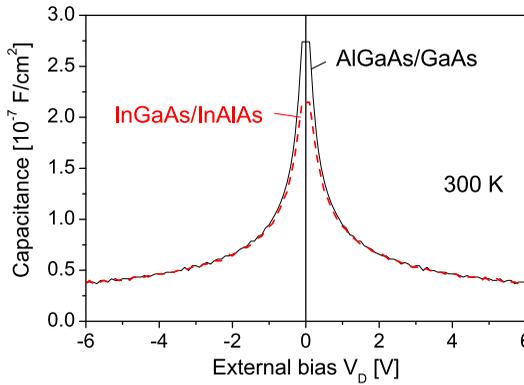


Fig. 4.19 $C - V$ characteristics at 300 K for an AlGaAs/GaAs (solid line) and InGaAs/InAlAs (dashed line) varactors. The AlGaAs/GaAs varactor consists of 3/3/3-nm $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}/\text{AlAs}/\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ heterostructure barrier. The heterostructure barrier in the InGaAs/InAlAs varactor is 8/3/8 nm $\text{In}_{0.52}\text{Al}_{0.48}\text{As}/\text{AlAs}/\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ (Reprinted with permission from Y. Fu, L. Dillner, J. Stake, M. Willander, and E. L. Kollberg, Capacitance analysis of AlGaAs/GaAs and InAlAs/InGaAs heterostructure barrier varactors, J. Appl. Phys., vol. 83, pp. 1457–1462, 1998. Copyright 1998, American Institute of Physics)

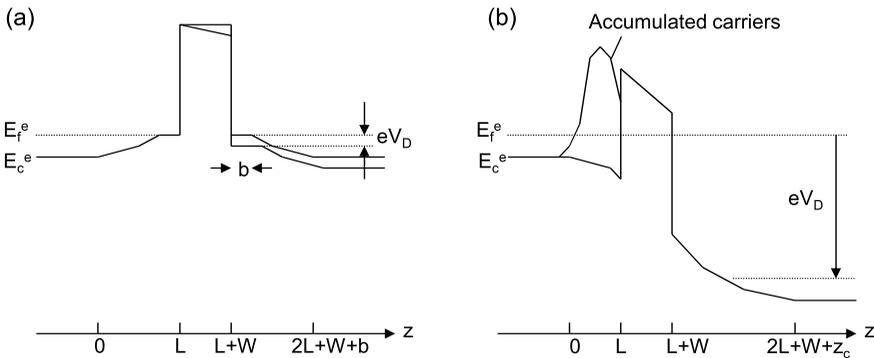


Fig. 4.20 Schematic energy band structure at (a) small and (b) at large external bias V_D

actor with a barrier structure of a 3-nm AlAs layer sandwiched between two 8-nm InAlAs layers.

4.4.2 $C - V$ Characteristics

The $C - V$ characteristics at 300 K are plotted in Fig. 4.19 calculated self-consistently from the Schrödinger and Poisson equations.

For further analysis, we approximate the energy band structure of the varactor shown in Fig. 4.17 at small and large bias by Fig. 4.20. Much alike the resonant tunneling diode, the varactor consists of an emitter and a collector, both are doped with

a same doping level N_D . There is a central barrier with a thickness W . Two undoped spacers (thickness L) separate the barrier from the doped emitter and collector. The conduction bandedge of the emitter is denoted as E_c^e , which is set to be the potential reference (thus $E_c^e = 0$). E_f^e denotes the Fermi level in the emitter. E_c^c and E_f^c denote the conduction bandedge and Fermi level in the collector, respectively. At zero external bias, $E_c^c = 0$ and $E_f^c = E_f^e$. With a nonzero V_D , $E_c^c = -eV_D$ and $E_f^c = E_f^e - eV_D$. And a Coulombic potential $\phi(z)$ distributes due to the distributions of dopants and electrons.

4.4.2.1 Small External Bias

At low temperature, $e^{(E_f^e - E_i)/k_B T} \gg 1$ when E_i is below E_f^e , thus the sheet density, see Eq. (4.27), of a one-dimensional electron state E_i in the emitter is

$$\frac{m_c^* k_B T}{2\pi \hbar^2} \ln [1 + e^{(E_f^e - E_i)/k_B T}] \approx \frac{m_c^* (E_f^e - E_i)}{2\pi \hbar^2} \quad (4.60)$$

We approximate the electron state E_i semiclassically so that its wave function ψ_i is constant in the classically allowed region ($E_i > E_c^e$). Moreover, we approximate all the states E_i below E_f^e are identical so that the distribution of electrons originated from the emitter is $n_s(E_f^e - E_c^e - e\phi)$, where n_s is an effective sheet density. n_s can be obtained by the charge neutral condition in the emitter such that, at $z = -\infty$, $\phi = 0$ and $E_c^e = 0$ so that $n_s E_f^e = N_D$. We then obtain the following Poisson equations for the emitter

$$\frac{d^2 \phi}{dz^2} = \begin{cases} \frac{e^2}{\epsilon} [N_D - n_s(E_f^e - e\phi)] & z < 0 \\ \frac{e^2}{\epsilon} [-n_s(E_f^e - e\phi)] & 0 \leq z < L \end{cases} \quad (4.61)$$

and for the collector,

$$\frac{d^2 \phi}{dz^2} = \begin{cases} \frac{e^2}{\epsilon} [N_D - n_s(E_f^c - e\phi)] & z > \beta \\ \frac{e^2}{\epsilon} [-n_s(E_f^c - e\phi)] & \beta_1 < z \leq \beta \end{cases} \quad (4.62)$$

where

$$\beta = 2L + W, \quad \beta_1 = L + W + b \quad (4.63)$$

and b is the width of the depletion layer, see Fig. 4.20.

The corresponding solution of the above equations is

$$\phi(z) = \begin{cases} \left(\frac{E_f^e}{2} + c_1\right)e^{\alpha z} & z < 0 \\ E_f^e + c_1 e^{\alpha z} - \frac{E_f^e}{2} e^{-\alpha z} & 0 < z < L \end{cases} \quad (4.64)$$

in the emitter, where $\alpha^2 = e^2 n_s / \epsilon$. For the collector,

$$\phi(z) = \begin{cases} -eV_D + \left(\frac{E_f^e}{2} + c_2\right)e^{\alpha(\beta-z)} & z > \beta \\ E_f^e - eV_D + c_2e^{\alpha(\beta-z)} - \frac{E_f^e}{2}e^{-\alpha(\beta-z)} & \beta_1 < z < \beta \end{cases} \quad (4.65)$$

c_1 and c_2 are coefficients to be determined via boundary conditions.

And the net charges in the emitter and collector are

$$\begin{aligned} Q_e &= \frac{en_s}{\alpha} \left(c_1 e^{\alpha L} + \frac{E_f^e}{2} e^{-\alpha L} \right) \\ Q_c &= \frac{en_s}{\alpha} \left[c_2 e^{\alpha(L-b)} + \frac{E_f^e}{2} e^{-\alpha(L-b)} \right] \end{aligned} \quad (4.66)$$

When $V_D = 0$, $Q_e = 0$ so that

$$c_1 = -\frac{E_f^e}{2} e^{-2\alpha L}, \quad \phi(L) = E_f^e (1 - e^{-\alpha L})$$

As long as L is finite, $\phi(L)$ will be smaller than E_f^e . In this case, b is zero when eV_D is lower than $E_f^e e^{-\alpha L}$. The above equations are then very much simplified.

At $z = L + W$, the continuities of the potential

$$\begin{aligned} E_f^e + c_1 e^{\alpha L} - \frac{E_f^e}{2} e^{-\alpha L} + \alpha W \left(c_1 e^{\alpha L} + \frac{E_f^e}{2} e^{-\alpha L} \right) \\ = E_f^e - eV_D + c_2 e^{\alpha L} - \frac{E_f^e}{2} e^{-\alpha L} \end{aligned} \quad (4.67)$$

and the electric field

$$c_1 e^{\alpha L} + \frac{E_f^e}{2} e^{-\alpha L} + c_2 e^{\alpha L} + \frac{E_f^e}{2} e^{-\alpha L} = 0 \quad (4.68)$$

result in the following relation between eV_D and Q_e

$$-eV_D = \frac{\alpha(2 + \alpha W)}{en_s} Q_e \quad (4.69)$$

In Eq. (4.67), the left side is the potential of Eq. (4.65) at $z = L + W$ in the collector region. The first three terms on the right side give the potential of Eq. (4.64) at $z = L$ in the emitter region, the rest is the potential drop across the barrier.

Thus, at small external bias, the differential capacitance is

$$C_{\max} = \frac{dQ_e}{dV_D} = \frac{\alpha}{\epsilon(2 + \alpha W)} \quad (4.70)$$

Note that Q_e is negative when deriving the above Eq. (4.70). The relation between C_{\max} and W in the above equation can be understood in a straightforward way. When the barrier becomes thinner, the potential drop across the barrier is reduced. The reduction is used to drive carriers so that the capacitance increases. When n_s is large, i.e., large α , e.g., in metal materials, Eq. (4.70) becomes

$$C = \frac{1}{\epsilon W}$$

which is the common expression of capacitance for a capacitor formed by two metal plates separated from each other at a distance of W .

For semiconductor materials of interest, Eq. (4.70) indicates that the capacitance is proportional to m_c^*/ϵ as clearly confirmed by Fig. 4.19 which shows a larger capacitance for an AlGaAs/GaAs varactor. For an InGaAs/InAlAs varactor, $m_c^* = 0.043$, while for an AlGaAs/GaAs varactor, $m_c^* = 0.065$, in the 300 nm doped regions.

The capacitance is usually measured by applying a small ac signal in addition to the dc bias. If the amplitude of the ac signal is large compared with the Fermi energy, b in Eq. (4.62) is no longer negligible. The spacer in the collector region is readily depleted so that

$$C_{\max} \propto \frac{1}{L + W} \quad (4.71)$$

is expected, as compared with Eq. (4.70).

In obtaining Eq. (4.70) we have calculated the carrier concentration by

$$n_s(E_f - \phi)$$

In a real varactor structure where the doped region is quite wide (300 nm in Table 4.5), a carrier density of states of the bulk material is to be used. Again we assume that all the states between E_f and conduction bandedge E_c are totally occupied and all states above E_f are totally empty (which is true when the temperature is rather low, and/or when the doping level is very high). For a three-dimensional material, the carrier density is then

$$n = \int_{E_c}^{E_f} \frac{2d\mathbf{k}}{(2\pi)^3} = n_{3D}(E_f - E_c)^{3/2} \quad (4.72)$$

where

$$n_{3D} = \frac{1}{3\pi^2} \left(\frac{2m_c^*}{\hbar^2} \right)^{3/2}$$

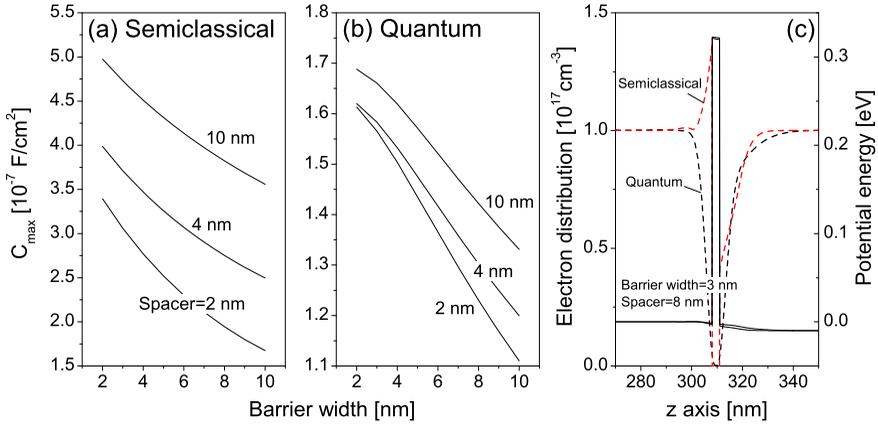


Fig. 4.21 (a) C_{\max} calculated by the Poisson equation of Eq. (4.73). (b) C_{\max} self-consistently calculated from Schrödinger and Poisson equations at 300 K. (c) Comparison of electron and potential energy distributions

And the following Poisson equation is to be studied

$$\frac{d^2\phi}{dz^2} = \begin{cases} \frac{e^2}{\epsilon} [N_D - n_{3D}(E_f - \phi)^{3/2}] & z < 0 \\ \frac{e^2}{\epsilon} [-n_{3D}(E_f - \phi)^{3/2}] & 0 < z < L \\ \frac{e^2}{\epsilon} [-n_{3D}(E_f - eV_D - \phi)^{3/2}] & L + W < z < \beta \\ \frac{e^2}{\epsilon} [N_D - n_{3D}(E_f - eV_D - \phi)^{3/2}] & z > \beta \end{cases} \quad (4.73)$$

The boundary conditions are

$$\phi(-\infty) = 0, \quad V(\infty) = -eV_D$$

$\phi(z)$ is obtained from the above Poisson equation. Knowing $\phi(z)$, $Q_e(V_D)$ is readily calculated. Figure 4.21(a) is the calculation result of C_{\max} as function of spacer thickness L , barrier width W for an $\text{Al}_{0.4}\text{Ga}_{0.6}\text{As}/\text{GaAs}$ varactors based on Eq. (4.73). A voltage step of 0.01 volt is applied. Equation (4.71) is clearly confirmed.

However, the self-consistent Schrödinger and Poisson equations give us a bit different picture about the C_{\max} which are depicted in Fig. 4.21(b) due to the electron wave nature. For a wave, its amplitude decreases when its distance to a barrier becomes comparable to its wave length. For the varactors under investigation, the distance of the first peak carrier distribution to the barrier edge is about 20 nm, see Fig. 4.21(c). Carriers are depleted within this range. It is then expected that the carrier distribution and energy band structure are almost unchanged for spacers whose layer thicknesses are comparable with this distance. On the other hand, a classical particle can actually touch the barrier edge so that the carrier distribution depends strongly on the spacer layer thickness.

4.4.2.2 Large External Bias

When the external bias is largely increased so that the conduction bandedge of the collector at $z = L$, $-e\phi(L)$ is below zero, a triangle quantum well forms in front of the barrier.

At 300 K, when $N_D = 10^{17} \text{ cm}^{-3}$, E_f is 16 meV in the InGaAs/InAlAs varactor below the conduction bandedge of the emitter. The sheet density is $2.0 \times 10^{11} \text{ cm}^{-2}$ when the sublevel energy is zero. When the sublevel is lowered by 30 meV, the sheet density becomes $4.6 \times 10^{11} \text{ cm}^{-2}$. At this stage, the first excited state is formed in the quantum well in addition to the ground state, so that the total sheet density is $6.6 \times 10^{11} \text{ cm}^{-2}$. For an AlGaAs/GaAs varactor with the same N_D , $E_f = -35 \text{ meV}$. The sheet density is $1.6 \times 10^{11} \text{ cm}^{-2}$ when the sublevel energy is zero. When the ground state is lowered by 20 meV, the first excited state appears and the total sheet density becomes $4.8 \times 10^{11} \text{ cm}^{-2}$.

The energy band structure is schematically represented by Fig. 4.20(b). The external bias largely applies on the depletion region in the collector because of the huge localized carrier concentration in the accumulation region in the emitter. The figure also represents the band structures when L is small or when the voltage step is comparable to the Fermi energy. For small L , the band bending in the emitter is small according to Eq. (4.65). Extra carriers readily accumulate in the quantum well in front of the barrier.

The electron distribution in the collector can then be well approached by an effective depletion length z_c . The potential ϕ and the electric field E at $z = \beta$ are then

$$\phi(\beta) = \frac{eN_D z_c^2}{2\epsilon}, \quad E(\beta) = -\frac{eN_D z_c}{\epsilon} \quad (4.74)$$

Because the net charge in the barrier region is negligibly small,

$$\phi(L) = \frac{eN_D [z_c^2 + 2(L + W)z_c]}{2\epsilon}, \quad E(L) = E(\beta) \quad (4.75)$$

As discussed earlier, lowering $V(L)$ increases very much the sheet density of the localized state. Due to the large sheet density of localized electrons, the triangle quantum well in front of the barrier is quite shallow. Its depth D is:

$$D = eV_D - e\phi(L) \quad (4.76)$$

The width of the triangle quantum well is however quite large because of the wave function extension of the localized sublevels in the emitter region. Energy separations among sublevels are reduced with the increased well width so that more sublevels exist in the quantum well. This is another reason for the high sheet density of localized states.

When the external bias is large, D can be neglected as compared with eV_D so that

$$eV_D = \frac{eN_D [z_c^2 + 2(L + W)z_c]}{2\epsilon}, \quad Q_c = eN_D z_c \quad (4.77)$$

it is then finally concluded that

$$C_{\min} = \frac{e\epsilon}{\sqrt{(L+W)^2 + \frac{2\epsilon V_D}{N_D}}} \quad (4.78)$$

Figure 4.19 directly confirms Eq. (4.78), as it is observed here that the capacitance becomes independent of the materials used in the varactor at large external bias. Thus, for a small C_{\min} low doping concentration is desired.

4.5 High-Electron-Mobility Transistor

In 1967 Stern and Howard first formulated the two-dimensional electron transport for the silicon inversion layers [30]. Dingle, Störmer, Gossard, Wiegmann [31] and Hess [32] in 1978 and 1979 investigated experimentally and theoretically the mobility enhancement due to the separation of carriers from their parent donors. In GaAs/AlGaAs heterostructures, electrons confined at the GaAs side of the heterointerface have exhibited enhanced mobilities [33]. Much attention then has been focused on modulation-doped heterostructures in order to realize ultra-high-speed field-effect devices, such as the high-electron-mobility transistor (HEMT) [34].

Walukiewicz et al. in 1984 considered the electron mobility in an AlGaAs/GaAs heterostructure with a triangular potential well on the GaAs side including the phonon, alloy, disorder and remote impurity scatterings [35, 36], while more complicated model for ionized impurity scattering was involved by Yokoyama and Hess in 1986 when analyzing the two-dimensional electronic transport in the AlGaAs/GaAs HEMT [37]. To describe the impurity scattering, Hirakawa and Sakaki [38] used the expression derived by Price [39] for phonon scattering.

Here we shall study the remote impurity scattering involving the realistic electron wave functions at the AlGaAs/GaAs heterointerface in an AlGaAs/GaAs HEMT. The AlGaAs/GaAs HEMT is usually grown by molecular beam epitaxy on semi-insulating (001) GaAs substrate. An n -type GaAs buffer layer is followed by an undoped AlGaAs spacer of thickness L , then n^+ -AlGaAs, and finally the cap. The existence of a band offset ΔE_c at the AlGaAs/GaAs heterointerface causes the electrons to transfer from the AlGaAs side to the GaAs side. The mobilities of these electrons are very high because of the spatial separation from parent dopants which are located in the n^+ -AlGaAs layer. A metal contact is mounted on top of the cap layer so that a gate bias V_G can be applied to modify the electron states.

Set the sample growth direction as the z direction and the perpendicular plane as the xy plane, the charge distribution and conduction bandedge, as functions of gate bias V_G , are calculated by the Schrödinger and Poisson equations and the results are plotted in Fig. 4.22. It is clear from Fig. 4.22 that electrons are transferred from the highly doped $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ layer to the undoped GaAs side, and the electron transfer is controllable by the gate, as expected. Since the spatial distribution of the electrons in the GaAs layer is very narrow (about 15 nm), the electrons are

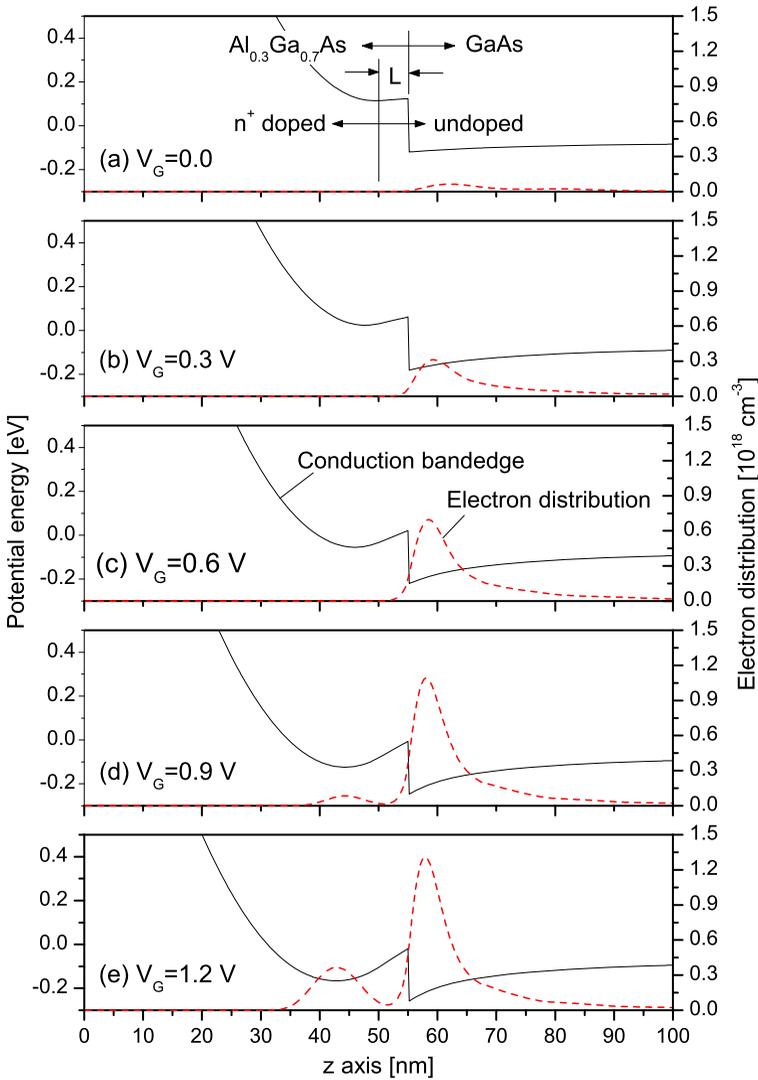


Fig. 4.22 Conduction bandedge (solid lines) and electron distribution (dashed lines) of an $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}/\text{GaAs}$ high-electron-mobility transistor at 300 K

usually denoted as the quasi-two-dimensional electron gas (2DEG). In Fig. 4.22 the n^+ -doped $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ layer is 50 nm thick and the thickness L of the undoped $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ spacer is 5 nm. Depending on the purpose of the device, the thickness of the n^+ -doped AlGaAs layer can be tuned accordingly. When a small depletion gate bias is needed, thinner n^+ -doped AlGaAs layer is to be grown.

4.5.1 Remote Impurity Scattering

We focus on an unscreened ionized impurity potential

$$V'(\mathbf{r}, \mathbf{r}_0) = \frac{-e^2}{4\pi\epsilon|\mathbf{r} - \mathbf{r}_0|} \quad (4.79)$$

to understand the fundamental mechanism of the enhanced electron mobility, where \mathbf{r}_0 and \mathbf{r} are spatial positions of the impurity and the electron. This simple impurity potential is actually implicitly implicated in Fig. 4.22 that at a moderate gate bias, the AlGaAs layer is free of electrons. Further increasing the gate bias introduces free electrons to the AlGaAs layer. When free electrons begin to accumulate in the AlGaAs layer, they start to screen the gate bias, and we will lose the proper control over the 2DEG at the AlGaAs/GaAs heterointerface. Therefore, the impurities in the AlGaAs layer should be largely ionized and their potentials not to be screened at device working conditions of the HEMT.

By discussions in Chap. 2, the probability $s_{k \rightarrow q}(\mathbf{r}_0)$ that the interaction due to dopant at \mathbf{r}_0 causes one electron to transit from state $\Psi_k(\mathbf{r})$ to state $\Psi_q(\mathbf{r})$ is

$$s_{k \rightarrow q}(\mathbf{r}_0) = \frac{2\pi}{\hbar} \left| \langle \Psi_k(\mathbf{r}) | V'(\mathbf{r}, \mathbf{r}_0) | \Psi_q(\mathbf{r}) \rangle \right|^2 \delta[E(k) - E(q)] \quad (4.80)$$

where the interaction potential $V'(\mathbf{r}, \mathbf{r}_0)$ of Eq. (4.79) is time-independent. In a three-dimensional system with translational symmetry, the Bloch theorem says that the total wave function of an electron state \mathbf{k} is

$$\Psi_{\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{N}} e^{i\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}}(\mathbf{r}), \quad u_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r} + \mathbf{R}) \quad (4.81)$$

In the case of a layered system grown along the z direction, the electron in the conduction band is described by

$$\Psi_{i\mathbf{k}_{xy}}(\mathbf{r}) = \psi_i(z) e^{i\mathbf{k}_{xy}\cdot\boldsymbol{\rho}} u_c(\mathbf{r}) \quad (4.82)$$

where \mathbf{k}_{xy} is the wave vector because of the translational symmetry in the xy plane. $\boldsymbol{\rho} = (x, y)$. $\psi_i(z)$ is the envelope function of state E_i determined by the one-dimensional Schrödinger equation along the z direction.

By inserting the above wave function form into Eq. (4.80), we obtain

$$s_{i\mathbf{k}_{xy} \rightarrow j\mathbf{q}_{xy}}(\mathbf{r}_0) = \frac{2\pi\delta}{\hbar} \left| \int \psi_i^*(z) V'(\mathbf{r}, \mathbf{r}_0) \psi_j(z) e^{i(\mathbf{q}_{xy} - \mathbf{k}_{xy})\cdot\boldsymbol{\rho}} d\mathbf{r} \right|^2 \quad (4.83)$$

where

$$\delta = \delta[E_i + E(\mathbf{k}_{xy}) - E_j - E(\mathbf{q}_{xy})]$$

and $E(\mathbf{k}_{xy}) = \hbar^2 k_{xy}^2 / 2m_c^*$ is the kinetic energy in the xy plane.

Since the density of atoms in GaAs is $4 \times 10^{22} \text{ cm}^{-3}$, while the concentration N_D of n^+ -doped impurities is normally less than 10^{18} cm^{-3} , the impurities can be treated as independent from each other and the scatterings from different impurities are incoherent. The total impurity scattering from state $i\mathbf{k}_{xy}$ to state $j\mathbf{q}_{xy}$ is

$$\begin{aligned} S_{i\mathbf{k}_{xy} \rightarrow j\mathbf{q}_{xy}} &= \int N_D(\mathbf{r}_0) s_{i\mathbf{k}_{xy} \rightarrow j\mathbf{q}_{xy}}(\mathbf{r}_0) d\mathbf{r}_0 \\ &= \frac{e^4 \delta}{8\pi \epsilon^2 \hbar} \int N_D(\mathbf{r}_0) \left| \int \frac{\psi_i^*(z) \psi_j(z)}{|\mathbf{r} - \mathbf{r}_0|} e^{i(\mathbf{q}_{xy} - \mathbf{k}_{xy}) \cdot \boldsymbol{\rho}} d\mathbf{r} \right|^2 d\mathbf{r}_0 \end{aligned} \quad (4.84)$$

Let $\lambda = |\mathbf{q}_{xy} - \mathbf{k}_{xy}|$ and

$$g(\lambda|z - z_0|) = \int \frac{e^{i(\mathbf{q}_{xy} - \mathbf{k}_{xy}) \cdot \boldsymbol{\rho}}}{|\mathbf{r} - \mathbf{r}_0|} d\boldsymbol{\rho} \quad (4.85)$$

It is easy to see that $g(\alpha)$ is small when $\alpha = \lambda|z - z_0|$ is large, i.e., when the electron has a large kinetic energy in the xy plane (large k and q) and/or the spatial separation between the electron and impurity ($|z - z_0|$) is large. Numerically $g(\alpha) \approx 1/\alpha$ so that

$$S_{i\mathbf{k}_{xy} \rightarrow j\mathbf{q}_{xy}} \approx \frac{e^4 \delta}{8\pi \epsilon^2 \lambda^2 \hbar} g_{ij} \quad (4.86)$$

where

$$g_{ij} = \int N_D(z_0) \left| \int \frac{\psi_i^*(z) \psi_j(z)}{|z - z_0|} dz \right|^2 dz_0 \quad (4.87)$$

Here $N_D(\mathbf{r}_0) = N_D(z_0)$ since the doping profile is one dimensional along the z direction, see Fig. 4.22. And the total electron-impurity scattering rate is

$$\begin{aligned} S_{\text{total}} &= \sum_{i\mathbf{k}_{xy}; j\mathbf{q}_{xy}} S_{i\mathbf{k}_{xy} \rightarrow j\mathbf{q}_{xy}} f(i, \mathbf{k}_{xy}) [1 - f(j, \mathbf{q}_{xy})] \\ &\quad - S_{j\mathbf{q}_{xy} \rightarrow i\mathbf{k}_{xy}} [1 - f(i, \mathbf{k}_{xy})] f(j, \mathbf{q}_{xy}) \end{aligned} \quad (4.88)$$

where we have taken into account the Pauli exclusion principle that the initial state is occupied and the final state empty. It is easy to see that numerically $S_{i\mathbf{k}_{xy} \rightarrow j\mathbf{q}_{xy}} = S_{j\mathbf{q}_{xy} \rightarrow i\mathbf{k}_{xy}}$. By denoting them simply as $S_{i\mathbf{k}_{xy} j\mathbf{q}_{xy}}$,

$$S_{\text{total}} = \sum_{i\mathbf{k}_{xy}; j\mathbf{q}_{xy}} S_{i\mathbf{k}_{xy} j\mathbf{q}_{xy}} [f(i, \mathbf{k}_{xy}) - f(j, \mathbf{q}_{xy})] \quad (4.89)$$

Since $S_{i\mathbf{k}_{xy} j\mathbf{q}_{xy}}$ is proportional to $1/\lambda^2$ for the same subband states of i and j , see Eq. (4.86), the transport of carriers with high kinetic energies in the xy plane will be less affected by ionized impurities than those with small kinetic energies.

Assuming that the ionized impurity scattering is the dominant factor in determining the transport properties of the 2DEG, we obtain the relaxation time as

$$\frac{1}{\tau} = S_{\text{total}} \quad (4.90)$$

and the carrier mobility in the xy plane is expressed as

$$\mu = \frac{e}{n} \int \tau v_{xy}^2(E_{xy}) \frac{\partial f}{\partial E_{xy}} N(E_{xy}) dE_{xy} \quad (4.91)$$

see Eq. (2.172). Here $v_{xy}(E_{xy})$ is the group velocity in the xy plane, $N(E_{xy})$ is the electronic density of states, and n is the sheet density of the 2DEG.

Because of the factor $\partial f / \partial E_{xy}$ in the above equation, the mobility is largely determined by the relaxation time and group velocity at the Fermi level.

To optimize the carrier mobility with respect to the thickness, L , of the undoped AlGaAs spacer, we first notice that the first excited state is very close to the Fermi level, thus having much smaller group velocity, its contribution to the mobility can be neglected. High-energy states are largely unoccupied so that they do not contribute to the total carrier mobility. We only need to concentration on the ground state to calculate the mobility of the 2DEG in the xy plane.

Secondly, since the first excited state is well separated from the ground state, inter-subband transitions from ground subband to excited subbands are not likely because of the energy conservation factor δ . We only need to consider the intra-subband transition within the ground state.

Transitions are thus only possible at the Fermi surface. Let $\Delta_0 = E_f - E_c$ be the kinetic energy at the Fermi level, where E_c is the conduction bandedge. Since

$$\frac{\hbar^2 \lambda^2}{2m_c^*} \approx E_f - E_c$$

we obtain the expression for the relaxation time

$$\tau \propto \frac{\Delta_0}{g_{00}} \quad (4.92)$$

and the mobility is

$$\mu = \frac{\Delta_0^2}{ng_{00}} \quad (4.93)$$

apart from a proportionality constant. In the above mobility expression, the first Δ_0 comes from the relaxation time τ and the second Δ_0 from the group velocity

$$\frac{1}{2} m_c^* v_{xy}^2(E_f) = \Delta_0$$

In Fig. 4.23 we have listed three principal factors that determine the mobility as well as the mobility itself as functions of the AlGaAs spacer thickness L . Here we

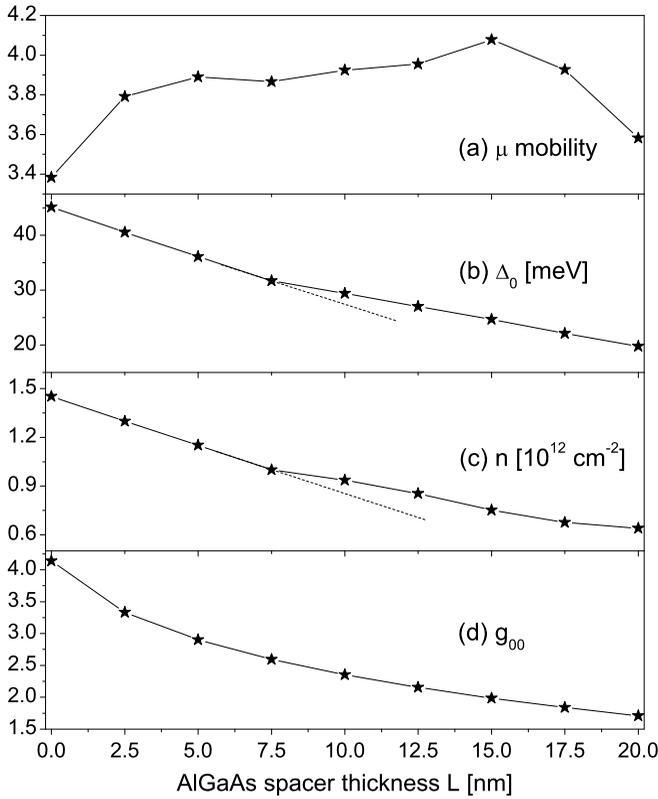


Fig. 4.23 Carrier mobility (a) and three principal factors Δ_0 (b), n (c) and g_{00} (d) that determine the carrier mobility as functions of the AlGaAs spacer thickness L in an $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}/\text{GaAs}$ high-electron-mobility transistor

see that the increase of the spacer layer thickness decreases g_{00} . And L basically represents the overlapping between 2DEG and dopants.

The AlGaAs/GaAs heterointerface is in principle an $n^+ - n^-$ junction with a built-in potential. Without the undoped AlGaAs spacer layer, the built-in potential applies completely on the GaAs side because of high doping level in AlGaAs side. Due to the conduction band offset between GaAs and AlGaAs, the quantum well thus created in the GaAs side is the deepest, resulting in the highest sheet density of the 2DEG confined in this quantum well. When an undoped AlGaAs spacer layer is introduced, it will share the built-in potential, and the depth of the quantum well in the GaAs side is decreased. Δ_0 and the sheet density of the 2DEG are consequently decreased as well. When the spacer is thick enough, the undoped AlGaAs spacer will be able to sustain the whole built-in potential. Further increasing L will no longer affect energy subband structure in the GaAs side. And the GaAs conduction bandedge becomes flat and there are no free carriers.

By Fig. 4.23, we see that n and Δ_0 decreases monotonically as function of L . However, at about $L = 10$ nm, there are shoulders in both the $n - L$ and $\Delta_0 - L$ relationships. Checking carefully at subband structures, we see that when $L < 10$ nm, the ground state of the 2DEG is lower than that of the carriers confined in the n^+ AlGaAs layer, while for $L > 10$ nm, the relationship is reversed.

Assuming the domination of the ionized impurity scattering in determining the carrier mobility, the mobility of the 2DEG is determined by the following four factors. (1) The overlapping of the carrier distribution with dopants, g_{00} ; (2) The energy of the carrier in the xy plane, k_{xy}^2 ; (3) The square of the group velocity at the Fermi level, v_{xy}^2 ; And finally, (4) the sheet density of the 2DEG, n . The square of the group velocity at the Fermi level is proportional to Δ_0 , which is the kinetic energy at the Fermi level in the xy plane. The relations of Δ_0 and n with L are the same, so that their effects on the mobility cancel with each other and the mobility is then mainly determined by the overlapping and the kinetic energy at the Fermi level. Since Δ_0 decreases while $1/g_{00}$ increases with increasing L , an optimal value of the mobility as a function of L is expected, as demonstrated in Fig. 4.23. In the present AlGaAs/GaAs HEMT structure, the optimal mobility is reached when the undoped AlGaAs spacer is about 15 nm thick.

4.5.2 δ -Doped Field-Effect Transistor

Modern semiconductor growth techniques have provided great possibilities to create thin doping layers with a thickness of a few atomic layers which is normally referred to as δ -doping. δ -doped layers in Si molecular beam epitaxy (MBE) can be achieved by solid-phase epitaxy [40] and by ion beam doping [41]. The charged carriers are, in this case, confined in a quantum well induced by the δ -doped impurity layers.

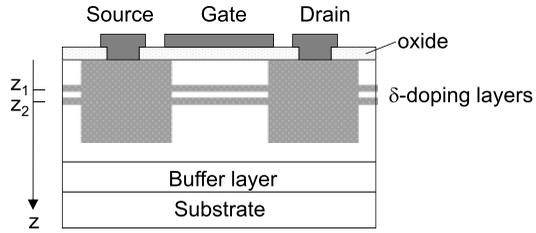
We assume that the z axis is along the sample growth direction. A buffer layer is followed by one (singly δ -doping) or two planar doped impurity layers (doubly δ -doping) whose doping profile is modeled as

$$N_D(\mathbf{r}) = \frac{N_{2D}}{\alpha\sqrt{2\pi}} e^{-\frac{(z-z_1)^2}{2\alpha^2}} \quad (4.94)$$

centered at $z = z_1$, where N_{2D} is the two-dimensional sheet doping density. A metal contact is mounted on top of the cap layer so that external bias can be applied to modify the electronic states in the doped regions. A typical δ -doped metal-oxide-semiconductor field-effect transistor (MOSFET) is schematically shown in Fig. 4.24.

We briefly discuss the carriers in the quantum well induced by the δ -doping layer. For the one-dimensional n -type Si system along the z axis, the description of the electron states is very simple by the effective-mass approximation, since the sample growth direction (001) is one of the principal axis on each of the six ellipsoids of constant energy. The six ellipsoids are grouped to:

Fig. 4.24 Schematic δ -doped metal-oxide-semiconductor field-effect transistor MOSFET. Here two δ -doping layers (centered at z_1 and z_2) are inserted



- (a) $m_z^* = 0.19, m_{xy}^* = 0.43, g = 4$
- (b) $m_z^* = 0.98, m_{xy}^* = 0.19, g = 2$

The effective masses are in the unit of free electron mass. g is the band degeneracy. As functions of the gate bias, the total charge distribution and the conduction bandedge are plotted in Fig. 4.25 for a doubly δ -doped Si field effect transistor. It is observed that only the lowest two sublevels are largely occupied, they are confined mainly around the δ -doped impurity layers. Remaining carriers are quite extended.

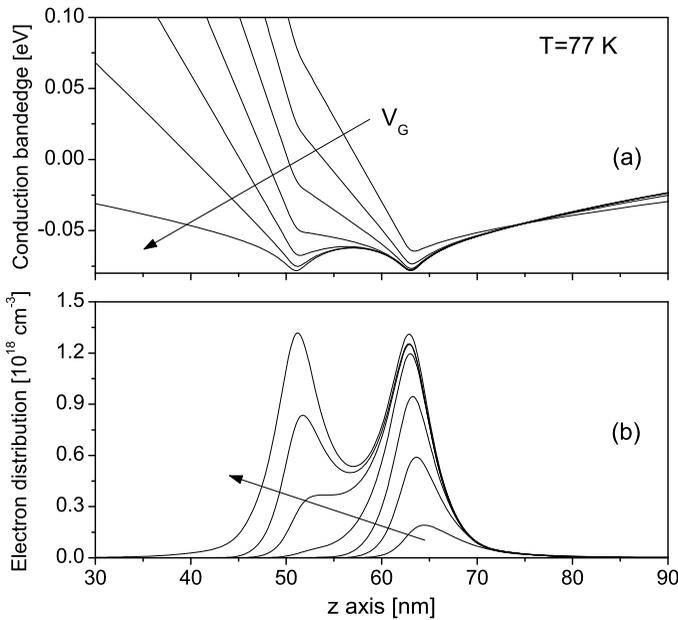


Fig. 4.25 Conduction bandedge (a) and electron distribution (b) of a doubly δ -doped Si field effect transistor calculated at 77 K. The arrows indicate the increase of the gate bias V_G (0.0, 0.25, 0.5, 0.75, 1.0, 1.25, 1.5 V). At $z = 0$ nm in the z axis is the oxide-semiconductor interface. The Fermi level is at -77.7 meV (Reprinted with permission from Y. Fu and M. Willander, Subband structure and ionized impurity scattering of the two dimensional electron gas in δ -doped field effect transistor, J. Appl. Phys., vol. 78, pp. 3504–3510, 1995. Copyright 1995, American Institute of Physics)

Figure 4.25 indicates that the depletion of the carriers confined in the doubly δ -doped region can be divided into two stages. There is only one quantum valley which is occupied in the conduction band of the doubly δ -doped field effect transistor when the gate bias $V_G = 0$. A gate bias of $V_G = 1.5$ V is needed in order to push down the second quantum valley close to the surface and get it occupied.

Let us focus on a simple picture about the impurity-scattering-limited carrier transport in the xy plan by considering the scattering of electrons by an unscreened Coulombic potential of an ionized impurity Eq. (4.79). Let $\lambda = |\mathbf{q}_{xy} - \mathbf{k}_{xy}|$,

$$\frac{g_0(\lambda|z - z_0|)}{\lambda} = \int \frac{e^{i(\mathbf{q}_{xy} - \mathbf{k}_{xy}) \cdot \boldsymbol{\rho}}}{|\mathbf{r} - \mathbf{r}_0|} d\boldsymbol{\rho} \quad (4.95)$$

$$g_1(\lambda, z_0) = \left| \int \psi_i^*(z) g_0(\lambda|z - z_0|) \psi_j(z) dz \right|^2$$

the scattering matrix element can be expressed as

$$s_{i\mathbf{k}_{xy} \rightarrow j\mathbf{q}_{xy}} = \frac{e^4 \delta}{8\pi \epsilon^2 \lambda^2 \hbar} \int N_D(\mathbf{r}_0) g_1(\lambda, z_0) d\mathbf{r}_0 \quad (4.96)$$

If the kinetic energy of the electron in the xy plane is rather small and/or the separation between impurities and carriers is small so that

$$\lambda|z - z_0| \approx 0$$

$$g_0 \approx 1$$

and

$$g_1 = \left| \int \psi_i^*(z) \psi_j(z) dz \right|^2 = \delta_{ij} \quad (4.97)$$

The last equality in the above expression comes from the wave function orthogonality. It is thus shown then that the inter-subband transitions are negligible if λ is small. The scattering matrix element becomes

$$s_{i\mathbf{k}_{xy} \rightarrow j\mathbf{q}_{xy}} \propto \int N_D(\mathbf{r}_0) d\mathbf{r}_0 \quad (4.98)$$

indicating that electrons with small kinetic energy or/and small spatial separation between electrons and impurities are completely under the influence of the impurities.

On the other hand, if the kinetic energy is very large (large λ),

$$g_0(\lambda|z - z_0|) \approx \delta_{z, z_0}$$

so that

$$g_1 = \psi_i^2(z_0) \psi_j^2(z_0) \quad (4.99)$$

and the scattering matrix element becomes

$$s_{ik_{xy} \rightarrow jq_{xy}} \propto \int N_D(\mathbf{r}_0) \psi_i^2(z_0) \psi_j^2(z_0) d\mathbf{r}_0 \quad (4.100)$$

i.e., the scattering probability is proportional to the overlapping of the electrons with impurities. Since the spatial extension of the doped impurities is much narrower (“ δ -doping”) than that of the electrons, the overlapping between electrons and impurities is expected to be very small. It is therefore concluded that the ionized impurity scattering probability is small when λ is large.

Assume that the ionized impurity scattering is the principal factor that determines the carrier mobility, the first advantage of the δ -doped field effect transistor over uniformly-doped materials is naturally the small overlapping between impurities and carriers. For a uniform material, the carriers overlap completely with impurities.

The energy band diagram of a homogeneous material is continuous, thus the averaged value of the \mathbf{k} vectors will not be very large, due to the large density of states along the z axis. However, along the z axis in the δ -doped field effect transistor, the electronic states become quantized so that the energy separations among subbands are very large. If the first excited state is to be occupied, the ground state must be occupied by electrons whose kinetic energies in the xy plane equal the energy difference between the first excited state and the ground state. By numerical calculations, it is easy to show that the energy difference can be 15 meV for singly δ -doped Si field-effect transistor and 5 meV for doubly δ -doped field-effect transistor. The Fermi level of a homogeneous 10^{20} cm^{-3} doped Si is very close to the conduction bandedge because of the large electron effective mass. From this theoretical analysis of ionized impurity scattering we thus expect an enhanced carrier mobility in a δ -doped field-effect transistor as compared with uniformly doped materials.

4.6 Nano-scale Field-Effect Transistor

By the fast development of the processing technology [42, 43], electronic devices are being made smaller and smaller [44], driven by the need to obtain larger and faster circuits (mostly CMOS) that consume less power [45]. In the early years, the MOSFETs had dimensions that made it possible for the device designer to rely on classical and semiclassical basic theories and fairly straightforward models. The evolution towards smaller sizes made these basic assumptions too crude, forcing the designer to take more and more special effects into account. Most important is the quantum effect in determining the small size device characteristics. Quantization effects in the inversion layer as well as along the channel can no longer be neglected.

4.6.1 Wave Characteristics and Threshold Voltage

We study the three-dimensional quantum effect by solving the three-dimensional Schrödinger equation

$$H(\mathbf{r})\psi_E(\mathbf{r}) = \left[-\frac{\hbar^2 \nabla^2}{2m_c^*} - e\phi(\mathbf{r}) + V_c(\mathbf{r}) \right] \psi_E(\mathbf{r}) = E\psi_E(\mathbf{r}) \quad (4.101)$$

and Poisson's equation

$$\nabla[\epsilon(\mathbf{r})\nabla\phi(\mathbf{r})] = -e[N_D(\mathbf{r}) - n(\mathbf{r})] \quad (4.102)$$

Here

$$n(\mathbf{r}) = \int N(E)f(E)|\psi_E(\mathbf{r})|^2 dE \quad (4.103)$$

is the carrier density at position \mathbf{r} . $N(E)$ is the density of states and $f(E)$ is the Fermi distribution function. $e > 0$ is the charge unit. N_D is the doping profile. The equations are written down for negatively charged electrons. $V_c(\mathbf{r})$ is the conduction bandedge profile. ϵ is the dielectric constant.

The calculation of the Poisson equation is straightforward if $n(\mathbf{r})$ is known. However, the three-dimensional Schrödinger equation, Eq. (4.101), is very complicated to the common knowledge. It is noticed that we do not need the complete information about the electron wave function, at least for the Poisson equation where only the amplitude of the wave function is needed when calculating $n(\mathbf{r})$. There is just one numerical method that serves perfectly our need. It is the recursion method developed by Haydock, Heine, and Kelly from the Green's function theory [46, 47]. The recursion method constructs a new set of normalized orthogonal basis, in which Hamiltonian $H(\mathbf{r})$ is turned into a tri-diagonal matrix. The diagonal elements of the Green's function of $H(\mathbf{r})$ are expressed as the continued fractions of the diagonal and off-diagonal elements of the tri-diagonalized Hamiltonian matrix, from which the amplitudes of wave functions and thus local densities of states (LDOS) are obtained.

To do numerical calculations, we first digitalize the electron coordinate space \mathbf{r} by a mesh structure. A simple mesh structure is cubic such $\mathbf{r} = (x, y, z) = (i\Delta, j\Delta, k\Delta)$, where i, j and k are integers, and Δ the cubic mesh size. For simplifying mathematical notations, we denote (i, j, k) by a single index ℓ , i.e.,

$$\ell = (i, j, k) \equiv (i\Delta, j\Delta, k\Delta) = (x, y, z) = \mathbf{r}$$

For example, for an electron in a space of $(i, j, k) \in (1 : 10, 1 : 20, 1 : 30)$, i.e., a rectangle of size $10 \times 20 \times 30 \Delta^3$, we can denote $\ell \in 1 : 10$ as $(i \in 1 : 10, j = 1, k = 1)$, $\ell \in 11 : 20$ as $(i \in 1 : 10, j = 2, k = 1), \dots$. We now construct a set of base func-

tions. The first base function $|1\rangle$ is

$$|1\rangle = \begin{pmatrix} \vdots \\ 0 \\ 1 \\ 0 \\ 0 \\ \vdots \end{pmatrix} \begin{matrix} \ell - 1 \\ \dots\dots \ell \\ \ell + 1 \\ \ell + 2 \end{matrix} \quad (4.104)$$

the second base function $|2\rangle$ and higher base function $|n\rangle$ are defined by

$$\begin{aligned} |2\rangle &= H|1\rangle - a_1|1\rangle \\ \dots \\ |n+1\rangle &= H|n\rangle - a_n|n\rangle - b_{n-1}|n-1\rangle \\ \dots \end{aligned} \quad (4.105)$$

In this set of base functions, Hamiltonian H is tri-diagonalized

$$H = \begin{pmatrix} a_1 & \sqrt{b_1} & 0 & 0 & 0 & 0 & \dots \\ \sqrt{b_1} & a_2 & \sqrt{b_2} & 0 & 0 & 0 & \dots \\ 0 & \sqrt{b_2} & a_3 & \sqrt{b_3} & 0 & 0 & \dots \\ 0 & 0 & \sqrt{b_3} & a_4 & \sqrt{b_4} & 0 & \dots \\ 0 & 0 & 0 & \sqrt{b_4} & a_5 & \sqrt{b_5} & \dots \\ \dots & & & & & & \dots \end{pmatrix} \quad (4.106)$$

where the diagonal and off-diagonal elements a_n and b_n are

$$a_n = \frac{\langle n|H|n\rangle}{\langle n|n\rangle}, \quad b_n = \frac{\langle n+1|n+1\rangle}{\langle n|n\rangle} \quad (4.107)$$

Since

$$G(\mathbf{r}, \mathbf{r}', E) = \langle \mathbf{r} | \frac{1}{E - H} | \mathbf{r}' \rangle$$

is the Green function of H , the local Green function $G(\mathbf{r}, \mathbf{r}, E)$ of H in the form of Eq. (4.106) is

$$G(\mathbf{r}, \mathbf{r}, E) = \frac{1}{E - a_1 - \frac{b_1}{E - a_2 - \frac{b_2}{E - a_3 - \dots}}} \quad (4.108)$$

we obtain the local density of states $\nu(\mathbf{r}, E)$ from the local Green function $G(\mathbf{r}, \mathbf{r}, E)$ at \mathbf{r} [48–50]

$$\nu(\mathbf{r}, E) = -\frac{1}{\pi} \lim_{\beta \rightarrow 0} \{\text{Im}[G(\mathbf{r}, \mathbf{r}, E + i\beta)]\} = \sum_n |\psi_n(\mathbf{r})|^2 \delta(E - E_n) \quad (4.109)$$

Modify the first base function in Eq. (4.104), for example, to

$$|1\rangle = \begin{pmatrix} \vdots \\ 0 \\ 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix} \begin{matrix} \ell - 1 \\ \ell \\ \dots \dots \ell + 1 \\ \ell + 2 \end{matrix} \quad (4.110)$$

we will find the local density of states at spatial location $\ell + 1$. And finally the carrier concentration at \mathbf{r} is

$$n(\mathbf{r}) = \int f(E)v(\mathbf{r}, E)dE \quad (4.111)$$

In this way, the self-consistent solution of Schrödinger and Poisson equations can be carried out without direct knowledge of the wave functions.

Note that the base functions $|n\rangle$ defined by Eqs. (4.105) are not normalized (they are orthogonal [46]). In numerical calculations it is however good to normalize the base functions. Denote X_n as the normalized function of $|n\rangle$, i.e.,

$$|X_n\rangle = \frac{|n\rangle}{\sqrt{\langle n|n\rangle}} \quad (4.112)$$

with some simple mathematical manipulations, we have the following recursive relationships

$$\begin{aligned} A_n &= \langle X_n|H|X_n\rangle \\ |n+1\rangle &= H|X_n\rangle - A_n|X_n\rangle - B_{n-1}|X_{n-1}\rangle \\ B_n &= \sqrt{\langle n+1|n+1\rangle} \\ |X_{n+1}\rangle &= \frac{|n+1\rangle}{B_n} \end{aligned} \quad (4.113)$$

and

$$a_n = A_n, \quad b_n = B_n^2 \quad (4.114)$$

In Fig. 4.26(a) we plot schematically a nano-scale transistor. The gate length is $L_G = 40$ nm, the thickness of the oxide layer between the gate and the Si layer is $t_{ox} = 3$ nm, the effective channel length is $L_{eff} = 25$ nm, and the ultra-shallow junction depth is $x_j = 10$ nm. Based on the experimental data, see Fig. 6 of [51, 52], the doping profile in the source and drain junctions is modeled by

$$N_D(x, z) = 10^{21-0.3z} \text{ cm}^{-3} \quad (4.115)$$

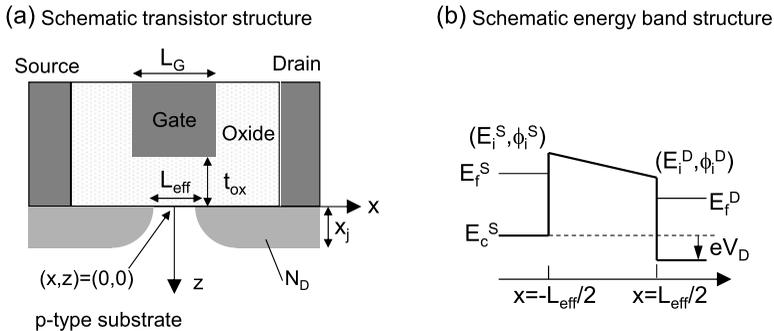


Fig. 4.26 (a) Schematic nano-scale transistor. The substrate doping level is 10^{18} cm^{-3} . (b) Energy diagram along the conduction channel

for $0 \leq z \leq 10$ and $|x| \geq 22.5$. For $0 \leq z \leq 10$, $12.5 \leq |x| \leq 22.5$ and $\sqrt{(x-22.5)^2 + z^2} \leq 10$,

$$N_D(x, z) = 10^{21-0.3\sqrt{(x-22.5)^2+z^2}} \text{ cm}^{-3} \quad (4.116)$$

Elsewhere, the doping level is p -type at a level of 10^{18} cm^{-3} . Here x and z are in the unit of nm. We set the potential energy of the substrate valence band edge as zero, and the gate as $|x| < 20$ nm and $z \leq 0$ so that $z \geq t_{\text{ox}}$ defines the Si region.

The calculated electron distribution is plotted in Fig. 4.27(a) based on Eqs. (4.101), (4.102). Using the classical model by setting $|\psi|^2$ as constant in classically allowed regions (total energy larger than the potential energy) and $|\psi|^2 = 0$ in classically forbidden region (total energy smaller than the potential energy), the calculated electron distribution is shown in Fig. 4.27(b). Because of the large built-in potential, there is a clear overlapping between electrons from the source and the drain even at zero gate bias in Fig. 4.27(b). From the usual depletion approximation, the width of the space-charge region is estimated to be a few hundred nanometers, much wider than the effective channel length of 25 nm designed in the experiment. The classical result of Fig. 4.27(b) is thus expected.

The difference between Fig. 4.27(a) and (b) is due to the wave property. At a boundary to a barrier region, the wave amplitude is small (for infinitely high barrier, the amplitude is zero). By the classical model, on the other hand, the amplitude is constant in the classically allowed region. Thus, the increase of the carrier concentration from a barrier region (classically forbidden region) is gradual in the wave picture. A distance from the barrier exists before the amplitude of the wave becomes significant.

The electron distributions at $V_G = 0.5$ volt have been calculated which show that at this gate bias, the electron waves from the source overlap with those from the drain, so that the conducting channel becomes open. When $V_G = 0.40$ volt, the electron concentration at the middle point between the source and the drain is about 10^{12} cm^{-3} , indicating the threshold voltage value. Experimentally, the extrapolated threshold voltage is 0.42 volt [51, 52].

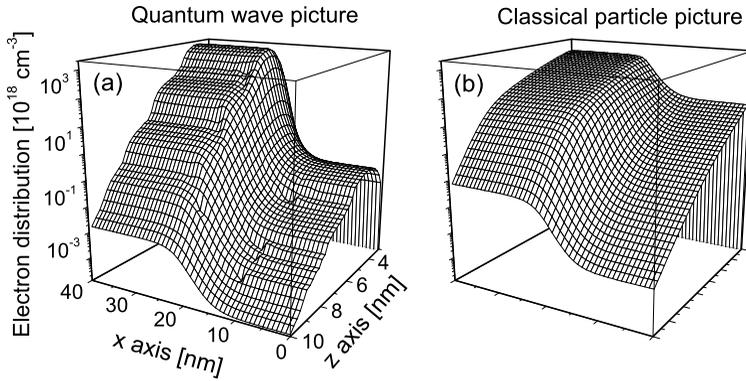


Fig. 4.27 Electron distributions based on (a) quantum mechanical wave model and (b) classical particle model. $V_G = 0$

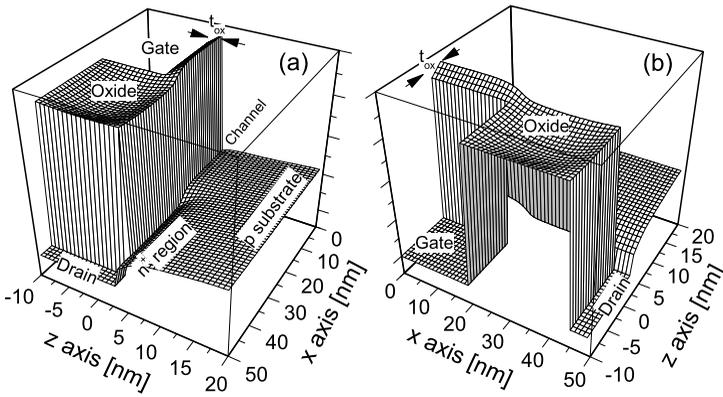


Fig. 4.28 Spatial distribution of the conduction band edge from two different viewing perspectives

Figures 4.28(a) and (b) show the conduction band edge profile when $V_G = V_D = 0$ from two different viewing perspectives.

4.6.2 Steady-State Wave Transport

Conductance quantization has been revealed experimentally when studying quantum ballistic transport of an electron through a narrow construction (quantum wire). As the number of transistors integrated on a circuit continues to increase, discrete device dimensions start to reach the nanometer region. Already in 1990 high-electron-mobility transistors with gate lengths as short as 25 nm were fabricated [53]. To analyze and design electronic devices on a sub-100 nm scale it is neces-

sary to go beyond the well established Boltzmann transport equations to develop simulation techniques based on quantum mechanical transport equations.

Experimental fabrication indicates clearly the normal transistor operation features in short channel transistors at room temperature [51, 52]. They also show quantum wave interference effect and conductance quantization at relatively high temperature [54–56]. Extensive theoretical works have been devoted in this fast developing device physics field [57], e.g., Monte Carlo simulation [58], non-equilibrium Green's function theory [59] and real-time Green's function formalism [60], Wigner function [61] and many more. Many theoretical investigations assume very idealized conduction channel and even zero electric field in the conduction channel. Here let us study the quantum effects in nano-scale electronic devices to connect the properties of quantum transport to the transistor operation characteristics.

We concentrate on the electron transport in the conduction channel in the form of a quantum wire, see Fig. 4.26(a). The source is normally grounded and the drain is biased at V_D so that its local potential energy is lowered by eV_D . Here we focus on the n -type field-effect transistor. The extension of the present discussion to the hole transport in a p -type transistor is straightforward. Because of the large spatial extensions of the source and the drain, quantum quantization effects are negligible in these junction areas.

Quasi equilibrium states are usually assumed in the junctions due to the high doping levels. The quasi equilibrium states are defined by two local Fermi levels, E_f^S and E_f^D , in the source and drain, respectively. When assuming the same doping levels in the source and drain,

$$E_f^S = E_f, \quad E_f^D = E_f - eV_D$$

where the source is grounded and V_D is the drain bias. See Fig. 4.26(b).

Quantum mechanical wave nature of the carrier is expected due to the reduced dimensions. The carrier is described by the Schrödinger equation

$$E\psi(\mathbf{r}) = \left[-\frac{\hbar^2 \nabla^2}{2m_c^*} + V(\mathbf{r}) \right] \psi(\mathbf{r}) \quad (4.117)$$

at steady state. Here $V(\mathbf{r})$ is the electric potential energy induced by electrons, the ionized impurities and the external gate and drain biases. m_c^* is the electron effective mass.

The boundary conditions for the above Schrödinger equation are given as: (i) electrons in the form of waves injected from the source moving toward the drain; (ii) electron waves injected from the drain moving toward the source. An electron from the source injected into the conducting channel is described by

$$\psi_i^S(y, z) e^{ik_i^S x} \quad (4.118)$$

where $\psi_i^S(y, z)$ and E_i^S are i th eigenfunction and eigenvalue in the yz plane in the source

$$E_i^S \psi_i^S(y, z) = \left[-\frac{\hbar^2}{2m_c^*} \left(\frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) + V_S(y, z) \right] \psi_i^S(y, z) \quad (4.119)$$

Here V_S is the potential energy at $x = -L_{\text{eff}}/2$. see Fig. 4.26b. The wave vector of the propagating wave, k_i^S must be real.

The electron wave becomes partially reflected from the conduction channel and the total wave function of the electron in the source region is

$$\psi(x, y, z)|_{x < -L_{\text{eff}}/2} = \psi_i^S(y, z)e^{ik_i^S x} + \sum_{\ell} r_{\ell} \psi_{\ell}^S(y, z)e^{-ik_{\ell}^S x} \quad (4.120)$$

The electron wave which is partially transmitted into the drain is described by

$$\psi(x, y, z)|_{x > L_{\text{eff}}/2} = \sum_{\ell} t_{\ell} \psi_{\ell}^D e^{ik_{\ell}^D x} \quad (4.121)$$

Here $(\psi_{\ell}^D, E_{\ell}^D)$ are eigenstates in the drain at $x = L_{\text{eff}}/2$

$$E_{\ell}^D \psi_{\ell}^D(y, z) = \left[-\frac{\hbar^2}{2m_c^*} \left(\frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) + V_D \right] \psi_{\ell}^D(y, z) \quad (4.122)$$

In the formalism of the quantum wave transport, the total electron energy, $E_{k^S, i}$, is conserved so that the sums over ℓ in Eqs. (4.120), (4.121) are restricted by the following equation

$$E_{k^S, i} = E_i^S + \frac{\hbar^2 (k_i^S)^2}{2m_c^*} = E_{\ell}^S + \frac{\hbar^2 (k_{\ell}^S)^2}{2m_c^*} = E_{\ell}^D + \frac{\hbar^2 (k_{\ell}^D)^2}{2m_c^*} \quad (4.123)$$

Here k_{ℓ}^S and k_{ℓ}^D may be imaginary to represent decaying waves in the source and drain.

The wave function ψ is thus obtained from Schrödinger equation, Eq. (4.117), by satisfying the wave function boundary conditions, Eqs. (4.120), (4.121), and the energy conservation Eq. (4.123).

The current density from the source to the drain,

$$i_S = e \sum_i \int \frac{2dk^S}{2\pi} f(E_{k^S, i}, E_f) \sum_{\ell} |t_{\ell}|^2 \frac{\hbar k_{\ell}^D}{m_c^*} \quad (4.124)$$

is obtained by inserting Eq. (4.121) into the standard quantum mechanical expression of probability flow density Eq. (2.10). Since the external electric field induced by the drain bias applies along the x direction, the component of the current in the yz plane is zero, only $\hbar k_{\ell}^D/m_c^*$ appears in Eq. (4.124).

A similar expression is obtained for the current density i_D transmitting from the drain to the source. The net current density I_D between the source and the drain is given by

$$I_D = i_S + i_D \quad (4.125)$$

The conductance G and the transconductance G_{tr} are defined as

$$G = \frac{\partial I_D}{\partial V_D}, \quad G_{tr} = \frac{\partial I_D}{\partial V_G} \quad (4.126)$$

Here V_G is the gate bias.

When the drain bias V_D is small as compared with the gate bias, its effect on the potential energy across the conduction channel in the yz plane can be approached as

$$V(x, y, z) = V(y, z) - eV_D \frac{x + L_{eff}/2}{L_{eff}}$$

where L_{eff} is the length of the conduction channel, Fig. 4.26(b). Since $V(y, z)$ is independent of x , the wave functions and the corresponding eigenenergies in the yz plane at position x in the conduction channel are

$$(\psi_\ell(y, z), E_\ell)|_x = \left(\psi_\ell^S(y, z), E_\ell^S - eV_D \frac{x + L_{eff}/2}{L_{eff}} \right)$$

It is then easy to see that because of the wave function orthogonality

$$\langle \psi_i(y, z) | \psi_\ell(y, z) \rangle = \delta_{i\ell} \quad (4.127)$$

In other words, sub-states in the yz plane do not mix with each other ($r_\ell = t_\ell = 0$ when $\ell \neq i$, where ψ_i is the incoming wave) during the wave transport through the conduction channel. The current density is then

$$i_S = -e \sum_i \int \frac{2dk_i^S}{2\pi} f(E_{k^S, i}, E_f) |t_i|^2 \frac{\hbar k_i^D}{m_c^*} \quad (4.128)$$

The wave transport through the conduction channel becomes effectively one dimensional along the x direction. In this case, an electron comes into the channel with wave

$$e^{ik_i^S x}$$

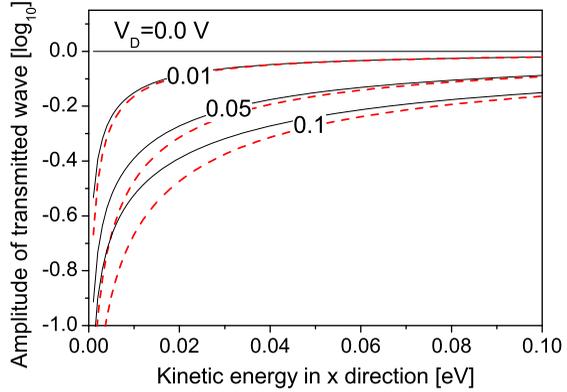
it is partially reflected

$$r_i e^{-ik_i^S x}$$

and is partially transmitted

$$t_i e^{ik_i^D x}$$

Fig. 4.29 Amplitudes of the transmitted waves $|t|^2$ through a conduction channel of $20 \times 100 \text{ nm}^2$ at $V_D = 0.00, 0.01, 0.05, 0.10 \text{ V}$. Dashed lines are obtained by Eq. (4.129)



We first consider a zero-length conduction channel so that we have the following boundary conditions for the above three waves due to the continuities of the waves and their first-order derivatives

$$1 + r_i = t_i, \quad k_i^S(1 - r_i) = k_i^D t_i \longrightarrow t_i = \frac{2}{1 + \frac{k_i^D}{k_i^S}} \quad (4.129)$$

Here we see that if $V_D = 0$, $t_i = 1$. When the bias is increased, t_i decreases because

$$\frac{\hbar^2(k_i^D)^2}{2m^*} = eV_D + \frac{\hbar^2(k_i^S)^2}{2m_c^*} \quad (4.130)$$

due to energy conservation. Equation (4.129) however is valid for a conduction channel of zero length. For realistic situations, we have to calculate the amplitude of the transmission wave, t_i , by numerical methods.

Numerical results of the amplitudes of the transmitted waves at different bias are plotted in Fig. 4.29, where we have assumed that the conduction channel is 20 nm thick in the y direction and 100 nm long in the x direction (the width is assumed to be wide enough so that the quantization effect in the z direction is neglected). When the waves are injected from the source, $|t|^2 \leq 1$; when the waves are injected from the drain, $|t|^2 \geq 1$, see Eq. (4.129). It is observed here that the amplitude of the transmitted wave is unity only when V_D is exactly zero. When V_D is increased, the amplitude of the transmitted wave decreases when the wave propagates from high potential energy region to low potential energy region, while it increases when the wave is transmitting in the opposite direction.

The amplitudes calculated by Eq. (4.129) are also plotted in Fig. 4.29 as dashed lines for comparison. Deviation becomes significant when the wave vector, either k^S or k^D is small. However, the current density is also very small when the wave vector is small so that the errors in calculating the conduction current introduced by using Eq. (4.129) are small. In the analytical analysis, we can approximate the amplitude of the transmitted wave, $|t|$, by Eq. (4.129).

We now need to study the relationship between E_f and E_i in a realistic device before further discussing the $I - V$ characteristics and conductances. We consider a typical high-electron-mobility transistor (HEMT). Since the HEMT is usually quite wide, the quantum confinement effect in the z direction is neglected,

$$\psi_i(y, z) = \psi_i(y)e^{ik_z z}, \quad E_i(y, z) = E_i(y) + \frac{\hbar^2 k_z^2}{2m_c^*} \quad (4.131)$$

where $\psi_i(y)$ and $E_i(y)$ are energy sublevels in the y direction.

By the relationship of Eq. (4.129), the total current density from the source to the drain becomes

$$I_D = \frac{8e}{h} \sum_i \int_0^\infty \frac{dk_z}{2\pi} \times \int_0^\infty dE_x \frac{\sqrt{E_x(E_x + eV_D)}}{(\sqrt{E_x} + \sqrt{E_x + eV_D})^2} [f(E, E_f) - f(E, E_f - eV_D)] \quad (4.132)$$

after necessary mathematical manipulations. Here

$$E = E_i(y) + E_x + \frac{\hbar^2 k_z^2}{2m_c^*}, \quad E_x = \frac{\hbar^2 k_x^2}{2m_c^*} \quad (4.133)$$

We first discuss the quantum transport at low temperature with an infinitely small drain bias $eV_D \ll E_x$ so that

$$\frac{\sqrt{E_x(E_x + eV_D)}}{(\sqrt{E_x} + \sqrt{E_x + eV_D})^2} \approx \frac{1}{4} \quad (4.134)$$

Moreover, at extremely low temperature,

$$f(E, E_f) = \begin{cases} 1 & E \leq E_f \\ 0 & \text{otherwise} \end{cases}$$

so that Eq. (4.132) becomes

$$I_D = \frac{2e}{h} \sum_i \int \frac{dk_z}{2\pi} \int dE_x [f(E, E_f) - f(E, E_f - eV_D)] \\ = \frac{2e^2 V_D}{h} \sum_i \sqrt{\frac{2m_c^*(E_f - E_i)}{h^2}} \quad (4.135)$$

where the sum over i in the above equation is restricted in such a way that $E_f \geq E_i$, as we discussed early for real k_i^S .

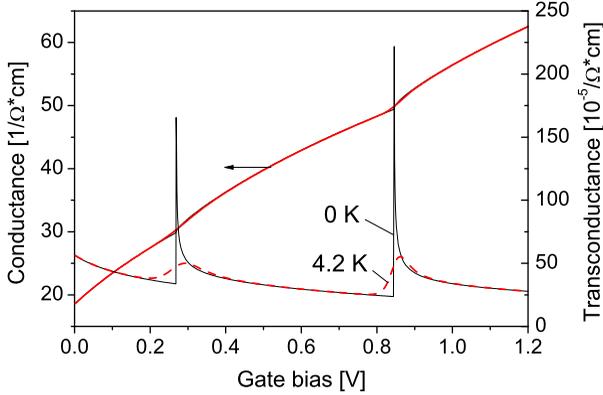


Fig. 4.30 Conductance and transconductance of quantum wave transports in an AlGaAs/GaAs HEMT. $V_D = 10^{-5}$ V and $T = 0.0$ (solid lines) and 4.2 K (dashed lines) (Reprinted from Y. Fu and M. Willander, Carrier transport characteristics of small size field effect transistors, *Physica E*, vol. 4, pp. 149–155, 1999, with permission from Elsevier)

Since

$$N(E_f) = \sum_i \sqrt{\frac{2m_c^*(E_f - E_i)}{h^2}} \quad (4.136)$$

is the total number of energy sublevels in the conduction channel at $x = 0$ which are occupied, we obtain

$$I_D = \frac{2e^2 V_D}{h} N(E_f) \quad (4.137)$$

It is easy to show that the above equation is rather general. It is valid when the width of the transistor is reduced along the z direction so that the electron states there become discrete.

This gives us rather simple expressions for conductance and transconductance at low temperature and low bias

$$G = \frac{2e^2}{h} N(E_f), \quad G_{tr} = \frac{2e^2 V_D}{h} \frac{\partial N(E_f)}{\partial V_G} \quad (4.138)$$

In Fig. 4.30 we plot the calculated conductance and transconductance as functions of the gate bias at a rather small drain bias ($V_D = 10^{-5}$ volt) for quantum wave transport in the AlGaAs/GaAs HEMT of Fig. 4.22. The discontinuities in the conductance and singularities in the transconductance correspond to new occupied sublevels in the conduction channel following the increase of the gate bias. It is observed that the thermal excitation easily smears the sharp structures in the transconductance.

When the gate bias is increased, there will be more occupied sublevels in the conduction channel, and thus more peaks in the $G_{tr} - V_G$ relationship. We have

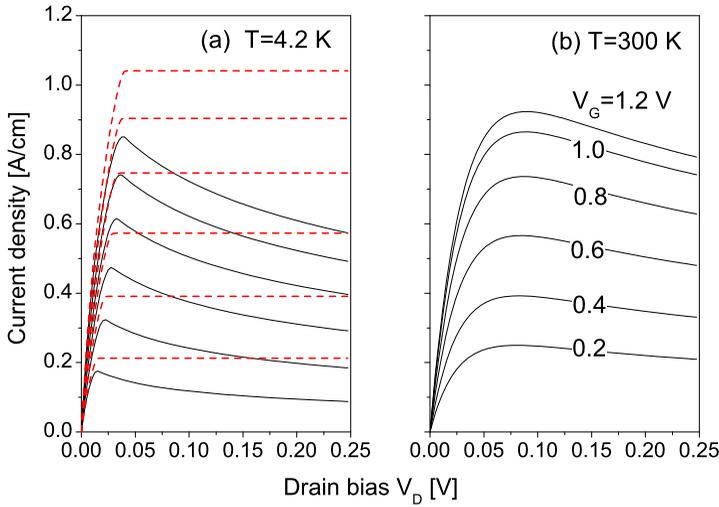


Fig. 4.31 Calculated $I_D - V_D$ characteristics of the HEMT at (a) 4.2 K and (b) 300 K. The dotted lines in (a) are obtained by setting $|t|^2 = 1$ (Reprinted from Y. Fu and M. Willander, Carrier transport characteristics of small size field effect transistors, Physica E, vol. 4, pp. 149–155, 1999, with permission from Elsevier)

thus obtained oscillations in the transconductance as a function of the gate bias. It has been shown that in addition to the Coulombic blockade effect in the small field-effect transistor, this quantum transport of electron waves through energy sublevels in the conduction channel also induces oscillations in the transconductance [62].

In Fig. 4.31 we present the numerically calculated $I_D - V_D$ characteristics of the HEMT at 4.2 K and 300 K. We first look at the dotted lines in Fig. 4.31(a) which are calculated by setting $|t|^2 = 1$. We observe the expected linear relationship between I_D and V_D at low drain bias. The current density becomes saturated when the drain bias is much increased to such a value that the local Fermi level in the drain junction is below the conduction bandedge of the source. At this moment, all the states occupied in the source contribute to the total current and all the occupied states in the drain become reflected from the conduction channel.

However, the current density (solid lines) never reaches its optimal value at perfect transmission ($|t|^2 = 1$). The wave reflection is enhanced by the large potential energy difference between the source and the drain due to the drain bias. When the drain bias is large enough,

$$I_D \propto \frac{1}{\sqrt{V_D}} \tag{4.139}$$

which is easily deduced from Eq. (4.132). The above equation is clearly demonstrated in Fig. 4.31(a).

When the temperature is increased, high kinetic energy states become occupied, thus the critical value of the drain bias at which Eq. (4.139) becomes valid is much increased, as shown in Fig. 4.31(b).

4.6.3 Interface Roughness

Molecular beam epitaxy was rapidly developed for heterostructures of various carrier confining potentials. It is however realized that due to the inherent nature of the growth kinetics and mechanisms, the width and depth of the confining potential are in fact functions of spatial coordinate in both the lateral (in the plane perpendicular to the growth direction) and vertical (along the growth direction) directions [63]. In other words, a certain degree of lateral and vertical intermixings of the two materials at their heterointerface is inevitable. The lateral intermixing is usually referred as the interface roughness, which is expected to influence directly the performance of an electron device based on the heterointerface [64, 65].

Consider an AlGaAs/GaAs heterointerface. Let the growth direction be the z axis and the lateral plane the xy plane. The interface roughness is mathematically expressed by a potential function as $\delta V(x, y, z = 0)$ ($z = 0$ is the position of the interface), which varies from zero to ΔE_c where an Al atom protrudes completely into the GaAs layer. The interface roughness also introduces a potential change along the z direction. However, since the lattice constant (5.6 Å) is very small compared with the electron extension (the wave extension along the z direction is about 15 nm in common AlGaAs/GaAs quantum wells), this potential change can be neglected.

Because of the complication of the exact nature of the interface roughness, the interface roughness is modeled here by three random numbers, namely, a_i , x_i and y_i in the following way

$$\delta V(x, y, 0) = \Delta E_c \sum_i a_i \theta(x - x_i + 5) \cdot \theta(x_i + 5 - x) \cdot \theta(y - y_i + 5) \cdot \theta(y_i + 5 - y) \quad (4.140)$$

where

$$\theta(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

The position of the protruding Al atom, (x_i, y_i) are randomly generated by the computer. The number 5 here represents the extension of the interface roughness in the unit of Å, which approximately is the size of the AlAs unit cell. The concentration of the protruding Al atoms varies from zero to its maximal value of the Al concentration in AlGaAs. The last random number, $a_i \in (0, 1)$, controls the depth of the Al atom protruding into the GaAs layer.

Let us set the conduction channel as 20 nm wide. When the width of the channel is increased, more sublevels in the y direction become occupied and the current increases accordingly while the physics about transport remains unchanged.

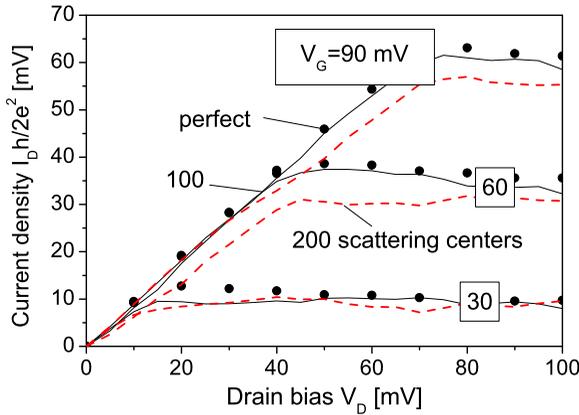


Fig. 4.32 $I_D - V_D$ characteristics through perfect conduction channel, a conduction channel with 100 and 200 scattering centers due to the interface roughness. From the low curve to the high one, the corresponding gate biases are $V_G = 30, 60$ and 90 mV (Reprinted from Y. Fu, Y. Mu, and M. Willander, Interface roughness effect on the $I - V$ relation of a AlGaAs/GaAs heterojunction field effect transistor, *Superlattices Microstruct.*, vol. 23, pp. 417–425, 1998. Copyright 1998, with permission from Elsevier)

Let us also set the channel length quite short. Without any scattering centers in the conduction channel, the channel length does not affect very much the current. When the interface roughness is introduced, short channel length can cause fluctuations in the $I_D - V_D$ characteristics because of the small number of scattering centers involved in the calculation. This problem is avoided by averaging the current values calculated for many different scattering center configurations.

Figure 4.32 shows the calculated $I_D - V_D$ characteristics as a function of the gate bias V_G for different scattering center concentrations. We have only plotted the results of the transport of carriers occupying the ground sublevel in the z direction. At high V_G , high sublevels in the z direction are occupied and their contributions to I_D are additive.

Now we consider in the similar manner the impurity scattering potential for carrier wave transmission in the same AlGaAs/GaAs high-electron-mobility transistor. The problem has been addressed in the last section where we have treated the remote ionized impurity scattering by the scattering matrix method, the electron mobility was derived from the scattering matrix element. The method was in principle semi-classical. Here let us study the wave transmission through the potential induced by the ionized impurity.

The spatial position of an ionized impurity, (x_ℓ, y_ℓ, z_ℓ) , is generated in the computer by random numbers. The impurities concentration is accounted by N_D . L is the thickness of the AlGaAs spacer layer.

Figure 4.33 shows the calculated $I_D - V_D$ characteristics for two different impurity concentrations and configurations. The results are configuration-dependent because of the small number of impurities present in the short channel that can be

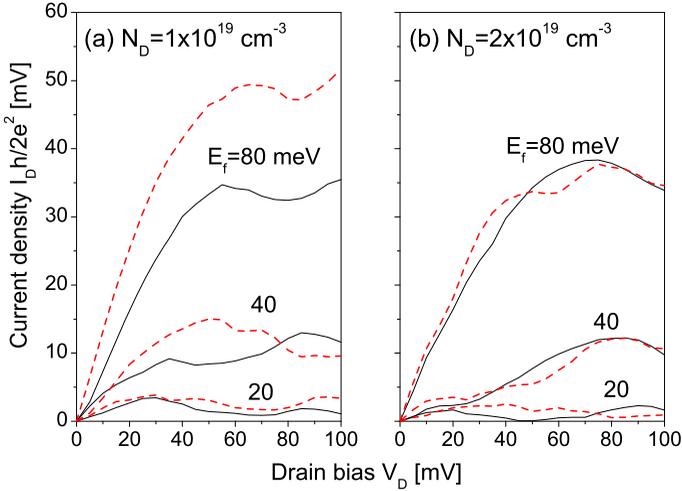


Fig. 4.33 $I_D - V_D$ characteristics for two different impurity concentrations and two different configurations (Reprinted with permission from Y. Fu and M. Willander, Near ballistic transport and current-voltage characteristics of a GaAs/AlGaAs heterojunction field effect transistor under the influence of impurity scattering, *J. Appl. Phys.*, vol. 82, pp. 5227–5230, 1997. Copyright 1997, American Institute of Physics)

numerically investigated in the computer system. In Figs. 4.33, the impurity concentrations are $N_D = 10^{19}$ and $2 \times 10^{19} \text{ cm}^{-3}$, respectively. They are about ten times higher than the real impurity doping concentration. High doping level has been chosen to avoid too strong numerical fluctuations in the current due to the small number of scattering centers experienced by the carriers.

Similar effect on the conduction current is observed here due to the ionized impurity scattering as the interface roughness effect.

In addition to the general feature of current reduction due to the interface roughness and ionized impurity scatterings, the effect of the current reduction caused by the drain bias ($I_D \propto 1/\sqrt{V_D}$ discussed in the last section) is reduced because of the higher scattering probabilities when the electronic kinetic energy is small (low V_D). Thus, the critical value of the drain bias at which I_D becomes significantly reduced by V_D is further pushed upwards.

We have thus shown that the ionized impurities and interface roughness are two very important factors in determining the electron transport properties in quantum-size electronic devices. Combining Figs. 4.32 and 4.33, it can be concluded that for the AlGaAs/GaAs heterojunction field-effect transistor, the source-drain current is optimal when the AlGaAs spacer is 10 nm thick. This is very close to the result derived in the last section.

4.6.4 Time-Dependent Wave Packet Transmission

We have thus far studied the carrier transport at steady state. Because of the reduced device size and increased device operation speed, we now study the time-dependent carrier transport from the source to the drain in the Si nano-scale field-effect transistor of Figs. 4.26(a), 4.27, 4.28 in Sect. 4.6.1.

In Sect. 4.3.2 we already discussed the response of the quantum transport in a double-barrier resonant-tunneling structure to a time-dependent perturbation. The superposition was performed by coupling the incoming plane wave at energy E with waves at $E \pm i\hbar\omega$, where ω is the frequency of the time-dependent perturbation and i is an integer. Furthermore, we have discussed the quantum wave transmission in Sect. 2.1. The quantum wave transmission scheme was adopted to study the carrier conduction through a heterostructure barrier varactor induced by an ac bias [66]. The problem we have now is described by the following time-dependent Schrödinger equation

$$H\psi(x, z, t) = i\hbar \frac{\partial \psi(x, z, t)}{\partial t} \quad (4.141)$$

$$H = -\frac{\hbar^2}{2m_c^*} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} \right) + V(x, z, t)$$

where m_c^* is the electron effective mass. It is reminded here that the single effective mass is only used to simplify the mathematical expressions, whereas in the numerical calculations we have to take into account the six ellipsoids in the Si conduction band. $V(x, z, t)$ is the potential energy. We have assumed translation symmetry in the y direction so that the y direction is not explicit in Eq. (4.141). However, the kinetic energy of the free motion in this direction has to be included (see later).

Because of the high doping concentrations in the contact areas, local quasi-equilibrium states are assumed with well defined local quasi Fermi levels, E_f^S and E_f^D in the source and drain, respectively. Electrons are injected to the conduction channel from both the source and the drain. We consider an initial wave packet at $t = 0$

$$\psi_0^S(z) \int_0^\infty \frac{2dk_x}{2\pi} \frac{e^{ik_x(x+40)}}{\sqrt{1 + \exp\left(\frac{E_x + E_y + E_0^S - E_f^S}{k_B T}\right)}} \quad (4.142)$$

which is centered at $x = -40$ nm. Here $E_x = \hbar^2 k_x^2 / 2m_c^*$, $E_y = \hbar^2 k_y^2 / 2m_c^*$. $[\psi_0^S(z), E_0^S]$ is the ground state along the z direction in the source. k_y in Eq. (4.142) is the wave vector of the y -direction free motion. It must be included since the occupation of electron states in the source is limited by the local Fermi level E_f^S . Therefore a fast motion in the y direction indicates in general a slow motion in the x direction. Moreover, it is noticed that the state of E_x is two-fold degenerate, one with k_x and the other $-k_x$. Only state k_x is included in the above wave packet since we consider the motion of a wave packet from the source to the drain along the x

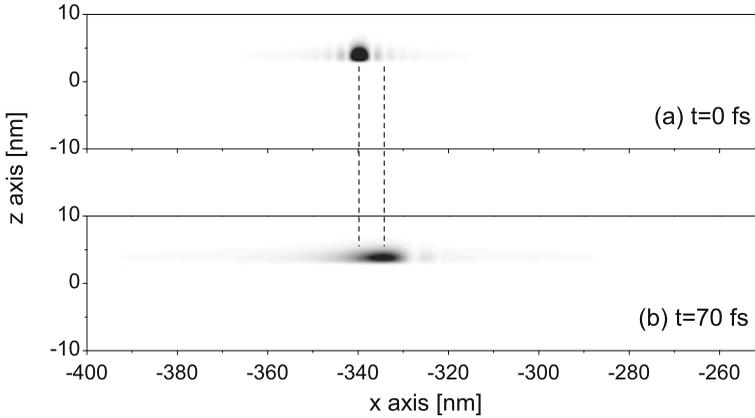


Fig. 4.34 Wave packet transmission in the shallow source junction. $t_{\text{ox}} = 3$ nm so that $z < 3$ nm is the oxide region

direction. Similar expressions can be written down for waves transmitting from the drain to the source.

Numerically it is obtained (Sect. 4.6.1) that $E_0^S = 0.2202$ and $E_f^S = 0.5126$ eV. Inserting them into Eq. (4.142) and setting $k_y = 0$, we find the wave function of an initial electron wave packet. First we put the wave packet deep in the source ($x = -340$ nm). The motion of this wave packet (its amplitude) in the source junction is shown in Fig. 4.34. The wave packet moves along the x direction. At the same time, it becomes diffused, as can be expected.

The speed of the wave packet is determined by its average kinetic energy. By taking an electron with a kinetic energy of $E_f^S - E_0^S = 0.2924$ eV (the maximum kinetic energy in the source along the x direction when neglecting thermal excitation) and an effective mass of $0.7m_0$, the distance that the electron moves in 70 fs is about 27 nm. In Fig. 4.34 we observe the travel distance of about 7.5 nm for the wave packet.

Now we put the electron wave packet close to the conduction channel, e.g., $x = -40$ nm. Figures 4.35 and 4.36 show the motion pictures of the amplitude of the wave packet in two different MOSFETs under the same gate bias (the drain bias is set to be $V_D = 0.5$ V). A gate leakage current is observed when the gate oxide layer becomes too thin. The gate leakage current can also be induced by a high gate bias ($|x| < 20$ nm and $z \leq 0$ is the gate region).

To show the wave function in the barrier regions including the conduction channel and the gate oxide layer, we have used the logarithmic scale in Figs. 4.35 and 4.36. The wave packet which transmits to the drain is however very diffusive due to the large drain bias of $V_D = 0.5$ V. Here we have neglected all kinds of scattering processes so that the transmitted wave packet gains 0.5 eV from the drain bias. The transmitted wave packet thus moves very fast, and its spatial profile becomes diffused. The diffusive distribution of the leakage wave in the gate is due to the

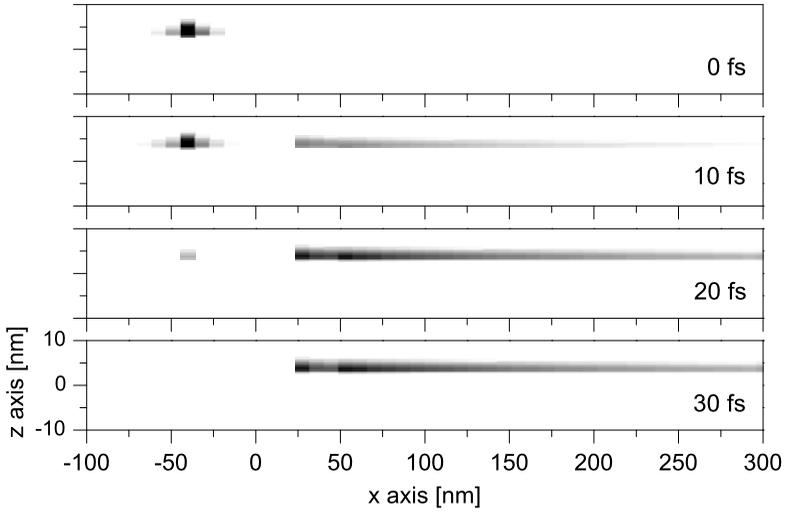


Fig. 4.35 Wave packet transmission. $t_{ox} = 3$ nm, $V_G = 0.8$ V

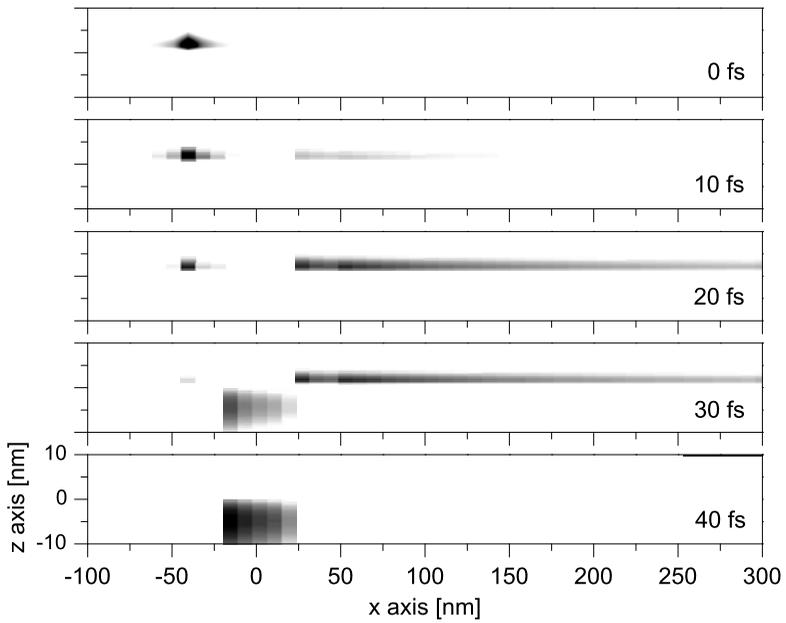
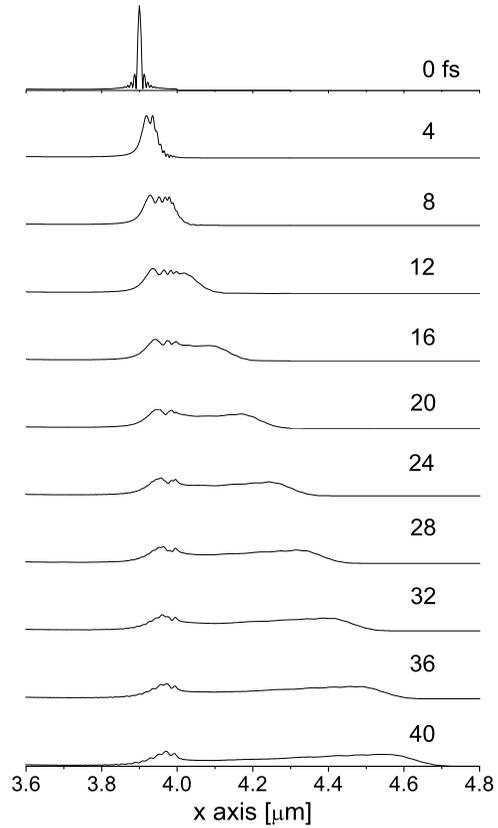


Fig. 4.36 Wave packet transmission. $t_{ox} = 1$ nm, $V_G = 0.8$ V

same reason. The results are more clearly reflected in a one-dimensional transport in Fig. 4.37.

Fig. 4.37 One-dimensional wave packet transmission in a potential field of $V(z) = 0$ when $z < 0.4 \mu\text{m}$; $V(z) = -0.5 \text{ eV}$ when $z > 0.5 \mu\text{m}$; for $0.4 < z < 0.5 \mu\text{m}$, $V(z)$ changes linearly from 0 to 0.5 eV (Reprinted with permission from Y. Fu and M. Willander, Electron wavepacket transport through nanoscale semiconductor device in time domain, *J. Appl. Phys.*, vol. 97, p. 094311(7), 2005. Copyright 2005, American Institute of Physics)



One remarkable effect of the tunneling transport characteristics is its speed. By Fig. 4.35 we see that the tunneling of the wave packet through the conduction channel (25 nm) is complete in about 40 fs. However, it must be indicated that such a speed is due to the 0.5-V drain bias. Letting $V_D = 0$ and re-doing the numerical calculation it is easy to show that the speed of the wave packet tunnel is almost identical to the free motion of the wave packet as Fig. 4.34.

By analyzing carefully Figs. 4.35 and 4.36, we conclude that for the nanoscale MOSFET under investigation (40-nm gate length, 25-nm channel length, and biased at a drain bias of 0.5 V), a significant gate current is induced when the gate bias exceeds 0.7 V ($t_{\text{ox}} = 3 \text{ nm}$). For $t_{\text{ox}} = 1 \text{ nm}$, the same gate current is induced when the gate bias is only 0.2 V; when the gate bias reaches 0.8 V, the whole wave packet tunnels to the gate bias. These results are obtained when $V_D = 0.5 \text{ V}$, they are expected to be rather different if the drain bias is much reduced.

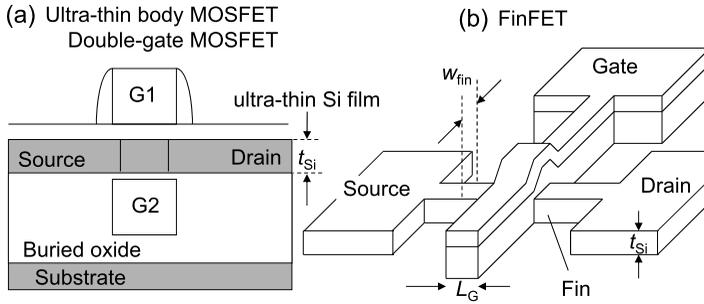


Fig. 4.38 (a) Without gate G2, the device is referred to as ultrathin body MOSFET. With G2, it is called the double-gate MOSFET. The thin body/film ensures that no conduction path is far from the gate(s) to suppress the leakages. (b) FinFET. Formation of the narrow fin, low-resistance source/drain contacts will be a primary challenge for $L_G < 20$ nm

4.6.5 Nanometer MOSFET Architectures

The effective channel length approaches 10 nm, and at the same time, the source and drain junction depths (x_j in Fig. 4.26) are also being drastically reduced. Moreover, the width of the conduction channel must be greatly reduced for the sake of total component miniaturization. Eventually the conduction channel is to take the form of a short quantum wire or, effectively, a three-dimensional quantum dot, with a nanoscale in its width, thickness and length.

The MOSFET structure of Fig. 4.26 is normally referred as the planar bulk MOSFET. Ultrathin body MOSFET and FinFET [67, 68] of Fig. 4.38, with ultrathin Si film on insulator (SOI) [69], are expected to constitute mainstream nanometer CMOS technology due to excellent scalability, superior immunity against short-channel effects, near-ideal subthreshold slope, high near-ballistic drive-current and transconductance, low subthreshold intrinsic capacitance, improved electrical isolation, the possibility of optimal operation with relatively low channel doping, and certain advantageous circuit characteristics. And, above all, they can be fabricated with a process flow that utilizes techniques and tools relatively similar to those currently used for fabricating planar bulk MOSFETs. The ultrathin body SOI may be either a single-gate or a double-gate MOSFET. The double-gate MOSFET has superior scalability (i.e., superior control of short-channel effects) for very small MOSFETs due to the electrical shielding action of the bottom gate for electric fields originating from charges in the source and drain. However, there are major issues with the complicated process flow required (hence high cost) and difficult manufacturability issues.

In addition to the nano MOSFET configurations of Fig. 4.38, there are other variations including silicon-on-nothing [70], vertical MOSFET [71, 72], and ballistic tunneling MOSFET (Fig. 4.39), when the dimension of the conduction channel becomes less than 10 nm [73].

The main issue that the MOSFET designer has to consider here is the distribution of the electric field in the nanoscale field-effect transistor. The electrodes in the

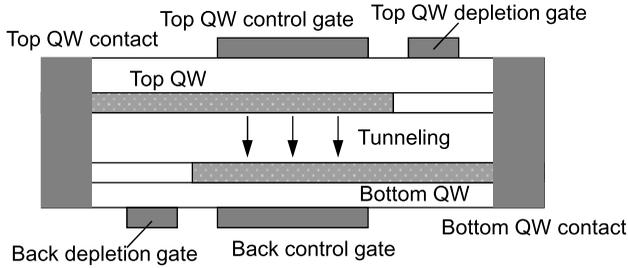


Fig. 4.39 Schematic of the double-electron-layer tunneling transistor

conventional-size MOSFET are rather extended in space so that they generate rather uniform electric fields which are used to control the electron flow in the device. When the feature size of the device reaches nanoscale, the electrodes become sheets and wires, and the electric fields from them are expected to be rather diffused. Figure 4.40(a) shows the geometric structure of the gates in a dual-gate MOSFET [74]. The lower gate was designed to create a conduction channel between the source and the drain and the upper gate was supposed to deplete carriers directly underneath the metal bars, thus cutting the conduction channel into a series of quantum islands. However, self-consistent calculation show that the electric fields from different metal bars in the upper gate merge into a rather smooth potential barrier, see Figs. 4.40(b) and (c).

The reason is very simple: The upper gate can be simplified as an array of metal cylinders extended along the y axis. Since the electric potential of a metal cylinder is in the form of $\ln \sqrt{(x - x_0)^2 + (z - z_0)^2}$, where (x_0, z_0) is the center of the cylinder, the solution for the corresponding Poisson equation of the array of metal cylinders is

$$\phi(x, z) = A \sum_{i=-3}^3 \ln \sqrt{(x - id)^2 + z^2} + B \quad (4.143)$$

where A and B are constant to be determined by boundary conditions. $d = 300$ nm is the distance between two adjacent metal cylinders, and there are 7 cylinders ($i = -3, -2, -1, 0, 1, 2, 3$). $\phi(x, z)$ shows significant spatial variations along the x direction when z is small (in terms of the spatial separation d between adjacent metal cylinders). However, when $z > d$, $\phi(x, z)$ is only a big potential barrier, as observed in Figs. 4.40(b) and (c). We thus see that the upper gate only makes the conduction channel, which is about 300 nm beneath the upper gate, narrower.

A good solution to the problem is the gate-all-around (GAA) MOSFET configuration. Figure 4.41 shows the device schematic of coaxially gated silicon nanowire transistor structure. The cross-section of the as-grown nanowire consists of a p -doped Si core (10 nm) with subsequent layers of i -Ge (10 nm), SiO_x (4 nm), and p -Ge (5 nm). The source and drain electrodes are in contact with the inner i -Ge core, while the gate electrode is in contact with the outer p -Ge shell and electrically

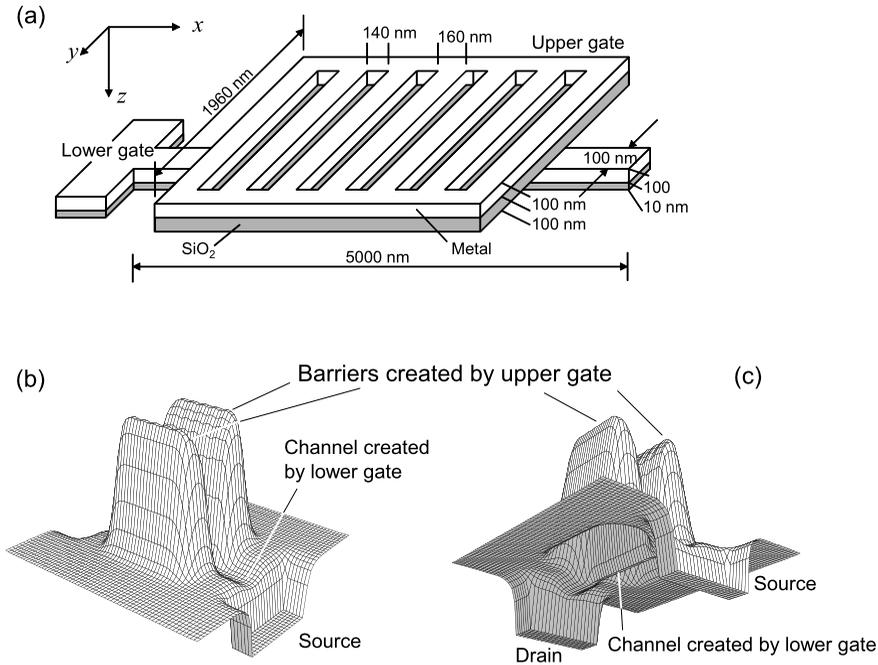


Fig. 4.40 (a) The dual-gate structure and geometric parameters of the Si metal-oxide-semiconductor field-effect transistor. (b) and (c): potential energy profile (different perspectives) at the Si side of the interface between oxide layer and Si. The height of the barriers induced by the upper gate is 0.81 eV, while the depth of the conducting channel is -0.145 eV. The upper gate is biased at 6 V while the lower gate at -3.5 V (Reprinted from Y. Fu and M. Willander, Energy band diagram of nanostructure field-effect transistor, Surf. Sci., vols. 361–362, pp. 500–504, 1996, with permission from Elsevier)

isolated from the core by the SiO_x layer [75]. However, the fabrication and processing of such a configuration is rather complicated. In addition, we have to consider the integration of the components into circuits [76].

Figure 4.42 shows the conduction current of the transmitting wave packet as a function of the kinetic energy along the quantum wire (defined as the x direction, see Fig. 4.41) with a diameter of 10 nm. Through a perfect quantum wire, the conduction current is proportional to the momentum of the wave packet, which in its turn is proportional to the square root of the kinetic energy, see curve (a). The other two lines in the figures represent the currents of the transmission electron under the influence of an ionized impurity located at (b) the center and (c) the edge of the silicon quantum wire, respectively. The quantum wire is $1.0 \mu\text{m}$ long coordinated along the x axis, $x \in (-0.5, 0.5) \mu\text{m}$, and the impurity locates at $x = 0$.

Moreover, a resonance is observed at about 85 meV in the current spectra, as can be further clarified by Fig. 4.43, which compares the wave packet transmissions at three electron energies, 40, 85 and 300 meV. A small resonance can be distinguished in curve (c) of Fig. 4.42 at about 20 meV.

Fig. 4.41 Schematic of coaxially gated nanowire transistors structure

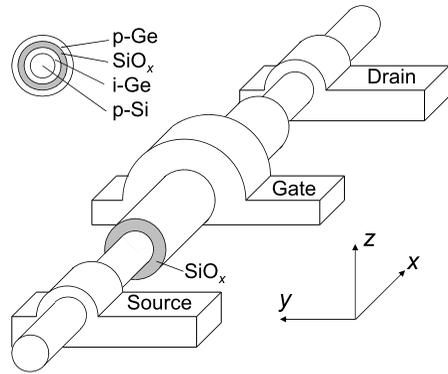
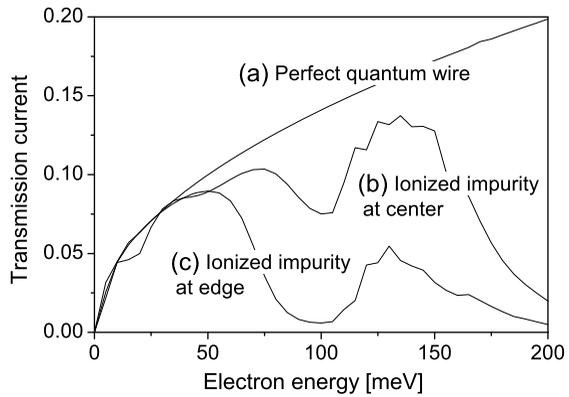
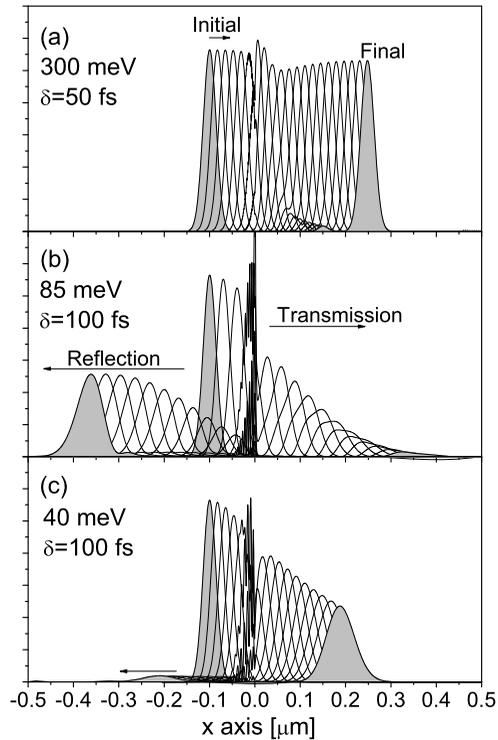


Fig. 4.42 Conduction current of the transmitting wave packet as a function of the electron energy E .
 (a) The perfect wavepacket transmission of \sqrt{E} .
 (b) There is an ionized impurity at the center of the silicon quantum wire.
 (c) The ionized impurity locates at the edge of the quantum wire



It has thus been shown that one single impurity considerably reduces the carrier conductance. The volume of a wire with the present dimensions (10 nm diameter, 1.0 μm length) is about $7.85 \times 10^{-17} \text{ cm}^3$ and the Si/SiO₂ interface area is about $3.14 \times 10^{-10} \text{ cm}^2$. To limit the number of impurities per wire to one, as assumed in the calculation, the concentration of bulk impurities must be limited to about $1.3 \times 10^{16} \text{ cm}^{-3}$, while the total concentration of surface states must not be higher than about $3.2 \times 10^9 \text{ cm}^{-2}$. The former value is higher than what would be expected from existing silicon processes and can probably be easily satisfied. However, the maximum surface state concentration required is in the lower range of what is available in the best planar Si/SiO₂ processes of today. The scattering potential of an ionized impurity at the edge used in the numerical calculation of Figs. 4.42 and 4.43 is the long-range Coulombic potential, whereas the scattering potential of many surface states is short range.

Fig. 4.43 Motion pictures of the wave packets transporting through quantum wires having one ionized impurity scattering center located at the center of the quantum wire. The time interval of the motion sequences is 50 fs. Electron energies are $E = 300, 85$ and 40 meV



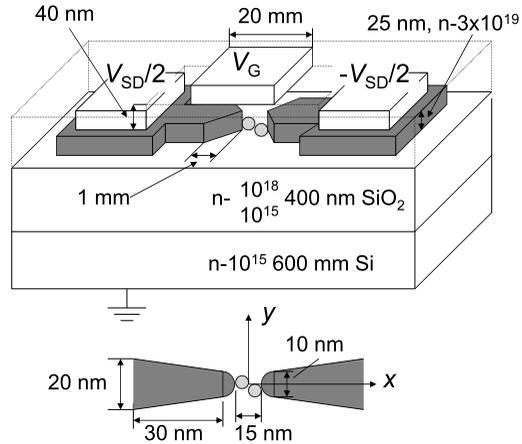
4.7 Coulombic Blockade and Single-Electron Transistor

We come to the quantum dot domain by further decreasing the quantum wire length. A quantum box defined by lateral potential barriers in a thin Si-film and by two vertical potential barriers due to the gate oxide and to the buried dielectric of the SON architecture leads consequently to Coulomb-blockade properties [77]. Conductance oscillations in the subthreshold regime were also reported in submicron silicon transistors at liquid helium temperatures [78]. Ionescu et al. has reviewed the CMOS evolution, where CMOS-SET (single-electron transistor) hybrid circuits were envisioned [79].

Silicon SETs have been fabricated in the form of point-contact MOSFETs with various channel widths using electron-beam lithography and anisotropic etching technique on SOI substrates [80–82]. The dot size, as small as 5.3 nm in a silicon point-contact channel, showed a negative differential conductance due to quantum levels up to 25 K [83, 84]. For a similar point-contact MOSFET and the gate oxide by chemical vapor deposition (CVD) instead of thermal oxidation, it was found that the single-electron addition energy is about 259 meV and the dot diameter is less than 4.4 nm [85]. Si SET with in-plane point contact metal gates was reported [86].

Si nanocrystals were deposited by very high frequency plasma decomposition of silane, resulting in an average dot diameter of 8 nm, the source-drain electrode

Fig. 4.44 Si-nanocrystal based single electron transistor (Reprinted from Y. Fu, M. Willander, A. Dutta, and S. Oda, Carrier conduction in Sinanocrystal-based single-electron transistor—I. Effect of gate bias, Superlattices Microstruct., vol. 28, pp. 177–187, 2000, with permission from Elsevier)



separation is 30 nm, and a gate electrode is employed so that the charge states in quantum dots can be controlled [87, 88] (see Fig. 4.44), Si SET memory using a multiple-tunnel junction was also fabricated by electron-beam direct writing technique [89], and the experimental works have been carefully analyzed theoretically [90–93]. The nanocrystals were then further processed into single electron non-volatile memory devices [94].

The most prominent effect appearing in a nanoscale single-electron transistor is of course the Coulombic blockade effect. When one electron arrives at the conduction channel, its Coulombic potential lifts up the local potential energy so that the later electrons become scattered. For a nanoscale conduction channel, the potential energy changes very significantly so that the scattering is very strong. In other words, the transport of other electrons becomes blockaded. Figure 4.45 shows a typical drain current as a function of the source-drain and gate voltages.

When extending the road beyond CMOS, we also see quantum dot cellular automata (QCA) along the logic line of the emerging technology sequence [95]. Ambitions of researchers have been to develop faster and denser integrated circuits and the field is sliding towards nanoelectronics and molecular electronics [96, 97], among them are cellular automata and cellular neural networks [98, 99].

The novel cellular automaton was proposed by C. Lent, W. Porod and their collaborators at the University of Notre Dame [100, 101]. The design depends on the dynamics of direct and local interactions between devices and their neighbors. The Lent-Porod quantum-dot cellular automata (QCA) scheme composes of many quantum dot cells seeking the lowest energy state for the entire assembly of cells. Refer to Fig. 4.46, a quantum cell consists of five quantum dots with two confined electrons. The wireless ground-state computing of the QCA is based on the transfer of the electric fields of electrons in the cell. A high energy state exists when two electrons in one cell are in close proximity to each other due to the repulsive forces of their electric fields.

Because of the necessary carrier transfer among quantum dots within one quantum cell, wave function overlappings and therefore the Coulombic exchange in-

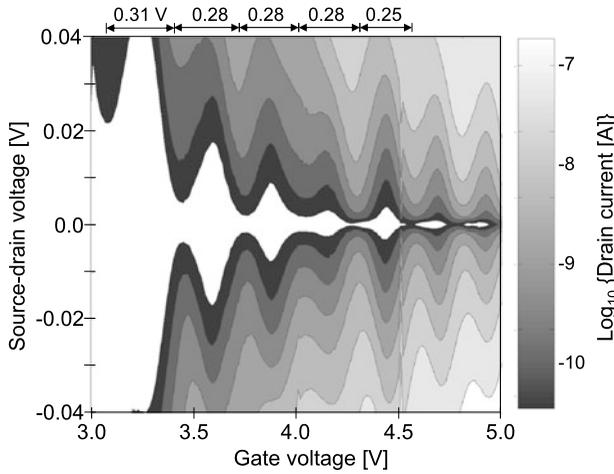
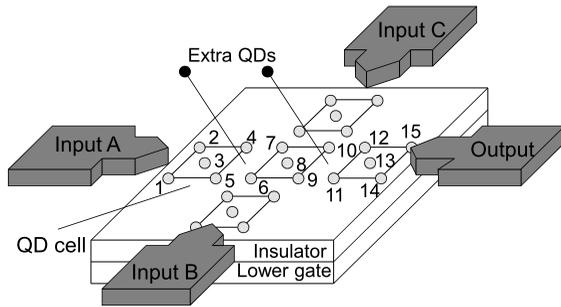


Fig. 4.45 Drain current contour plot of a device at the source-drain voltage and gate voltage plane. The measurement temperature was 20 K (Reprinted from Y. Fu, M. Willander, A. Dutta, and S. Oda, Carrier conduction in Si-nanocrystal-based single-electron transistor—I. Effect of gate bias, Superlattices Microstruct., vol. 28, pp. 177–187, 2000, with permission from Elsevier)

Fig. 4.46 QCA majority logic gate. Signal *A*, *B*, and *C* arrive through QCA wires and are fixed as the result of previous operations. The cell state is determined by the state of the majority of the three inputs. The output proceeds out to the right down a QCA wire



interactions are inevitable. Consider the three QD cells marked with QD index in Fig. 4.46, each quantum dot has one localized electron state available to accommodate one electron, but the electron is missing (e.g., due to thermal excitation). We now introduce one, or two electrons at the left cell by the input A.

We denote sublevels in the QD as E_{nm} and ψ_{nm} , where n is the QD index and m the sublevel index. When there is only one electron injected into the system from the contact,

$$f(\mathbf{r}) = \sum A_{nm} \psi_{nm} \tag{4.144}$$

and the Schrödinger equation becomes

$$i\hbar \frac{dA_{ij}}{dt} = \sum_{nm} A_{nm} H_{ijnm} \tag{4.145}$$

where $H_{ijnm} = \langle \psi_{ij} | H | \psi_{nm} \rangle$. Let

$$A_{i1}(t=0) = \delta_{i,1} \quad (4.146)$$

the calculated result shows that due to the wave function overlapping, this electron moves from one quantum cell to the other and gets bounced back from the other end of the quantum dot array. The spatial distribution of the electron extends over every quantum dot after 20 ps.

When there are two electrons, because of the antisymmetric property, a convenient choice of the pair-state basis function is

$$|n_1 n_2 m_1 m_2, \mathbf{r}_1 \mathbf{r}_2\rangle = \frac{1}{\sqrt{2}} \begin{vmatrix} \psi_{n_1 m_1}(\mathbf{r}_1) & \psi_{n_1 m_1}(\mathbf{r}_2) \\ \psi_{n_2 m_2}(\mathbf{r}_1) & \psi_{n_2 m_2}(\mathbf{r}_2) \end{vmatrix} \quad (4.147)$$

where the electrons are expected to have the same spins at low temperature, and

$$f(\mathbf{r}_1, \mathbf{r}_2) = \sum A_{n_1 n_2 m_1 m_2} |n_1 n_2 m_1 m_2, \mathbf{r}_1 \mathbf{r}_2\rangle \quad (4.148)$$

The matrix element of the Coulombic interaction between the two electrons is

$$\begin{aligned} V_{i_1 j_1 i_2 j_2 n_1 m_1 n_2 m_2} &= \langle \psi_{i_1 j_1}(\mathbf{r}_1) \psi_{i_2 j_2}(\mathbf{r}_2) | \frac{e^2}{4\pi\epsilon|\mathbf{r}_1 - \mathbf{r}_2|} | \psi_{n_1 m_1}(\mathbf{r}_1) \psi_{n_2 m_2}(\mathbf{r}_2) \rangle \\ &\quad - \langle \psi_{i_1 j_1}(\mathbf{r}_1) \psi_{i_2 j_2}(\mathbf{r}_2) | \frac{e^2}{4\pi\epsilon|\mathbf{r}_1 - \mathbf{r}_2|} | \psi_{n_1 m_1}(\mathbf{r}_2) \psi_{n_2 m_2}(\mathbf{r}_1) \rangle \end{aligned} \quad (4.149)$$

The first term is the direct Coulombic interaction between two electrons and the second term is the exchange Coulombic interaction due to antisymmetric property of the many-electron state. This term becomes negligible when the overlapping of the wave functions at different quantum dots is small.

The final two-electron Schrödinger equation is

$$\begin{aligned} i\hbar \frac{dA_{i_1 j_1 i_2 j_2}}{dt} &= \sum_{n_1 m_1 n_2 m_2} A_{n_1 m_1 n_2 m_2} (\delta_{i_1 j_1 n_1 m_1} H_{i_2 j_2 n_2 m_2} \\ &\quad + \delta_{i_2 j_2 n_2 m_2} H_{i_1 j_1 n_1 m_1} + V_{i_1 j_1 i_2 j_2 n_1 m_1 n_2 m_2}) \end{aligned} \quad (4.150)$$

By setting $A_{1121} = 1$ at time $t = 0$, Fig. 4.47(a) shows the spatial distribution of the two electrons in the time domain. The two electrons are well confined in the first quantum cell. Checking the values of interactions, it is observed that when the electrons stay with each other as the first nearest neighbor, the direct Coulombic interaction is 3.0 meV and the Coulombic exchange interaction is -1.5 meV, so that the total energy of the pair electrons is: $2E_0 + 3.0 - 1.5$ meV. When the two electrons become second nearest neighbors, the total energy becomes $2E_0 + 2.1$ meV! Thus, the Coulombic exchange interaction reduces the total energy of the pair electrons when they stay close.

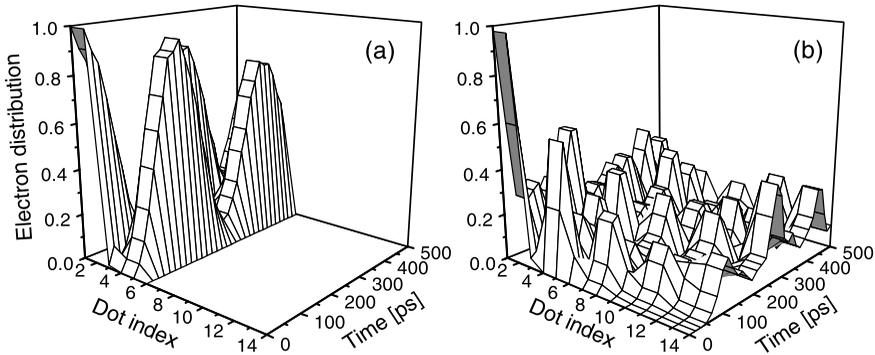


Fig. 4.47 (a) Two electrons in the QCA in the time-domain. (b) Two electrons in the QCA with extra quantum dots are inserted between two adjacent quantum cells (Reprinted with permission from Y. Fu and M. Willander, Modeling and design of quantum-dot cellular automata, *J. Appl. Phys.*, vol. 83, pp. 3186–3191, 1998. Copyright 1998, American Institute of Physics)

Adding the extra two QDs between the cells, see Fig. 4.46, the electrons extend very quickly to all cells, Fig. 4.47(b).

The important conclusion we have thus derived is the importance of the many-body interaction, i.e., the Coulombic exchange interaction in determining the carrier transport property of the quantum cellular automata. This many-body interaction facilitates the fabrication of the quantum dot cellular automata in a positive way that it provides a kind of self-binding force between the two electrons in one quantum cell so that the two electrons become better isolated from other electrons, since it is the polarization of the electron distributions in the quantum cells that is transferring across the cellular automata, instead of the charge transfer in normal electronic devices.

4.8 Future Perspectives

4.8.1 Carbon Nanotubes

Electronics of single-wall carbon nanotubes (SWNT) was discussed [102, 103], a 0.4 μm long semiconducting single-wall carbon nanotube reveals single-electron charging at temperatures up to 160 K [104], similar works were reported [105]. A single-electron transistor (SET) was fabricated using the single-wall carbon nanotube grown directly on to the Si atomic-force-microscopic (AFM) tip and TiO_x lines as the tunnel junctions which showed room-temperature Coulombic oscillation with periods of 1 V [106], Coulombic blockade effect was observed in a metal-on-tube metal/multi-wall carbon nanotube/metal structure (metal = Ti or Pt/Au) [107], and multi-wall carbon nanotube between two gold electrodes at SiO_2 surface [108].

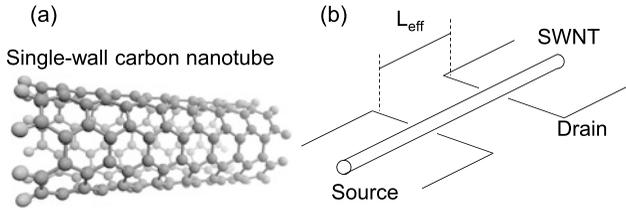


Fig. 4.48 (a) atomic structure of a short single-wall carbon nanotube (SWNT), (b) schematic drawing of a SWNT device

Electrical transport properties of intramolecular $p - n - p$ junctions formed on individual semiconducting carbon nanotubes were reported, well-defined and highly reproducible SETs with much smaller size than the geometrical length of the nanotube were obtained [109]. Single-electron tunneling through amorphous carbon dots array was reported [110].

Figure 4.48(a) shows the geometric structure of a short single-wall carbon nanotube which consists of only 150 C atoms, with 10 hydrogen atoms attached to the left end and another 10 hydrogen atoms to the right end, and Fig. 4.48(b) shows schematically a single-wall carbon nanotube based device.

4.8.2 Molecular Devices

The possibility of atom/molecule switching devices, Atom Relay Transistor (ART) and Molecular Single Electron Switching (MOSES) with total dimensions of a few nm and an operation speed of more than THz to supersede the present metal-oxide-semiconductor devices and to establish the new era of Atom Electronics has been extensively discussed and exploited [111], and devices based on ART will be ultimate in its performance of switching speed [112]. Quantum computing devices are implemented into the same substrate as the widely used large-scale-integrated circuits are discussed [113]. A variety of molecular organic semiconductors have been investigated for electronic applications [114, 115]. Figure 4.49 shows schematically a single molecular device connected to two atomically sharp gold electron reservoirs.

4.8.3 Metallic Devices

Single electron memory utilizing a Schottky in-plane gate quantum wire transistor with nano Schottky metal dots whose positions and sizes can be precisely controlled was realized [116]. The offset charge noise in metallic single electron tunneling (SET) devices fabricated on dielectric substrates was experimentally studied [117, 118]. Background charge rearrangements in metallic SETs were modeled in

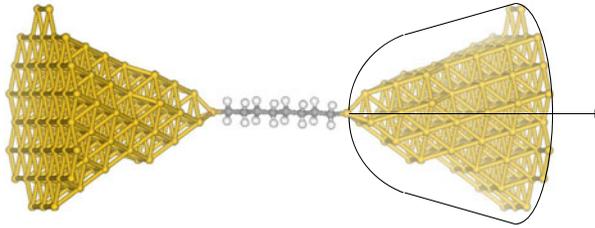


Fig. 4.49 Structure of a single molecular (1,8-octanedithiol) device connected to two atomically sharp gold electron reservoirs

two-level tunneling systems as a Poisson process [119]. A SET was made from a cadmium selenide nanocrystal [120].

Metal based SETs were fabricated by the step edge cut off process [121]. An aluminum SET was bonded to an InP HEMT for high frequency applications [122]. There were multilevel SET devices with Au-SiO-Al and Al-AlO_x-SiO-Al overlap capacitors [123]. And a self-aligned metallic SETT was fabricated on a separated-by-implanted-oxygen substrate, where an array of 10–50 gold islands of 1–3 nm diameter was isolated between source and drain electrodes on a silicon nanowire [124]. Fabrication of metallic SETTs has been developed using different lithography steps for the preparation of the different layers [125]. A few tens of nm wide oxidized metal line was formed using a scanning tunneling microscopy (STM) tip and an AFM cantilever as a cathode, which works as an energy barrier for an electron in SET [126].

4.8.4 Ferromagnetic Devices

There was also ferromagnetic dot structure embedded in split-gate quantum wires on AlGaAs/InGaAs/GaAs pseudomorphic high-electron-mobility transistor (PM-HEMT) by two-step surface modification with STM [127].

References

1. Selberherr S (1984) Analysis and simulation of semiconductor devices. Springer, Wien, p 8
2. Esaki L (1958) New phenomenon in narrow germanium $p-n$ junctions. Phys Rev 109:603–604
3. Capasso F (ed) (1990) Physics of quantum electron devices. Springer, Berlin
4. Esaki L (1985) Semiconductor superlattices and quantum wells. In: Chadi JD, Harrison WA (eds) Proc 17th int conf on the physics of semiconductors, San Francisco, 1984. Springer, New York, pp 473–483
5. Heiblum M, Nathan MI, Thomas DC, Knoedler CM (1985) Direct observation of ballistic transport in GaAs. Phys Rev Lett 55:2200–2203
6. Bonnefoi AR, Chow DH, McGill TC (1985) Inverted base-collector tunnel transistors. Appl Phys Lett 47:888–890

7. Luryi S, Capasso F (1985) Resonant tunneling of two-dimensional electrons through a quantum wire: a negative transconductance device. *Appl Phys Lett* 47:1347–1349
8. Yoshimura H, Schulman JN, Sakaki H (1990) Charge accumulation in a double-barrier resonant-tunneling structure studied by photoluminescence and photoluminescence-excitation spectroscopy. *Phys Rev Lett* 64:2422–2425
9. Luryi S (1985) Frequency limit of double-barrier resonant-tunneling oscillators. *Appl Phys Lett* 47:490–492
10. Sollner TCLG, Brown ER, Goodhue WD, Le HQ (1990) In: Capasso F (ed) *Physics of quantum electron devices*. Springer, Berlin, p 145
11. Goldman VJ, Tsui DC, Cunningham JE (1987) Observation of intrinsic bistability in resonant-tunneling structures. *Phys Rev Lett* 58:1256–1259
12. Sollner TCLG, Goldman VJ, Cunningham JE (1987) Comment on ‘Observation of intrinsic bistability in resonant-tunneling structures’. *Phys Rev Lett* 59:1622–1623
13. Boric O, Tolmunen TJ, Kollberg E, Frerking MA (1992) Anomalous capacitance of quantum well double-barrier diodes. *Int J Infrared Millim Waves* 13:799–814
14. Sollner TCLG, Goodhue WD, Tannenwald PE, Parker CD, Peck DD (1983) Resonant tunneling through quantum wells at frequencies up to 2.5 THz. *Appl Phys Lett* 43:588–590
15. Hou Y, Wang W-P, Li N, Lu W, Fu Y (2008) Effects of series and parallel resistances on the current-voltage characteristics of small-area air-bridge resonant tunneling diode. *J Appl Phys* 104, 074508
16. Zhu B, Chao KA (1987) Phonon modes and Raman scattering in GaAs/Ga_{1-x}Al_xAs. *Phys Rev B* 36:4906–4914
17. Goldman VJ, Tsui DC, Cunningham JE (1987) Evidence for LO-phonon-emission-assisted tunneling in double-barrier heterostructures. *Phys Rev B* 36:7635–7637
18. Wingreen NS, Jacobsen KW, Wilkins JW (1988) Resonant tunneling with electron-phonon interaction: an exactly solvable model. *Phys Rev Lett* 61:1396–1399
19. Rydberg A, Grönqvist H, Kollberg E (1990) Millimeter- and submillimeter-wave multipliers using quantum-barrier-varactor (QBV) diodes. *IEEE Electron Device Lett* 11:373–375
20. Kollberg E, Stake J, Dillner L (1996) Heterostructure barrier varactors at submillimeter waves. *Philos Trans R Soc A, Math Phys Eng Sci* 354:2383–2398
21. Reddy VK, Neikirk DP (1993) High breakdown voltage AlAs/InGaAs quantum barrier varactor diodes. *Electron Lett* 29:464–466
22. Hui S, Zhang WM, Domier CW, Luhmann NC Jr, Sjogren LB, Liu XLH (1995) Novel concept for improved nonlinear transmission line performance. *IEEE Trans Microw Theory Tech* 43:780–789
23. Lieneweg U, Hancock BR, Maserjian J (1987) Barrier-intrinsic-N⁺ (BIN) diodes for near-millimeter wave generation. In: *Conference digest: 20th int conf infrared and millimeter waves*. IEEE Press, New York, pp 6–7
24. Liu HXL, Qin XH, Sjogren LB, Chung E, Domier CW, Luhmann NC Jr (1992) Monolithic high-power millimeter-wave quasi-optical frequency multiplier arrays using quantum barrier devices. *IEEE Trans Electron Devices* 39:2668
25. Rahal A, Bosisio RG, Boch E, Rogers C, Ovey J (1996) Planar V-band frequency tripler for indoor communication systems. *Proc SPIE* 2842:209–214
26. Batey J, Wright SL (1986) Energy band alignment in GaAs:(Al, Ga)As heterostructures: the dependence on alloy composition. *J Appl Phys* 59:200–209
27. Landheer D, Liu HC, Buchanan M, Stoner R (1989) Tunneling through AlAs barriers: Gamma-X transfer current. *Appl Phys Lett* 54:1784–1786
28. Krishnamurthi K, Nilsen SM, Harrison RG (1994) GaAs single-barrier varactors for millimeter-wave triplers: guidelines for enhanced performance. *IEEE Trans Microw Theory Tech* 42:2512–2516
29. Dillner L, Stake J, Kollberg E (1997) Analysis of symmetric varactor frequency multipliers. *Microw Opt Technol Lett* 15:26–29
30. Stern F, Howard W (1967) Properties of semiconductor surface inversion layers in electric quantum limit. *Phys Rev* 163:816–835

31. Dingle R, Störmer HL, Gossard AC, Wiegmann W (1978) Electron mobilities in modulation-doped semiconductor superlattices. *Appl Phys Lett* 33:665–667
32. Hess K (1979) Impurity and phonon scattering in layered structures. *Appl Phys Lett* 35:484–486
33. Hiyamizu S, Saito J, Nanbu K, Ishikawa T (1983) Improved electron mobility higher than 10^6 cm²/Vs in selectively doped GaAs/N-AlGaAs heterostructures grown by MBE. *Jpn J Appl Phys* 22:L609–L611
34. Mimura T, Hiyamizu S, Fujii T, Nanbu K (1980) A new field-effect transistor with selectively doped GaAs/*n*-Al_xGa_{1-x}As heterojunctions. *Jpn J Appl Phys* 19:L225–L227
35. Walukiewicz W, Ruda HE, Lagowski J, Gatos HC (1984) Electron mobility limits in a two-dimensional electron gas: GaAs-GaAlAs heterostructures. *Phys Rev B* 29:4818–4820
36. Walukiewicz W, Ruda HE, Lagowski J, Gatos HC (1984) Electron mobility in modulation-doped heterostructures. *Phys Rev B* 30:4571–4582
37. Yokoyama K, Hess K (1986) Monte Carlo study of electronic transport in Al_{1-x}Ga_xAs/GaAs single-well heterostructures. *Phys Rev B* 33:5595–5606
38. Hirakawa K, Sakaki H (1986) Mobility of the two-dimensional electron gas at selectively doped *n*-type Al_xGa_{1-x}As/GaAs heterojunctions with controlled electron concentrations. *Phys Rev B* 33:8291–8303
39. Price PJ (1981) Two-dimensional electron transport in semiconductor layers. I. Phonon scattering. *Ann Phys* 133:217–239
40. Zeindl HP, Wegehaupt T, Eisele I, Oppolzer H, Reisinger H, Tempel G, Koch F (1987) Growth and characterization of a delta-function doping layer in Si. *Appl Phys Lett* 50:1164–1166
41. Ni W-X, Hansson GV, Sundgren J-E, Hultman L, Wallenberg LR, Yao J-Y, Markert LC, Greene JE (1992) Delta-function-shaped Sb-doping profiles in Si(001) obtained using a low-energy accelerated-ion source during molecular-beam epitaxy. *Phys Rev B* 46:7551–7558
42. Sze SM (ed) (1990) High-speed semiconductor devices. Wiley, New York
43. Manasreh MO (ed) (1993) Semiconductor quantum wells and superlattices for long-wavelength infrared detectors. Artech House, Boston
44. Iwai H (1993) CMOS device architecture and technology for the 0.25 micron to 0.025 micron generations. In: Borel J, Gentil P, Noblance JP, Nouailhat A, Verdone M (eds) Proceedings of the 23rd European solid state device research conference, Grenoble, France, 1993. Frontières, Gif-sur-Yvette, pp 513–520
45. Chandrakasan AP, Sheng S, Brodersen RW (1992) Low-power CMOS digital design. *IEEE J Solid-State Circuits* 27:473–484
46. Haydock R, Heine V, Kelly MJ (1972) Electronic structure based on the local atomic environment for tight-binding bands. *J Phys C, Solid State Phys* 5:2845–2858
47. Fu Y, Xu W, Zheng Z-B (1987) Impurity induced vibrations in light doped silicon. *Solid State Commun* 62:163–167
48. Friedel J (1954) Electronic structure of primary solid solutions in metals. *Adv Phys* 3:446–507
49. Kittel C (1963) Quantum theory of solids. Wiley, New York, p 339
50. Heine V, Weaire D (1970) Pseudopotential theory of cohesion and structure. *Solid State Phys* 24:249–463
51. Ono M, Saito M, Yoshitomi T, Fiegna C, Ohguro T, Iwai H (1993) Sub-50 nm gate length *n*-MOSFETs with 10 nm phosphorus source and drain junctions. In: Proceeding of the international electron devices meeting, pp 119–122
52. Ono M, Saito M, Yoshitomi T, Fiegna C, Ohguro T, Iwai H (1995) A 40 nm gate length *n*-MOSFET. *IEEE Trans Electron Devices* 42:1822–1830
53. Han J, Ferry D, Newman P (1990) Ultra-submicrometer-gate AlGaAs/GaAs HEMTs. *IEEE Electron Device Lett* 11:209–211
54. Hashizume T, Okada H, Hasegawa H (1996) Quantum transport in a Schottky in-plane-gate controlled GaAs/AlGaAs quantum well wires. *Physica B* 227:42–45

55. Omura Y, Kurihara K, Takahashi Y, Ishiyama T, Nakajima Y, Izumi K (1997) 50-nm channel nMOSFET/SIMOX with an ultrathin 2- or 6-nm thick silicon layer and their significant features of operations. *IEEE Electron Device Lett* 18:190–193
56. Pelouard JL, Teissier R, Matine N, Pardo F (1997) Dynamic behaviour of the metal heterojunction bipolar transistor. In: International conference on indium phosphide and related materials. IEEE Press, New York, pp 169–172
57. Ando T (1996) Mesoscopic transport in low dimensional systems. In: 23rd international conference on the physics of semiconductors. World Scientific, Singapore, pp 59–68
58. Dollfus P (1997) Si/Si_{1-x}Ge_x heterostructures: electron transport and field effect transistor operating using Monte Carlo simulation. *J Appl Phys* 82:3911–3916
59. Lake R, Klimeck G, Bowen RC, Jovanovic D (1997) Single and multiband of quantum electron transport through layered semiconductor devices. *J Appl Phys* 81:7845–7869
60. Vasileska D, Eldridge T, Ferry DK (1996) Quantum transport: silicon inversion layers and InAlAs-InGaAs heterostructures. *J Vac Sci Technol B* 14:2780–2785
61. Nedjalkov M, Dimov I, Bordone P, Brunetti R, Jacoboni C (1997) Using the Wigner function for quantum transport in device simulation. *Math Comput Model* 25:33–53
62. Fu Y, Mu Y, Willander M (1996) Quantum ballistic transport in a dual-gate Si transistor. *IEEE Trans Electron Devices* 43:2030–2032
63. Madhukar A (1990) The nature of molecular beam epitaxy and consequences for quantum microstructures. In: Capasso F (ed) *Physics of quantum electron devices*. Springer, Berlin, Chap. 2
64. Ando T, Fowler AB, Stern F (1982) Electronic properties of two-dimensional systems. *Rev Mod Phys* 54:437–672
65. Fu Y, Willander M (1991) Lateral-nonuniformity effect on the I-V spectrum in a double-barrier resonant-tunneling structure. *Phys Rev B* 44:13631–13634
66. Fu Y, Willander M, Stake J, Dillner L, Kollberg EL (2000) Carrier conduction through the quantum barrier region in a heterostructure barrier varactor induced by an AC bias. *Superlattices Microstruct* 28:135–141
67. Bohr M (2001) MOS transistor scaling challenges. In: *Proceedings of the international symposium ULSI process integration II*. ECS proceedings, vol 2001-2, pp 463–473
68. Lindert N, Chang L, Choi Y-K, Anderson EH, Lee W-C, King T-J, Bokor J, Hu C (2001) Sub-60-nm quasi-planar FinFETs fabricated using a simplified process. *IEEE Electron Device Lett* 22:487–489
69. Celler GK, Cristoloveanu S (2003) Frontiers of silicon-on-insulator. *J Appl Phys* 93:4955–4978
70. Jurczak M, Skotnicki T, Paoli M, Tormen B, Martins J, Regolini JL, Dutartre D, Ribot P, Lenoble D, Pantel R, Monfray S (2000) Silicon-On-Nothing (SON)—an innovative process for advanced CMOS. *IEEE Trans Electron Devices* 47:2179–2187
71. Plummer JD (2000) Silicon MOSFETs (conventional and non-traditional) at the scaling limit. In: *Proc of device research conference*, pp 3–7
72. Schultz T, Rosner W, Risch L, Korbel A, Langmann U (2001) Short-channel vertical sidewall MOSFETs. *IEEE Trans Electron Devices* 48:1783–1788
73. Simmons JA, Blount MA, Moon JS, Baca WE, Reno JI, Hafich MJ (1997) Unipolar complementary bistable memories using gate-controlled negative differential resistance in a 2D-2D quantum tunneling transistor. In: *Electron devices meeting, IEDM technical digest*, pp. 755–758 (cat no 97CH36103)
74. Matsuoka H, Ichiguchi T, Yoshimura T, Takeda E (1994) Coulomb blockade in the inversion layer of a Si metal-oxide-semiconductor field-effect transistor with a dual-gate structure. *Appl Phys Lett* 64:586–588
75. Lauhon LJ, Gudiksen MS, Wang D, Lieber CM (2002) Epitaxial core-shell and core-multishell nanowire heterostructures. *Nature* 420:57–61
76. Iwai H CMOS Downsizing toward sub-10 nm. www.iwai.ae.titech.ac.jp/pdf/iwaironbun/ulis03.pdf

77. Monfray S, Souifi A, Boeuf F, Ortolland C, Poncet A, Militaru L, Chanemougame D, Skotnicki T (2003) Coulomb-blockade in nanometric Si-film silicon-on-nothing (SON) MOS-FETs. *IEEE Trans Nanotechnol* 2:295–300
78. Peters MG, den Hartog SG, Dijkhuis JI, Buyk OJA, Molenkamp LW (1998) Single electron tunneling and suppression of short-channel effects in submicron silicon transistors. *J Appl Phys* 84:5052–5056
79. Ionescu AM, Declercq MJ, Mahapatra S, Banerjee K, Gautier J (2002) Few electron devices: towards hybrid CMOS-SET integrated circuits. In: Proceedings of the 39th conference on design automation, New Orleans, Louisiana, USA, pp 88–93
80. Ishikuro H, Hiramoto T (1997) Energy spectrum of the quantum-dot in a Si single-electron-device. In: 55th annual device research conference digest, Fort Collins, CO, USA, 23–25, June 1997, pp 84–85 (cat no 97TH8279)
81. Ishikuro H, Hiramoto T (1997) Quantum mechanical effects in the silicon quantum dot in a single-electron transistor. *Appl Phys Lett* 71:3691–3693
82. Ishikuro H, Hiramoto T (1999) Fabrication of nano-scale point contact metal-oxide-semiconductor field-effect-transistors using micrometer-scale design rule. *Jpn J Appl Phys* 38:396–398
83. Saitoh M, Hiramoto T (2001) Suppression of series parasitic resistance and observation of quantum effects in a silicon single-electron transistor. In: Proceedings of the 2001 1st IEEE conference on nanotechnology, IEEE-NANO 2001, Maui, HI, USA, 28–30 October 2001, pp 243–247 (cat no 01EX516)
84. Saitoh M, Saito T, Inukai T, Hiramoto T (2001) Transport spectroscopy of the ultrasmall silicon quantum dot in a single-electron transistor. *Appl Phys Lett* 79:2025–2027
85. Saitoh M, Takahashi N, Ishikuro H, Hiramoto T (2001) Large electron addition energy above 250 meV in a silicon quantum dot in a single-electron transistor. *Jpn J Appl Phys* 40:2010–2012
86. Wang TH, Li HW, Zhou JM (2001) Si single-electron transistors with in-plane point-contact metal gates. *Appl Phys Lett* 78:2160–2162
87. Dutta A, Kimura M, Honda Y, Otobe N, Itoh A, Oda S (1997) Fabrication and electrical characteristics of single electron tunneling devices based on Si quantum dots prepared by plasma processing. *Jpn J Appl Phys* 36:4038–4041
88. Dutta A, Lee SP, Hayafune Y, Hatatani S, Oda S (2000) Single-electron tunneling devices based on silicon quantum dots fabricated by plasma process. *Jpn J Appl Phys* 39:264–267
89. Dutta A, Lee SP, Hatatani S, Oda S (1999) Silicon-based single-electron memory using a multiple-tunnel junction fabricated by electron-beam direct writing. *Appl Phys Lett* 75:1422–1424
90. Fu Y, Willander M, Dutta A, Oda S (2000) The gate bias vs. the number of electrons confined in Si dot based single electron transistor. *Proc SPIE* 3975(1–2):1027–1032
91. Dutta A, Oda S, Fu Y, Willander M (2000) Electron transport in nanocrystalline Si based single electron transistors. *Jpn J Appl Phys* 39:4647–4650
92. Fu Y, Willander M, Dutta A, Oda S (2000) Carrier conduction in Si-nanocrystal-based single-electron transistor—I. Effect of gate bias. *Superlattices Microstruct* 28:177–187
93. Fu Y, Willander M, Dutta A, Oda S (2000) Carrier conduction in Si-nanocrystal-based single-electron transistor—II. Effect of drain bias. *Superlattices Microstruct* 28:189–198
94. Dutta A, Hayafune Y, Oda S (2000) Single electron memory devices based on plasma-derived silicon nanocrystals. *Jpn J Appl Phys* 39:L855–L857
95. Hutchby JA, Bourianoff GI, Zhirnov VV, Brewer JE (2002) Extending the road beyond CMOS. *IEEE Circuits Devices Mag* 18:28–41
96. Montemerlo MS, Love JC, Opiteck GJ, Goldhaber-Gordon D, Ellenbogen JC (1996) Technologies and designs for electronic nanocomputers. MITRE, McLean
97. Goldhaber-Gordon D, Montemerlo MS, Love JC, Opiteck GJ, Ellenbogen JC (1997) Overview of nanoelectronic devices. *Proc IEEE* 85:521–540
98. Toffoli T, Margolus N (1987) Cellular automata machines: a new environment for modeling. MIT, Cambridge

99. Tanamoto T, Katoh R (1996) The possibility of higher temperature operation in quantum cellular automata (QCA). *IEICE Trans Electron* E79-C:1550–1556
100. Lent CS, Tougaw PD, Porod W (1993) Bistable saturation in coupled quantum dots for quantum cellular automata. *Appl Phys Lett* 62:714–716
101. Lent CC, Tougaw PD, Porod W, Bernstein GH (1993) Quantum cellular automata. *Nanotechnology* 4:49–57
102. Hu J, Ouyang M, Yang P, Lieber CM (1999) Controlled growth and electrical properties of heterojunctions of carbon nanotubes and silicon nanowires. *Nature* 399:48–51
103. Johnson AT (1999) Electronics of single-wall carbon nanotubes. In: Wuorinen JH (ed) IEEE international solid-state circuits conference, ISSCC. Digest of technical papers, San Francisco, CA, USA, 15–17 February 1999, 1st edn. pp 210–211
104. Kong J, Zhou C, Yenilmez E, Dai H (2000) Alkaline metal-doped n-type semiconducting nanotubes as quantum dots. *Appl Phys Lett* 77:3977–3979
105. Yao Z, Postma HWC, Balents L, Dekker C (1999) Carbon nanotube intramolecular junctions. *Nature* 402:273–276
106. Matsumoto K, Gotoh K (2001) Nano-processing using carbon nano tube probes and its device applications. In: International semiconductor device research symposium. Symposium proceedings, Washington, DC, USA, 5–7 December 2001, pp 354–357 (cat no 01EX497)
107. Kanda A, Ootuka Y, Tsukagoshi K, Aoyagi Y (2001) Electron transport in metal/multiwall carbon nanotube/metal structures (metal = Ti or Pt/Au). *Appl Phys Lett* 79:1354–1356
108. Roschier L, Penttila J, Martin M, Hakonen P, Paalanen M, Tapper U, Kauppinen EI, Journet C, Bernier P (1999) Single-electron transistor made of multiwalled carbon nanotube using scanning probe manipulation. *Appl Phys Lett* 75:728–730
109. Kong J, Cao J, Dai H, Anderson E (2002) Chemical profiling of single nanotubes: intramolecular p-n-p junctions and on-tube single-electron transistors. *Appl Phys Lett* 80:73–75
110. Miura N, Numaguchi T, Yamada A, Konagai M, Shirakashi J-I (1997) Single-electron tunneling through amorphous carbon dots array. *Jpn J Appl Phys* 36:1619–1621
111. Wada Y (1995) A proposal of atom/molecule switching devices. *Optoelectron, Dev Technol* 10:205–220
112. Ahmad S (1998) Semiconductor switching devices-future trends. *Def Sci J (India)* 48:45–59
113. Tanamoto T (2000) Quantum gates by coupled quantum dots and measurement procedure in field-effect-transistor structure. *Fortschr Phys* 48:1005–1021
114. Schon JH (2001) High mobilities in organic semiconductors: basic science and technology. *Synth Met* 122:157–160
115. Schon JH, Kloc Ch, Batlogg B (2001) Ambipolar organic devices for complementary logic. *Synth Met* 122:195–197
116. Okada H, Hasegawa H (2001) Novel single electron memory device using metal nano-dots and Schottky in-plane gate quantum wire transistors. *Jpn J Appl Phys* 40:2797–2800
117. Ahlers F-J, Krupenin VA, Lotkhov SV, Niemeyer J, Presnov DE, Scherer H, Weimann T, Wolf H, Zorin AB (1996) Investigation of the offset charge noise in single electron tunneling devices. In: Braun A (ed) Conference on precision electromagnetic measurements digest, Braunschweig, Germany, 17–21 June 1996, pp 507–508 (cat no 96CH35956)
118. Krupenin VA, Presnov DE, Savvateev MN (1998) Noise in Al single electron transistors of stacked design. *J Appl Phys* 84:3212–3215
119. Furlan M, Heinzl T, Jeanneret B, Lotkhov SV (2000) Coulomb blockade peak statistics influenced by background charge configuration. *J Low Temp Phys* 118:297–306
120. Klein DL, Roth R, Lim AKL, Alivisatos AP, McEuen PL (1997) A single-electron transistor made from a cadmium selenide nanocrystal. *Nature* 389:699–701
121. Altmeyer S, Hamidi A, Spangenberg B, Kurz H (1997) 77 K single electron transistors fabricated with 0.1 μm technology. *J Appl Phys* 81:8118–8120

122. Pettersson J, Wahlgren P, Delsing P, Haviland DB, Claeson T, Rorsman N, Zirath H (1996) Extending the high-frequency limit of a single-electron transistor by on-chip impedance transformation. *Phys Rev B* 53:R13272–R13274
123. Visscher EH, Verbrugh SM, Lindeman J, Hadley P, Mooij JE (1995) Fabrication of multilayer single-electron tunneling devices. *Appl Phys Lett* 66:305–307
124. Ford EM, Ahmed H (1998) Fabrication of self-aligned metallic Coulomb blockade devices on Si nanowires. *J Vac Sci Technol B* 16:3800–3803
125. Weimann T, Scherer H, Wolf H, Krupenin VA, Niemeyer J (1998) A new technology for metallic multilayer single electron tunneling devices. *Microelectron Eng* 41–42:559–562
126. Matsumoto K (1998) Room temperature single electron transistor made by STM/AFM nano-oxidation process. In: Hou HQ, Sah RE, Pearton SJ, Ren F, Wada K (eds) Proceedings of the symposium on light emitting devices for optoelectronic applications and twenty-eighth state-of-the-art program on compound semiconductors, San Diego, CA, USA, 3–8 May 1998, pp 68–77
127. Kikutani T, Aoki N, Hong CU, Hori H, Yamada S (1998) Quantum transport in ferromagnetic dot structure embedded in semiconductor quantum wires. *Physica B* 249–251:513–517

Chapter 5

Nanostructured Optoelectronics

Abstract Detailed aspects about light-matter interactions in nanostructure materials are presented and discussed in terms of optoelectronics performance. We begin with quantum selection rules about optical intraband and interband transitions and thereafter the optical grating required for photodetector. We further discuss the functions of solar cell, light-emitting diode, and nanostructure laser. Quantum-dot-based biomarkers for bioimaging applications are introduced by the end of the chapter to demonstrate the extension of information-communication-technology (ICT)-predominant solid-state nanotechnologies to many other fields.

5.1 Optical Transition and Quantum Selection Rule

Infrared absorption by intersubband transitions in quantum wells and superlattices have been extensively investigated in recent years as a result of the applications of these structures as infrared photodetectors and other optoelectronic devices. Absorption coefficient studies are of great importance for these devices because of their direct connections to the design of the infrared detector, to the detecting wavelength and to the detecting efficiency. The measurement of the absorption coefficient is a commonly used method to study the quantum well structures. Here, we start from basic quantum mechanics to calculate the absorption coefficient of intersubband transition in both the quantum well and superlattice structures.

We begin the discussions about optical transition in nanostructures from our previous study in Sect. 3.3, namely, Eq. (3.84) about the steady-state optical coefficient $g(\hbar\omega_s)$ of nanostructure

$$g(\hbar\omega_s) = \sum_{\mathbf{qk}} \frac{e^2 \hbar^2 |\langle \mathbf{q} | \mathbf{e}_s \cdot \nabla | \mathbf{k} \rangle|^2}{2m_0^2 c \epsilon \omega_s \Omega} \frac{\Gamma_k}{(E_q - E_k \pm \hbar\omega_s)^2 + \Gamma_k^2} [f(E_k) - f(E_q)] \tag{5.1}$$

Here the electromagnetic field is described by photon energy $\hbar\omega_s$, photon momentum $\hbar s$, and polarization vector \mathbf{e}_s . c and ϵ are speed of light in the nanostructure and dielectric constant of the nanostructure. Ω is the volume of the nanostructure. $[\mathbf{k}, E_k, f(E_k)]$ and $[\mathbf{q}, E_q, f(E_q)]$ describe the initial and final states of the electron before and after optical transition.

To make the discussions easy, let us consider optical excitations in a quantum well structure grown along the z direction. The electron states are then in the form

$$e^{i(k_x x + k_y y)} \psi_i(z)$$

where k_x and k_y are wave vectors in the x and y directions, respectively, and $\psi_i(z)$ is the wave function in the z direction. There are two types of optical transitions, the first one is between conduction-band and valence-band states, referred to as interband transition, Eq. (3.100),

$$g_{\text{inter}}(\hbar\omega_s) = \frac{e^2 \Gamma |\mathbf{e}_s \cdot \mathbf{p}_{cv}|^2}{2m_0^2 c \epsilon \omega_s L_{\text{QW}}} \times \sum_{ji} \iint \frac{|\langle \phi_j(z) | \psi_i(z) \rangle|^2 [f(E_{ik_x k_y}) - f(E_{jk_x k_y})]}{(E_{jk_x k_y} - E_{ik_x k_y} \pm \hbar\omega_s)^2 + \Gamma^2} \frac{2dk_x dk_y}{(2\pi)^2} \quad (5.2)$$

Here $\psi_i(z)$ denotes the electron state in the conduction band, and ϕ_j the hole state in the valence band. The second is referred to be intraband optical transition, Eq. (3.103). For conduction band,

$$g_{\text{intra}}(\hbar\omega_s) = \frac{e^2 \hbar^2 \Gamma e_z^2}{2m_0^2 c \epsilon \omega_s L_{\text{QW}}} \times \sum_{ji} \iint \frac{|\langle \psi_j(z) | \frac{\partial}{\partial z} | \psi_i(z) \rangle|^2 [f(E_{ik_x k_y}) - f(E_{jk_x k_y})]}{(E_{jk_x k_y} - E_{ik_x k_y} \pm \hbar\omega_s)^2 + \Gamma^2} \frac{2dk_x dk_y}{(2\pi)^2} \quad (5.3)$$

Similar expression can be written down for intraband optical transition in the valence band.

Figure 5.1(a) shows schematically the energy band structure and optical transitions of a quantum well embedded in an infinitely high barrier. Because of the infinite high barrier, the electron and hole states along the z axis in the conduction band and valence band are totally confined in the quantum well region and can be mathematically described as

$$\psi_i(z) = \phi_i(z) = \sqrt{\frac{2}{L_{\text{QW}}}} \sin\left(\frac{i\pi z}{L_{\text{QW}}}\right) \quad (5.4)$$

where the quantum index of the state $i = 1, 2, \dots$ is a positive integer and $z \in (0, L_{\text{QW}})$, L_{QW} is the quantum well width.

The first set of the requirements for nonzero optical transitions, which are normally referred to as the quantum selection rules, or simply quantum selections, is that the wave vector of the final electron state in the xy plane must equal to the

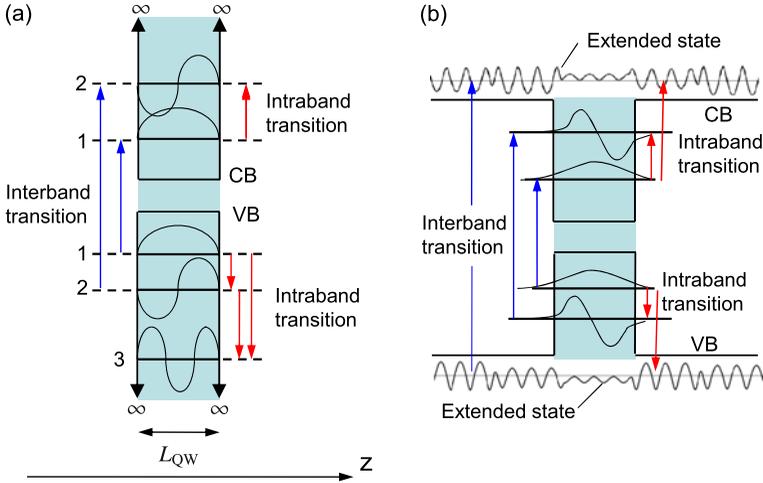


Fig. 5.1 (a) Interband and intraband optical transitions due to photon absorption in a quantum well embedded in infinitely high barrier. (b) Quantum well in a finite high barrier

one of the initial electron state. This is simply due to the translational symmetry in the xy plane of the quantum well. Exactly the same requirement is needed when we consider the optical transition in bulk material where translation symmetry exists in all three dimensions. And by Sect. 3.3, we can state further that k_z is to be conserved for the optical transition in a quantum well which is extended along the z axis. Thus, optical transition is also referred to be the vertical transition in the k space. This leads to the well-known fact that the optical properties of Si are not as good as most III–V materials since Si is an indirect material while common III–V materials are direct.

For interband optical transitions, the quantum indices of the conduction band state and the valence band state must be the same, i.e., $i = j$, since

$$\langle \psi_i(z) | \phi_j(z) \rangle = \delta_{ij} \tag{5.5}$$

when $\psi_i(z)$ and $\phi_i(z)$ are expressed as Eq. (5.4), while for intraband optical transitions,

$$\left\langle \psi_j(z) \left| \frac{\partial}{\partial z} \right| \psi_i(z) \right\rangle = \frac{i}{L_{QW}} \left\{ \frac{1 - \cos[(j+i)\pi]}{j+i} + \frac{1 - \cos[(j-i)\pi]}{j-i} \right\} \tag{5.6}$$

If $i = 1$, we can see that transitions to $j = 2, 4, \dots$ are possible, while it is not allowed to $j = 3, 5, \dots$. And the largest transition probability is when $j - i = \pm 1$.

On the other side, there is a distinct requirement about the electromagnetic field. For the interband optical transitions, Eq. (5.2) contains $|\mathbf{e}_s \cdot \mathbf{p}_{cv}|^2$, where \mathbf{e}_s is the polarization vector of the electric field, while e_z , the component of the electric field in the z direction, is contained in Eq. (5.3). We know by the fundamental electromagnetic field theory that \mathbf{e}_s is perpendicular to the propagation direction of the

electromagnetic field, i.e., s . For a quantum well along the z direction, we therefore realize that interband optical transition occurs for any polarization of the electric field, but intraband optical transition only occurs when $e_z \neq 0$. More specifically, a quantum well along the z direction will not emit a photon along the z direction when one electron transits from a high-energy state in the conduction band to a low-energy state in the same conduction band. The emitted photon due to intraband transition will propagate in a direction which is perpendicular to the z axis. Similarly, the quantum well along the z direction will not absorb a photon that incidents along the z direction to induce an intraband transition either in the conduction band or in the valence band. This generates a profound impact on the design of quantum well based optical devices. We will discuss it in more detail a bit later.

In reality we do not normally work with nanostructure systems that are embedded in infinitely high barriers. The barriers are almost always finite, especially when we work on optoelectronic systems where either we inject electrons to get photons, light-emitting diodes and lasers, or we use photons to generate electrons for photodetector and solar cell applications. In all these cases, electrons, either in the initial state or the final state of an optical transition, need to be extended so that they can be reached by external circuits. Thus, we actually deal with nanostructures as depicted in Fig. 5.1(b) where the barrier height is finite so that the envelope functions of the electron states are more or less extended. The diffusion of the envelope functions across heterointerfaces makes the theoretical treatment of the optical spectrum more complicated, while we expect similar major features of interband and intraband optical transitions as in the case of a quantum well in infinitely high barriers.

One complication in the theoretical analysis of the optical spectrum is the variation of the material parameters in the heterostructures. We discussed the effective mass in multiple quantum wells in Sect. 1.8, Eq. (1.159), that for a one-dimensional heteromaterial along the z direction, the kinetic energy is expressed as

$$\hbar^2 k_z \frac{1}{2m^*(z)} k_z$$

It is easy to see that a generalization of the above expression to a three-dimensional case results in

$$\frac{1}{2} \mathbf{p} \bar{w} \mathbf{p} \quad (5.7)$$

where \bar{w} is the 3×3 inverse effective-mass tensor taking into accounting the anisotropic effect of the band structure, see Eqs. (1.73), and its elements $w_{mn} = w_{nm} = 1/m_{mn}^*$, where $m, n = x, y, z$. Here $\mathbf{p} = \hbar \mathbf{k}$, \mathbf{k} is the wave vector of the electron in the crystal. And the light-matter interaction, i.e., the interaction between an electromagnetic field and the electron now becomes

$$e \mathbf{A} \bar{w} \mathbf{p} \quad (5.8)$$

See Sect. 3.2.

5.2 Intraband Optical Transition

Now we are going to calculate the absorption coefficient due to intraband optical transitions in a quantum well along the z axis, the plane perpendicular to this direction is defined as the xy plane. Here a radiation is switched on at time $t = 0$. Since the quantum well is translationally symmetric in the xy plane, the envelope function there is in the form of plane wave and the total wave function can be expressed as

$$\psi_{i\mathbf{k}}(\mathbf{r}) = \psi_i(z)u(\mathbf{r})e^{i\mathbf{k}\cdot\boldsymbol{\rho}} \quad (5.9)$$

where $u(\mathbf{r})$ is the Bloch function, which is the same for all i and \mathbf{k} since we focus only on intraband optical transitions in this section. $\mathbf{k} = (k_x, k_y)$ and $\boldsymbol{\rho}$ are the wave vector and coordinate in the xy plane, respectively. $\mathbf{r} = (\boldsymbol{\rho}, z) = (x, y, z)$. $\psi_i(z)$ is the normalized envelope function in the z direction.

Since $u(\mathbf{r})$ is periodic with a period of lattice unit cell, the integration over $\boldsymbol{\rho}$ can be approximated as

$$\int e^{i(\mathbf{q}-\mathbf{k})\cdot\boldsymbol{\rho}} |u(\mathbf{r})|^2 d\boldsymbol{\rho} \approx \int e^{i(\mathbf{q}-\mathbf{k})\cdot\boldsymbol{\rho}} d\boldsymbol{\rho} \int_{\text{cell}} |u(\mathbf{r})|^2 d\boldsymbol{\rho} \quad (5.10)$$

Moreover,

$$\int e^{i(\mathbf{q}-\mathbf{k})\cdot\boldsymbol{\rho}} d\boldsymbol{\rho} = \delta_{\mathbf{k}\mathbf{q}} \quad (5.11)$$

we obtain the optical transition matrix element

$$W_{jki\mathbf{q}} = \langle \psi_{jk}(\mathbf{r}) | \mathbf{A} \bar{w} \mathbf{p} | \psi_{iq}(\mathbf{r}) \rangle = \delta_{\mathbf{k}\mathbf{q}} A \sum_m e_m w_{mz} \langle \psi_j(z) | p_z | \psi_i(z) \rangle \quad (5.12)$$

where $\mathbf{A} = A\mathbf{e}_s$, $\mathbf{e}_s = (e_x, e_y, e_z)$, $m = x, y, z$. We have approximated in Eq. (5.12) that the envelope function does not vary much within one unit cell so that

$$\int \psi_j^*(z) p_z \psi_i(z) |u(\mathbf{r})|^2 dz \approx \int \psi_j^*(z) p_z \psi_i(z) dz \int_{\text{cell}} |u(\mathbf{r})|^2 dz \quad (5.13)$$

The reader can find similar but more detailed mathematical manipulations in Sect. 3.3.

Equation (5.12) indicates that the intraband optical transition occurs only when the momentum in the xy plane is conserved. And the transition is between the envelope functions in the z direction. These have been discussed in the previous section. Note that

$$\delta_{\mathbf{k}\mathbf{q}} = \begin{cases} 1, & \text{if } \mathbf{k} = \mathbf{q} \\ 0, & \text{otherwise} \end{cases} \quad (5.14)$$

so that we can simplify the notation of the optical transition matrix element of Eq. (5.12) to be $W_{ji}(\mathbf{k})$.

Table 5.1 Material parameters of four commonly used semiconductors. The unit of the inverse effective mass elements is $1/m_0$, where m_0 is the free electron mass

Material	Major axis	Dielectric constant ϵ	w_t	w_ℓ
AlAs	[001]	10.1	5.263	0.909
AlGaSb	[111]	13.0	6.304	0.766
Si	[001]	11.8	5.263	1.020
Ge	[111]	16.0	12.195	0.610

As discussed briefly in the previous section, and is further shown by Eq. (5.12), the optical transition matrix element $W_{ji}(\mathbf{k})$ will be zero for an electromagnetic field which is propagating along the z direction ($e_z = 0$) if $w_{xz} = w_{yz} = 0$. For spherical bands, $w_{mn} = w_{nm} = 0$ when $m \neq n$, see Eq. (1.74), so that $W_{ji}(\mathbf{k}) = 0$ when $e_z = 0$. This implies that the quantum well made of spherical bands will not absorb light which incidents normally to the quantum well surface. We will discuss such a case in the next section.

On the other hand, it is a new story for an ellipsoidal band described by Eq. (1.75). We know from Chap. 1 that the conduction band of Si consists of six ellipsoids described by a longitudinal effective mass $m_\ell = 0.9163$ along the major axis of the ellipsoid and a transverse effective mass $m_t = 0.1905$ in the plane perpendicular to the major axis, expressed in the unit of free electron mass m_0 . For almost all the indirect-gap semiconductor materials, the conduction bands have either six symmetrical ellipsoidal constant energy surfaces (usually the X valley electrons), or eight symmetrical half ellipsoidal constant energy surfaces (usually the L valley electrons). Table 5.1 lists material parameters of four common semiconductors AlAs, AlGaSb, Si and Ge. For X valley electrons, there are six ellipsoids whose major axes are oriented along three crystal axes, [100], [010] and [001], which form the principal coordinate system XYZ . The inverse effective mass tensor \bar{w}_p is diagonal in this principal coordinate system. For example, the \bar{w}_p matrix, for the two degenerate ellipsoids whose major axes are in the [001] direction, is in the form of

$$\bar{w}_p = \begin{pmatrix} w_t & 0 & 0 \\ 0 & w_t & 0 \\ 0 & 0 & w_\ell \end{pmatrix} \quad (5.15)$$

where $w_t = 1/m_t$, $w_\ell = 1/m_\ell$. The equation of the constant energy surface of these two ellipsoids is

$$\frac{\hbar^2}{2} (w_t k_X^2 + w_t k_Y^2 + w_\ell k_Z^2) = E \quad (5.16)$$

Table 5.2 lists degeneracy g of the lowest energy valleys. By Eq. (5.12) we can expect the similar quantum selection rule that the quantum wells made from ellipsoidal bands can not absorb radiations that incident normally to the quantum well surfaces if the quantum wells are grown along one of the major axes.

We now discuss a quantum well whose growth direction, i.e., the z axis, does not coincide with any of the three major axes. Let us denote the quantum well growth

Table 5.2 Degeneracy g of the lowest energy valleys $[XYZ]$

X valley			L valley		
$[XYZ]$	g	Occupied valleys	$[XYZ]$	g	Occupied valleys
$X = Y = Z$	6	all	$X = Y = 0$	4	all
$X = Y > Z$	4	$[100][010]$	$Z = 0$	2	$[111][11\bar{1}]$
$X > Y > Z$	2	$[100]$	otherwise	1	$[111]$

direction by its crystal direction notation $[\ell mn]$. The choice of the x and y axes is arbitrary as long as the two axes are perpendicular to $[\ell mn]$ as well as perpendicular to each other. In order to simplify the mathematics, we choose the y axis to be perpendicular to not only the already defined z axis, but also to one major axis of the ellipsoid. The x axis is then uniquely determined by its orthogonality with both the y and z axes. In this xyz coordinate system, the inverse effective mass tensor has nonzero off-diagonal elements, which will enable this type of quantum well to detect normal incident radiation. Let us analyze the situation step by step.

The general equation of the constant energy surface is

$$\frac{\hbar^2}{2}(w_{xx}k_x^2 + w_{yy}k_y^2 + w_{zz}k_z^2 + 2w_{xy}k_xk_y + 2w_{yz}k_yk_z + 2w_{zx}k_zk_x) = E \quad (5.17)$$

which results in the following effective-mass Schrödinger equation

$$\left[-\frac{\hbar^2}{2} \left(w_{xx} \frac{d^2}{dx^2} + w_{yy} \frac{d^2}{dy^2} + w_{zz} \frac{d^2}{dz^2} + 2w_{xy} \frac{d^2}{dx dy} + 2w_{yz} \frac{d^2}{dy dz} + 2w_{zx} \frac{d^2}{dz dx} \right) + V(z) \right] \psi_i(x, y, z) = E \psi_i(x, y, z) \quad (5.18)$$

see Sect. 1.8. Here $V(z)$ is the potential energy due to the heterostructure of the quantum well along the z axis.

The first-order derivatives with respect to z in the above equation can be eliminated by letting

$$\psi_i(x, y, z) = \xi_i(z) \exp \left[\frac{-iz(w_{xz}k_x + w_{yz}k_y)}{w_{zz}} \right] e^{i(k_x x + k_y y)} \quad (5.19)$$

so that the partial envelope function $\xi_i(z)$ in the z direction is determined by

$$\left[\frac{-w_{zz}\hbar^2}{2} \frac{d^2}{dz^2} + V(z) \right] \xi_i(z) = E_i \xi_i(z) \quad (5.20)$$

We call $\xi_i(z)$ as the partial envelope function in the z direction to distinguish the complete envelope function

$$\psi_i(z) = \xi_i(z) \exp \left[\frac{-iz(w_{xz}k_x + w_{yz}k_y)}{w_{zz}} \right]$$

The total energy of Eq. (5.18) is

$$E_{ik} = E_i + E_k \quad (5.21)$$

where E_k is the kinetic energy in the xy plane.

$$E_k = \frac{\hbar^2}{2} (w_{xx}k_x^2 + 2w_{xy}k_xk_y + w_{yy}k_y^2) \quad (5.22)$$

In many practical samples, the barrier is quite high and the effective mass is large, so that the wave function penetration into the barrier region can be neglected and the barrier can be approximated as infinitely high. Take L_{QW} as the quantum well width,

$$\xi_{i-1}(z) = \sqrt{\frac{2}{L_{\text{QW}}}} \sin\left(\frac{i\pi z}{L_{\text{QW}}}\right), \quad E_{i-1} = \frac{i^2\pi^2\hbar^2w_{zz}}{2L_{\text{QW}}^2} \quad (5.23)$$

where i is an integer. It is noticed here that since the sublevel is normally indexed as $0, 1, 2, \dots$ where E_0 denotes the ground sublevel, E_1 the first excited sublevel and so on, whereas i in Eq. (5.23) must be a non-zero integer, we denote the wave function and the corresponding eigen energy by $(i - 1)$.

If we only consider the transition between the ground and the first excited sublevels,

$$E_1 - E_0 = \frac{3\pi^2\hbar^2w_{zz}}{2L_{\text{QW}}^2}, \quad \left| \left\langle \psi_1(z) \left| \frac{\partial}{\partial z} \right| \psi_0(z) \right\rangle \right|^2 = \frac{64}{9L_{\text{QW}}^2} \quad (5.24)$$

For the second relationship, see Eq. (5.6).

We obtain the absorption coefficient

$$\alpha(\hbar\omega_s) = \frac{32e^2\hbar^2\Gamma n_s}{9c\epsilon\omega_s L_{\text{QW}}^3 \left[\left(\frac{3\pi^2\hbar^2w_{zz}}{2L_{\text{QW}}^2} - \hbar\omega_s \right)^2 + \Gamma^2 \right]} \left| \sum_m e_m w_{mz} \right|^2 \quad (5.25)$$

by Eq. (5.3). Γ is the lifetime of the excited state, ϵ and c are the dielectric constant and the speed of light, respectively. $\mathbf{e}_s = (e_x, e_y, e_z)$ is the polarization vector of the electromagnetic field and $\hbar\omega_s$ is the photon energy, n_s is the sheet density of electrons occupying E_0 in the quantum well which will be discussed a bit later.

It is obvious that the absorption is maximal if $E_1 - E_0 = \hbar\omega_s$, the corresponding optimal absorption α_{max} is (apart from constants)

$$\alpha_{\text{max}} \propto \frac{1}{L_{\text{QW}}w_{zz}} \left| \sum_m e_m w_{mz} \right|^2 \quad (5.26)$$

Note Eq. (5.24) so that $E_1 - E_0 = \hbar\omega_s$ means

$$\frac{3\pi^2\hbar^2w_{zz}}{2L_{\text{QW}}^2} = \hbar\omega_s \quad (5.27)$$

In other words, L_{QW} , w_{zz} and $\hbar\omega_s$ are inter-correlated. Device design normally asks for a quantum well structure with the best absorption coefficient for a fixed optical energy. Thus, for a fixed $\hbar\omega_s$, the above equation indicates that we need to modify L_{QW} proportional to $\sqrt{w_{zz}}$ when w_{zz} is changed due to the change of the growth direction $[\ell mn]$. Thus, for a fixed $\hbar\omega_s$, the optical absorption coefficient becomes

$$\alpha_{\max} \propto \frac{1}{w_{zz}^{3/2}} \left| \sum_m e_m w_{mz} \right|^2 \quad (5.28)$$

When considering the normal incident and non-polarized radiation, i.e., $e_x = e_y$ and $e_z = 0$ so that

$$\begin{aligned} e_z &= 0 \\ \langle e_x w_{xz} + e_y w_{yz} \rangle^2 &= \langle e_x \rangle^2 w_{xz}^2 + \langle e_y \rangle^2 w_{yz}^2 \\ \langle e_x \rangle^2 &= \langle e_y \rangle^2 = \frac{1}{2} \end{aligned} \quad (5.29)$$

we obtain

$$\alpha_{\max} \propto \frac{w_{xz}^2}{w_{zz}^{3/2}} \quad (5.30)$$

We need to calculate w_{xz} and w_{zz} in order to estimate and optimize the absorption coefficient of Eq. (5.30). Knowing $[\ell mn]$ and thereafter coordinate xyz , we are able to calculate the inverse effective mass elements w_{mn} in Eq. (5.18) from \bar{w}_p by coordinate transformation. The coordinate transformation matrix \bar{T} between coordinate system XYZ and xyz is unitary, i.e., $\bar{T}^{-1} = \bar{T}^\tau$, where \bar{T}^{-1} and \bar{T}^τ are the inversion and transpose of matrix \bar{T} , respectively. The inverse effective mass tensor \bar{w} in coordinate system xyz is

$$\bar{w} = \bar{T}^\tau \bar{w}_p \bar{T} \quad (5.31)$$

For the ellipsoid whose major axis is in the $[001]$ direction of the principal coordinate system,

$$\bar{T} = \begin{pmatrix} \frac{\ell n}{r\sqrt{\ell^2+m^2}} & \frac{-m}{\sqrt{\ell^2+m^2}} & \frac{\ell}{r} \\ \frac{mn}{r\sqrt{\ell^2+m^2}} & \frac{\ell}{\sqrt{\ell^2+m^2}} & \frac{m}{r} \\ \frac{-(\ell^2+m^2)}{r\sqrt{\ell^2+m^2}} & 0 & \frac{n}{r} \end{pmatrix} \quad (5.32)$$

where $r^2 = \ell^2 + m^2 + n^2$. For other ellipsoids, the \bar{T} matrices can be obtained in the similar way.

For the L valley material, the calculation is a bit more complicated since the principal axes do not coincide with any of $[100]$, $[010]$, $[001]$ directions. What we do is first transfer \bar{w}_p , expressed in the principal coordinate system XYZ , to the one

in [100], [010], [001] system, and then to the system xyz determined by the growth direction. Therefore, Eq. (5.31) should be modified as

$$\bar{w} = \bar{T}^\tau \bar{B} \bar{w}_\rho \bar{B}^\tau \bar{T} \quad (5.33)$$

whose elements are

$$w_{mn} = \sum_k T_{km} \sum_{k'} B_{kk'} w_{k'} \sum_{k''} B_{k''k'} T_{k''n} \quad (5.34)$$

For the ellipsoid whose major axis is in the [111] direction,

$$\bar{B} = \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{-2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \end{pmatrix} \quad (5.35)$$

Moreover, since the y axis of coordinate system xyz is set to be perpendicular to the major axis of the ellipsoid, $w_{yy} = w_t$ and $w_{xy} = w_{yz} = 0$. Thus,

$$\bar{T} = \begin{pmatrix} \frac{(m+n)\ell - (m^2+n^2)}{rr_1} & \frac{n-m}{r_1} & \frac{\ell}{r} \\ \frac{(n+\ell)m - (n^2+\ell^2)}{rr_1} & \frac{\ell-n}{r_1} & \frac{m}{r} \\ \frac{(\ell+m)n - (\ell^2+m^2)}{rr_1} & \frac{m-\ell}{r_1} & \frac{n}{r} \end{pmatrix} \quad (5.36)$$

where $r^2 = \ell^2 + m^2 + n^2$, as defined before, and $r_1^2 = (n-m)^2 + (\ell-n)^2 + (m-\ell)^2$.

w_{mn} in coordinate system xyz can then be obtained by the above transformations. However, we hereby present an alternative and simple method instead of the direct coordinate transformation. Obviously w_{xz} and w_{zz} are linear functions of w_t and w_ℓ , see Eq. (5.34). And it is known that if $w_t = w_\ell$, the inverse effective mass tensor in the principal coordinate system is a numerical matrix, i.e., a unit matrix multiplied by a constant. Such a matrix is invariable under any unitary transformation. It is thus expected that for any w_t and w_ℓ ,

$$w_{zz} = (1-a)w_t + aw_\ell \quad (5.37)$$

$$w_{xz} = b(w_t - w_\ell)$$

In other words, in the expression of w_{zz} , the sum of the coefficients of w_t and w_ℓ equals one, while for w_{xz} , it is zero. Thus, the calculation of w_{xz} and w_{zz} becomes extremely simplified. By Eq. (5.34), we can get a and b of Eq. (5.37) as

$$a = \left(\sum_k T_{kz} B_{kz} \right)^2, \quad b = - \left(\sum_k T_{kx} B_{kz} \right) \left(\sum_k T_{kz} B_{kz} \right) \quad (5.38)$$

Equations (5.37, 5.38) are also valid for X valleys after setting \bar{B} as an unit matrix. Finally we have

$$\begin{aligned} w_{zz} &= (1 - T_{zz}^2)w_t + T_{zz}^2 w_\ell \\ w_{xz} &= -T_{zx}T_{zz}w_t + T_{zx}T_{zz}w_\ell \end{aligned} \quad (5.39)$$

This method is of great convenience to determine the elements of the inverse effective mass tensor in the new coordinate system whose z axis is along the quantum well growth direction. However, this calculation of inverse mass tensor elements can be further simplified by a more detailed investigation.

In the transformation of the coordinate systems, there are four mathematically invariable quantities

$$\begin{aligned} \begin{vmatrix} w_{xx} & w_{xy} & w_{xz} & 0 \\ w_{xy} & w_{yy} & w_{yz} & 0 \\ w_{xz} & w_{yz} & w_{zz} & 0 \\ 0 & 0 & 0 & -E \end{vmatrix} &= \begin{vmatrix} w_t & 0 & 0 & 0 \\ 0 & w_t & 0 & 0 \\ 0 & 0 & w_\ell & 0 \\ 0 & 0 & 0 & -E \end{vmatrix} = -E w_t^2 w_\ell \\ \begin{vmatrix} w_{xx} & w_{xy} & w_{xz} \\ w_{xy} & w_{yy} & w_{yz} \\ w_{xz} & w_{yz} & w_{zz} \end{vmatrix} &= \begin{vmatrix} w_t & 0 & 0 \\ 0 & w_t & 0 \\ 0 & 0 & w_\ell \end{vmatrix} = w_t^2 w_\ell \end{aligned} \quad (5.40)$$

$$w_{xx} + w_{yy} + w_{zz} = 2w_t + w_\ell$$

$$w_{xx}w_{yy} + w_{yy}w_{zz} + w_{zz}w_{xx} - w_{xy}^2 - w_{yz}^2 - w_{zx}^2 = w_t^2 + 2w_t w_\ell$$

In above equations we have chosen $w_{XX} = w_{YY} = w_t$ and $w_{ZZ} = w_\ell$. Other choices of \bar{w}_p elements give the same results as shown by the above equation.

By these four equations it is easy to see

$$w_{xx}w_{zz} - w_{xz}^2 = w_t w_\ell \quad (5.41)$$

$$w_{xx} + w_{zz} = w_t + w_\ell$$

since $w_{yy} = w_t$ and $w_{xy} = w_{yz} = 0$ because of the choice of the y axis. These two equations are universal and thus are very useful to simplify the calculation of the elements of the inverse mass tensor and the discussion of the limits of the normal incident radiation absorption as well as its optimization. From these two equations and Eqs. (5.37), we can write w_{xz} in a much simpler form

$$w_{xz}^2 = a(1-a)(w_t - w_\ell)^2 \quad (5.42)$$

where a is defined by Eq. (5.38). We can therefore write the following equations

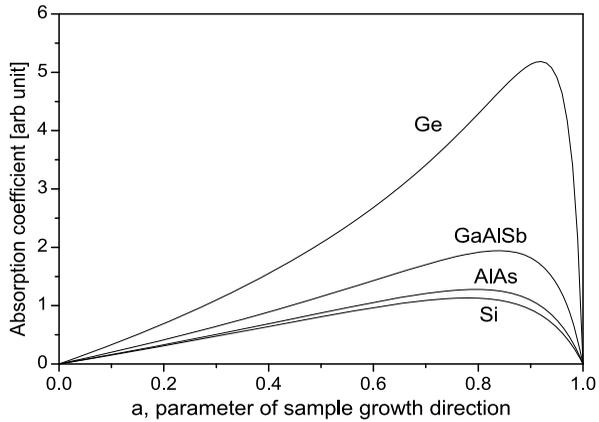
$$w_{zz} = (1-a)w_t + aw_\ell \quad (5.43)$$

$$w_{xz} = \sqrt{a(1-a)}(w_t - w_\ell)$$

Table 5.3 The parameter a for X and L valleys. $[XYZ]$ denotes the principal coordinate system of the major axis, while $[\ell mn]$ is the quantum well growth direction. $r^2 = \ell^2 + m^2 + n^2$

X valley			L valley		
Major axis	ar^2		Major axis	$3ar^2$	
$[XYZ] =$	$[100]$	ℓ^2	$[XYZ] =$	$[111]$	$(\ell + m + n)^2$
	$[010]$	m^2		$[\bar{1}11]$	$(-\ell + m + n)^2$
	$[001]$	n^2		$[1\bar{1}\bar{1}]$	$(\ell - m + n)^2$
				$[11\bar{1}]$	$(\ell + m - n)^2$

Fig. 5.2 Absorption coefficient of the normal incident radiation based on the intersubband transition as a function of the parameter a for Ge, GaAlSb, AlAs and Si quantum well materials (W. Xu, Y. Fu, M. Willander, and S. C. Shen, Theory of normal incident absorption for the intersubband transition in n -type indirect-gap semiconductor quantum wells, Phys. Rev. B, vol. 49, pp. 13760–13766, 1994)



In other words, we have transformed the calculation of elements of the inverse effective mass tensor into the calculation of one parameter a , which is evidently much simpler. In Table 5.3, we list the expressions of the parameter a for X - and the L -valley electrons.

Back to the absorption equation, Eq. (5.30) now becomes

$$\alpha_{\max} \propto \frac{a(1-a)(w_t - w_\ell)^2}{[(1-a)w_t + aw_\ell]^{3/2}} \tag{5.44}$$

by inserting Eq. (5.43).

The well-known quantum mechanical selection rule about optical transition is again clearly reflected in the above equation. Namely, if the quantum well is grown along one of the major axes of the semiconductor material, $a = 1$ so that $\alpha_{\max} = 0$ which means that optical transition is not possible when the propagation direction of the incident radiation coincides with the quantum well growth direction.

Figure 5.2 shows the absorption coefficient α as a function of parameter a for four quantum well materials which have been widely investigated.

Table 5.4 Maximal absorption coefficient and optimal growth directions. The photon energy $\hbar\omega_s$ is in the unit of eV and the quantum-well width L_{QW} in Å. The unit of the inverse effective mass elements is $1/m_0$, where m_0 is the free electron mass

Material	Valley	Major axis	w_t	w_ℓ	a'	$\frac{\alpha_{\text{max}}}{\sqrt{\hbar\omega_s}}$	$L_{\text{QW}}\sqrt{\hbar\omega_s}$	$\sqrt{w_t}$	Optimal $[\ell mn]$
AlAs	X	[001]	5.263	0.909	0.795	1.278	14.17	2.294	[102][113][203]
GaAsSb	L	[111]	6.304	0.766	0.839	1.941	13.25	2.511	[123][133][112]
Si	X	[001]	5.263	1.020	0.780	1.132	14.52	2.294	[102][113][203]
Ge	L	[111]	12.195	0.610	0.918	5.182	12.87	3.492	[122][112][133]

It is easy to obtain the condition of a maximal $\alpha_{\text{max}}(a)$ as a function of a with respect to a such

$$\left. \frac{d\alpha_{\text{max}}(a)}{da} \right|_{a=a'} = 0 \quad (5.45)$$

from which we obtain

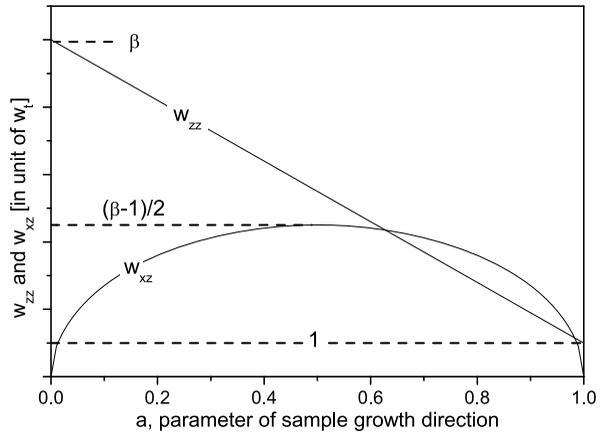
$$a' = \frac{3w_t + w_\ell - \sqrt{w_t^2 + w_\ell^2 + 14w_t w_\ell}}{2(w_t - w_\ell)} \quad (5.46)$$

Using the formulas listed in Table 5.3, we can calculate the a parameter for the different growth directions of the quantum well. By comparing the calculated a parameter with a' of Eq. (5.46), the optimal growth direction for the absorption of the normal incident radiation can be obtained.

For X electrons in the energy valley with an [100] major axis and L electrons with an [111] major axis, we present the calculation results in Table 5.4 for the four different quantum well materials. The value of $\sqrt{w_t}$ is also given, which will be useful for later discussions. It is clear that if the calculated optimal growth direction is $[\ell mn]$, the direction of $[m\ell n]$ for the [001] X valley, and the directions obtained by the index rotation for the [111] L valley are also optimal because of the symmetry. The optimal growth directions for other valleys can be obtained in the similar way. It should be pointed out that in Table 5.4 only low-index optimal growth directions of the quantum well detector have been listed, where we have not taken into account the problems of the real material growth. In addition, the results are obtained for only a single valley whose major axis is listed in the table.

Figure 5.2 and Table 5.4 clearly demonstrate that for maximal absorption, a large ratio between w_t and w_ℓ is desirable, and among these four quantum well materials studied, Ge is the best candidate for the largest absorption for the normal incident radiation based on the intersubband transitions. This conclusion agrees with the experimental results [1]. It is of interest to compare the absorption coefficients of indirect-gap systems with well-known direct-gap systems, e.g., GaAs quantum well. Still keeping $\hbar\omega_s$ constant, it is easy to obtain from Eq. (5.25) that for a non-polarized incident radiation $\alpha = 0.48$ at 45 degrees incident illumination for GaAs well where $w_t = w_\ell = 15.0$.

Fig. 5.3 Elements of inverse effective-mass tensor w_{zz} and w_{xz} (in unit of w_t) vs parameter a (W. Xu, Y. Fu, M. Willander, and S. C. Shen, Theory of normal incident absorption for the intersubband transition in n -type indirect-gap semiconductor quantum wells, Phys. Rev. B, vol. 49, pp. 13760–13766, 1994)



Let us compare the absorption of the normal incident radiation with that of the parallel incident radiation for same quantum well structure based on the intraband optical transitions. For the parallel incident radiation, $e_x = 0$ and $e_y = e_z$, and by Eq. (5.25), the absorption coefficient α_{\parallel} can be written as

$$\alpha_{\parallel} \propto \frac{w_{zz}^2}{w_{zz}^{3/2}} = \sqrt{w_{zz}} \quad (5.47)$$

It is then more convenient to compare w_{xz} with w_{zz} instead of comparing Eq. (5.30) with Eq. (5.47). From Eqs. (5.37), (5.42), assuming $w_t = \beta w_{\ell}$, we have

$$\begin{aligned} w_{zz} &= [(1-a)\beta + a]w_{\ell} \\ w_{xz} &= \sqrt{a(1-a)(1-\beta)}w_{\ell} \end{aligned} \quad (5.48)$$

Obviously, the parameter β is related with the anisotropic property of the effective mass tensor of the quantum well material ($0 < \beta \leq 1$). $\beta = 1$ means that the effective mass is isotropic so that $w_{xz} = 0$. The absorption coefficient of the intersubband transition for the normal incident radiation vanishes, similar to the Γ electrons in a GaAs/GaAlAs quantum well. The parameter a , defined in Eq. (5.38), represents the orientation characteristics of the quantum well ($0 \leq a \leq 1$). Both $a = 0$ and $a = 1$ imply that the growth direction of the quantum well coincides with one of the principal axes so that $w_{xz} = 0$, and thus the absorption coefficient of the intraband transition for the normal incident radiation vanishes though the quantum well is made of the indirect gap semiconductor material. Figure 5.3 shows w_{zz} and w_{xz} as functions of parameter a . By a very simple calculation, it is found out that the straight line of w_{zz} and the curve of w_{xz} have two cross-points only if

$$\beta > \beta_0 = 3 + \sqrt{8} = 5.828$$

In this case, when

$$\frac{3\beta - 1 - \sqrt{\beta^2 - 6\beta + 1}}{4(\beta - 1)} < a < \frac{3\beta - 1 + \sqrt{\beta^2 - 6\beta + 1}}{4(\beta - 1)} \quad (5.49)$$

we will have $w_{xz} > w_{zz}$. In other words, the absorption of the normal incident radiation is larger than that of the parallel incident radiation. Thus, for the same quantum well, the conditions for a large absorption of the normal incident radiation are as follow: First, the anisotropy effect of the effective mass tensor of the quantum well material should be large enough, so that the longitudinal mass m_ℓ is at least 5.828 times larger than the transverse mass m_t . Secondly, the growth direction of the quantum well must satisfy Eq. (5.49). For example, for electrons in GaAlSb L valley, $\beta = 8.229 > \beta_0$. If one chooses [213] or [233] as the growth direction $[\ell mn]$, the parameter a is 0.8571 or 0.9607, which satisfies Eq. (5.49). Thus, the absorption of the normal incident radiation is larger than that of the parallel incident radiation in this quantum well. For a quantum well made of Ge, the same analysis can be done and a similar conclusion is obtained. However, for Si quantum well, since $\beta = 5.16 < \beta_0$, the absorption of the normal incident radiation is always smaller than that of the parallel incident radiation for any growth direction.

Because the maximal value of the absorption coefficient of the parallel incident radiation based on the intersubband transitions is $\alpha_{\parallel\max} \propto \sqrt{w_t}$ (when the growth direction coincides with one major axis), as listed in Table 5.4, it is noticed that for a quantum well, the maximal value of the absorption coefficient of the parallel incident radiation is not always greater than that of the normal incident radiation.

Now we discuss the sheet density n_s of electrons occupying E_0 in Eq. (5.25). The total absorption coefficient is the sum of the ones from all occupied valleys, it is thus necessary to investigate the degeneracies and the occupancies of different energy valleys. For many practical applications, the total sheet density of electrons in the quantum well is about $5 \times 10^{12} \text{ cm}^{-2}$. Let us examine the position of the Fermi level E_f with respect to the ground sublevel E_0 , for which we need to know the density of the electron states N_2 in the xy plane, which can be obtained by straightforward calculation similar to Sect. 1.9. However there is a simpler way.

The three-dimensional density of electron states for an ellipsoidal energy valley with transverse and longitudinal effective masses m_t and m_ℓ is

$$N_3 \propto m_t^2 m_\ell \quad (5.50)$$

in the principal coordinate system, see Eqs. (1.82), (1.83). It remains the same for the ellipsoidal valley in any Cartesian coordinate system xyz by the method of invariables during coordinate transformation, see Eqs. (5.40). Most important of all is the fact that N_3 is a physical quantity and thus is independent of the choice of the coordinate system. On the other hand, we understand by Sect. 1.9 that the density of states of a one-dimensionally extended system $N_1 \propto \sqrt{m^*}$, where m^* is the effective mass along the extended dimension. For the quantum well under investigation, the effective mass along the z direction is $1/w_{zz}$, see Eq. (5.20). It is thus easy see that

the two-dimensional density of states of the ellipsoidal valley in the xyz coordinate system is

$$N_2 = \frac{1}{\pi \hbar^2 L_{QW}} \sqrt{\frac{w_{zz}}{w_l^2 w_\ell}} \quad (5.51)$$

Let us examine the position of the Fermi level E_f with respect to the ground sublevel E_0 . We know from previous chapters that the sheet density of electrons occupying state E_0 is

$$n_s = g N_2 \int_{E_0}^{\infty} \frac{dE}{1 + \exp[(E_0 - E_f)/k_B T]} = g N_2 k_B T \ln \left[1 + \exp\left(\frac{E_f - E_0}{k_B T}\right) \right] \quad (5.52)$$

which becomes

$$n_s = g N_2 (E_f - E_0) \quad (5.53)$$

at low temperatures. Here g is the degenerate factor which is listed in Table 5.2 for X and L valley electrons. The six X valleys are degenerate in bulk materials. For a quantum well grown along the $[\ell mn]$ direction, the six valleys are still degenerate if $\ell = m = n$. If $\ell = m > n$, two of the six X valleys are lifted up so that the degeneracy of the lowest energy valleys becomes 4, while if $\ell > m > n$, the lowest energy valleys are only two-fold degenerate. Similar considerations can be carried out for L valleys. For a sheet density of carriers of $n_s = 5 \times 10^{12} \text{ cm}^{-2}$, it is only the lowest energy valleys which are occupied. For example, for an n -type Si quantum well grown on the $[110]$ direction, $g = 4$,

$$E_f - E_0 = 9.0 \text{ meV}$$

while

$$E_1 - E_0 = 27.9 \text{ meV}$$

Among the four commonly used heterostructure materials (AlAs, AlGaSb, Si and Ge), $(E_f - E_0)$ of AlGaSb grown along the $[111]$ direction is the largest, which is 75.0 meV, $(E_1 - E_0)$ is however even larger (265.6 meV).

Table 5.5 shows results of the optimal growth directions of different quantum wells. It is found that in the listed quantum well structures, the most common optimal growth direction is $[110]$. The Ge/SiGe quantum well which has the largest anisotropy of the effective mass has the largest absorption coefficient for normal incident radiation. For high doping concentration or other detecting wavelengths, the conclusions may be different, but the physical principle is the same.

For four most promising n -type quantum wells, the absorption coefficient is plotted as a function of parameter a (characterizing the sample growth direction) in Fig. 5.2. We also list in Table 5.4 the optimal growth directions. We have taken $c = c_0/\sqrt{\epsilon}$, where ϵ is in the unit of ϵ_0 . The permeability of the sample is assumed to have the value of the free space. Since the dielectric constant of GaSb is 16.0, the dielectric constant of AlGaSb is approximated as 13.0, the average between GaSb and AlAs.

Table 5.5 Quantum well growth directions and the total absorption coefficient

Material	Growth direction	Occupied valley	α_{total} (Arb. unit)
AlAs	[110]	[100][010]	3.50
	[111]	all	3.40
AlGaSb	[110]	[111][11 $\bar{1}$]	3.23
	[100]	all	2.90
	[102]	[111][1 $\bar{1}$ 1]	2.86
Si	[110]	[100][010]	3.23
	[111]	all	3.18
Ge	[110]	[111][11 $\bar{1}$]	6.31
	[203]	[111][1 $\bar{1}$ 1]	5.93
	[102]	[111][1 $\bar{1}$ 1]	5.26

It is noticed that in addition to the dependence of the absorption coefficient on the sample growth direction, through the parameter of w_{zz} in Eq. (5.24), the quantum well width L_{QW} is another critical factor in determining the response wavelength of the photodetector. This is the reason that $L_{\text{QW}}\sqrt{\hbar\omega_s}$ and $\alpha_{\text{max}}/\sqrt{\hbar\omega_s}$ are listed in Table 5.4. Table 5.4 thus provides a guild line for designing the optical quantum well photodetector: the optimal growth direction and the quantum well width.

We now extend the discussions to the optical transition between mini bands. Semiconductor superlattice provides us with a good way to increase the mobility of the photocarriers. For a single quantum well or multiple quantum well systems, the photo-excited carriers are still largely localized in the quantum well. For a superlattice, the ground states in each quantum well can be quite well separated from one another, while the excited states form delocalized mini bands so that the photocarriers are more mobile.

When forming mini bands in the z direction, the quantum designation m consists of the subband index m and the wave vector q_z in the form of

$$\begin{aligned}\xi_{m,q_z}(z) &= \frac{1}{\sqrt{M}} v_{m,q_z}(z) e^{iq_z z} \\ v_{m,q_z}(z) &= v_{m,q_z}(z+d)\end{aligned}\tag{5.54}$$

where d is the sum of the well and barrier widths (i.e., the period of the superlattice), M is the total sample thickness. $v_{m,q_z}(z)$ is normalized within one unit period of the superlattice.

$$\begin{aligned}\left\langle \xi_{m,q_z} \left| \frac{\partial}{\partial z} \right| \xi_{n,k_z} \right\rangle \\ = \frac{1}{M} \int v_{m,q_z}^* \frac{\partial v_{n,k_z}}{\partial z} e^{i(k_z - q_z)z} dz + \frac{ik_z}{M} \int v_{m,q_z}^* v_{n,k_z} e^{i(k_z - q_z)z} dz\end{aligned}$$

$$= \frac{\delta_{k_z, q_z}}{d} \int_d v_{m, q_z}^* \frac{\partial v_{n, k_z}}{\partial z} dz \quad (5.55)$$

where q_z and k_z are limited within the first Brillouin zone in the z direction. Momentum conservation in the xy plane is required. In addition, we see by the above equation that the momentum in the z direction is also conserved. Since the momentum of a photon is negligibly small as compared with that of the electron, the optical transition is always vertical in the \mathbf{k} space.

The total transition probability from ground state to excited state m is

$$W_m = \int \frac{2d\mathbf{Q}}{(2\pi)^3} \left| \sum_n A_n w_{nz} \langle v_m | p_z | v_0 \rangle \right|^2 \frac{2e^2\gamma}{\hbar^2(\Omega_m^2 + \gamma^2)} f[E_0(\mathbf{Q})] \{1 - f[E_m(\mathbf{Q})]\} \quad (5.56)$$

where $\mathbf{Q} = (q_z, \mathbf{q})$,

$$E_m(\mathbf{Q}) = E_{\mathbf{q}} + E_m(q_z)$$

$$\hbar\Omega_m = E_m(\mathbf{Q}) - E_0(\mathbf{Q}) - \hbar\omega$$

$E_m(q_z)$ is the energy dispersion in the z direction because of the translational symmetry of the superlattice.

Let us calculate and compare the absorption coefficient of a superlattice with a quantum well. For simplicity we assume that the energy band in the xy plane is parabolic so that $E_m(\mathbf{q}) = E_0(\mathbf{q})$. We consider a n -type Si superlattice for which the lowest energy levels are six X valley states. The sample is grown along [102], as optimized in Table 5.4. The sheet density of carriers in *one* quantum well is set to 10^{12} cm^{-2} in the superlattice (with this sheet density only the ground subband is occupied). We focus on the optical transition from the ground subband to the first excited subband.

Figure 5.4(a) shows the energy dispersion in the z direction. The barrier height is assumed to be 0.125 eV. The energy dispersion shown in Fig. 5.4(a) is calculated by the standard transfer matrix method [2], see also Sect. 6.2 in the next chapter. The Si quantum well thickness is set as 50 Å while two values of barrier thickness, 10 and 20 Å are chosen to compare the absorption coefficients of the superlattice detector with a quantum well detector. Figure 5.4(a) shows that the dispersion in the z direction of the superlattice can hardly be approached as parabolic.

Figure 5.4(b) presents the absorption coefficient numerically calculated from Eq. (5.56) for a normal unpolarized incident radiation. We observe two major effects. The first effect is the energy dispersion in the z axis which reduces the number of carriers at one particular value of q_z (which is far less than n_s). The energy dispersion is induced by the overlappings of electron waves from different quantum wells (the wave penetration into the barrier region). Thus the absorption coefficient is decreased because of the reduction of the carrier numbers and the spectrum of the absorption coefficient becomes wider due to the energy dispersion. The matrix element $\langle v_m | p_z | v_0 \rangle$ is also reduced in the superlattice as compared with the quantum well. It is however the energy dispersion in the z direction that is the principal cause for the optical absorption reduction.

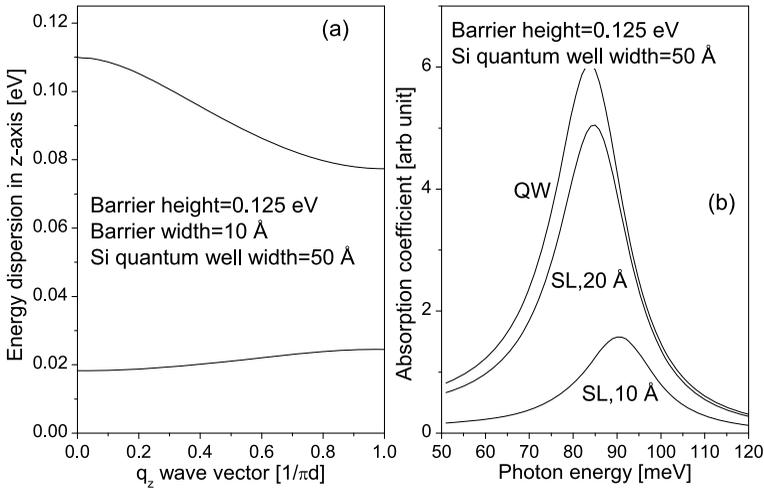


Fig. 5.4 (a) Energy dispersion of a *n*-type Si superlattice grown along [102] (the *z* axis). (b) Absorption coefficient of normal incident radiation as a function of barrier thickness for a *n*-type Si superlattice grown along [102] (Reprinted with permission from Y. Fu, M. Willander, and Wenlan Xu, Optical absorption coefficients of semiconductor quantum-well infrared detectors, *J. Appl. Phys.*, vol. 77, pp. 4648–4654, 1995. Copyright 1995, American Institute of Physics)

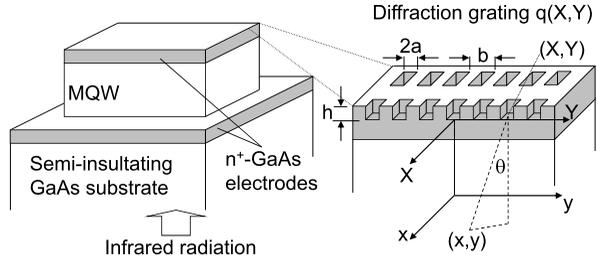
The absorption of the quantum well can be easily retrieved by increasing the barrier thickness. In the case of the *n*-type Si [102] superlattice, a barrier thickness of larger than 3 nm is enough to retrieve the absorption coefficient of the corresponding quantum well detector. For a barrier thickness of 2 nm, the Fermi level is only about 5 meV above the ground subband, far less than the distance between the first excited subband and the ground subband.

In Fig. 5.4(b), the absorption coefficient is calculated for only one quantum well in the superlattice. The total absorption coefficient of the superlattice is the number of quantum wells in the structure times the one we have calculated. This of course will increase very much the absorption coefficient of the superlattice. But the half width of the absorption peak is not improved.

5.3 Optical Grating and Crosstalk

For the most commonly used AlGaAs/GaAs materials, normal incident radiation absorption is not possible because of their isotropic effective masses of the active Γ electrons. Optical grating is required by such quantum-well infrared photodetectors (QWIPs) to diffract the normal incident radiation into non-normal transverse electromagnetic modes. One of the key requirements for the focal plane array (FPA) fabrication is to optimize the diffraction grating. In order to improve the grating coupling efficiency for the quantum-well infrared photodetectors, a lot of works have been done, such as the studies on symmetric grating [3]; long period gratings [4];

Fig. 5.5 Geometric structures of the quantum well infrared photodetector and the diffraction grating structure for optical coupling. “MQW” denotes multiple quantum wells



a waveguide combined grating coupler [5–7] and metallic diffraction grating [8–10]. Random gratings [11, 12] and quantum-well-infrared-photodetector structure contained grating have also been proposed to improve the detector performance.

Here we try to give an outline of the grating theory for quantum-well infrared photodetectors based on the well-known Huygen’s principle and Kirchhoff formula to calculate the absorption coefficient for a general shaped reflection grating coupled quantum-well infrared photodetector.

The AlGaAs/GaAs quantum well infrared photodetector is prepared normally as follows: semi-insulating (100) GaAs substrate, GaAs contact layer (1.0 μm , Si-doped 10^{18} cm^{-3}), 50 periods of quantum wells (thus denoted as multiple quantum wells), GaAs top contact layer (2.0 μm Si-doped 10^{18} cm^{-3}). The quantum well is made of $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ barrier (45 ~ 50 nm thick) and GaAs quantum well (4 ~ 5 nm). The structure is schematically shown in Fig. 5.5.

We consider a two-dimensional diffraction grating shown in Fig. 5.5(b). The size of one photodetector pixel is $L \times L$. The diffraction grating structure consists of $2a \times 2a$ rectangular apertures, or circular apertures with radius a periodically arranged on the optical pixel, at distance b from each other. The depth of the apertures is denoted as h .

For electromagnetic waves propagating in semiconductor materials with negligible electric conductivity, the wave equation for the electric field \mathbf{E} , Eq. (3.18) becomes

$$\nabla^2 \mathbf{E} = \epsilon \mu \frac{\partial^2 \mathbf{E}}{\partial t^2} \quad (5.57)$$

where ϵ is the permittivity or dielectric constant, μ is the magnetic permeability of the medium. In our case we simply assume $\mu = 1$. The velocity of the light waves in free space is $c = \sqrt{\epsilon_0 \mu_0}$, it is $v = c/n$ in the medium, $n = \sqrt{\epsilon/\epsilon_0}$ is the refractive index.

For the diffraction radiation detection experiments with unpolarized incident radiation, the vector nature of the wave amplitude will not be important. The only consequence of the vector properties is multiplication of the diffracted intensity by a polarization factor, depending only on the angle of diffraction. Hence we may assume that we are dealing with a scalar function $\phi(\mathbf{r})$ which is a solution to the wave equation [13]

$$\nabla^2 \phi = \epsilon \mu \frac{\partial^2 \phi}{\partial t^2} \quad (5.58)$$

For pure elastic scattering, we need consider only a single frequency ω so that the wave equation is of the form

$$\nabla^2 \phi + 4\pi^2 k^2 \phi = 0 \quad (5.59)$$

where $2\pi k = \sqrt{\epsilon\mu}\omega$ is the wave number.

An important solution to Eqs. (5.59) is

$$\phi_0 = \frac{e^{-i2\pi kr_Q}}{r_Q} \quad (5.60)$$

which corresponds to a spherical wave radiating from a point source Q .

Huygen's principle indicates that the propagation of waves through space involves the generation of secondary waves at each point of a wave front so that the envelope of the secondary waves becomes a new wave front. By this the Kirchhoff formula is written in the form of: The wave field ϕ at point Q due to any source wave ϕ_0 is given by integrating over any closed surface S containing Q as [14, 15]

$$\phi = \frac{1}{4\pi} \oint_S \left[\frac{e^{-i2\pi kr}}{r} \cdot \nabla \phi_0 - \phi_0 \cdot \nabla \left(\frac{e^{-i2\pi kr}}{r} \right) \right] dS \quad (5.61)$$

where r is the distance from point Q to the surface S .

The optical wave function for the refractive diffraction grating defined by its reflection coefficient $q(X, Y)$ which is perpendicular to the direction z of radiation incidence, becomes

$$\begin{aligned} \phi(x, y, R) = & \frac{1}{4\pi} \iint q(X, Y) \frac{e^{-i2\pi kr}}{r} \cdot \frac{e^{-i2\pi kr_Q}}{r_Q} \\ & \times \left[\left(\frac{1}{r} + i2\pi k \right) \cos \theta - \left(\frac{1}{r_Q} + i2\pi k \right) \cos \theta_Q \right] dXdY \end{aligned} \quad (5.62)$$

where $r^2 = (x - X)^2 + (y - Y)^2 + R^2$, R is the distance of the observation plane xy to the diffraction grating plane XY (in our quantum well infrared photodetector, this is the thickness of the top n^+ -GaAs contact). Here the surface S is the XY plane plus a closing surface at infinity. r_Q is the distance from the source Q to the point (X, Y) . θ and θ_Q are the angles between the z axis and vectors \mathbf{r} and \mathbf{r}_Q .

When the light source is far away from the diffraction plane XY (the illuminating light is in the form of plane wave propagating along the z axis),

$$r_Q = \infty \gg k^{-1}, \quad \theta_Q = 0$$

so that Eq. (5.62) becomes

$$\phi(x, y, R) = \beta(k) \iint q(X, Y) \frac{e^{-i2\pi kr}}{r} \left[\left(\frac{1}{r} + i2\pi k \right) \cos \theta - i2\pi k \right] dXdY \quad (5.63)$$

where $\beta(k)$ is the normalization factor depending on the photon spectrum of the radiation source. If the intensity of the radiation is independent of its wave vector,

$$\int \int |\phi|^2 dx dy = \text{constant} \quad (5.64)$$

The intensity distribution of the optical field is

$$I(x, y, R) = |\phi(x, y, R)|^2 \quad (5.65)$$

Let us make simple analyses about Eq. (5.63). It is easy to observe that if $q(X, Y)$ is constant in the rather extended area, $\phi(x, y, R)$ is also constant, independent of (x, y) ,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{e^{-i2\pi kr}}{r} \left[\left(\frac{1}{r} + i2\pi k \right) \cos \theta - i2\pi k \right] dXdY = \text{constant} \quad (5.66)$$

In other words, when we simply cover the photodetector having a large surface area with a uniform refractive metal film, mirror reflection of the source radiation is obtained.

In general the extension of $q(X, Y)$ is limited. In a focal plane array system, $q(X, Y)$ is determined by the refractive grating structure on one optical pixel. In other words, $q(X, Y) = 0$ outside the optical pixel when considering only one single optical pixel. When the radiation wavelength is very small (large k , or high photon energy) as compared with the geometric dimension of the optical pixel and the distance between the observation plane and the diffraction grating plane (R),

$$|\phi(x, y, R)|^2 = |q(X, Y)|^2 \quad (5.67)$$

apart from a constant. The above result is due to the rapid oscillation factor $e^{-i2\pi kr}$ so that only the radiation from the source at $(X, Y) = (x, y)$ is observed at (x, y) . We thus obtain a mirror reflection from the reflective grating structure.

Now we apply Eq. (5.63) for the infrared radiation of $1/k = 10 \mu\text{m}$ to calculate the optical field diffracted from a reflective grating structure of $40 \times 40 \mu\text{m}^2$ ($X, Y \in -20, 20 \mu\text{m}$), corresponding to one optical pixel in the focal plane array. The grating consists of 7×7 apertures arranged in the form of Fig. 5.5. $a = 2$, $b = 8$, $h = 1.5$ and $R = 2 \mu\text{m}$ (R is the thickness of the top n^+ GaAs contact layer which separates the grating from the absorbing GaAs quantum wells). The calculation results ($x, y \in 0, 50 \mu\text{m}$) of circular and square apertures are presented in Fig. 5.6.

Periodicity is observed in the optical field ϕ in the xy plane. The period is determined by the wavelength of the incident radiation. The optical field along the y axis is presented for different radiation wavelength in Fig. 5.7. We observe the periodicity of $1/k$. Such a periodicity is understood by considering one-dimensional grating structure so that the oscillation factor becomes

$$e^{-i2\pi kr} = e^{-i2\pi kx} e^{i2\pi kX}$$

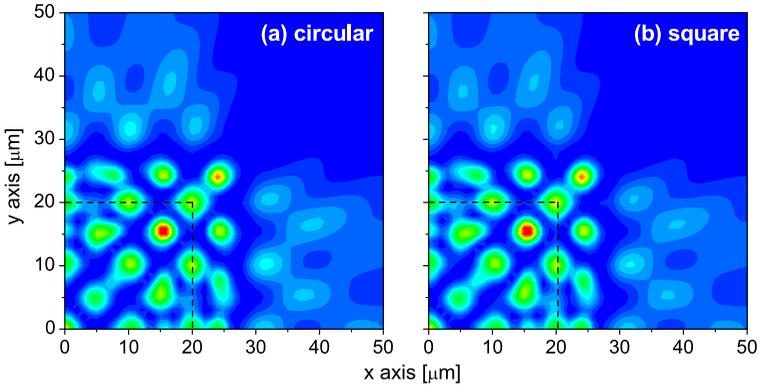
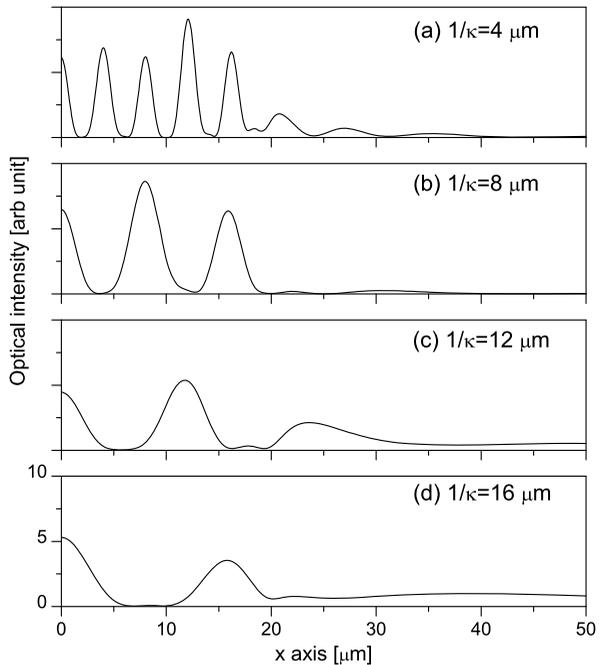


Fig. 5.6 Intensity distributions of the optical field for $1/k = 10 \mu\text{m}$ with (a) circular diffraction apertures; (b) square diffraction apertures. *Dashed lines* mark the edge of the optical pixel (Reprinted with permission from Y. Fu, M. Willander, W. Lu, and Wenlan Xu, Optical coupling in quantum well infrared photodetector by diffraction grating, *J. Appl. Phys.*, vol. 84, pp. 5750–5755, 1998. Copyright 1998, American Institute of Physics)

Fig. 5.7 Optical intensity distributions along the y axis of the optical fields for different radiation wave vectors (Reprinted with permission from Y. Fu, M. Willander, W. Lu, and Wenlan Xu, Optical coupling in quantum well infrared photodetector by diffraction grating, *J. Appl. Phys.*, vol. 84, pp. 5750–5755, 1998. Copyright 1998, American Institute of Physics)



Fringing fields, see Fig. 5.6 when $x, y > 20 \mu\text{m}$ and Fig. 5.7 for $x > 20 \mu\text{m}$, on the other hand, are very strong due to the long wavelength of the incident radiation which causes the crosstalk between adjacent optical pixels on a focal plane array.

The problem can be solved by a new design of the optical grating, see more in the following section.

The solution of Eq. (5.59) can be expressed as

$$\phi(x, y, z) = e^{i2\pi kz} + \sum_{\mathbf{q}} t_{\mathbf{q}} e^{i2\pi(\mathbf{q}\cdot\rho - q_z z)} \quad (5.68)$$

where $e^{i2\pi kz}$ is the incident light in the form of plane wave along the z direction,

$$t_{\mathbf{q}} = \iint \phi(x, y, z) e^{-i2\pi\mathbf{q}\cdot\rho} dx dy \quad (5.69)$$

is the amplitude of the reflected wave diffracted from the grating surface which propagates along $-z$ direction with wave vector $\mathbf{q} = (q_x, q_y)$ in the xy plane,

$$q_x^2 + q_y^2 + q_z^2 = k^2 \quad (5.70)$$

when considering elastic diffraction.

The design of the diffraction grating is to optimize the optical intensities $t_{\mathbf{q}}$ with small q_z . The larger can be these optical intensities, the larger will be A_z , the component of the optical field along the z direction, and thus a larger optical coupling between the incident photons and active electrons in the photodetector.

By considering Fig. 5.7 and the responding wave vector of the quantum well infrared photodetector ($1/k = 10 \mu\text{m}$), we can conclude that for $R \approx 1/k$,

$$t_{q_z} \approx \begin{cases} 1 & \text{if } q_z = 0 \\ 0 & \text{otherwise} \end{cases} \quad (5.71)$$

so that $A_z \approx A/2$, where A is the amplitude of the vector potential of the incident infrared radiation and A_z is the z component of the vector potential of the radiation diffracted from the grating diaphragm.

We have thus learned that optical grating formed by a metal film on the top contact layer of the photodetector is based on the reflection and diffraction of the light by the metal due to the large dielectric constant of the metal. We simply assumed thus far a huge dielectric constant for the metal grating structure so that light is perfectly reflected from the metal. Most fundamentally, the large dielectric constant of metal materials is the result of free electrons in metals. Here let us exploit the motions of free electrons in metal in an electromagnetic field, and vice versa, i.e., the evolution of the electromagnetic field in the presence of free electrons, which has been studying and developing into an exciting field of plasmonics. We will then use the concept of plasmonics to design optical gratings.

In the Maxwell's microscopic equations

$$\begin{aligned} \nabla \cdot \mathbf{E} &= \frac{n_{\text{tot}}}{\epsilon_0} \\ \nabla \cdot \mathbf{B} &= 0 \end{aligned} \quad (5.72)$$

$$\begin{aligned}\nabla \times \mathbf{B} &= \mu_0 \mathbf{j}_{\text{tot}} + \epsilon_0 \mu_0 \frac{\partial \mathbf{E}}{\partial t} \\ \nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t}\end{aligned}$$

n_{tot} represents the total, possibly both time- and space-dependent, electric charges, i.e., free (n_{free}) as well as induced (n_{ind}) (polarization) charges, and \mathbf{j}_{tot} represents the total, possibly also both time- and space-dependent, electric current, i.e., conduction currents (motion of free electrons) as well as all atomistic (polarization, magnetization) currents.

To include situations involving matter, it is useful to distinguish external fields and fields produced by the response of the material. By denoting \mathbf{P} as the local electric polarization of matter and \mathbf{M} the local magnetization, the electric displacement field \mathbf{D} and magnetic field strength \mathbf{H} are given by

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}, \quad \mathbf{H} = \frac{1}{\mu_0} \mathbf{B} - \mathbf{M} \quad (5.73)$$

where \mathbf{E} and \mathbf{B} are the electric field and magnetic flux density, respectively. μ_0 and ϵ_0 are the permeability and permittivity of free space, respectively. The sources of \mathbf{D} and \mathbf{H} are n_{free} and \mathbf{j}_{free} , respectively. The Maxwell's equations now read

$$\begin{aligned}\nabla \cdot \mathbf{D} &= n_{\text{free}} \\ \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{H} &= \mathbf{j}_{\text{free}} + \frac{\partial \mathbf{D}}{\partial t} \\ \nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t}\end{aligned} \quad (5.74)$$

In order to solve real problems, relations describing the responses of the materials, the so-called constitutive relations, are needed

$$\mathbf{j} = \sigma \mathbf{E}, \quad \mathbf{B} = \mu \mu_0 \mathbf{H}, \quad \mathbf{P} = \epsilon_0 \chi \mathbf{E} \quad (5.75)$$

with phenomenological coefficients σ , μ , and χ (conductivity, permeability and electric susceptibility, respectively). They depend on the medium under consideration, but are normally assumed to be independent of the electromagnetic field strength, in other words, the medium is linear. The coefficients can also be frequency dependent. The time dependence of the involved fields can be assumed to be harmonic, e.g., $\mathbf{E}(\mathbf{r}, t) = \mathbf{E}(\mathbf{r})e^{-i\omega t}$. More complicated time-dependent fields can be written as superposition of such fields (which is normally referred to as Fourier synthesis). Introducing all these equations into the Maxwell's equations,

$$\begin{aligned}\nabla \cdot (\epsilon_0 \epsilon \mathbf{E}) &= n_{\text{free}} \\ \nabla \cdot \mathbf{B} &= 0\end{aligned} \quad (5.76)$$

$$\nabla \times \mathbf{H} = -i\omega \left(\epsilon + \frac{i\sigma}{\epsilon_0\omega} \right) \epsilon_0 \mathbf{E}$$

$$\nabla \times \mathbf{E} = i\omega\mu\mu_0 \mathbf{H}$$

with the dielectric function $\epsilon = 1 + \chi$.

We now study free electrons in a metal. Compared with semiconductors and insulators, the extraordinary optical properties of metals are due to the existence of a gas of free electrons n_{free} . In the Drude-Sommerfeld model which is naturally simplified but still rather effective, these electrons are approximated as independent and point-like. In between independent collisions with various collision centers (such as lattice ions, other electrons, defects, phonons, etc.), which occur with an average rate of $1/\tau$, these electrons move freely, or, are accelerated under the influence of an external electromagnetic field, resulting in a drift motion. Each collision leads to a complete loss of directional information and results in a random orientation of the electron velocity afterwards. In reality, only electrons near the Fermi level contribute because the Pauli exclusion principle does not allow deeper lying electrons to change their electronic state. Band-structure corrections lead to a modification of this motion. These corrections are customarily incorporated into an effective mass m^* , which is in general different from the free-electron mass m_0 .

Many (but not all) properties of real metals, including their optical properties as described by the frequency-dependent dielectric function $\epsilon(\omega)$ (see below), are surprisingly well predicted from this simple model.

If n_{free} electrons per unit volume all move with velocity \mathbf{v} , the current density is

$$\mathbf{j}_{\text{free}} = -n_{\text{free}}e\mathbf{v} \quad (5.77)$$

Assume τ as the relaxation time (or the average time between successive collisions), the average velocity acquired in the presence of a dc field \mathbf{E} is $-e\mathbf{E}\tau/m^*$ so that

$$\mathbf{j}_{\text{free}} = \frac{n_{\text{free}}e^2\tau}{m^*} \mathbf{E} \quad (5.78)$$

where m^* is the effective mass of the free electron in the metal.

Treating the electron as a semiclassical particle and letting $\mathbf{p} = m^*\mathbf{v}$ be the momentum of the electron, the equation of motion is

$$\frac{d\mathbf{p}}{dt} = -\frac{\mathbf{p}}{\tau} + \mathbf{F} \quad (5.79)$$

where $\mathbf{F} = -e(\mathbf{E} + \mathbf{v} \times \mathbf{B})$ is the Lorentz force the electron feels in the electromagnetic field.

To calculate the current in a metal driven by a time-dependent electric field, we write the field in the form

$$\mathbf{E}(t) = \Re \{ \mathbf{E}(\omega) e^{-i\omega t} \}$$

where ω is the angular frequency, then seek a steady-state solution of the form

$$\mathbf{p}(t) = \Re\{\mathbf{p}(\omega)e^{-i\omega t}\}$$

for the equation of motion, Eq. (5.79). We find that $\mathbf{p}(\omega)$ must satisfy

$$-i\omega\mathbf{p}(\omega) = -\frac{\mathbf{p}(\omega)}{\tau} - e\mathbf{E}(\omega) \quad (5.80)$$

i.e.,

$$\mathbf{p}(\omega) = -\frac{e\tau}{1 - i\omega\tau}\mathbf{E}(\omega)$$

so that the current density is $\mathbf{j}(t) = \Re\{\mathbf{j}(\omega)e^{-i\omega t}\}$,

$$\mathbf{j}_{\text{free}}(\omega) = -\frac{n_{\text{free}}e\mathbf{p}(\omega)}{m^*} = \frac{n_{\text{free}}e^2\tau/m^*}{1 - i\omega\tau}\mathbf{E}(\omega) \quad (5.81)$$

One customarily writes this result as

$$\mathbf{j}_{\text{free}}(\omega) = \sigma(\omega)\mathbf{E}(\omega) \quad (5.82)$$

where $\sigma(\omega)$, known as the frequency-dependent conductivity, is given by

$$\sigma(\omega) = \frac{\sigma_0}{1 - i\omega\tau}, \quad \sigma_0 = \frac{n_{\text{free}}e^2\tau}{m^*} \quad (5.83)$$

Assume now that these electrons are subject to a time harmonic electric field (optical wave) of the form $\mathbf{E}(t) = E_0e^{-i\omega t}\mathbf{y}_0$, where \mathbf{y}_0 is the unit vector along the y axis. The equation of motion of the electron is

$$\frac{d^2y}{dt^2} + \frac{1}{\tau}\frac{dy}{dt} = -\frac{e}{m^*}E_0e^{-i\omega t} \quad (5.84)$$

by Eq. (5.79), where $p_y = m^*dy/dt$, m^* is the mass of the electron and e is the elementary charge. The above equation has the solution $y(t) = y_0e^{-i\omega t}$, and upon substitution we find the position of the electron at time t to be

$$\mathbf{y}(t) = \frac{e\mathbf{E}(t)}{m^*\omega(\omega + i/\tau)} \quad (5.85)$$

The polarization \mathbf{P} is defined as the induced dipole moment ($-e\mathbf{y}$ for one electron at position \mathbf{y}) per unit volume

$$\mathbf{P} = -n_{\text{free}}e\mathbf{y} = -\frac{n_{\text{free}}e^2}{m^*\omega(\omega + i/\tau)}\mathbf{E} \quad (5.86)$$

As Eqs. (5.75), we also define the polarization for linear dielectrics in terms of the permittivity of free space ϵ_0 , and the electrical susceptibility χ as

$$\mathbf{P} = \epsilon_0\chi\mathbf{E} = \epsilon_0[\epsilon(\omega) - 1]\mathbf{E} \quad (5.87)$$

$\epsilon(\omega)$, the relative permittivity, or dielectric constant of the material under investigation, is found to be

$$\epsilon(\omega) = 1 - \frac{\omega_p^2}{\omega(\omega + i/\tau)} = 1 - \frac{\omega_p^2}{\omega^2 + 1/\tau^2} + i \frac{\omega_p^2 \tau}{\omega(1 + \omega^2 \tau^2)} \quad (5.88)$$

where $\omega_p = \sqrt{n_{\text{free}} e^2 / m^* \epsilon_0}$ is known to be the plasma frequency. This is commonly called the Drude model for the dielectric constant of a metal, since it is based on the Drude theory of electrical conductivity.

If the positive ion core background has a dielectric constant labeled ϵ_∞ which is essentially constant up to frequencies well above ω_p for many common metal materials, then Eq. (5.88) becomes

$$\epsilon(\omega) = \epsilon_\infty \left[1 - \frac{\tilde{\omega}_p^2}{\omega(\omega + i/\tau)} \right] \quad (5.89)$$

where $\tilde{\omega}_p = \omega_p / \epsilon_\infty$.

Often it is useful to work backwards from Eq. (5.88) and find the plasma frequency and relaxation time in terms of the dielectric constant $\epsilon = \epsilon' + i\epsilon''$, resulting in

$$\begin{aligned} \tau &= \frac{1 - \epsilon'}{\omega \epsilon''} \\ \omega_p &= \sqrt{(1 - \epsilon') \left(\omega^2 + \frac{1}{\tau^2} \right)} \end{aligned} \quad (5.90)$$

Note that a heavily doped semiconductor will optically have the same behavior as a metal and a similar derivation can be done for its dielectric constant. It turns out to be similar to that of a metal, with the only alteration being the replacement of the '1' in the first real term in Eq. (5.88) by the relative permittivity of the undoped semiconductor. The reader can refer to Sect. 3.7 about the derivation of the effective permittivity of quantum-dot exciton polariton.

Now back to the crosstalk issue in the optical grating. Figure 5.8 shows the basic fabrication process and final functioning of a quantum-well infrared photodetector focal plan array. In common photodetection array structure, an epoxy underfill is used between pixels for mechanic support. The refractive index of epoxy is about 1.5 (which is much smaller than that of GaAs) and the infrared absorption by the epoxy is characterized by a few narrow peaks. SiO₂ or Si₃N₄ thin films are also used to cover on the pixels as passivation materials whose optical absorptions are very small [16]. Thus we can neglect the optical effect of the underfill material in the quantum-well infrared photodetector.

Knowing the geometric structure thereafter the spatial distribution of the dielectric constant, the spatial and temporal development of an incident electromag-

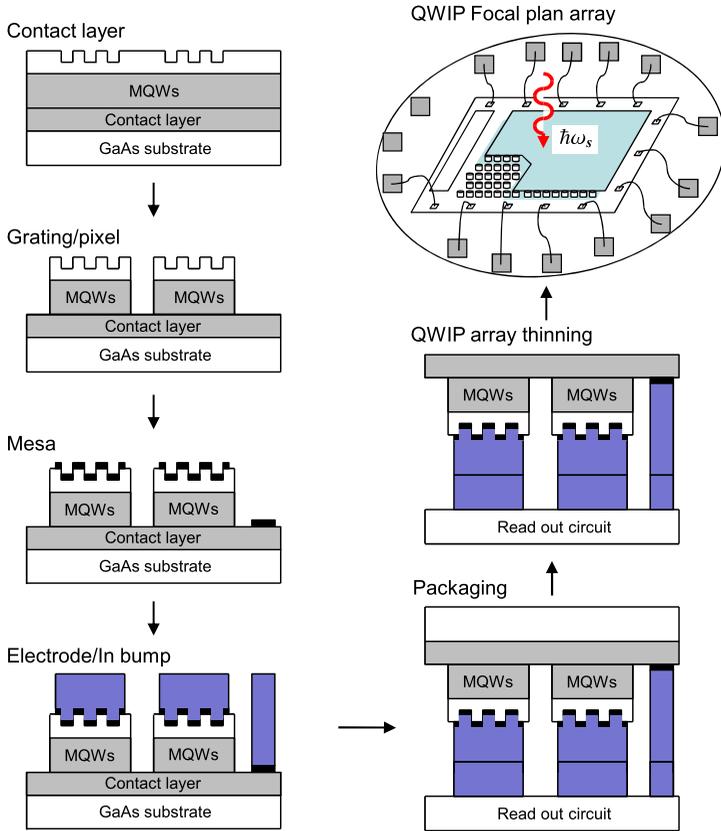


Fig. 5.8 Basic fabrication process and functioning quantum well infrared photodetector (QWIP) focal plan array

netic field can be calculated by the so-called finite-difference time-domain (FDTD) method, a direct time-domain solution of Maxwell differential equations on spatial grids or lattices [17, 18]. FDTD is beyond the scope of this book, while Refs. [17, 18] are recommended for introduction and numerical operation about FDTD. In the following we present the numerical results of electric fields in optical gratings.

We have applied Huygen’s principle and found out that in the conventional reflective diffraction gratings lithographically patterned on the tops of photodetectors, i.e., Figs. 5.5 and 5.8, the fringing electromagnetic field outside of the optical pixel is very strong, as shown in Fig. 5.6. Moreover, Fig. 5.9 shows strong diffraction from the corners. The part that diffracts into the space between pixels propagates much faster in the underfill than in the semiconductor material, resulting in a direct crosstalk. The crosstalk caused by the fringing field in Fig. 5.6 can be reduced by

Fig. 5.9 Corner diffraction and crosstalk in the grating structure. (a) Schematics, (b) spatial distribution of the E_z amplitude excited by a radiation pulse (Reprinted with permission from Z.-H. Chen, S. Hellström, Z.-Y. Yu, M. Qiu, and Y. Fu, Time-resolved photocurrents in quantum well/dot infrared photodetectors with different optical coupling structures, Appl. Phys. Lett., vol. 100, p. 043502(5), 2012. Copyright 2012, American Institute of Physics)

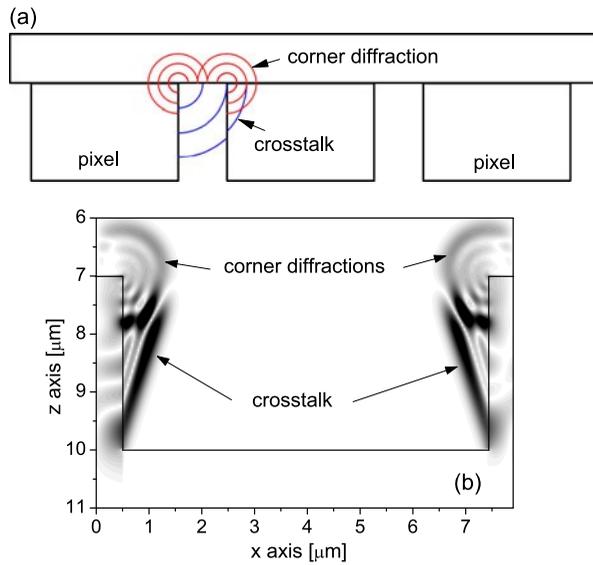
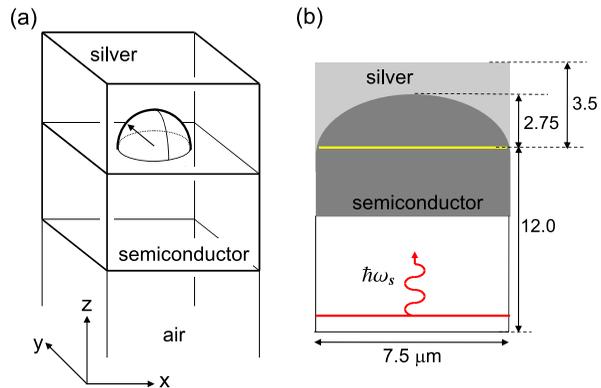


Fig. 5.10 Concave reflector structure. The length unit is μm



putting quantum wells close to the metal grating surface, while the corner diffraction induced crosstalk is rather difficult to be reduced in common pixel structured quantum well infrared photodetector. Furthermore, the wavelength of the optimal diffraction strongly depends on the grating geometry.

Figure 5.10 shows a concave metallic reflector that can effectively diffract and focus the electric field components to deep absorbing layers without causing inter-pixel crosstalk [19].

Figure 5.11 shows the distributions of E_x and E_z in the concave pixel structure at three different times. A strong focusing effect is observed in E_x , which also results in a very strong and similarly focused E_z component. Similar focusing effect and strong E_z are also obtained when continuous wave radiations are used

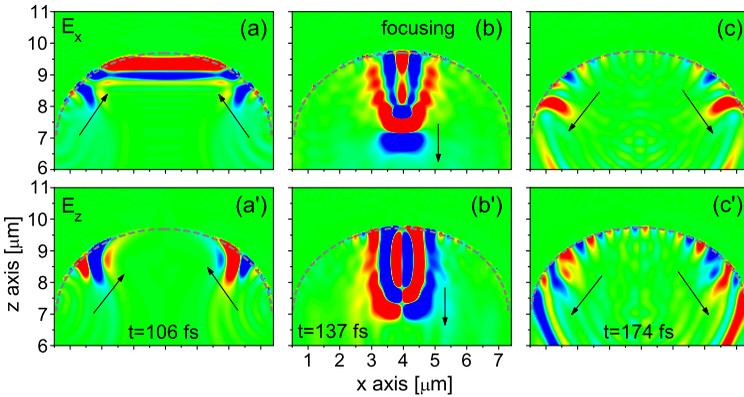


Fig. 5.11 Spatial distributions of (a, b, c) E_x and (a', b', c') E_z in the concave pixel structure at three different times excited by a radiation pulse (wavelength $2 \sim 10 \mu\text{m}$). Arrows mark the motion of the electric field (Reprinted with permission from Z.-H. Chen, S. Hellström, Z.-Y. Yu, M. Qiu, and Y. Fu, Time-resolved photocurrents in quantum well/dot infrared photodetectors with different optical coupling structures, *Appl. Phys. Lett.*, vol. 100, p. 043502(5), 2012. Copyright 2012, American Institute of Physics)

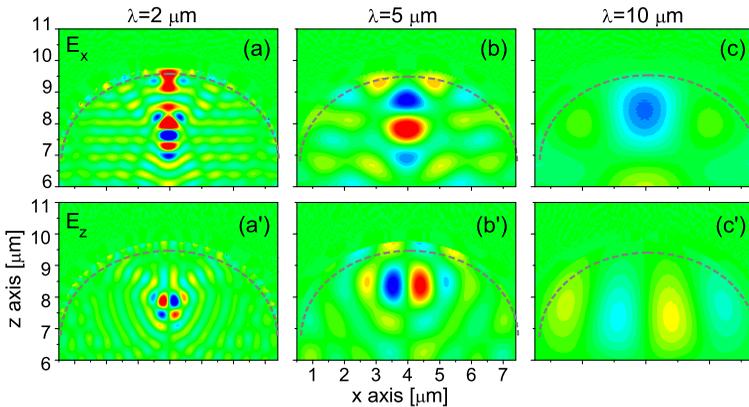
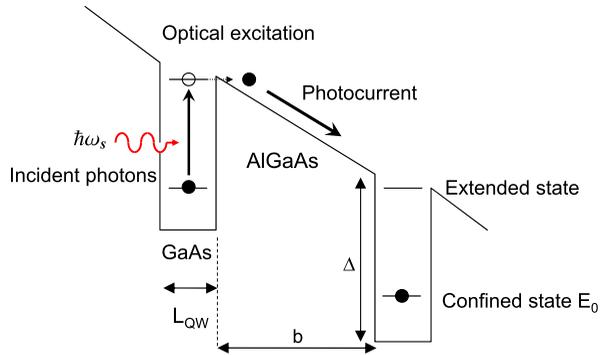


Fig. 5.12 Spatial distributions of (a, b, c) E_x and (a', b', c') E_z in the concave pixel structure excited by three continuous wave radiations at wavelengths 2, 5, and $10 \mu\text{m}$

($\lambda = 2 \sim 10 \mu\text{m}$ in the same concave structure), see Fig. 5.12. By the concave pixel structure, we not only eliminate the crosstalk but also generate a huge diffracted and focused electric field which significantly increases the photocurrents in broad-band photodetectors.

Fig. 5.13 Basic structure and function of a GaAs/AlGaAs quantum well infrared photodetector. L_{QW} is the GaAs quantum well width, normally about 5 nm, b the AlGaAs barrier width (50 nm), Δ the barrier height



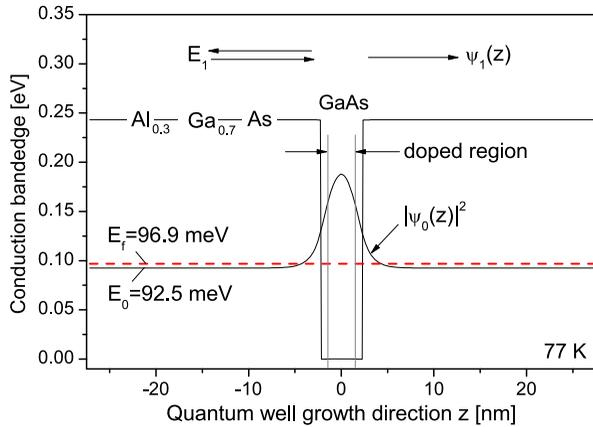
5.4 Nanostructure Infrared Photodetector

5.4.1 Quantum Well Infrared Photodetector

After the determination of the z component of the electric field of the incident radiation from the previous section, we can now calculate the optical absorption coefficient for our quantum well infrared photodetector. Figure 5.13 shows the key structure and function of a GaAs/AlGaAs multiple-quantum-well based infrared photodetector. GaAs quantum wells embedded in AlGaAs barriers are sandwiched between two n -type doped GaAs contacts, see Fig. 5.5. The GaAs quantum well width L_{QW} and the AlGaAs barrier height Δ are designed in such a way that there is only one confined state in the GaAs quantum well. The reason of one confined state is to obtain a narrow response wavelength range. As we will see a bit later that the minimal photon energy required to photoexcite an electron confined in the GaAs quantum well is the distance from the confined state to the bandedge of the AlGaAs conduction band. And in theory there will be no up limit for the photon energy but numerical calculations show a decreased optical transition matrix between the confined state and the extended state when the energy distance between the two states increase too much. The GaAs quantum wells are n -type doped to such a level that only the confined state is occupied. An incident photon $\hbar\omega_s$ will be absorbed by one electron occupying the confined state which will transit to an extended state above the bandedge of the AlGaAs conduction band. AlGaAs barrier regions are not doped for two reasons. The first one is that we need an electric field to drive the photogenerated electrons in the extended states to the electrodes. According to Ohm's law, electric fields will be strong in the areas with low charge concentrations. Secondly, the photocurrent will experience weak scatterings due to impurities when the AlGaAs region is as pure as possible. Moreover, in practical devices, only the central part of the GaAs quantum well is doped (δ doping) in order to restrict the spatial distribution of the dopants.

Normal GaAs/AlGaAs-based quantum well infrared photodetector consists of 50 GaAs quantum wells to increase the possibility of absorbing incident photons that the photons will have further chances to interact with electrons in the second

Fig. 5.14 Conduction bandedge [eV] and electron wave function distribution $|\psi_0(z)|^2$ in one period of the GaAs/AlGaAs quantum well infrared photodetector at 77 K



GaAs quantum well if they miss to interact with electrons in the first GaAs quantum well. Too many quantum wells are however difficult to grow uniformly due to eventual accumulations of various defects at GaAs/AlGaAs heterointerfaces. We consider a common GaAs/AlGaAs quantum well infrared photodetector which consists of $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ barrier with thickness 50 nm, GaAs quantum well thickness $L_{\text{QW}} = 4.5$ nm whose the central 2.5 nm thick layer is doped at 10^{18} cm^{-3} . The energy band structure of one well-barrier period, at zero bias and 77 K, is shown in Fig. 5.14, which is self-consistently calculated from Poisson and Schrödinger equations. There is one confined state $E_0 = 92.5$ meV in the quantum well and the Fermi level is $E_f = 96.9$ meV at 77 K, where the conduction bandedge of GaAs is taken to be the energy reference. The conduction bandedge of the $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ barrier is 243.2 meV so that the response wavelength of the photodetector is $8.2 \mu\text{m}$ ($\hbar\omega_s = 243.2 - 92.5 \text{ meV} = 150.7 \text{ meV}$).

We can extract the electron concentration from Fig. 5.14 in the following way. By the numerical scale of the vertical axis, the peak value of the electron envelope wave function $|\psi_0(z)|^2$ is 0.188, while the zero is positioned at 0.093 (E_0 in eV) so that the net peak height is $0.188 - 0.093 = 0.095$, which corresponds to a peak electron concentration of $0.47 \times 10^{18} \text{ cm}^{-3}$. Comparing with the doping distribution (doped region marked by two horizontal arrows in Fig. 5.14), we see that the electron distribution is a bit broad when including the penetrations of electrons from the GaAs quantum well to the surrounding $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ barriers. But the charges, both the dopants and the electrons, are largely confined in the GaAs quantum well, resulting in an almost zero electric field in the GaAs quantum well. In Fig. 5.14, the electric field in $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ barrier regions is also zero since we have set the external bias zero. A straightforward deduction is that the external bias will apply mainly in the $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ barrier regions since the charges are localized in the GaAs quantum well regions, as schematically shown in Fig. 5.13.

Since the electric field in the GaAs quantum well is very small, the quantum well can be approximated as square. Let the GaAs quantum well width be L_{QW} , the AlGaAs barrier width b , the barrier height Δ , and the growth direction as the z axis, see

Fig. 5.13. Since the Hamiltonian of the conduction-band electron is translationally symmetric in the xy plane, the total wave function can be expressed as

$$\Psi_m(\mathbf{k}, \mathbf{r}) = \psi_m(z)u(\mathbf{r})e^{i\mathbf{k}\cdot\boldsymbol{\rho}} \quad (5.91)$$

where the envelope wave function $\psi_m(z)$ is determined by

$$\left[\frac{-\hbar^2}{2m^*} \frac{\partial^2}{\partial z^2} + V(z) \right] \psi_m(z) = E_m \psi_m(z) \quad (5.92)$$

$$V(z) = \begin{cases} \Delta & -\frac{L_{\text{QW}}}{2} - b \leq z < -\frac{L_{\text{QW}}}{2} \\ 0 & -\frac{L_{\text{QW}}}{2} \leq z < \frac{L_{\text{QW}}}{2} \\ \Delta & \frac{L_{\text{QW}}}{2} \leq z < \frac{L_{\text{QW}}}{2} + b \end{cases} \quad (5.93)$$

which repeats periodically along the z axis

$$V[z + \ell(b + L_{\text{QW}})] = V(z) \quad (5.94)$$

where ℓ is an integer.

The total energy of the electron is

$$E_m(\mathbf{k}) = E_m + E_k, \quad E_k = \frac{\hbar^2(k_x^2 + k_y^2)}{2m^*} \quad (5.95)$$

where $\mathbf{k} = (k_x, k_y)$ and $\boldsymbol{\rho}$ are the wave vector and coordinate in the xy -plane.

The boundary conditions for the ground state $\psi_0(z)$ are unambiguous that the corresponding wave function is zero deep inside the AlGaAs barriers

$$\psi_0(z) = \begin{cases} A_0 e^{k_{0b}z} & z < -\frac{L_{\text{QW}}}{2} \\ B_0 \cos(k_{0w}z) & -\frac{L_{\text{QW}}}{2} \leq z \leq \frac{L_{\text{QW}}}{2} \\ A_0 e^{-k_{0b}z} & z > \frac{L_{\text{QW}}}{2} \end{cases} \quad (5.96)$$

where

$$\begin{aligned} \frac{\hbar^2 k_{0w}^2}{2m^*} &= E_0, & \frac{\hbar^2 k_{0b}^2}{2m^*} &= \Delta - E_0 \\ B_0 &= \sqrt{2} \left[L_{\text{QW}} + \frac{1}{k_{0b}} + \frac{\sin(k_{0w}L_{\text{QW}})}{k_{0w}} + \frac{\cos(k_{0w}L_{\text{QW}})}{k_{0b}} \right]^{-1/2} \\ A_0 &= B_0 \cos\left(\frac{k_{0w}L_{\text{QW}}}{2}\right) \exp\left(\frac{k_{0b}L_{\text{QW}}}{2}\right) \end{aligned} \quad (5.97)$$

The approximation we made for $\psi_0(z)$ is that since the barrier width b is rather wide compared with the well width L_{QW} , $\psi_0(z)$ in one GaAs quantum well does

not penetrate through the AlGaAs barriers to reach the neighboring GaAs quantum wells. The reader can refer to the numerical result in Fig. 5.14.

The determination of the extended states whose energies are higher than Δ is not straightforward. First of all, the conduction bandedge of the AlGaAs barrier will not be flat when an external bias applies. To avoid complications we assume that the external bias is zero, we then calculate the photocurrent from the left to the right as an estimation about the photocurrent under an external bias in a real device. The next question about the extended state is its boundary conditions, i.e., where comes the electron and where goes the electron. There are three theoretical sets of boundary conditions, while the reality can be a mixture of them.

Since the barrier width b is rather wide compared with the well width L_{QW} , we can approximate each GaAs quantum well as embedded in infinitely extended AlGaAs barriers so that

$$\psi_{\text{box}} = \begin{cases} A_1 \sin(k_{1b}z - \delta) & z < -\frac{L_{\text{QW}}}{2} \\ B_1 \sin(k_{1w}z) & -\frac{L_{\text{QW}}}{2} \leq z \leq \frac{L_{\text{QW}}}{2} \\ A_1 \sin(k_{1b}z + \delta) & z > \frac{L_{\text{QW}}}{2} \end{cases} \quad (5.98)$$

Here

$$\begin{aligned} \frac{\hbar^2 k_{1w}^2}{2m^*} &= E_1, & \frac{\hbar^2 k_{1b}^2}{2m^*} &= E_1 - \Delta \\ A_1 &= \sqrt{\frac{1}{L_{\text{QW}} + b}} \\ B_1 &= A_1 \left[\sin^2\left(\frac{k_{1w}L_{\text{QW}}}{2}\right) + \left(\frac{k_{1w}}{k_{1b}}\right)^2 \cos^2\left(\frac{k_{1w}L_{\text{QW}}}{2}\right) \right]^{-1/2} \\ \delta &= \sin^{-1} \left[\frac{B_1}{A_1} \sin\left(\frac{k_{1w}L_{\text{QW}}}{2}\right) \right] - \frac{k_{1b}L_{\text{QW}}}{2} \end{aligned} \quad (5.99)$$

However, due to the consideration of the carrier transport in the quantum well infrared photodetector under normal device working conditions, we must envisage the conduction of the electrons in the extended states from one GaAs quantum well to the next. We first consider the running wave

$$\psi_{\text{running}}(z) = \begin{cases} e^{ik_{1b}z} + r e^{-ik_{1b}z} & z < -\frac{L_{\text{QW}}}{2} \\ A e^{ik_{1w}z} + B e^{-ik_{1w}z} & -\frac{L_{\text{QW}}}{2} \leq z < \frac{L_{\text{QW}}}{2} \\ t e^{ik_{1b}z} & z \geq \frac{L_{\text{QW}}}{2} \end{cases} \quad (5.100)$$

For the square quantum well described by Eq. (5.93), it is easy to obtain an analytical expression for the amplitude of the transmitted wave

$$t = \frac{2\alpha}{2\alpha \cos(k_{1w}L_{\text{QW}}) - i(1 + \alpha^2) \sin(k_{1w}L_{\text{QW}})} \quad (5.101)$$

by matching the wave function and its first-order derivative at interfaces of $z = \pm L_{\text{QW}}/2$. Here $\alpha = k_{1b}/k_{1w}$.

An alternative form for the conducting electrons in the extended states is the Bloch wave form

$$\psi_{\text{Bloch}}(z) = u(z)e^{iKz}, \quad u(z) = u[z + \ell(L_{\text{QW}} + b)] \quad (5.102)$$

as in the context of standard semiconductor material and device physics, where $u(z)$ accounts the periodic boundary conditions of the GaAs/AlGaAs multiple quantum wells. ℓ in the above equation is an integer.

We are now able to calculate the absorption spectrum of the photodetector after we have determined the electron states. Since the effective mass of the Γ electron is isotropic, the interaction between the incident radiation and the Γ electrons in the quantum well infrared photodetector is

$$\frac{e}{m^*} \mathbf{A} \cdot \mathbf{p} \quad (5.103)$$

for an incident light of $Ae^{i\omega_s t}$. We repeat a few equations from previous sections. The optical transition matrix element is

$$\langle \psi_1(\mathbf{k}, \mathbf{r}) | \mathbf{A} \cdot \mathbf{p} | \psi_0(\mathbf{q}, \mathbf{r}) \rangle = i A_z \hbar \delta_{\mathbf{k}, \mathbf{q}} \left\langle \psi_1(z) \left| \frac{\partial}{\partial z} \right| \psi_0(z) \right\rangle \quad (5.104)$$

Again we have observed that the momentum is conserved in the xy plane, the transition is between the envelope functions in the z direction. The transition probability from state (E_0, \mathbf{k}) to (E_1, \mathbf{k}) is

$$w(\mathbf{k}) = \left| \frac{e A_z \hbar}{m^*} \left\langle \psi_1(z) \left| \frac{\partial}{\partial z} \right| \psi_0(z) \right\rangle \right|^2 \frac{2\Gamma}{\hbar[(E_1 - E_0 - \hbar\omega)^2 + \Gamma^2]} \quad (5.105)$$

It is easy to see that $w(\mathbf{k})$ is actually \mathbf{k} -independent because of the momentum conservation and the isotropic effective mass. The probability of an electron transiting from state E_0 to E_1 is

$$W = \int w(\mathbf{k}) \{ f[E_0(\mathbf{k})] - f[E_1(\mathbf{k})] \} \frac{2d\mathbf{k}}{(2\pi)^2} \quad (5.106)$$

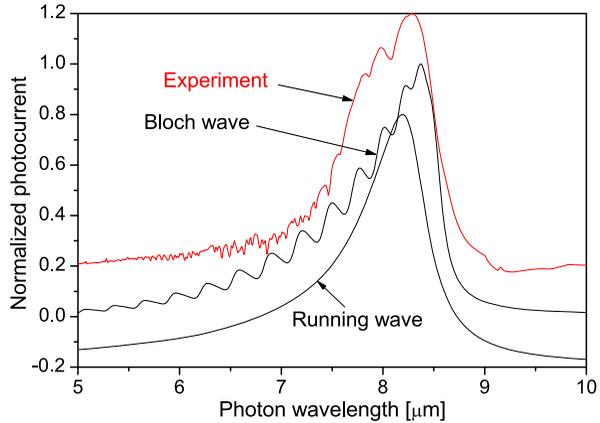
By Fig. 5.14, E_f is far below the barrier height, i.e., Δ , in our photodetector so that $f[E_1(\mathbf{q})] = 0$ and

$$n = \int f[E_0(\mathbf{k})] \frac{2d\mathbf{k}}{(2\pi)^2} \quad (5.107)$$

is the carrier density in the GaAs quantum well, the optical absorption coefficient of the photodetector is obtained as

$$\alpha = \frac{nw}{n_{\text{ph}}c}, \quad n_{\text{ph}} = \frac{2\epsilon\omega_s |\mathbf{A}|^2}{\hbar} \quad (5.108)$$

Fig. 5.15 Calculated optical absorption coefficient and a typical photocurrent spectrum of GaAs/AlGaAs quantum well infrared photodetector measured at 77 K (Reprinted from Y. Fu, Boundary conditions of continuum states in characterizing photocurrent of GaAs/AlGaAs quantum well infrared photodetector, *Superlattices Microstruct.*, vol. 30, pp. 69–74, 2001, with permission from Elsevier)



Here v is the speed of light in the photodetector.

Here we see that due to the isotropic effective mass of the active Γ electrons, normal incident photon absorption is not possible. A component of the optical electric field along the quantum well growth direction, A_z is required. For fabricating large arrays of optical photodetectors based on Γ electrons in AlGaAs/GaAs materials, diffraction gratings are needed for achieving large optical coupling, as we discussed already in the previous section.

Numerical results are presented in Fig. 5.15, where the experimental spectrum (spectrum with running-wave boundary conditions) is vertically shifted up (down). The experimental photocurrent spectrum was obtained at 77 K. Figure 5.15 suggests the Bloch-state boundary conditions as the most proper ones for continuum states in characterizing the photocurrent spectrum of the quantum well infrared photodetector consisting of multiple quantum wells. The discrepancy between the theoretical and experimental spectra at low wavelength range can be the results of the complicated conduction band structure at higher electron energies, where the simple effective mass approximation needs to be modified.

5.4.2 Quantum Wire Infrared Photodetector

We have thus far focused on the optical properties of quantum wells, especially the quantum well infrared photodetector. One major issue for the quantum well based optical devices is the quantum selection rule that electrons in the quantum well cannot interact with light that propagates along the quantum well growth direction. This constrains overwhelmingly the application of these devices such as photodetection since the plane perpendicular to the growth direction is the natural plane to receive incident radiation. We discussed before the solution of optical grating, another solution is to use quantum wires, and/or quantum dots as the active photon absorbers. Typical structure of quantum wire based infrared photodetector is

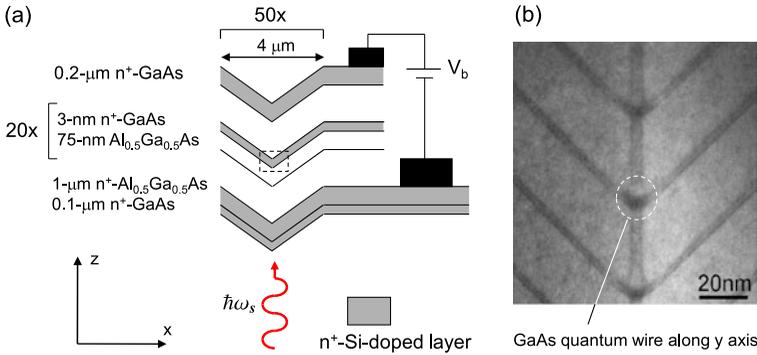


Fig. 5.16 (a) Schematic structure of the quantum wire infrared photoconductor. (b) Transmission electron microscopy (TEM) image of the quantum wire cross section

schematically shown in Fig. 5.16(a) and a transmission electron microscopy (TEM) image of stacked GaAs quantum wires which are extended along the y axis, see Fig. 5.16(b) [20], grown on V grooved GaAs substrate. The structure does not require optical grating.

5.4.3 Quantum Dot Infrared Photodetector

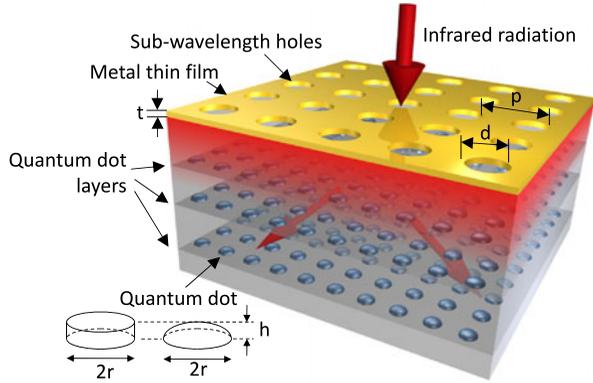
Quantum dots are three-dimensional nanostructures which can interact with photons incident from any direction. For quantum-dot-based infrared photodetector (QDIP), surface coating is still needed, not because of any quantum selection rule about light-matter interaction, but mostly due to the consideration of reducing optical reflection at the air-device surface, in other words, anti-reflection. A schematic quantum-dot infrared photodetector is shown in Fig. 5.17.

In recent years sub-wavelength hole arrays in metallic films have received considerable attention after the experimental findings of Ebbesen in 1998 which showed that the transmission of light through these structures is orders of magnitude higher than expected by the standard aperture theory [21, 22].

In this section we study the transmission of an incident electromagnetic field through a sub-wavelength hole array then calculate the photocurrent enhancement in the quantum-dot infrared photodetector underneath the grating. The grating is schematically shown by the inset in Fig. 5.18 where we consider a sub-wavelength hole array in a thin metal film, d denotes the diameter of circular holes, P the period of the hole array, and t the thickness of the metal film. t is normally in the order of 100 nm, $p = 3 \sim 4 \mu\text{m}$, and $d = 1.0 \mu\text{m}$. The photon wavelength of interest ranges between 2 to 12 μm .

Note that a major difference between the optical diffraction grating in Sect. 5.3 and here is that the optical diffraction grating is a reflection grating so that light incidents from the bottom of the quantum well infrared photodetector, see Fig. 5.5,

Fig. 5.17 Sub-wavelength hole perforated thin metal film as the optical diffraction grating for quantum-dot-based infrared photodetector. Quantum dots are here dome shaped with a circular base of radius r and a height h [3D color illustration by Staffan Hellström]



which will make the device structure more complicated. Here in Fig. 5.17, light comes vertically down into the quantum dot infrared photodetector which is integrated directly on a read-out circuit at the bottom, so that the device structure is much simplified, in addition to the large optical transmission (see below) through the sub-wavelength hole perforated thin metal film.

We first present results of optical transmission and diffraction through the grating by the three-dimensional finite-difference time-domain (FDTD) numerical simulation method. Refer to Fig. 5.18(a), a simulation volume with the size of $3 \mu\text{m}$ in x and y directions and $10 \mu\text{m}$ in the z direction is used, containing three parts, air (a) $0 < z < 5 \mu\text{m}$, a 150-nm-thick gold film (m) $5 \leq z < 5.15 \mu\text{m}$, and GaAs substrate (s) $z \geq 5.15 \mu\text{m}$. Pulsed transverse-electric-polarized (TE polarized) plane waves in a broad wavelength range from 2 to $12 \mu\text{m}$ are injected into the simulation volume along the z axis. Periodic boundary conditions are used in both x and y directions to produce the periodic array structure. The relative complex dielectric constant of Au is given by the lossy Drude model discussed in the previous section

$$\epsilon_{\text{Au}}(\omega) = \epsilon_{\infty} - \frac{(\epsilon_s - \epsilon_{\infty})\omega_p^2}{\omega(\omega + i\gamma)} \quad (5.109)$$

where $\epsilon_{\infty} = 12.18$ and $\epsilon_s = 12.75$ are the relative permittivities at infinite and zero frequency, $\omega_p = 1.545 \times 10^{16}$ Hz is the Au bulk plasma frequency, $\gamma = 9.01 \times 10^{13}$ Hz is the damping constant.

The transmission results for a few hole sizes and also for the case of a thicker metal film and the case without any metal film are shown in Fig. 5.18(b), where the plasmonic features and their mode numbers are identified by

$$\lambda_{\text{SPP}} = \frac{P}{\sqrt{i^2 + j^2}} \sqrt{\frac{\epsilon_{\text{Au}}\epsilon_d}{\epsilon_{\text{Au}} + \epsilon_d}} \quad (5.110)$$

where ϵ_{Au} are ϵ_d are dielectric constants of the Au film and dielectric (either semiconductor or air), respectively.

Figure 5.18(b) shows that the optical transmission can be improved by the photonic crystal and surface plasmon effects. However, the average transmission

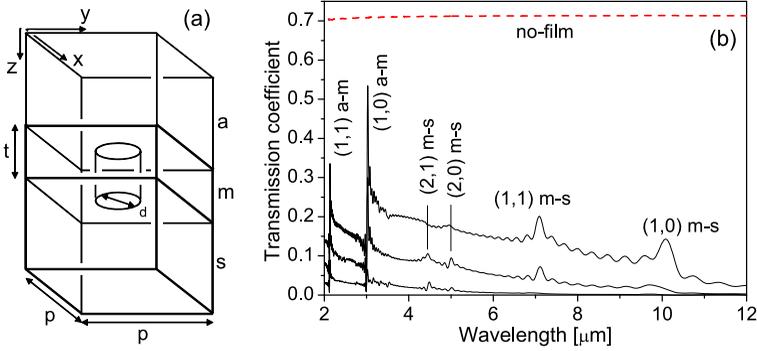


Fig. 5.18 (a) Geometric structure of the perforated Au thin film. “a” denotes air, “m” metal and “s” semiconductor. t is the film thickness and d the diameter of the hole. (b) Transmission coefficient of the optical grating. Plasmon modes (i, j) are labeled, “a-m” and “m-s” denote air-metal and metal-semiconductor interfaces. From low to high transmission curves: $d = 800, 1200, 1600$ nm. The dashed curve at about 0.7 denotes the case without any metal film (Reprinted with permission from S. Hellström, Z.-H. Chen, Y. Fu, M. Qiu, R. Soltanmoradi, Q. Wang, and J. Y. Andersson, Increased photocurrent in quantum dot infrared photodetector by subwavelength hole array in metal thin film, Appl. Phys. Lett., vol. 96, p. 231110(3), 2010. Copyright 2010, American Institute of Physics)

through the grating is merely $10 \sim 20$ % as compared with 70 % for the case of no film, whereas the measured photocurrent enhancement achieved in the experiments are often much higher so the reason for the enhancement is not due to the transmission enhancement.

In Fig. 5.19, the temporal developments of E_x and E_z of the optical field are shown using logarithmic scale, represented by the surface integral of the optical field components on planes subdividing the z axis. The reflection and transmission of the pulse are easily seen as well as the diffraction of the pulse.

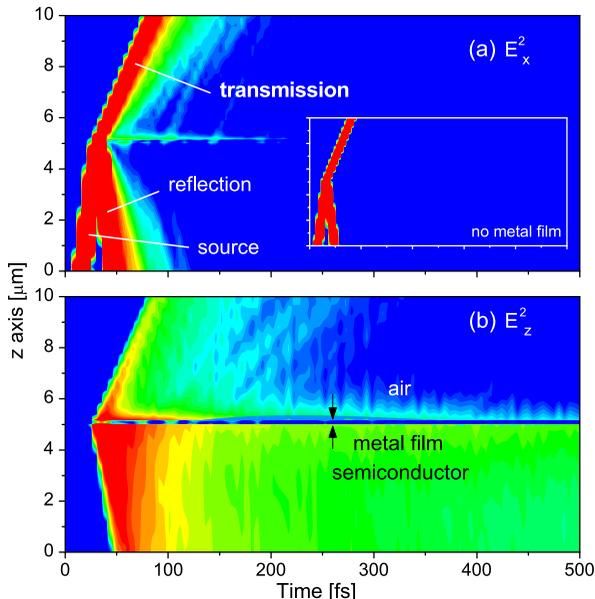
Most quantum dots grown by molecular beam epitaxy and metal-organic chemical vapor deposition methods are dome shaped with a circular base of radius r and a height h , see Fig. 5.17. We model such a quantum dot as a cylinder with a circular cross section (radius r) in the xy plane and a height h along the z axis. We consider only the low-energy levels in the conduction band for which we can approximate the confinement potential as infinitely high so that we can separately solve the Schrödinger equation along the z direction as a square quantum well with a width h and in the xy plane a cylindrical quantum well disk with radius r . A cylindrical quantum wire positioned along the z axis is described by

$$\left[-\frac{\hbar^2}{2m^*} \left(\frac{d^2}{dx^2} + \frac{d^2}{dy^2} \right) + V(x, y) \right] \psi(x, y) = E \psi(x, y) \quad (5.111)$$

where m^* is the effective mass of electron,

$$V(x, y) = \begin{cases} -\Delta & \text{when } x^2 + y^2 \leq r^2 \\ 0 & \text{when } x^2 + y^2 > r^2 \end{cases} \quad (5.112)$$

Fig. 5.19 Temporal developments and spatial distributions of (a) E_x^2 and (b) E_z^2 . Inset in (a) shows E_x^2 in the absence of a metal film in which case the E_z field is always zero (Reprinted with permission from S. Hellström, Z.-H. Chen, Y. Fu, M. Qiu, R. Soltanmoradi, Q. Wang, and J. Y. Andersson, Increased photocurrent in quantum dot infrared photodetector by subwavelength hole array in metal thin film, *Appl. Phys. Lett.*, vol. 96, p. 231110(3), 2010. Copyright 2010, American Institute of Physics)



$\Delta \geq 0$ is the confinement energy of the conduction-band electron. In the cylindrical coordinate of $(x, y) = (\rho, \phi)$ and by writing $\psi(x, y) = e^{im\phi} J_m(\rho)$,

$$\frac{d^2 J_m(\rho)}{d\rho^2} + \frac{1}{\rho} \frac{dJ_m(\rho)}{d\rho} + \left\{ \frac{2m^*[E_m - V(\rho)]}{\hbar^2} - \frac{m^2}{\rho^2} \right\} J_m(\rho) = 0 \quad (5.113)$$

where m is an integer (negative or positive so the state of $|m|$ is two-fold degenerate). Inside the quantum wire, $q_m^2 = 2m^*(E_m + \Delta)/\hbar^2$, the above equation becomes

$$\frac{d^2 J_m(\rho)}{d\rho^2} + \frac{1}{\rho} \frac{dJ_m(\rho)}{d\rho} + \left(q_m^2 - \frac{m^2}{\rho^2} \right) J_m(\rho) = 0 \quad (5.114)$$

The solutions to the above equation are called Bessel functions, see Sect. 3.6. For ground state $m = 0$,

$$J_0(\rho) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{iq_0\rho \sin \tau} d\tau \quad (5.115)$$

Including the wave function in the z direction which is confined between $z = 0$ and $z = h$, the ground-state envelope function is then

$$\psi_0(\rho, \phi, z) = \sqrt{\frac{2}{h}} \sin\left(\frac{\pi z}{h}\right) \frac{1}{\beta\sqrt{\pi r^2}} \int_{-\pi}^{\pi} e^{iq_0\rho \sin \tau} d\tau \quad (5.116)$$

where β is a normalization factor. The exact numerical value of q_0 and thereafter E_0 of the ground state are obtained when we set the wave function to zero at the

edge of the quantum dot ($\rho = r$, where $2r$ is the diameter of the base, see Fig. 5.17) which gives $q_0r = 2.405$, see Table 3.7.

The optical matrix element between ψ_0 expressed by the above equation and an extended plane wave in the conduction band

$$\psi_{\mathbf{k}_\rho k_z}(\rho, \phi, z) = \frac{1}{\sqrt{\Omega}} e^{i(k_\rho \rho + k_z z)} \quad (5.117)$$

becomes

$$\alpha = \mathbf{E} \cdot \langle \psi_0 | \nabla | \psi_{\mathbf{k}_\rho k_z} \rangle = (E_\rho k_\rho + E_z k_z) \langle \psi_0 | \psi_{\mathbf{k}_\rho k_z} \rangle \quad (5.118)$$

where E_ρ and E_z are components of the electric field \mathbf{E} in the ρ and z directions, respectively. Ω denotes the space where the plane wave extends which equals the volume (i.e., $\pi r^2 h$) of the quantum dot (wave function normalization). In writing down Eq. (5.117) for the extended state, we have completely neglected the effect of the quantum-dot potential Eq. (5.112). The reason behind such an approximation is based on the fact that the extended state is expected to extend in a spatial region much larger than the quantum dot. For $q_0 \neq 0$, and after integration over ϕ which gives 2π , the optical matrix element becomes

$$\alpha = \frac{2\sqrt{2}(E_\rho k_\rho + E_z k_z)}{\beta h r^2} \int_0^h \sin\left(\frac{\pi z}{h}\right) e^{i k_z z} dz \int_0^r \int_{-\pi}^{\pi} e^{i(k_\rho \rho + q_0 \rho \sin \tau)} d\tau \rho d\rho \quad (5.119)$$

Including the energy conservation and the density of states for free electrons in the GaAs material, the photocurrent density spectrum is

$$i(\hbar\omega) = \int |\alpha|^2 v_z \delta(\hbar\omega - E_{\mathbf{k}} - E_0) \frac{2d\mathbf{k}}{(2\pi)^3} \quad (5.120)$$

where $\mathbf{k} = (\mathbf{k}_\rho, k_z)$, $E_{\mathbf{k}} = \hbar^2 k^2 / 2m^*$, E_0 is the energy of the ground electron state $\psi_0(\rho, \phi, z)$ in the quantum dot. v_z is the speed along the z direction of the electron described by the plane wave $e^{i(k_\rho \rho + k_z z)}$, i.e., $\hbar k_z / m^*$. Here we assume the normal quantum dot photodetection configuration with an external bias applied along the z direction.

Figure 5.20 shows the photocurrent density spectra of the four grating structures ($d = 800, 1200, 1600$ nm and the no-film structure) with the quantum dot positioned at $2 \mu\text{m}$ away from the metal-semiconductor surface. We can observe that for quantum dots with large aspect ratios (large $h/2r$), extra optical grating does not help since the total radiation transmission through the grating is far less than the no-film case (see Fig. 5.18). However, for flat quantum dots with small aspect ratios (small $h/2r$), the z component E_z of the electric field is critically required for efficient light-matter interactions inside quantum dots. For them, the mid-infrared photocurrents of two grating structures exceed the one without film even though the total transmission rate is as low as only 20 % (it is about 70 % without film), as clearly demonstrated experimentally [23–26].

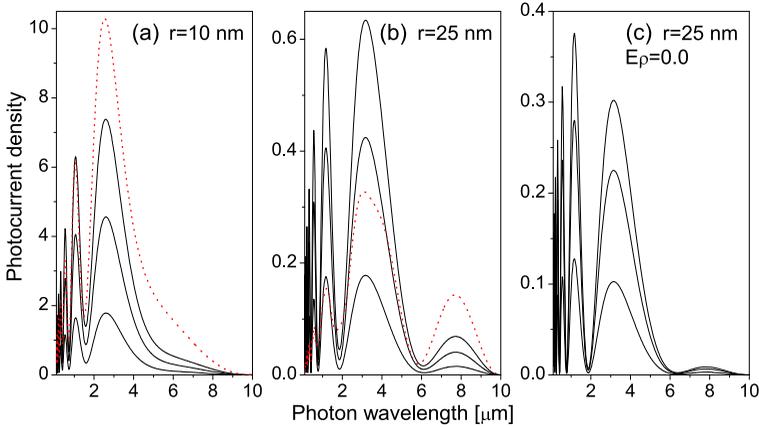


Fig. 5.20 Photocurrent density spectra of the four grating structures under investigation. *Red dotted lines* correspond to the no-film case. From low to high photocurrent spectra: $d = 800, 1200, 1600$ nm. $h = 5$ nm. **(a)** $r = 10$ nm; **(b)** $r = 25$ nm; **(c)** $r = 25$ nm with E_ρ excluded, showing that the E_z contribution is large (Reprinted with permission from S. Hellström, Z.-H. Chen, Y. Fu, M. Qiu, R. Soltanmoradi, Q. Wang, and J. Y. Andersson, Increased photocurrent in quantum dot infrared photodetector by subwavelength hole array in metal thin film, *Appl. Phys. Lett.*, vol. 96, p. 231110(3), 2010. Copyright 2010, American Institute of Physics)

We emphasize that a large contribution of the photocurrent for small aspect-ratio quantum dots comes from the coupling to E_z , and Fig. 5.20(c) shows the same structure as (b) but with the E_ρ component set to zero.

Peaks and nodes in Fig. 5.20 are due to the resonances between plane waves $e^{ik_\rho\rho}$ and the radially confined states in the xy plane.

5.5 SiGe Heterostructure Internal Emission Infrared Photodetector

We have thus far discussed two principal factors about the photodetector, namely, the optical field (to be detected) and the light-matter interaction that converts the light signal into electric signal. There are many other factors, both experimental (such as material growth and device processing) as well theoretical (energy relaxation processes, effect of external biases, etc.), have to be carefully studied in order to understand and eventually design and optimize the device operation. In this section we study an infrared photodetector based on a p -type SiGe heterostructure, as an example to demonstrate critical issues related to practical device design.

The photodetector is a SiGe/Si heterointerface composed of a heavily doped (normally higher than 10^{20} cm $^{-3}$) p^+ -type Si $_{1-x}$ Ge $_x$ layer grown on an (100) p -Si substrate. The value of Ge fraction x is usually between 0.2 and 0.4 [27, 28]. Valence band offset creates a Schottky barrier so that very few thermally excited carriers

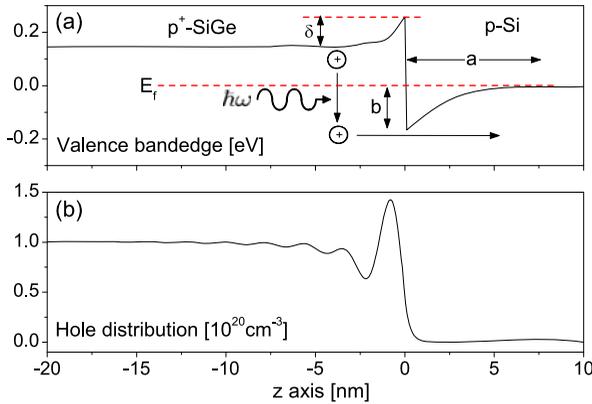


Fig. 5.21 (a) Valence band structure of the $p\text{-Si}/p^+\text{-Si}_{1-x}\text{Ge}_x$ heterostructure internal emission infrared detector. (b) Hole distribution at 300 K (Reprinted with permission from Y. Fu, S. C. Jain, M. Willander, and J. J. Loferski, Valence band structure of heavily doped strained $\text{Ge}_x\text{Si}_{1-x}$ layers, *J. Appl. Phys.*, vol. 74, pp. 402–407, 1993. Copyright 1993, American Institute of Physics)

can go over the barrier at low temperatures. However infrared photons can be efficiently absorbed by the free carriers. This causes an internal photon emission of the holes over the barrier which provides us with the photodetection mechanism. Figure 5.21(a) shows the valence bandedge of such an internal emission infrared photodetector, where b denotes the minimal barrier height for the holes in p^+ -type $\text{Si}_{1-x}\text{Ge}_x$ to transport to the Si side. Since the effective mass of valence-band holes is normally quite large (close to free electron mass m_0 , see Tables 1.4 and 1.5 and convert band parameters $\gamma_1, \gamma_2, \gamma_3, L, M$ and N to valence-band hole effective masses) so that the tunneling probability through the barrier is small. b thus represents the minimal photon energy required to photoexcite a hole from p^+ $\text{Si}_{1-x}\text{Ge}_x$ to Si. It is referred to be the cut-off wavelength of the photodetector (above which the photodetector will not response).

For the photodetector of Fig. 5.21(a), the experimental cut-off wavelength was shown to be 125 meV [27] (people use both the wavelength and/or photon energy to denote the cut-off wavelength, which are interchangeable $\lambda = 2\pi c/\omega$, λ is the wavelength and $\hbar\omega$ is the photon energy). Lin et al. [28] have calculated the cut-off wavelength of this detector by assuming that the barrier height is equal to the valence band offset ΔE_0 between $\text{Si}_{1-x}\text{Ge}_x$ and (001)-Si which is given by [29–31]

$$\Delta E_0 = 840x \text{ meV} \quad (5.121)$$

For $x = 0.42$, this gives a barrier height of 353 meV. Bandedge shift ΔE_{hd} due to heavy doping [32, 33] is quite large here because of the high doping concentration in p^+ $\text{Si}_{1-x}\text{Ge}_x$ and must be included in the calculations. Neglecting the effect of carrier redistribution across the interface, the barrier height b is given by [32]

$$b = \Delta E_0 + \Delta E_{\text{hd}} - E_f \quad (5.122)$$

The above expression gives still a value which is much higher than the experimental one. The next consideration is the band bending due to the charge redistribution at the heterointerface. Refer to the band diagram in Fig. 5.21(a), the one-dimensional Poisson equation can be written as

$$\frac{d^2\phi}{dz^2} = \begin{cases} -\frac{e}{\epsilon}(N_A^1 - p) & z < 0 \\ -\frac{e}{\epsilon}N_A^2 & 0 < z < a \\ 0 & z > a \end{cases} \quad (5.123)$$

Here we have set the sample growth direction as the z axis and have neglected the difference in the dielectric constants between Si and $\text{Si}_{1-x}\text{Ge}_x$. a is the spatial region where the holes in the p -Si are totally depleted due to the band bending. $N_A^1 = 10^{20} \text{ cm}^{-3}$ and $N_A^2 = 5 \times 10^{17} \text{ cm}^{-3}$ are the acceptor doping concentrations in p^+ -SiGe and p -Si, respectively. Let the Fermi level E_f of the whole system as the energy reference point, the boundary conditions for Eq. (5.123) are:

$$\begin{aligned} e\phi(z) &= E_1, & z \gg 0 \\ e\phi(0^-) - e\phi(0^+) &= \Delta E_0 + \Delta E_{\text{hd}}, & z = 0 \\ e\phi(z) &= E_2, & z \ll 0 \end{aligned} \quad (5.124)$$

E_1 and E_2 refer to the valence bandedges on the p -Si side and p^+ - $\text{Si}_{1-x}\text{Ge}_x$ side, respectively. Since the energy is being measured from the Fermi level, E_1 and E_2 are numerically equal to the Fermi energies E_{f1} and E_{f2} measured from the valence bandedges on the two sides.

At zero temperature, free carrier states are filled between the Fermi level and the valence bandedge. Below the Fermi level the hole states are empty. When $N_A^2 = 10^{17} \text{ cm}^{-3}$, $N_A^1 = 10^{20} \text{ cm}^{-3}$, $x = 0.4$, we obtain $E_{f1} = 4 \text{ meV}$ and $E_{f2} = 141 \text{ meV}$.

It is easy to solve Poisson equation Eq. (5.123) which gives a value of about 10 meV for the band bending δ . (δ is defined in Fig. 5.21(a) which is obtained quantum mechanically and will be explained in the following text.)

Such a small band bending effect is expected because of the large doping concentration in the p^+ -SiGe side. However, the space variation of the free carriers due to the quantum mechanical effect indicates that the classical approximation is not valid in the region around the heterointerface (we learned this in the previous chapter when we studied the nano-scale transistor). A simple effective-mass Schrödinger equation can be written for the valence-band holes:

$$\left(-\frac{\hbar^2}{2m_{cc}^*} \frac{d^2}{dz^2} + V_v \right) \psi(z) = E \psi(z) \quad (5.125)$$

where V_v is the valence bandedge and m_{cc}^* is the carrier-concentration effective mass. The hole distribution is now given by

$$p(z) = A \sum_E \ln \left[1 + \exp \left(\frac{E}{k_B T} \right) \right] |\psi(z)|^2 \quad (5.126)$$

Table 5.6 Different factors [meV] that determine the effective barrier height of the 10^{17}-cm^{-3} -doped Si/ 10^{20}-cm^{-3} -doped $\text{Si}_{0.58}\text{Ge}_{0.42}$ heterostructure internal emission infrared detector

ΔE_0 (heterojunction offset)	353
ΔE_{hd} (heavy doping effect)	76
E_f (Fermi level)	165
δ (quantum effect)	110
Cal. effective barrier height	155
Observed barrier height	125
Total discrepancy	30

A is a constant which can be determined by the charge neutrality condition for $z \ll 0$. Eq. (5.123) is to be revised to include the wave nature of the hole wave functions

$$\frac{d^2\phi}{dz^2} = \begin{cases} -\frac{e}{\epsilon}[N_A^1 - p(z)] & z < 0 \\ -\frac{e}{\epsilon}[N_A^2 - p(z)] & 0 < z < a \\ 0 & z > a \end{cases} \quad (5.127)$$

Equations (5.125), (5.127) are to be self-consistently solved for both the valence band energy diagram V_v and free carrier distribution $p(z)$. The numerical results are plotted in Fig. 5.21. From V_v we can readily obtain the band bending δ . We have used the carrier-concentration effective mass m_{cc}^* of 0.6 for 10^{20} cm^{-3} p^+ -type doping in $\text{Si}_{0.58}\text{Ge}_{0.42}$ and obtained $\delta = 110\text{ meV}$.

Different factors that determine the effective barrier height are listed in Table 5.6 together with the measured value for comparison. Now the total calculated effective barrier height becomes 155 meV, it is still larger than the experimentally determined value of 125 meV. There are other important factors that lower the effective barrier height such as the precise determination of the p^+ doping concentration and heterointerface roughness. In view of these considerations, the discrepancy of 30 meV between the current theory and experimental result is not serious.

5.6 Quantum Dot Solar Cell

Solar cell is mostly based on photovoltaic effect, and semiconductor $p - n$ junction based solar cell is the most popular structure, which is schematically shown in Fig. 5.22. At equilibrium, free electrons provided by donors in the n region and free holes provided by acceptors in the p region will recombine with each other at the junction interface, resulting in a depletion region, where donor ions and acceptor ions form the built-in electric field. Fermi level E_f is flat everywhere and there is no net current and voltage output. Under light illumination, excess electrons and holes are generated after absorbing photons in the depletion region, which will be driven by the built-in field to drift into opposite regions. In the short-circuit case when the Fermi level is still flat across the structure, the photocurrent output reaches the maximal value I_{sc} . Without external connection, i.e., the open-circuit case, the

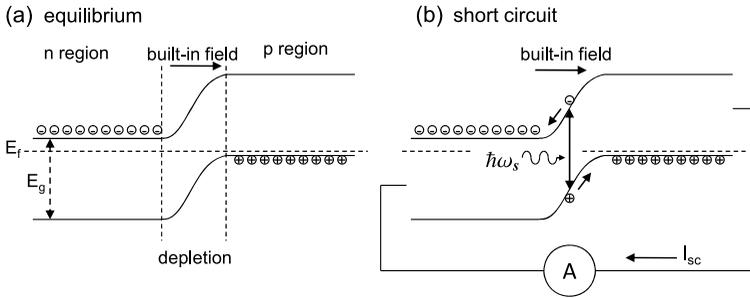


Fig. 5.22 (a) Energy band structure of a $p-n$ junction at equilibrium. (b) Short circuit of a $p-n$ junction under illumination. I_{sc} denotes the short-circuit current

excess carriers will accumulate, resulting in an open-circuit voltage V_{oc} . Here subscript “sc” and “oc” stand for short-circuit and open-circuit, respectively. Between the open-circuit and short-circuit cases, the relationship between the output voltage and current forms the $I-V$ characteristics, in which there is a point corresponding to the maximal output power output and the maximum photovoltaic efficiency η_{max} .

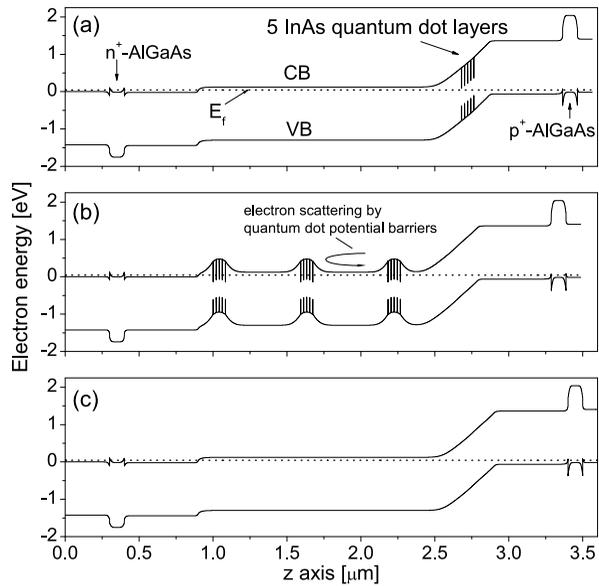
In Fig. 5.22(b) we only show photogenerations of electrons and holes in the depletion region. Photogenerations of electrons and holes also occur in the n and p regions, but with a much reduced contribution to the photovoltaic effect. First of all, the optical transition rates are relatively low since the electron (hole) states in the conduction (valence) band of the n (p) region are already occupied due to the dopings (Pauli exclusion principle). Secondly, the average electric fields in the two regions are very low (Ohm’s law). Thus, photogeneration of electrons and holes in the depletion region is the major contribution to the photovoltaic effect.

Based on the detailed balance principle, Shockley and Queisser deduced the bandgap-related ideal η_{max} of a single-junction solar cell, which is now known as the Shockley-Queisser limit which is 31 % for bandgap range 0.95 ~ 1.6 eV [34]. Si, GaAs, InP and CdTe are thus good candidates for solar cell applications. To further increase the photovoltaic efficiency, novel structures are to be introduced. One structure is a multi-junction tandem solar cell. Several semiconductor $p-n$ junctions with different bandgaps, from the top wide-bandgap junction to the bottom narrow-bandgap junction, are connected together in series, in order to absorb different-energy photons in different junctions. Another intensely studied novel structure is to use nanostructures, such as quantum wells and quantum dots, in a single junction solar cell to improve η_{max} . The main idea is that except the above-bandgap photon absorption in the bulk host semiconductor, the embedded nanostructures absorb sub-bandgap photons. Moreover, a huge potential benefit of using nanostructures is the impact ionization we have studied in Sect. 3.9 that a high-energy photon can generate more than one electron-hole pairs, i.e., multiple exciton generation (MEG). If we can extract the electrons and/or holes before they recombine, we will reach a quantum efficiency of more than 100 % (one photon in, more than one carrier out). And finally, compared to multi-junction tandem cell structures, nanostructure-embedded cells are more flexible in terms of material selection and device design.

Table 5.7 Basic GaAs $p-n$ solar cell structure

Layers	Doping [cm^{-3}]	Thickness [nm]
p GaAs	2×10^{17}	500
i GaAs	undoped	140
n^- GaAs	2×10^{16}	1860
n^+ GaAs	2×10^{18}	500

Fig. 5.23 One-dimensional energy band structures of the three $p-n$ GaAs solar cells. (a) Sample A: 5 layers of InAs quantum dots separated by 20 nm thick GaAs layers embedded in the depletion region; (b) Sample B: 3 groups of the same five InAs quantum dot layers embedded in the base region; (c) Sample C: reference solar cell (X.-J. Shang, J.-F. He, H.-L. Wang, M.-F. Li, Y. Zhu, Z.-C. Niu, and Y. Fu, Effect of built-in electric field on photovoltaic effect of InAs quantum dots in GaAs solar cells, *Appl. Phys. A*, vol. 103, pp. 335–341, 2011, with kind permission from Springer Science and Business Media)

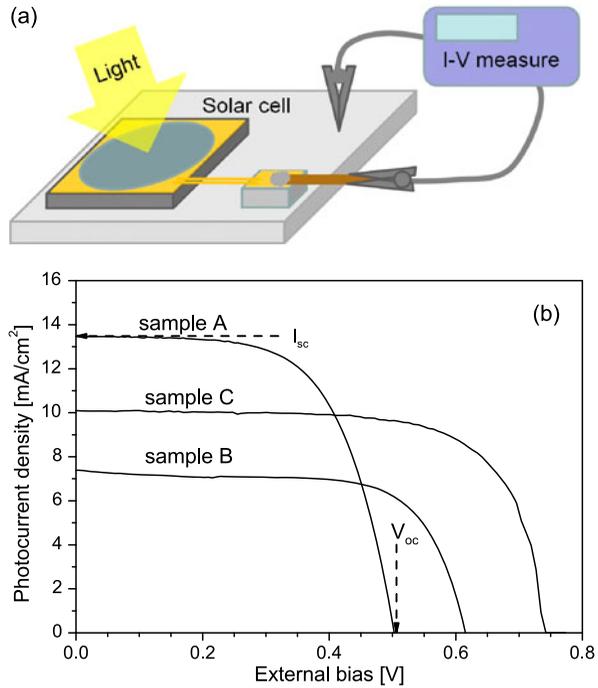


In this section let us study a common GaAs $p-n$ junction solar cell embedded with InAs quantum dots. We actually study three cells, one with InAs quantum dots embedded in the depletion region, denoted as sample A, one with InAs quantum dots embedded in the n region, i.e., sample B, and a reference cell without embedding any nanostructures, sample C. Table 5.7 lists the basic GaAs $p-n$ solar cell structure. There is also a top 100 nm p -doped $\text{Al}_{0.85}\text{Ga}_{0.15}\text{As}$ window layer and a back n -doped $\text{Al}_{0.2}\text{Ga}_{0.8}\text{As}$ surface field layer to prevent minority carriers from diffusing into the surfaces, see Fig. 5.23.

2.3 monolayers (ML) of InAs for sample A and B, which eventually form one quantum dot layer, were directly grown on GaAs at a rate of 0.1 ML/s and a 10-s growth interruption after every 0.1 ML growth. The quantum dot layer was then capped by GaAs spacer at a rate of 1 ML/s, followed by the next quantum dot layer growth. The thickness of GaAs spacer is 20 nm in sample A and B. The sheet quantum dot density on one quantum dot layer is approximately $4 \times 10^{10} \text{ cm}^{-2}$.

We approximate each quantum dot layer by an effective InAs quantum well layer with a thickness of 2.3 ML (0.7 nm, i.e., InAs quantum dot growth condition) along the sample growth direction (the z direction), while the density of electron and hole

Fig. 5.24 (a) Photovoltaic measurement setup. The back electrode is attached conductively to a metal base, while the top electrode is extended with wires which are insulated from the metal base. (b) $I - V$ characteristics under a solar simulator (1 sun AM1.5G) at room temperature (Reprinted with permission from X.-J. Shang, J.-F. He, M.-F. Li, F. Zhan, H.-Q. Ni, Z.-C. Niu, H. Pettersson, and Y. Fu, Quantum-dot-induced optical transition enhancement in InAs quantum-dot-embedded p-i-n GaAs solar cells, *Appl. Phys. Lett.*, vol. 99, p. 113514, 2011. Copyright 2011, American Institute of Physics)



states in such a quantum well is approached by a single δ function to account for the discrete nature of the three-dimensional confined sublevels. Energy band structures of the three solar cells at equilibrium are presented in Fig. 5.23 which were obtained self-consistently (we have used the self-consistent calculations many times before).

$I - V$ characteristics of the three cells under 1 sun AM1.5G were obtained by photovoltaic measurement setup in Fig. 5.24(a) and are presented in Fig. 5.24(b). The light intensity of the solar simulator (1 sun AM1.5G) is 100 mW/cm^2 calibrated by a standard Si cell. Compared with the reference sample C, both quantum-dot-embedded solar cells, A and B, have lower open-circuit voltages (V_{oc}). Moreover, sample A has the highest short-circuit photocurrent I_{sc} while sample B has the lowest I_{sc} .

The understanding of the experimental results in Fig. 5.24(b) is basically straightforward by the physics we have so far learned. The photocurrent in sample B is reduced due to the scattering of electrons by the quantum-dot-induced potential barriers, see Fig. 5.23, which, on the other hand, provides a better accumulation of the photogenerated excess carriers in the n and p regions, resulting in a higher open-circuit voltage. The photocurrent I_{sc} of sample A consists of not only the contribution of above-GaAs-bandgap photon absorption in the GaAs region, i.e., the major photocurrent of the reference sample C, it also includes the sub-GaAs-bandgap photon absorption in the InAs quantum dots. The reduced open-circuit voltage in sample A can be caused by the interband tunneling in the depletion region.

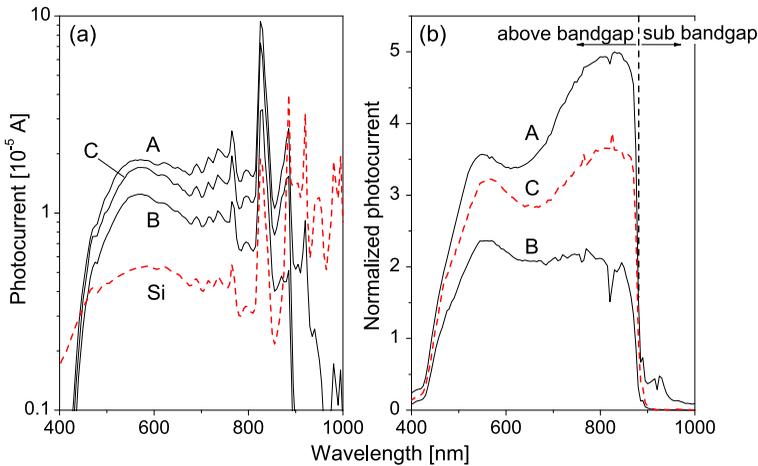


Fig. 5.25 (a) Photocurrent spectra at 300 K under a xenon lamp. (b) Photocurrent spectra normalized with respect to the Si photodiode (Reprinted with permission from X.-J. Shang, J.-F. He, M.-F. Li, F. Zhan, H.-Q. Ni, Z.-C. Niu, H. Pettersson, and Y. Fu, Quantum-dot-induced optical transition enhancement in InAs quantum-dot-embedded p-i-n GaAs solar cells, *Appl. Phys. Lett.*, vol. 99, p. 113514, 2011. Copyright 2011, American Institute of Physics)

Further experimental characterization of the three samples reveals more physics. Photocurrent spectra shown in Fig. 5.25 were measured by a Keithley multimeter under a xenon lamp with a monochromator, where a Si photodiode was used as reference. Two main spectral differences are observed between the three samples. The first one is the sub-GaAs-bandgap photocurrent ($\lambda > 900$ nm), and the second is the photocurrent enhancement in sample A and reduction in sample B above the GaAs bandgap ($\lambda < 900$ nm).

The decrease of sample B's photocurrent can be largely attributed to the reduced carrier mobility in the n region, as aforementioned. The most interesting thing about Fig. 5.25(b) is the enhancement of photocurrent in the above-GaAs-bandgap wavelength range of $\lambda < 900$ nm. As shown in Fig. 5.24, the short-circuit current I_{sc} of sample A is about 30 % higher than sample C. The same 30 % enhancement is also obtained here in Fig. 5.25(b) by integrating and comparing the relative photocurrents. Contribution to the observed photocurrent enhancement from sub-GaAs-bandgap absorptions due to InAs quantum dots is very small, see Fig. 5.25(b) when $\lambda > 900$ nm.

The most possible mechanics of the photocurrent enhancement above the GaAs bandgap in sample A is the increased photon absorption due to quantum-dot induced reflections of the wave functions of an incident conduction-band electron from one side of the depletion region and an incident valence-band hole from the other side. Refer to Fig. 5.26, for simplicity, we model the quantum-dot potential variations in the conduction and valence bands as one-dimensional quantum wells. The incident conduction-band electron, with an energy above the GaAs conduction band edge,

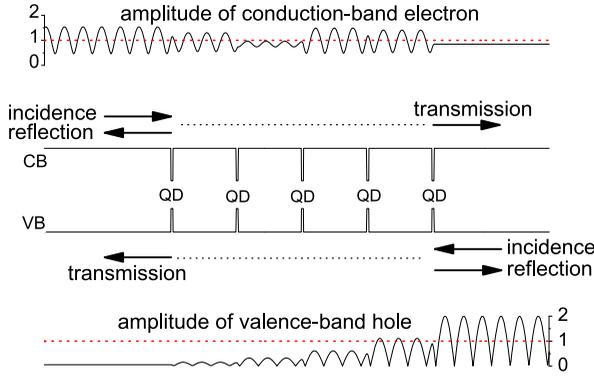


Fig. 5.26 Wave functions of conduction-band (CB) electron and valence-band (VB) hole with a transition energy $\hbar\omega = 1.593$ eV (above the GaAs bandgap). Also presented are the schematic CB and VB of GaAs embedded with InAs quantum dots modeled as one-dimensional quantum wells (Reprinted with permission from X.-J. Shang, J.-F. He, M.-F. Li, F. Zhan, H.-Q. Ni, Z.-C. Niu, H. Pettersson, and Y. Fu, Quantum-dot-induced optical transition enhancement in InAs quantum-dot-embedded p-i-n GaAs solar cells, *Appl. Phys. Lett.*, vol. 99, p. 113514, 2011. Copyright 2011, American Institute of Physics)

comes to the quantum-dot layers in a plane wave form

$$e^{ik_1z}$$

it is partially reflected

$$r_1 e^{-ik_1z}$$

and partially transmitted

$$t_1 e^{ik_1z}$$

Similarly the wave function components for the valence-band hole are

$$e^{-ik_2z}, \quad r_2 e^{ik_2z}, \quad t_2 e^{-ik_2z}$$

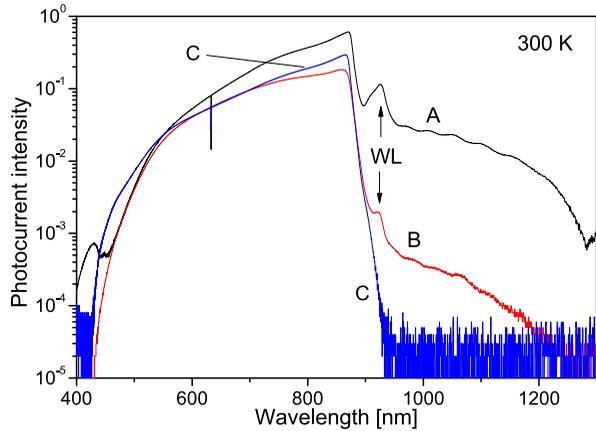
Note that there are other transmission cases such as both the conduction-band electron and the valence-band hole incident to the quantum-dot regions from the same side but they do not contribute to the photocurrent. Without quantum dots,

$$r_1 = r_2 = 0, \quad t_1 = t_2 = 1.0$$

The amplitudes of such wave functions are shown as dotted lines in Fig. 5.26. The optical interband transition matrix element between these two wave functions is proportional to

$$W_C = \langle e^{ik_1z} | e^{-ik_2z} \rangle = \int e^{-i(k_1+k_2)z} dz \quad (5.128)$$

Fig. 5.27 Room-temperature photocurrent spectra of the three samples (Reprinted with permission from X.-J. Shang, J.-F. He, M.-F. Li, F. Zhan, H.-Q. Ni, Z.-C. Niu, H. Pettersson, and Y. Fu, Quantum-dot-induced optical transition enhancement in InAs quantum-dot-embedded p-i-n GaAs solar cells, Appl. Phys. Lett., vol. 99, p. 113514, 2011. Copyright 2011, American Institute of Physics)



The wave functions are modified when QDs are embedded. As shown in Fig. 5.26, the conduction-band electron transmits relatively well due to the small electron effective mass, while the much heavier valence-band hole is greatly affected. For the two wave functions presented in Fig. 5.25, we can neglect the transmitted valence-band hole so that its wave function on the right side of the structure is

$$e^{-ik_2z} + e^{ik_2z}$$

We further approximate the conduction-band electron as a perfect transmission so that its wave function on the right side of the structure is

$$e^{ik_1z}$$

The optical transition matrix element becomes now

$$W_A = \langle e^{ik_1z} | e^{-ik_2z} + e^{ik_2z} \rangle = W_C + \int e^{-i(k_1-k_2)z} dz \quad (5.129)$$

W_C oscillates while the second term in W_A can be very large when $k_1 = k_2$. The condition of $k_1 = k_2$ can easily be fulfilled which explains the photocurrent enhancement of sample A over the whole optical range above the GaAs bandgap. Note that the W_A enhancement in the above analysis is very large as compared with experimental data due to the approximation of the three-dimensionally confined quantum-dot potentials by one-dimensionally confined quantum-well potentials, leading to a strongly exaggerated wave function reflection.

The full photocurrent spectra of all samples at room temperature are compared in Fig. 5.27. Sample A and B show sub-GaAs-bandgap photocurrents (900 ~ 1300 nm) related to quantum-dot absorption. The peak around 920 nm comes from the absorption in quantum-dot wetting layers (WL) (the general self-assembled quantum dot structure is shown in Fig. 1.16); the extended profile in 950 ~ 1300 nm comes from the absorption in quantum dots. Sample C has no such features.

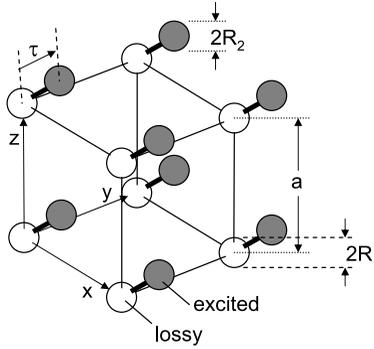


Fig. 5.28 Schematic quantum-dot dimer lattice for lossless negative epsilon in a primitive cubic lattice structure. R_1 and R_2 denote the radii of the quantum dots and a is the lattice constant of the primitive cubic lattice (Reprinted with permission from Y. Fu, Photonic energy band structure of excitonic quantum dot dimer system, J. Appl. Phys., vol. 106, p. 054302(5), 2009. Copyright 2009, American Institute of Physics)

Sample A has the highest quantum-dot induced photocurrent while sample B has the lowest one. It is related to the quantum-dot location in the $p - n$ structure. Sample A has five closely-stacked quantum-dot layers (spaced by 20 nm GaAs) in the depletion region, the light absorption in InAs quantum dots and the photocarrier extraction from quantum dots and transport (mostly drift under the built-in electric field) are effective. Sample B has totally fifteen 20-nm GaAs-spaced quantum dot layers in the flat-band n region, photocarrier extraction from quantum dots is a big problem due to potential variation.

5.7 Exciton-Polariton Photonic Crystal

As we learned in Sect. 3.7 that the electric polarization of an exciton in a nanostructure will modify the optical properties of the nanostructure significantly in terms of exciton polariton. The dielectric coefficient of the nanostructure in which the exciton is confined will be different due to the exciton polariton from its surrounding material, resulting in a dielectric contrast. In this section, we construct a quantum-dot photonic crystal by utilizing such a dielectric contrast. We position two types of spherical quantum dots, having radii R_1 and R_2 , respectively, in a face-centered cubic (fcc) lattice having a lattice constant a , see Fig. 5.28. Note that Fig. 5.28 shows a primitive cubic lattice for the sake of simple visualization. Type-I quantum dots (exciton energies are $\hbar\omega_1$), which are lossy, occupy the normal fcc lattice sites, while the already excited quantum dots, i.e., type-II quantum dots ($\hbar\omega_2$) are displaced from the type-I quantum dots by τ .

Let c_{ai} be the occupation of the ground state of quantum dot i positioned at lattice site \mathbf{a} , i.e., the valence band is completely filled and the conduction band is completely empty, the contribution of an excited exciton state i in this quantum dot

denoted by the spatial position of the quantum dot and the exciton state index i , \mathbf{a}_i , to the dielectric polarization is given by

$$\mathbf{P}_{ai}(\mathbf{r}) = \frac{e^2 \mathbf{p}_{cv} (2|c_{ai}|^2 - 1)}{\hbar(\omega_{ai} - \omega + i\gamma)\omega_{ai}^2 m_0^2} \psi_{ai}(\mathbf{r}, \mathbf{r}) \int \psi_{ai}^*(\mathbf{r}', \mathbf{r}') \mathbf{p}_{cv} \cdot \mathbf{E}(\mathbf{r}') d\mathbf{r}' \quad (5.130)$$

see Sect. 3.7, where $\psi_{ai}(\mathbf{r}, \mathbf{r}) = \psi_{ai}(\mathbf{r}_e, \mathbf{r}_h)|_{\mathbf{r}_e=\mathbf{r}_h=\mathbf{r}}$, $\psi_{ai}(\mathbf{r}_e, \mathbf{r}_h)$ is the exciton wave function, \mathbf{r}_e and \mathbf{r}_h denote the electron and hole, respectively. $\hbar\omega_{ai}$ is the excitonic energy and γ denotes the damping rate of the exciton state. We first introduce a few physical parameters relevant to the quantum-dot structure

$$\begin{aligned} \mu &= \frac{m_c^* m_v^*}{m_c^* + m_v^*}, & a_B &= \frac{4\pi\epsilon\hbar^2}{\mu e^2}, & \epsilon\omega_{LT} a_B^3 &= \frac{e^2 p_{cv}^2}{\pi \hbar \omega_0^2 m_0^2} \\ T_{ai}(\omega) &= \frac{\epsilon\omega_{LT} \pi a_B^3 (2|c_{ai}|^2 - 1)}{\omega_{ai} - \omega + i\gamma} \end{aligned} \quad (5.131)$$

where m_c^* and m_v^* are electron and hole effective masses so that μ is the reduced effective mass of the exciton. a_B and ω_{LT} are the exciton Bohr radius and the longitudinal-transverse splitting, respectively. For all the quantum dots under discussion, R_i is in the order of a_B so that one may neglect the internal motion of the electron-hole pair inside the quantum dot and the exciton ground state is described by as

$$\psi_{a1}(\mathbf{r}_e, \mathbf{r}_h) = \frac{1}{\sqrt{2\pi R_1} |\mathbf{r} - \mathbf{a}|} \sin\left(\frac{\pi |\mathbf{r} - \mathbf{a}|}{R_1}\right) \frac{1}{\sqrt{\pi a_B^3}} e^{-\frac{|\mathbf{r}_e - \mathbf{r}_h|}{a_B}} \quad (5.132)$$

for $|\mathbf{r} - \mathbf{a}| \leq R_1$, where $\mathbf{r} = (m_c^* \mathbf{r}_e + m_v^* \mathbf{r}_h)/(m_c^* + m_v^*)$ is the exciton center of mass, see Eq. (3.245). For type-II quantum dots of $i = 2$, the wave function of the exciton is expressed similarly by replacing \mathbf{a} with $\mathbf{a} + \boldsymbol{\tau}$ and R_1 with R_2 . Using these physical quantities, the Maxwell equations describing the incident electromagnetic field in such a quantum dot dimer lattice which is free of charges and free of drift-diffusion current are written as

$$\begin{aligned} \nabla \times [\nabla \times \mathbf{E}(\mathbf{r})] &= \mu_0 \omega^2 \mathbf{D}(\mathbf{r}) \\ \nabla \cdot \mathbf{D}(\mathbf{r}) &= 0 \end{aligned} \quad (5.133)$$

in the MKS unit system [units based on meters (m), kilograms (kg), and seconds (s)], where $\mathbf{E}(\mathbf{r})$ is the electric field, $\mathbf{D}(\mathbf{r})$ is the displacement vector, and ω is the angular frequency of the electromagnetic field. Here we have assumed that the quantum dot is composed of uniform isotropic linear media, for which $\mathbf{D} = \epsilon \mathbf{E}$ and $\mathbf{B} = \mu_0 \mathbf{H}$. \mathbf{H} is the magnetizing field and \mathbf{B} is the magnetic field. μ_0 is the magnetic constant (permeability of free space). The nonlocal material equation relating $\mathbf{D}(\mathbf{r})$ and $\mathbf{E}(\mathbf{r})$ is

$$\mathbf{D}(\mathbf{r}) = \epsilon \mathbf{E}(\mathbf{r}) + \sum_{ai} \mathbf{P}_{ai}(\mathbf{r}) \quad (5.134)$$

We have neglected the overlap of exciton envelope functions centered at different quantum dots so that excitons in different quantum dots are assumed to be coupled only via the electromagnetic field.

It follows from the second Eq. (5.133) and Eq. (5.134) that

$$\nabla \cdot \mathbf{E}(\mathbf{r}) = -\frac{1}{\epsilon} \nabla \cdot \mathbf{P}(\mathbf{r}) \quad (5.135)$$

so that the first Eq. (5.133) can be rewritten as

$$\nabla^2 \mathbf{E}(\mathbf{r}) + k^2 \mathbf{E}(\mathbf{r}) = -\frac{k_0^2}{\epsilon_0} \left\{ \mathbf{P}(\mathbf{r}) + \frac{1}{k^2} \nabla [\nabla \cdot \mathbf{P}(\mathbf{r})] \right\} \quad (5.136)$$

where $k_0 = \omega/c$, $k = k_0 n = \omega n/c$ and $n = \sqrt{\epsilon/\epsilon_0}$.

Equations (5.130) and (5.136) are our master equations about the electric field in the excitonic quantum-dot lattice.

We now seek for Bloch-like solutions of Eq. (5.136) satisfying

$$\mathbf{E}_q(\mathbf{r} + \mathbf{a}) = e^{i\mathbf{q} \cdot \mathbf{a}} \mathbf{E}_q(\mathbf{r}) \quad (5.137)$$

By this, we, in principle, assume that the system is lossless. This is artificially achieved by letting $\gamma = 0$ in Eq. (5.130). We expand the vector function $\mathbf{E}_q(\mathbf{r})$ in the Fourier series

$$\mathbf{E}_q(\mathbf{r}) = \sum_{\mathbf{g}} e^{i(\mathbf{q}+\mathbf{g}) \cdot \mathbf{r}} \mathbf{E}_{q+\mathbf{g}} \quad (5.138)$$

\mathbf{g} are the reciprocal lattice vectors. Denote $a_1 = \sqrt{2|c_1|^2 - 1}$, $a_2 = \sqrt{2|c_2|^2 - 1}$, $v_0 = a^3/4$ is the volume of the primitive unit cell of the fcc lattice, $\phi_i = a_i \psi_i$ and $t = \epsilon \omega_{\text{LT}} \pi a_B^3$, the excitonic polarization becomes

$$\mathbf{P}(\mathbf{r}) = \sum_{ai} \frac{t}{\omega_i - \omega + i\gamma} \phi_{ai}(\mathbf{r}, \mathbf{r}) \int \phi_{ai}^*(\mathbf{r}', \mathbf{r}') \mathbf{E}(\mathbf{r}') d\mathbf{r}' \quad (5.139)$$

Note that the excitonic wave function is real so that we have dropped off the complex conjugation of the wave function in the integrals in Eq. (5.130). The two integrals in the above equation can be transformed into

$$\begin{aligned} \int \phi_{1,a}(\mathbf{r}) \mathbf{E}(\mathbf{r}) d\mathbf{r} &= e^{i\mathbf{q} \cdot \mathbf{a}} \sum_{\mathbf{g}} I_{1,q+\mathbf{g}} \mathbf{E}_{q+\mathbf{g}} \equiv e^{i\mathbf{q} \cdot \mathbf{a}} \mathbf{\Lambda}_1 \\ \int \phi_{2,a+\boldsymbol{\tau}}(\mathbf{r}) \mathbf{E}(\mathbf{r}) d\mathbf{r} &= e^{i\mathbf{q} \cdot \mathbf{a}} \sum_{\mathbf{g}} I_{2,q+\mathbf{g}} \mathbf{E}_{q+\mathbf{g}} \equiv e^{i\mathbf{q} \cdot \mathbf{a}} \mathbf{\Lambda}_2 \end{aligned} \quad (5.140)$$

where

$$I_{1,q+\mathbf{g}} = \int \phi_{1,0}(\mathbf{r}) e^{i(\mathbf{q}+\mathbf{g}) \cdot \mathbf{r}} d\mathbf{r} \quad (5.141)$$

$$I_{2,\mathbf{q}+\mathbf{g}} = \int \phi_{2,0}(\mathbf{r}) e^{i(\mathbf{q}+\mathbf{g})\cdot(\mathbf{r}+\boldsymbol{\tau})} d\mathbf{r}$$

The sums $\sum_a \phi_{1,a}(\mathbf{r}) e^{i\mathbf{q}\cdot\mathbf{a}}$ and $\sum_{a+\boldsymbol{\tau}} \phi_{2,a+\boldsymbol{\tau}}(\mathbf{r}) e^{i\mathbf{q}\cdot\mathbf{a}}$ satisfy the translational symmetry and can be presented as

$$\begin{aligned} \sum_a \phi_{1,a}(\mathbf{r}) e^{i\mathbf{q}\cdot\mathbf{a}} &= \sum_g e^{i(\mathbf{q}+\mathbf{g})\cdot\mathbf{r}} \frac{I_{1,\mathbf{q}+\mathbf{g}}^*}{v_0} \\ \sum_{a+\boldsymbol{\tau}} \phi_{2,a+\boldsymbol{\tau}}(\mathbf{r}) e^{i\mathbf{q}\cdot\mathbf{a}} &= \sum_g e^{i(\mathbf{q}+\mathbf{g})\cdot\mathbf{r}} \frac{I_{2,\mathbf{q}+\mathbf{g}}^*}{v_0} \end{aligned} \quad (5.142)$$

The linear equations for the space harmonics $\mathbf{E}_{\mathbf{q}+\mathbf{g}}$ can thus be written

$$\begin{aligned} &(|\mathbf{q} + \mathbf{g}|^2 - k^2) \mathbf{E}_{\mathbf{q}+\mathbf{g}} \\ &= \frac{k_0^2 t}{v_0} \left[1 - \frac{1}{k^2} (\mathbf{q} + \mathbf{g})^2 \right] \left[\frac{I_{1,\mathbf{q}+\mathbf{g}}^*}{\omega_1 - \omega + i\gamma} \boldsymbol{\Lambda}_1 + \frac{I_{2,\mathbf{q}+\mathbf{g}}^*}{\omega_2 - \omega + i\gamma} \boldsymbol{\Lambda}_2 \right] \end{aligned} \quad (5.143)$$

Let $S(\mathbf{Q})_{\alpha\beta} = \delta_{\alpha\beta} - Q_\alpha Q_\beta / k^2$, $\alpha, \beta = x, y, z$, $\delta_{\alpha\beta}$ is the Kronecker symbol. Dividing Eq. (5.143) by $(|\mathbf{q} + \mathbf{g}|^2 - k^2)$, multiplying it by $I_{1,\mathbf{q}+\mathbf{g}}$ and summing over \mathbf{g} :

$$\begin{aligned} \sum_g I_{1,\mathbf{q}+\mathbf{g}} \mathbf{E}_{\mathbf{q}+\mathbf{g}} &= \boldsymbol{\Lambda}_1 \\ &= \frac{k_0^2 t}{v_0} \sum_g \frac{\widehat{S}(\mathbf{q} + \mathbf{g})}{|\mathbf{q} + \mathbf{g}|^2 - k^2} \left[\frac{|I_{1,\mathbf{q}+\mathbf{g}}|^2}{\omega_1 - \omega + i\gamma} \boldsymbol{\Lambda}_1 + \frac{I_{1,\mathbf{q}+\mathbf{g}} I_{2,\mathbf{q}+\mathbf{g}}^*}{\omega_2 - \omega + i\gamma} \boldsymbol{\Lambda}_2 \right] \\ &= \widehat{M}_1(\omega, \mathbf{q}) \boldsymbol{\Lambda}_1 + \widehat{M}_2(\omega, \mathbf{q}) \boldsymbol{\Lambda}_2 \end{aligned} \quad (5.144)$$

where

$$\begin{aligned} \widehat{M}_1(\omega, \mathbf{q}) &= \frac{k_0^2 t}{(\omega_1 - \omega + i\gamma)v_0} \sum_g \frac{\widehat{S}(\mathbf{q} + \mathbf{g})}{|\mathbf{q} + \mathbf{g}|^2 - k^2} |I_{1,\mathbf{q}+\mathbf{g}}|^2 \\ \widehat{M}_2(\omega, \mathbf{q}) &= \frac{k_0^2 t}{(\omega_2 - \omega + i\gamma)v_0} \sum_g \frac{\widehat{S}(\mathbf{q} + \mathbf{g})}{|\mathbf{q} + \mathbf{g}|^2 - k^2} I_{1,\mathbf{q}+\mathbf{g}} I_{2,\mathbf{q}+\mathbf{g}}^* \end{aligned} \quad (5.145)$$

Similarly, dividing Eq. (5.143) by $(|\mathbf{q} + \mathbf{g}|^2 - k^2)$, multiplying it by $I_{2,\mathbf{q}+\mathbf{g}}$ and summing over \mathbf{g} ,

$$\begin{aligned} \sum_g I_{2,\mathbf{q}+\mathbf{g}} \mathbf{E}_{\mathbf{q}+\mathbf{g}} &= \boldsymbol{\Lambda}_2 \\ &= \frac{k_0^2 t}{v_0} \sum_g \frac{\widehat{S}(\mathbf{q} + \mathbf{g})}{|\mathbf{q} + \mathbf{g}|^2 - k^2} \left[\frac{I_{2,\mathbf{q}+\mathbf{g}} I_{1,\mathbf{q}+\mathbf{g}}^*}{\omega_1 - \omega + i\gamma} \boldsymbol{\Lambda}_1 + \frac{|I_{2,\mathbf{q}+\mathbf{g}}|^2}{\omega_2 - \omega + i\gamma} \boldsymbol{\Lambda}_2 \right] \\ &= \widehat{M}_3(\omega, \mathbf{q}) \boldsymbol{\Lambda}_1 + \widehat{M}_4(\omega, \mathbf{q}) \boldsymbol{\Lambda}_2 \end{aligned} \quad (5.146)$$

where

$$\begin{aligned}\widehat{M}_3(\omega, \mathbf{q}) &= \frac{k_0^2 t}{(\omega_1 - \omega + i\gamma)v_0} \sum_{\mathbf{g}} \frac{\widehat{S}(\mathbf{q} + \mathbf{g})}{|\mathbf{q} + \mathbf{g}|^2 - k^2} I_{2, \mathbf{q} + \mathbf{g}} I_{1, \mathbf{q} + \mathbf{g}}^* \\ \widehat{M}_4(\omega, \mathbf{q}) &= \frac{k_0^2 t}{(\omega_2 - \omega + i\gamma)v_0} \sum_{\mathbf{g}} \frac{\widehat{S}(\mathbf{q} + \mathbf{g})}{|\mathbf{q} + \mathbf{g}|^2 - k^2} |I_{2, \mathbf{q} + \mathbf{g}}|^2\end{aligned}\quad (5.147)$$

We arrive at the vector equations

$$\begin{aligned}[\mathbf{I} - \widehat{M}_1(\omega, \mathbf{q})] \mathbf{A}_1 &= \widehat{M}_2(\omega, \mathbf{q}) \mathbf{A}_2 \\ [\mathbf{I} - \widehat{M}_4(\omega, \mathbf{q})] \mathbf{A}_2 &= \widehat{M}_3(\omega, \mathbf{q}) \mathbf{A}_1\end{aligned}\quad (5.148)$$

where \mathbf{I} is a 3×3 unit matrix. We can rewrite the above equations as

$$D = \begin{pmatrix} 1 - M_{1,11} & -M_{1,12} & -M_{1,13} & -M_{2,11} & -M_{2,12} & -M_{2,13} \\ -M_{1,21} & 1 - M_{1,22} & -M_{1,23} & -M_{2,21} & -M_{2,22} & -M_{2,23} \\ -M_{1,31} & -M_{1,32} & 1 - M_{1,33} & -M_{2,31} & -M_{2,32} & -M_{2,33} \\ M_{3,11} & M_{3,12} & M_{3,13} & M_{4,11} - 1 & M_{4,12} & M_{4,13} \\ M_{3,21} & M_{3,22} & M_{3,23} & M_{4,21} & M_{4,22} - 1 & M_{4,23} \\ M_{3,31} & M_{3,32} & M_{3,33} & M_{4,31} & M_{4,32} & M_{4,33} - 1 \end{pmatrix}\quad (5.149)$$

$$\begin{aligned}M_{1,\alpha\beta}(\Omega, \mathbf{q}) &= \frac{8}{\pi v_0} \frac{\omega_{\text{LT}} R_1^3 |a_1|^2}{(\omega_1 - \omega + i\gamma)} \sigma_{11,\alpha\beta}(\Omega, \mathbf{q}) \\ M_{2,\alpha\beta}(\Omega, \mathbf{q}) &= \frac{8}{\pi v_0} \frac{\omega_{\text{LT}} (R_1 R_2)^{3/2} a_1 a_2^*}{(\omega_2 - \omega + i\gamma)} \sigma_{12,\alpha\beta}(\Omega, \mathbf{q}) \\ M_{3,\alpha\beta}(\Omega, \mathbf{q}) &= \frac{8}{\pi v_0} \frac{\omega_{\text{LT}} (R_1 R_2)^{3/2} a_1^* a_2}{(\omega_1 - \omega + i\gamma)} \sigma_{21,\alpha\beta}(\Omega, \mathbf{q}) \\ M_{4,\alpha\beta}(\Omega, \mathbf{q}) &= \frac{8}{\pi v_0} \frac{\omega_{\text{LT}} R_2^3 |a_2|^2}{(\omega_2 - \omega + i\gamma)} \sigma_{22,\alpha\beta}(\Omega, \mathbf{q}) \\ \sigma_{st,\alpha\beta}(\Omega, \mathbf{q}) &= \sum_{\mathbf{g}} \frac{f(|\mathbf{g} + \mathbf{q}| R_s) f(|\mathbf{g} + \mathbf{q}| R_t) S_{\alpha\beta}(\mathbf{g} + \mathbf{q})}{\Omega^2(\mathbf{g} + \mathbf{q}) - \Omega^2} e^{i(s-t)(\mathbf{g} + \mathbf{q}) \cdot \boldsymbol{\tau}}\end{aligned}\quad (5.150)$$

$$f(x) = \frac{\pi^2 \sin x}{x(\pi^2 - x^2)}$$

$$\Omega = \frac{\omega}{\omega_1}$$

$$\Omega(\mathbf{Q}) = \frac{c|\mathbf{Q}|}{\omega_1 n}$$

where $s, t = 1, 2$. The exciton-polariton dispersion $\omega(\mathbf{q})$ satisfies the equation

$$\text{Det}|D| = 0 \quad (5.151)$$

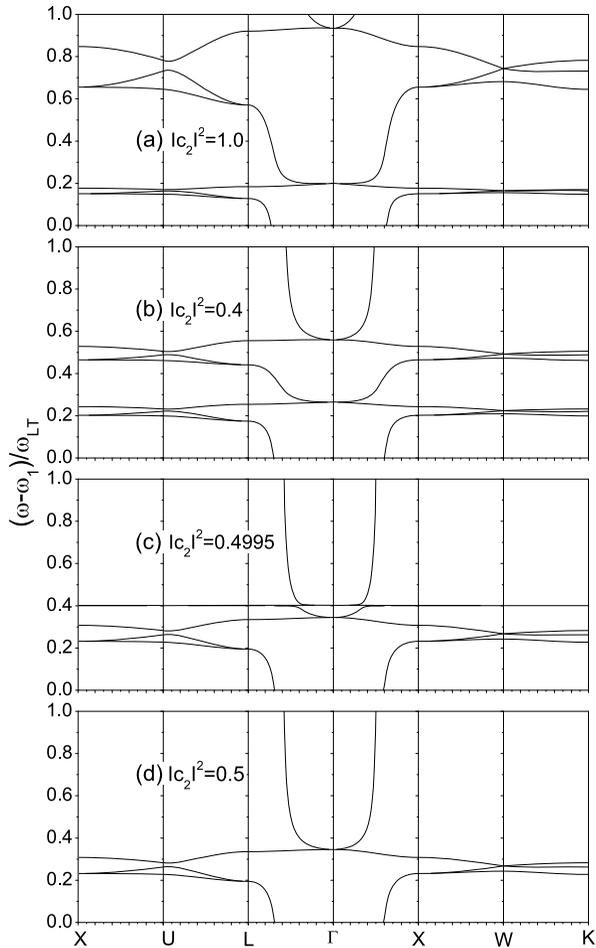
Note that by setting $|c_1|^2 = 0$ and $|c_2|^2 = 0.5$ (type-II quantum dots are totally transparent) we retrieve the case of the quantum dot lattice composed of only type-I quantum dots. By setting $|c_1|^2 = |c_2|^2 = 0$ and $\omega_1 = \omega_2$ we retrieve the case of a compound quantum-dot lattice.

The dispersion relationship $\omega(\mathbf{q})$ for the quantum-dot fcc lattice is presented in Fig. 5.29. The lattice is denoted by its lattice constant a , which is set to be $0.95a_{\text{Bragg}}$ (which is 116 nm for GaAs), where $a_{\text{Bragg}} = c\pi/\omega_1 n$. The lossy type-I quantum dots ($\hbar\omega_1 = 1.5$ eV) occupy the normal fcc lattice sites, while the excited type-II quantum dots ($\hbar\omega_2 = 1.503$ eV) are displaced by $\boldsymbol{\tau} = (a/2, a/2, a/2)$. $\hbar\omega_{\text{LT}} = 5$ meV. Here we observe the modification of the photonic band structure of the quantum-dot dimer lattice by pumping one type of the quantum dots (type-II quantum dots), which evolves from the one of only type-I quantum dots ($|c_1|^2 = 0$, i.e., at their ground exciton state) when type-II quantum dots are transparent [$|c_2|^2 = 0.5$, Fig. 5.29(d)] to the compound system [$|c_2|^2 = 1.0$, Fig. 5.29(a)]. We can observe modified but still characteristic features of the photonic dispersions of individual type-I and type-II quantum dots in their separate lattice formats in the compound system. More specifically, the resonance features of type-I quantum dots around $(\omega - \omega_1)/\omega_{\text{LT}} = 0.3$ in Fig. 5.29(d) becomes compressed by the radiative interaction between type-I quantum dots and type-II quantum dots, they are also shifted down to around 0.18 in Fig. 5.29(a).

The numerical solutions of Eq. (5.151) are symmetric with respect to $|c_2|^2$ and $1 - |c_2|^2$ when $c_1 = 0$. Thus, Fig. 5.29(a) represents also the photonic dispersion of the quantum dot dimer lattice when type-I and type-II quantum dots are all initially at their ground exciton states, i.e., $|c_1|^2 = |c_2|^2 = 0$. It is concluded that the modification of the excitonic state (from ground state to excited state) in one type of quantum dots in the dimer lattice does not affect the feature of the dispersion structure of the dimer lattice. Moreover, Fig. 5.29 shows that the effect of the excited type-II quantum dots at $(\mathbf{a} + \boldsymbol{\tau})$ persists even when the excitation rate is as close as to be 0.4995. This is due to the periodic boundary conditions. Furthermore, Fig. 5.29(d) remains exactly the same when type-II quantum dots located at $(\mathbf{a} + \boldsymbol{\tau})$ are physically removed.

Figure 5.29 suggests that the propagation of the electromagnetic field is strongly affected, especially in the wavelength ranges where the dispersion curves are flat, such as $(\omega - \omega_1)/\omega_{\text{LT}} \in (0.2, 0.4)$ in Fig. 5.29(d). In these wavelength ranges, the relationship between k and ω is very dispersive and light will be strongly diffracted. One simple way to check such a dispersion relationship is to measure the transmission and reflection of electromagnetic field through a quantum-dot embedded dielectric layer schematically shown in Fig. 5.30. The quantum dot has a radius of R , the square array has a period of L , and the dielectric film has a thickness of $2R$. Periodic boundary conditions are imposed on the four surfaces perpendicular to the dielectric film, while perfect matched layers are imposed at the top and bottom sur-

Fig. 5.29 Photonic dispersion relationship of a quantum-dot dimer fcc lattice. $R_1 = a/4$, $R_2 = a/5$, $\hbar\omega_1 = 1.5$ eV, $\hbar\omega_2 = 1.503$ eV, $\hbar\omega_{LT} = 5$ meV, $a = 0.95a_{\text{Bragg}}$, $\boldsymbol{\tau} = (a/2, a/2, a/2)$. $|c_1|^2 = 0$. (a) $|c_2|^2 = 1.0$ (complete excitation), (b) $|c_2|^2 = 0.4$, (c) $|c_2|^2 = 0.4995$, (d) $|c_2|^2 = 0.5$ (type-II quantum dots are totally transparent) (Reprinted with permission from Y. Fu, Photonic energy band structure of excitonic quantum dot dimer system, *J. Appl. Phys.*, vol. 106, p. 054302(5), 2009. Copyright 2009, American Institute of Physics)



faces. An incident electromagnetic field is initiated at one side of the dielectric layer and its transmission is detected at the other site.

We can calculate the transmission and reflection of the electromagnetic field through this dielectric layer very similar to the electron transmission through electronic structure such that we import an incident radiation, which will be reflected and transmitted. The electromagnetic field inside the dielectric layer consists of eigen states, i.e., states in Fig. 5.29 if the quantum dots in the dielectric layer form a fcc lattice. Here we try to calculate the transmission by adopting the FDTD method briefly introduced in Sect. 5.3, which requires a few steps of physical and mathematical treatments. As we learned before, the dielectric polarization of the exciton induced by the electromagnetic field is expressed as

$$\mathbf{P}(\mathbf{r}, t) = T(\omega)\psi(\mathbf{r}, \mathbf{r}) \int \psi^*(\mathbf{r}', \mathbf{r}')\mathbf{E}(\mathbf{r}', t)d\mathbf{r}' \quad (5.152)$$

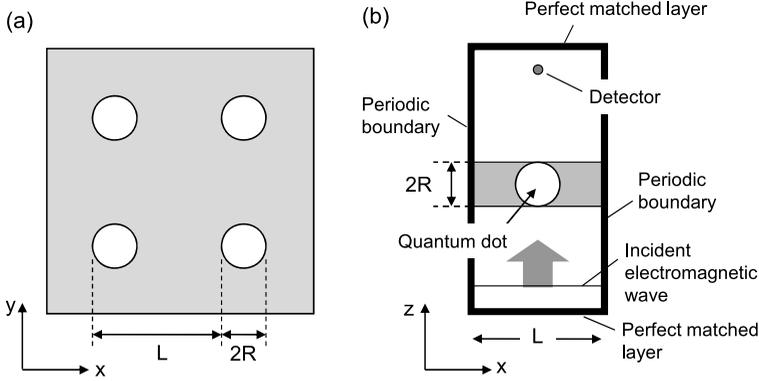


Fig. 5.30 (a) Schematic drawing of a dielectric film embedded with a square array of quantum dots. (b) A cross section of the FDTD computational domain consisting of a single unit cell of the quantum-dot array

see Eq. (3.237). where $\psi(\mathbf{r}_e, \mathbf{r}_h)|_{r_e=r_h}$ is the ground-state wave function of the exciton excited in the quantum dot centered at \mathbf{a}

$$\psi(\mathbf{r}_e, \mathbf{r}_h) = \frac{1}{|\mathbf{r} - \mathbf{a}| \sqrt{2\pi R}} \sin\left(\frac{\pi |\mathbf{r} - \mathbf{a}|}{R}\right) \frac{1}{\sqrt{\pi a_{\text{Br}}^3}} e^{-\frac{|\mathbf{r}_e - \mathbf{r}_h|}{a_{\text{Br}}}} \quad (5.153)$$

Here

$$\mathbf{r} = \frac{\mathbf{r}_e m_c^* + \mathbf{r}_h m_v^*}{m_c^* + m_v^*}$$

In addition, the coefficient $T(\omega)$ is given by

$$T(\omega) = \frac{2\pi\epsilon_0\epsilon\omega_{\text{LT}}\omega_0 a_{\text{Br}}^3}{\omega_0^2 - \omega^2 - 2i\omega\delta}. \quad (5.154)$$

Here ω_{LT} and a_{Br} are the exciton longitudinal-transverse splitting and exciton Bohr radius, respectively. ϵ is the dielectric index of the semiconductor material that makes the quantum dot, δ is a phenomenological parameter describing the decay of the quantum dot exciton ω_0 is the ground-state exciton resonance frequency of the quantum dots.

The interaction between light and nonlocal quantum dots is described by the time-dependent Maxwell's equations that are coupled to an equation for the light-induced excitonic polarization current in the quantum dot

$$\nabla \times \mathbf{E} = -\mu_0 \frac{\partial \mathbf{H}}{\partial t}, \quad \nabla \times \mathbf{H} = \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} + \mathbf{J}, \quad \mathbf{J} = \frac{\partial \mathbf{P}}{\partial t} \quad (5.155)$$

Here \mathbf{E} is the electric field, \mathbf{H} is the magnetic field, \mathbf{J} is a current density due to the nonlocal polarization \mathbf{P} of the quantum-dot exciton. To solve the above curl equations, Yee's discretization scheme is here employed. All field variables are defined

on a cubic grid. Electric and magnetic fields are temporally separated by one-half time-step and spatially interlaced by half a grid cell. Based on this scheme, center differences in both space and time are applied to approximate Maxwell's equations.

In order to obtain the relation between the current \mathbf{J} and the polarization \mathbf{P} , Eq. (5.152) is rewritten as

$$\mathbf{P}(\mathbf{r}, \omega) = \frac{A}{\omega_0^2 - \omega^2 - 2i\omega\delta} \mathbf{E}_{\text{new}}(\mathbf{r}, \omega) \quad (5.156)$$

where $A = \pi\epsilon_0\epsilon\omega_{\text{LT}}\omega_0$, and the new variable $\mathbf{E}_{\text{new}}(\omega)$ is defined as

$$\mathbf{E}_{\text{new}}(\omega) \equiv \text{sinc}\left(\frac{\pi|\mathbf{r} - \mathbf{a}|}{R}\right) \int \text{sinc}\left(\frac{\pi|\mathbf{r}' - \mathbf{a}|}{R}\right) \mathbf{E}(\mathbf{r}', \omega) \frac{d\mathbf{r}'}{R^3} \quad (5.157)$$

with $\text{sinc}(x) = \sin(x)/x$. It should be emphasized that such a dielectric permittivity is very similar to that of a Lorentz-type medium. The polarization current density $\mathbf{J}(\omega)$ is introduced as

$$\mathbf{J}(\mathbf{r}, \omega) \equiv -i\omega\mathbf{P}(\mathbf{r}, \omega) = \frac{-Ai\omega}{\omega_0^2 - \omega^2 - 2i\omega\delta} \mathbf{E}_{\text{new}}(\mathbf{r}, \omega) \quad (5.158)$$

Fourier transforming the above equation, its time-domain analog can be written as

$$\omega_0^2 \mathbf{J}(t) + 2\delta \frac{d\mathbf{J}(t)}{dt} + \frac{d^2 \mathbf{J}(t)}{dt^2} = A \frac{d\mathbf{E}_{\text{new}}(t)}{dt} \quad (5.159)$$

With the discrete time step Δt , and notation $\mathbf{J}^n \equiv \mathbf{J}(n\Delta t)$, a time-difference expression is then found

$$\mathbf{J}^{n+1} = a\mathbf{J}^n + b\mathbf{J}^{n-1} + c[\mathbf{E}_{\text{new}}^{n+1/2} - \mathbf{E}_{\text{new}}^{n-1/2}] \quad (5.160)$$

where

$$a = \frac{2 - \omega_0^2 \Delta t^2}{1 + \delta \Delta t}, \quad b = \frac{\delta \Delta t - 1}{1 + \delta \Delta t}, \quad c = \frac{A \Delta t}{1 + \delta \Delta t} \quad (5.161)$$

On the other hand, following Ampere's law,

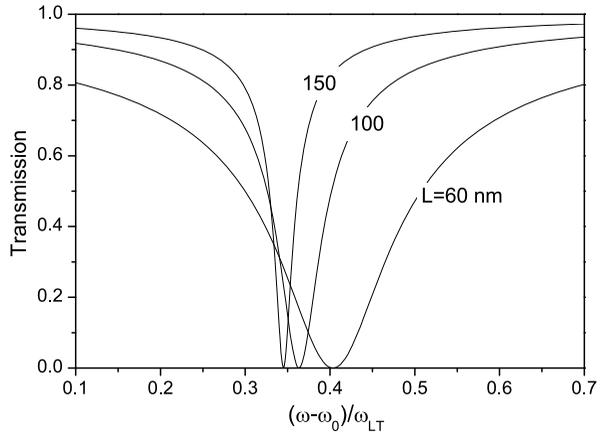
$$\nabla \times \mathbf{H} = \epsilon_0 \epsilon \frac{d\mathbf{E}(t)}{dt} + \mathbf{J} \quad (5.162)$$

the finite-difference expression can be written as

$$\mathbf{E}^{n+3/2} = \mathbf{E}^{n+1/2} + \frac{\Delta t}{\epsilon_0 \epsilon} [\nabla \times \mathbf{H}^{n+1} - \mathbf{J}^{n+1}] \quad (5.163)$$

Equations (5.160), (5.163) can therefore be utilized to simulate the nonlocal polarization of the quantum dots in two steps: (1) From \mathbf{J}^n , \mathbf{H}^n and $\mathbf{E}^{n-1/2}$ to obtain $\mathbf{E}^{n+1/2}$; (2) From $\mathbf{E}^{n-1/2}$ and $\mathbf{E}^{n+1/2}$ to obtain $\mathbf{E}_{\text{new}}^{n-1/2}$ and $\mathbf{E}_{\text{new}}^{n+1/2}$ and

Fig. 5.31 Transmission spectrum through quantum dot layer with different period L . The dielectric film is 40 nm thick (Y. Zeng, Y. Fu, M. Bengtsson, X.-S. Chen, W. Lu, and H. Ågren, Finite-difference timedomain simulations of exciton-polariton resonances in quantum-dot arrays, *Opt. Express*, vol. 16, pp. 4507–4519, 2008)



thereafter \mathbf{J}^{n+1} , meanwhile, from $\mathbf{E}^{n+1/2}$ and \mathbf{H}^n to obtain \mathbf{H}^{n+1} . Figure 5.31 shows the transmission spectrum through quantum dot embedded dielectric film of Fig. 5.30. A strong reflection is observed due to the exciton-polariton resonance close to the quantum-dot exciton state, as can be expected by referring to Fig. 5.29.

We now turn to experiments in order to verify Fig. 5.31, a multiple-layer quantum dot structure was formed, where each layer consisted of a 2 nm $\text{In}_{0.15}\text{Ga}_{0.85}\text{As}$ buffer layer, InAs quantum dots, 6 nm $\text{In}_{0.15}\text{Ga}_{0.85}\text{As}$ cap layer, and a 33 nm GaAs barrier. From atomic-force microscopic (AFM) measurement the quantum dot density was estimated to $9.4 \times 10^{10} \text{ cm}^{-2}$, and the average width and height to be 16 and 3.5 nm, respectively. Edge polished samples were used for the reflection measurement, enabling multiple reflection in the sample, and the transmitted intensity was measured at the second polished edge of the sample. An absorption layer was added to the quantum dot sample in order to minimize the signal from the light which is not reflected by the quantum dot layers but transmitted and reflected at the GaAs/air interface. An edge polished GaAs sample with identical geometry as the quantum dot sample (but the quantum dot layers and the absorption layer were omitted) was used as reference. The transmitted intensity through the quantum dot sample was divided by the transmitted intensity through the reference sample in order to account for the variation in the light source spectrum and the reflection at the first interface.

The reflection measurement was performed at room temperature with a Fourier transform spectrometer using a quartz halogen lamp as excitation source, a quartz beam splitter, and an InGaAs -detector. For the room temperature photoluminescence measurements a HeNe laser (632.8 nm) was used as the excitation source and the photoluminescence was detected by an InGaAs detector with conventional lock-in technique. The results are presented in Fig. 5.32 where a broad reflectance peak is observed at the wavelength of the photoluminescence peak. Figure 5.32 confirms well the theoretical expectation presented by Fig. 5.31.

As shown in Fig. 3.7, ϵ' of the quantum dot exciton polariton can be negative, which can be utilized for many potential applications such as left-handed photonic materials. However, a practical problem is ϵ'' , which is a measure of optical loss,

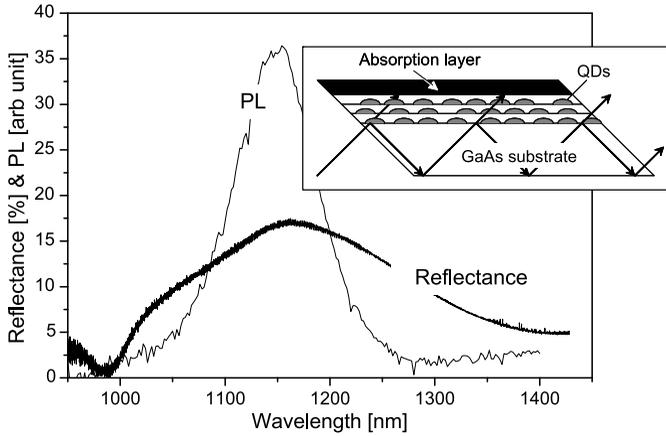


Fig. 5.32 Reflectance and photoluminescence (PL) spectra of the ten-layer InAs/InGaAs/GaAs quantum dot sample. *Inset* shows the geometry of the edge polished sample (Reprinted with permission from Y. Fu, H. Ågren, L. Höglund, J. Y. Andersson, C. Asplund, M. Qiu, and L. Thylen, Optical reflection from excitonic quantum-dot multilayer structure, Appl. Phys. Lett., vol. 93, p. 183117(3), 2008. Copyright 2008, American Institute of Physics)

see Sect. 3.7. For perfect waveguiding or imaging, we require $\epsilon'' = 0$ for which we consider distributions of two types of quantum dots, one providing an absorptive resonance and the other gain through either optical or electrical pumping. Consider a PbSe/ZnSe quantum-dot assembly immersed in a medium of dielectric constant ϵ_i . The macroscopic dielectric constant ϵ for the quantum-dot ensemble can be described by the dielectric theory of Maxwell-Garnet [35–37], which for two quantum-dot species can be written as

$$\frac{\epsilon - \epsilon_i}{\epsilon + 2\epsilon_i} = x_1 \frac{\epsilon_{\text{QD1}} - \epsilon_i}{\epsilon_{\text{QD1}} + 2\epsilon_i} + x_2 \frac{\epsilon_{\text{QD2}} - \epsilon_i}{\epsilon_{\text{QD2}} + 2\epsilon_i} \quad (5.164)$$

where x_i is the volume fraction of the i th quantum-dot species.

We consider two types of the background medium, semiconductor material with ϵ_i about 10 and polymer around 2. Furthermore, we assume a quantum-dot radius of 10 nm. Consider first the distribution of one type of quantum dots (i.e., $x_1 \neq 0$ and $x_2 = 0$), the spectra of ϵ for two quantum-dot densities and two background dielectric constants are shown in Fig. 5.33. Both ϵ' and ϵ'' increase with increasing quantum-dot density and decreasing ϵ_i . A low background dielectric constant is thus important in reducing the quantum-dot concentration.

For the combination of two types of quantum dots, one type is lossy, and the other has been pumped for optical gain, immersed in $\epsilon_i = 1.8$, the optical spectrum is presented in Fig. 5.34. Figure 5.34(a) depicts the spectra of ϵ'_{QD} and ϵ''_{QD} of the two quantum dots. Note that gain is at the higher energy, a condition we have found necessary in order to achieve $\epsilon' < 0$ and $\epsilon'' = 0$, which are given in Fig. 5.34(b). By combining lossy quantum dots with pumped quantum dots, the macroscopic

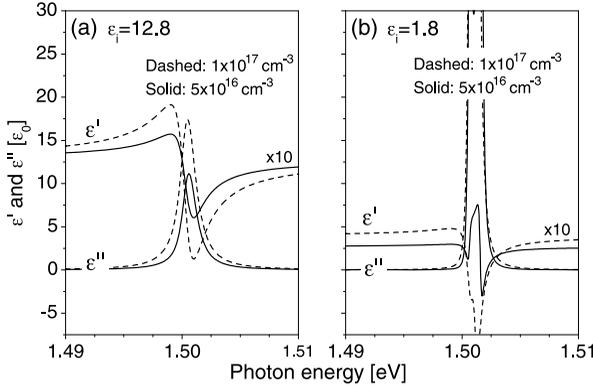


Fig. 5.33 Effective dielectric constant $\epsilon(\omega) = \epsilon'(\omega) + i\epsilon''(\omega)$ for an ensemble of quantum dots immersed in a medium of dielectric constant $\epsilon_i(\omega)$. The quantum-dot density are 5×10^{16} and $1 \times 10^{17} \text{ cm}^{-3}$, respectively, correspond to volume percentages of 20.5 % and 41 %. **(a)** Quantum dots embedded in semiconductor substrate $\epsilon_i = 12.8$. **(b)** Quantum dots immersed in a medium with $\epsilon_i = 1.8$, e.g., a conducting polymer (Reprinted with permission from Y. Fu, L. Thylen, and H. Ågren, A lossless negative dielectric constant from quantum dot exciton polaritons, *Nano Lett.*, vol. 8, pp. 1551–1555, 2008. Copyright 2008, American Chemical Society)

permittivity of our structure is approximately in the form of

$$\frac{i}{\omega - \omega_r - i\gamma} = \frac{-\gamma}{(\omega - \omega_r)^2 + \gamma^2} + i \frac{\omega - \omega_r}{(\omega - \omega_r)^2 + \gamma^2} \quad (5.165)$$

where ω_r denotes the macroscopic resonance frequency of the dimer lattice and γ is the damping rate. The loss is minimal when ω approaches ω_r . The expression follows exactly the Kramers-Kronig relations as required by the stability of the system. Such a spectrum is fundamentally different from

$$\frac{1}{\omega_r - \omega - i\gamma} = \frac{\omega_r - \omega}{(\omega_r - \omega)^2 + \gamma^2} + i \frac{\gamma}{(\omega_r - \omega)^2 + \gamma^2} \quad (5.166)$$

for which γ is to be reduced in the form of $\gamma = 0$, $\partial\gamma/\partial\omega = 0$ and $\partial^2\gamma/\partial\omega^2 > 0$ in order to become lossless, which results in a pole in the upper half-plane and hence violates causality [38]. $\epsilon' = -8.5$ and $\epsilon'' = 0$ is achieved in Fig. 5.34(b) at $\hbar\omega_r = 1.5015 \text{ eV}$ with a total quantum-dot density of $1.4 \times 10^{17} \text{ cm}^{-3}$, which is 58 % in terms of volume fraction.

Figure 5.29 is obtained under the periodic boundary condition of Eq. (5.137). In other words, the photonic dispersion relationship is rigorous for a three-dimensionally extended system, thus optical loss or gain has to be infinitely small. The quantum-dot dimer lattice in Fig. 5.34 for lossless negative dielectric constant by using the Maxwell-Garnett formalism, could amount to a mean-field approximation. The combination of the two figures indicates that at the wavelength of zero loss, the propagation of the electromagnetic field in the quantum-dot dimer lattice

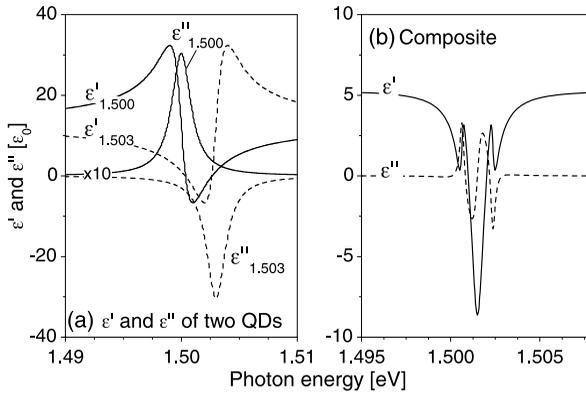


Fig. 5.34 (a) One type of quantum dots at 1.50 eV and $|a^{(0)}|^2 = 1.0$ (solid lines), the other quantum-dot type at 1.503 eV and $|a^{(0)}|^2 = 0.0$ (dashed lines). (b) The two types of quantum dots are immersed in $\epsilon_i = 1.8$. The densities of two type of quantum dots are $7 \times 10^{16} \text{ cm}^{-3}$ (Reprinted with permission from Y. Fu, L. Thylen, and H. Ågren, A lossless negative dielectric constant from quantum dot exciton polaritons, Nano Lett., vol. 8, pp. 1551–1555, 2008. Copyright 2008, American Chemical Society)

will be normal even the dielectric constant of the dimer lattice becomes effectively negative.

5.8 Resonant Tunneling Light Emitting Diode

Electron transitions occur due to the fact that electrons occupy stationary (eigen values and eigen functions of the system). When exposed to radiation, interaction between the photons of the radiation field and electrons occurs, resulting in radiative or non-radiative electron transitions between two energy levels. Absorption is a transition induced by incident photons that the electron is raised from a lower state E_1 to a higher state E_2 after absorbing a photon having an energy that equals the energy difference between the two electron states, see Fig. 5.35(a). This forms the basis for photodetection that has been studied in the previous sections. If an electron is initially in an excited state E_2 , it may fall to the lower energy state E_1 spontaneously, see Fig. 5.35(b). When doing so, the electron releases its excess energy in the form of a photon of random direction, i.e. the light emitted is incoherent. The energy of the emitted photon corresponds to the energy difference between the two energy states is

$$\hbar\omega_s = E_2 - E_1 \tag{5.167}$$

The process is called spontaneous emission because the transition is not triggered by external controlling intervention. This forms the basis for the light emitting device, which will be studied closely in this section. The third transition is the stimulated emission. Initially the electron is in an excited state. As it interacts with an incident

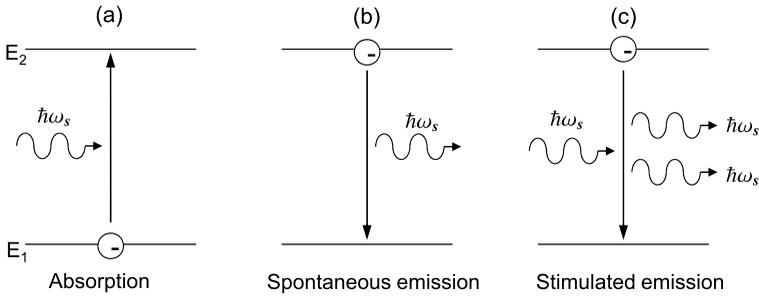


Fig. 5.35 Electron transitions between two energy levels E_1 and E_2 . (a) Absorption, (b) spontaneous emission, (c) stimulated emission. In all cases $\hbar\omega_s = E_2 - E_1$

photon, it is stimulated to emit a photon when it undergoes a downward transition, see Fig. 5.35(c). This phenomenon forms the basis of laser action that the emitted photon has exactly the same characteristics as the stimulating photon; it has the same energy, phase, polarization and direction of propagation, which will be studied in the next section.

All of the processes in Fig. 5.35 take place when the electron interacts with the radiation with appropriate energy. Which process is dominating depends on the intensity of the radiation and on whether the electron is at thermal equilibrium. At thermal equilibrium, the number of electrons, N_2 , in the excited state E_2 is smaller than the number of electrons N_1 at a lower energy state E_1 . The occupation of state E_i is in the form of Fermi-Dirac distribution

$$f(E_i) = \frac{1}{1 + \exp\left(\frac{E_i - E_f}{k_B T}\right)} \quad (5.168)$$

where E_f is the Fermi level. At this thermal equilibrium, a photon is more likely to be absorbed by electrons at the low energy levels.

For spontaneous and stimulated emissions, a population inversion is needed in which N_2 is larger than N_1 . In this situation, the total emission rate becomes larger than the absorption rate.

When a p -type and an n -type semiconductor material are brought in contact, a $p-n$ junction is formed. At thermal equilibrium, the Fermi level must be aligned across the $p-n$ junction by the diffusions of electrons from the n side to the p side and of holes in the opposite direction. When a steady state is reached, further diffusion is prevented by the so-called built-in electric field. See Sect. 4.1 about the functioning of the $p-n$ junction.

When applying an external positive voltage to the p side of the junction, the built-in field is reduced, and the junction is said to be forward biased. The electric field reduction in the depletion region induces further diffusion of electrons and holes across the junction and an electric current begins to flow. In the narrow depletion region of the junction, both electrons and holes are present simultaneously. They can recombine either radiatively or non-radiatively. The photons, that are emitted

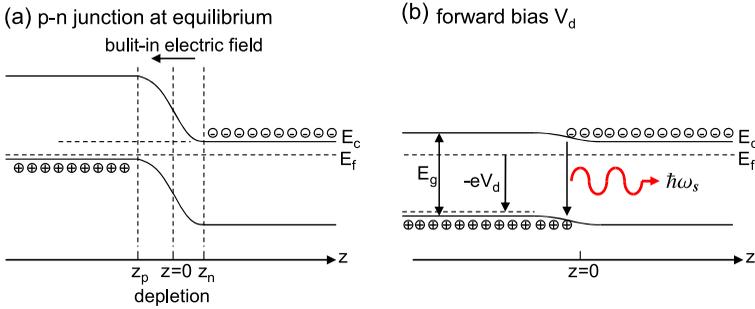


Fig. 5.36 A $p - n$ homojunction (a) at thermal equilibrium, and (b) forward biased

during radiative emission (spontaneous or stimulated emission), have an energy of $\hbar\omega \approx E_g$, where E_g is the energy bandgap, see Fig. 5.36. These photons can be absorbed through a reverse process that generates electron-hole pairs. When the external voltage exceeds a critical value a sufficient population inversion is achieved in the depletion region and the rate of photon emission exceeds that of absorption. Under these circumstances the $p - n$ junction is capable of optically emitting radiation at certain wavelength.

The $p - n$ junction is a so-called homojunction, which means that the junction is formed between p and n variants of the same semiconductor material. Because the excess electrons and holes distribute extensively in the depletion region which is relatively wide, it is thus difficult to achieve the high carrier densities in order to obtain a sufficiently high optical gain.

Moreover, the optical spectrum of the light emitted from a $p - n$ homojunction can be rather wide. Since the carrier concentrations in the depletion region is very low, let us first study the optical transitions in an intrinsic or lightly-doped bulk semiconductor from a valence-band state (E_v, \mathbf{k}, m_v^*) to a conduction-band state (E_c, \mathbf{k}, m_c^*) in the forms of

$$E_c(\mathbf{k}) = E_c + \frac{\hbar^2 k^2}{2m_c^*}, \quad E_v(\mathbf{k}) = E_v - \frac{\hbar^2 k^2}{2m_v^*} \quad (5.169)$$

where $n = c, v$, E_n is the bandedge, $E_c - E_v = E_g$ is the bandgap, and m_n^* is the effective mass of carriers in band n . The occupations of the two states can be expressed by the Fermi distribution function

$$f[E_n(\mathbf{k})] = \frac{1}{\exp\left[-\frac{E_n(\mathbf{k}) - E_f^n}{k_B T}\right] + 1} \quad (5.170)$$

E_f^n is the Fermi level of band n . At thermal equilibrium and without external excitation, $E_f^c = E_f^v = E_f$. In an n -type material, the number of conduction-band electrons is larger than the valence-band holes so that E_f is close to the conduction bandedge. E_f is close to the valence bandedge in a p -type material since the number

of valence-band holes is larger than conduction-band electrons. Under an optical excitation, electrons in the valence band will transit to the conduction band to become photoexcited electrons, leaving holes (known as photoexcited holes) in the valence band. Electrons in the conduction band and holes in the valence band will reach their respective quasi equilibrium states, characterized by quasi Fermi levels E_f^c and E_f^v . In intrinsic or lightly-doped materials, the densities of electrons and holes are all low so that E_f locates close the middle of the bandgap. Under weak photoexcitation, E_f splits into E_f^c and E_f^v and the distance between them is much smaller than E_g . Under these conditions, $E_c(\mathbf{k}) - E_f^c \gg k_B T$, and $E_f^v - E_v(\mathbf{k}) \gg k_B T$ so that the Fermi distribution can be approximated by Boltzmann distribution

$$\begin{aligned} f[E_c(\mathbf{k})] &= \exp\left[-\frac{E_c(\mathbf{k}) - E_f^c}{k_B T}\right] \\ 1 - f[E_v(\mathbf{k})] &= 1 - \frac{1}{\exp\left[-\frac{E_v(\mathbf{k}) - E_f^v}{k_B T}\right] + 1} \\ &= \frac{1}{\exp\left[-\frac{E_f^v - E_v(\mathbf{k})}{k_B T}\right] + 1} \approx \exp\left[-\frac{E_f^v - E_v(\mathbf{k})}{k_B T}\right] \end{aligned} \quad (5.171)$$

By Eq. (5.169),

$$E_c(\mathbf{k}) - E_v(\mathbf{k}) = E_g + \frac{\hbar^2 k^2}{2m_r} \equiv E \quad (5.172)$$

is the optical transition energy, $1/m_r = 1/m_c^* + 1/m_v^*$ which is normally referred to as the reduced effective mass. As mentioned before, the difference between E_f^c and E_f^v is very small under weak photoexcitation, i.e., $E_f^c - E_f^v \approx 0$ so that

$$\begin{aligned} \{1 - f[E_v(\mathbf{k})]\} f[E_c(\mathbf{k})] &= \exp\left[-\frac{E_c(\mathbf{k}) - E_f^c + E_f^v - E_v(\mathbf{k})}{k_B T}\right] \\ &\approx \exp\left(-\frac{E_g + \hbar^2 k^2 / 2m_r}{k_B T}\right) \end{aligned} \quad (5.173)$$

Insert it into the expression of optical transition

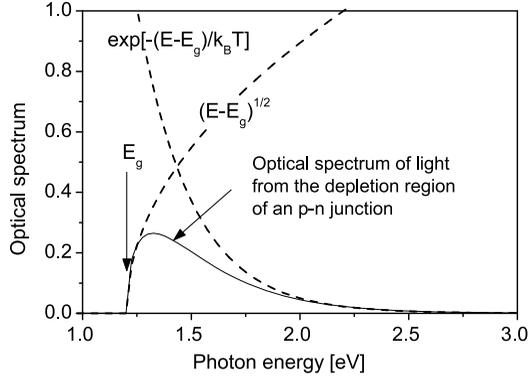
$$g(\hbar\omega) = \frac{\pi e^2 |\mathbf{e}_s \cdot \mathbf{p}_{cv}|^2}{m_0^2 \omega \epsilon} \int \delta\left[E_g + \frac{\hbar^2 k^2}{2m_r} - \hbar\omega\right] \exp\left(-\frac{E_g + \hbar^2 k^2 / 2m_r}{k_B T}\right) \frac{2d\mathbf{k}}{(2\pi)^3} \quad (5.174)$$

Moreover, by Eq. (5.172)

$$\frac{2d\mathbf{k}}{(2\pi)^3} = \frac{1}{4\pi^3} 2\pi k^2 dk = \frac{1}{2\pi^2} \left(\frac{2m_r}{\hbar^2}\right)^{3/2} \sqrt{E - E_g} dE \equiv N_3(E - E_g) dE \quad (5.175)$$

where $N_3(E)$ is the three-dimensional density of states. The condition for the validity of the above equation is $E \geq E_g$, i.e., $|\mathbf{k}|$ be real. We introduce a step function

Fig. 5.37 Optical spectrum of intrinsic or lightly-doped materials which shows that the optical bandwidth can be rather wide



$\theta(x)$ such that when $x \geq 0, \theta = 1$, otherwise $\theta = 0$. By this, the three-dimensional density of states can be expressed as

$$N_3(E - E_g) = \frac{1}{2\pi^2} \left(\frac{2m_r}{\hbar^2} \right)^{3/2} \sqrt{E - E_g} \theta(E - E_g) \quad (5.176)$$

And finally, the optical spectrum of the inter-band transition is

$$\begin{aligned} g(\hbar\omega) &= \frac{\pi e^2 |\mathbf{e}_s \cdot \mathbf{p}_{cv}|^2}{m_0^2 \omega \epsilon} \int \delta(E - \hbar\omega) \exp\left(-\frac{E - E_g}{k_B T}\right) N_3(E - E_g) dE \\ &= \frac{\pi e^2 |\mathbf{e}_s \cdot \mathbf{p}_{cv}|^2}{m_0^2 \omega \epsilon} \exp\left(-\frac{\hbar\omega - E_g}{k_B T}\right) N_3(\hbar\omega - E_g) \theta(\hbar\omega - E_g) \\ &\propto \sqrt{\hbar\omega - E_g} \exp\left(-\frac{\hbar\omega - E_g}{k_B T}\right) \theta(\hbar\omega - E_g) \end{aligned} \quad (5.177)$$

The optical spectrum of the above equation is presented in Fig. 5.37, indicating that the bandwidth can be rather wide. The reader may also refer back to the optical spectra of Fig. 3.4 which shows that the bandwidth of a three-dimensionally extended system is the widest, and the bandwidth decreases with the dimension of the electron system.

The problems can be partially solved by using heterojunctions, which are junctions formed by different semiconductor materials. A large amount of both theoretical and experimental works has been published concerning the resonant tunneling diode leading to a broad range of electric [39, 40] as well as optical [41, 42] applications. Photoluminescence characterization of the resonant-tunneling light-emitting diode (RTLED) consists of the recombination of electrons and holes that each tunnel from the opposite contact layers into the central active layer. There are two types of commonly investigated light-emitting tunneling diode.

1. Electrons or holes. If the two contact layers are both n^+ -type, the electrons are majority carriers and holes are photogenerated minority carriers (if the contacts

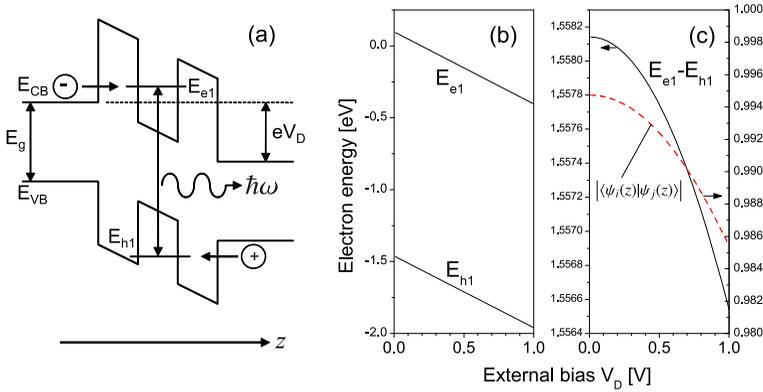


Fig. 5.38 (a) Schematic resonant tunneling light emitting diode. E_g is the energy bandgap of GaAs. (b) The first electron and hole quasi states E_{e1} and E_{h1} . (c) Transition energy $E_{e1} - E_{h1}$ and transition matrix element $\langle \psi_{e1}(z) | \psi_{h1}(z) \rangle$ as functions of the external bias V_D

are p^+ -type, the roles of electrons and holes are exchanged) [43]. The sample is prepared in the sequence of an n^+ -type (p^+ -type) GaAs substrate as the emitter, followed by an n -type (p -type) or not-intentionally doped GaAs spacer, AlAs barrier, GaAs well, AlAs barrier, n -type (p -type) or not-intentionally-doped GaAs spacer, and finally n^+ -type (p^+ -type) GaAs cap as the collector.

2. Electrons and holes. if one contact is n^+ -type and the other p^+ -type, both electrons and holes are majority carriers [44]. The emitter is an n^+ -type GaAs substrate followed by an n -type or not-intentionally-doped GaAs spacer, then AlAs barrier, GaAs well, AlAs barrier, p -type or not-intentionally-doped GaAs spacer, and finally p^+ -type GaAs cap as the collector.

The energy band structure of the resonant tunneling light-emitting diode is described by the combination of the Schrödinger and Poisson equations which was presented early when we describe the double barrier resonant tunneling diode. The general features are exactly the same as in Fig. 4.10. The idea of light emitting here is to inject electrons (holes) from the emitter (collector) to the electron (hole) quasi state in the central GaAs well between the two barriers where the injected electrons and holes will radiatively recombine to emit photons, see Fig. 5.38(a).

Let us discuss the radiative recombination of the electrons and holes in the quasi states between the two barriers. As in Sect. 4.3, let the sample growth direction be the z direction. As mentioned in the previous chapter, the wave functions of carriers in both the emitter and the collector are described by propagating waves in the z direction, and the envelope function in the xy plane is in the form of $e^{i\mathbf{k}_i \cdot \boldsymbol{\rho}}$ because of the translational symmetry there, where \mathbf{k}_i and $\boldsymbol{\rho}$ are the wave vector and coordinate in the xy plane, respectively. In the central active region including the double barriers the envelope wave function of the quasi states is denoted as $\psi_i(z)$, where i is the quantum index of the quasi states, and the total wave function

is

$$\psi_i(z)u_i(\mathbf{r})e^{i\mathbf{k}_i \cdot \boldsymbol{\rho}} \quad (5.178)$$

where $u_i(\mathbf{r})$ and \mathbf{k}_i are Bloch function and wave vector in the xy plane associated with quasi state i . $\boldsymbol{\rho}$ is the spatial coordinate of the electron (or hole) in the xy plane. Note that for all quasi states in the conduction band (valence band), $u_i(\mathbf{r})$ will be the same $u_c(\mathbf{r})$ ($u_v(\mathbf{r})$).

We have studied light-matter interaction in Sect. 3.2 where the light-matter interaction is generally described as

$$\frac{e\mathbf{A} \cdot \mathbf{p}}{m_0}$$

where \mathbf{A} is the vector field of the electromagnetic field, \mathbf{p} the momentum of the electron, and m_0 the free electron mass. Things become more complicated when studying electrons in nanostructures with photons. Following detailed steps in Chap. 3, we can obtain the following expression to describe electron and hole recombination,

$$W \propto \sum_{ij} \frac{\Gamma}{(E_j - E_i + \hbar\omega)^2 + \Gamma^2} |\langle \psi_i(z) | \psi_j(z) \rangle|^2 (f_i - f_j) \quad (5.179)$$

It is easy to understand that when wave function overlapping is zero, $W = 0$. By Fig. 4.10 we see that the wave function overlapping in the region outside the quantum well between the two barriers can then be neglected so that we only need to focus our attention on the quantum well between the two barriers.

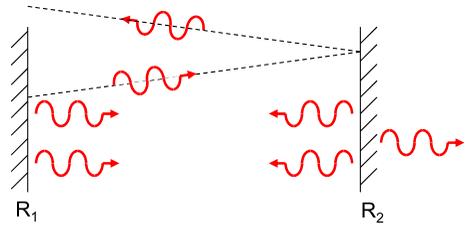
For the GaAs quantum well embedded between two AlAs barriers, the electron and hole states in the quantum well are well approximated as

$$\begin{aligned} \left[-\frac{\hbar^2}{2m_e^*} \frac{d^2}{dz^2} + V_e(z) \right] \psi_{ei}(z) &= E_{ei} \psi_{ei}(z) \\ \left[\frac{\hbar^2}{2m_h^*} \frac{d^2}{dz^2} + V_h(z) \right] \psi_{hj}(z) &= E_{hj} \psi_{hj}(z) \end{aligned} \quad (5.180)$$

where $V_e(z)$ and $V_h(z)$ are potential energies of electrons and holes, which are functions of the external bias V_D . In Figs. 5.38(b) and (c) we show the numerical results of E_{e1} , E_{h1} and the matrix element $|\langle \psi_{e1}(z) | \psi_{h1}(z) \rangle|$.

The energy difference between the ground electron and hole states slightly decreases when we increases V_D , so does the recombination probability $\langle \psi_i(z) | \psi_j(z) \rangle$. However, the modifications are relatively small. As compared with homojunction, the tunneling light-emitting diode offers a better control over the bandwidth of the emitted photons. Moreover, the photon wavelength can be easily tuned by varying the width of the central GaAs quantum well.

Fig. 5.39 Fabry-Perot plane-parallel resonator. Light perpendicular to the mirrors with reflectance R_1 and R_2 , will be repeatedly reflected within the cavity. Any inclined light will eventually escape the resonator



5.9 Nanostructure Laser

The laser structure is basically a light-emitting device, whose function principle has been discussed in the previous section. However, the function of a laser is much more complicated than a common light-emitting device, since the light from a laser is more specific. The laser, which is an acronym of “Light Amplification by Stimulated Emission of Radiation”, is an optical resonator containing a medium that amplifies light generated by stimulated emission. The laser beams have several unique features compared with other light beams:

1. they produce intense coherent monochromatic induced radiation;
2. the incoherent spontaneous radiation component in the beams, called noise, is negligible;
3. the beams can be highly collimated;
4. the beams can be focused very sharply.

A laser consists principally of three key elements. One is the laser medium that generates and amplifies the light. The second is the power supply, which delivers energy to the laser medium to excite light emission. The third is an optical cavity, or resonator, which acts as a spectral filter and concentrates the light to the laser medium.

Laser amplification can take place in a great variety of media including solids, gases, liquids and plasmas. What material is used depends mainly on which wavelengths to be produced and for which application. To obtain laser radiation, stimulated emission must occur in the laser medium. This means that if a radiation with appropriate energy passes through the material, the radiation that leaves, as a result of stimulated emission, has more photons than the incident radiation. The result is thus a coherent amplification of radiation of this frequency, since the stimulated emission has the same characteristics as the stimulating radiation, i.e., the same frequency, phase, polarization and direction of propagation.

When the medium, in which a population inversion has been created, is placed in an optical cavity (resonator), radiations at certain resonance frequencies will be confined and amplified in the cavity. There are several construction configurations for laser resonators. The simplest is the plane-parallel or Fabry-Perot resonator, in which two flat mirrors are placed at opposite ends of the cavity aligned parallel to each other and perpendicular to the cavity axis, see Fig. 5.39. One of the mirrors, or both, is slightly transparent so that part of the radiation can transmit through it.

Light rays perpendicular to the mirrors are reflected back and forth without escaping, but if they are inclined they will eventually escape. Most of the radiation that comes from spontaneous emission and thus travels in all directions will therefore escape from the open sides of the cavity. The exception is the radiation that travels along the direction normal to the cavity mirrors. This radiation will be amplified until the power emitted as laser radiation is balanced by the power input to sustain the population inversion. Since the amplification occurs only for one specific direction, the emitted radiation is confined to a very narrow, coherent and uni-directional beam, so that the unique features of a laser beam are fulfilled.

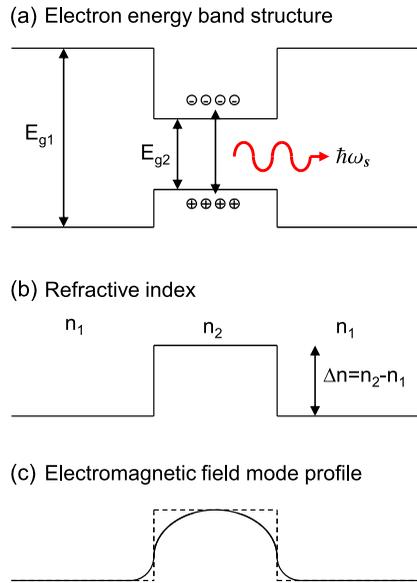
Semiconductor laser was invented 1962. Today it is widely used in a broad spectrum of applications. It serves a central role in fiber-optic communication and data storage, primarily because of major advances in laser power, spectral purity, efficiency, wavelength range, ability to be modulated at extremely rapid rates and long lifetimes despite its pinhead size.

In the so-called double heterojunction structure, a thin layer of a semiconductor material with a narrower bandgap is grown between the p -type and n -type layers. Because of the bandgap difference among the layers, the injected charge carriers are confined within the thin layer of the narrower bandgap material. This structure makes it easier to obtain sufficient inversion of the carrier population in order to produce optical gain by the stimulated radiative recombination. Because of the light generation within the thin narrow-bandgap layer, this layer is normally denoted as the active layer, whereas the surrounding layers of larger-bandgap materials are called cladding layers. The refractive index of the active layer is larger than that of the cladding layers because of its narrower bandgap. Due to the refractive index difference, the active layer acts as a wave guide. The light emitted parallel to the active layer will be guided along the active layer, so that the light intensity will be the highest in the middle of the active layer, where the gain is the largest. This mechanism is often referred as the index guiding. Consequently, the use of a double heterojunction structure results in both the carrier and the optical confinement, see Fig. 5.40.

As a result of the bandgap difference, electrons that are injected from the n -type region during the forward bias is prevented from diffusing beyond the energy barrier at the opposite heterojunction. Similarly, holes injected from the p side can not diffuse beyond their opposite heterojunction. The charge carriers will thus be trapped in the active region where they recombine either radiatively or non-radiatively. The light emitted along the active layer will be guided along the layer because of the difference in refractive index, and a spatially confined beam of radiation is created. Because the semiconductor laser is based on a diode structure, it is often called a diode laser.

The heterojunction structure is crucially important for the development of semiconductor lasers. The first person who successfully produced a heterojunction structure with clear borders among layers was a Russian professor Zhores I. Alferov in 1969. In 1970 the research group submitted its report about a double heterostructure laser operating at room temperature. Alferov and a German professor Herbert Kroemer were awarded the Nobel Prize in Physics 2000, for the development of semiconductor heterostructures for high-speed electronics and optoelectronics.

Fig. 5.40 Simultaneous confinements of charge carriers and generated light by the use of a double heterojunction structure. The active layer having a narrower bandgap has a higher refractive index than the cladding layers



A buried heterostructure semiconductor laser consists of a core active layer with cladding material around it. Normally the two materials are lattice-matched to avoid strain, since strain causes defects and dislocations which cause nonradiative recombinations. The cladding material has a wide bandgap and a low refractive index than the active layer in order to confine carriers and light respectively to the active layer.

In Fig. 5.41 we show a combination of InGaAsP active layer which is lattice-matched to the InP cladding. When we forward bias the $p-i-n$ structure it behaves a little like a forward biased $p-n$ junction with a hole in the middle where the carriers are trapped for recombination. It is much due to the confinement of the carriers in the i region that makes it possible to get inversion and net gain at a moderate forward bias.

The resonant-tunneling-injection hot-electron laser was proposed by Tolstikhin and Willander [45]. It is just a resonant-tunneling hot-electron transistor, with an ac-

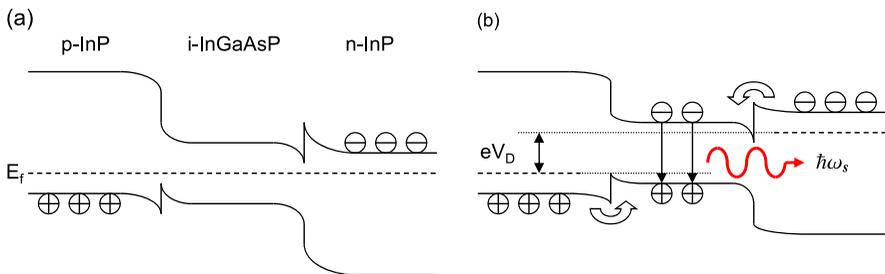
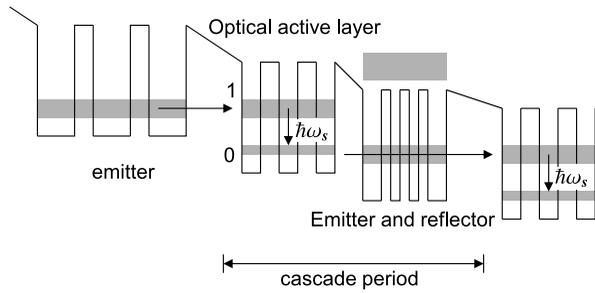


Fig. 5.41 Energy band structure of InP/InGaAsP/InP (a) without external bias and (b) with a bias V_D

Fig. 5.42 Schematic conduction band profile of a portion of the GaInAs/AlInAs quantum cascade laser under positive bias



tive layer and a waveguiding structure inserted between the base and collector. The basic idea is that we shall be able to control both the number of injected electrons via V_{be} , the base-emitter bias (which is proportional to the intensity of the optical field) and the kinetic energy that the electrons carry into the active layer via V_{cb} , the collector-base bias (the output optical energy is the sum of this kinetic energy and the energy bandgap).

In a quantum cascade laser, the population inversion between states of laser action is designed by tailoring the electron intersubband transition [46]. Since the initial report of the quantum cascade laser in 1994, the emission wavelengths in the 4–8.5 μm range have been demonstrated using AlInAs/GaInAs heterostructures [47]. Figure 5.42 schematically represents the energy band profile of a quantum cascade laser. To prevent electron escape into the continuum, which will largely reduce the vertical transition, a superlattice is placed as an effective reflector for electrons in the excited state while simultaneously ensuring a swift escape from the low state to the high state in the next cascade period via the miniband.

More recently, improvement of the quantum cascade laser has been reported with a thin barrier, inserted between the injector layer and the optical active region, to increase tunneling injection efficient. This improvement and substituting the AlInAs cladding layers with InP layer of higher thermal conductivity, have led to room-temperature high peak power of 200 nW pulse-operation at 5.2 μm [48]. Continuous wave single-mode operation was also reported up to 140 K, as well as room-temperature pulse-operation at 8.5 μm . This is the first semiconductor laser operating at room-temperature in the mid-infrared regime. The current performance of the quantum cascade laser makes them a candidate for applications such as environmental sensing and pollution monitoring in the 3–5 and 8–13 μm atmospheric windows.

5.10 Light Emission from Highly Strained $\text{In}_{0.3}\text{Ga}_{0.7}\text{As}/\text{GaAs}$ Quantum Wells by Dipole δ Doping

GaAs-based telecom lasers emitting at 1.3 and 1.55 μm , aiming at replacing the commercial InP-based counterparts, have attracted great attention. The common

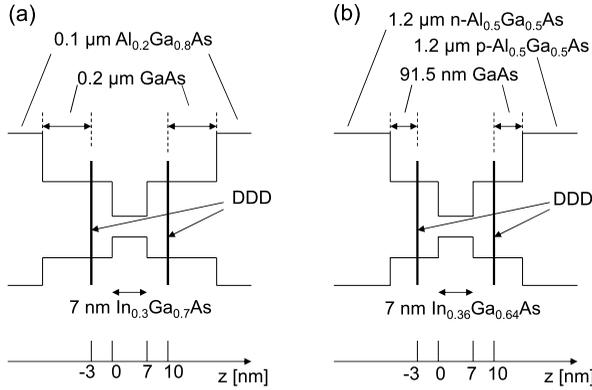


Fig. 5.43 (a) Schematic dipole δ doping (DDD) structure for photoluminescence measurement. Except the two DDD layers, the sample is basically intrinsic. (b) DDD laser structure (Reprinted with permission from Y. Fu, S.-M. Wang, X.-D. Wang, and A. Larsson, Redshift of the light emission from highly strained $\text{In}_{0.3}\text{Ga}_{0.7}\text{As}/\text{GaAs}$ quantum wells by dipole d-doping, *J. Appl. Phys.*, vol. 98, p. 043501(5), 2005. Copyright 2005, American Institute of Physics)

strategy is to find gain materials like InGaAs quantum dots or $\text{GaInNAs}(\text{Sb})$ quantum wells. One way to reduce the interband transition energy below the bandgap of the host materials is the dipole δ doping (DDD) because of the Stark shift induced by the internal field of the δ -doped layers [49, 50]. Figures 5.43 show schematically the sample structures, which contains a 7-nm-thick $\text{In}_{0.3}\text{Ga}_{0.7}\text{As}$ quantum well embedded in the center of the 0.4- μm -thick GaAs barrier further sandwiched between two 0.1- μm -thick $\text{Al}_{0.2}\text{Ga}_{0.8}\text{As}$ barriers. The n - and p -type δ -doping are placed 3 nm from each side of the InGaAs quantum well, respectively, using Si and Be as dopants.

The spatial distributions of dopants in the two δ -doped layers are modeled as

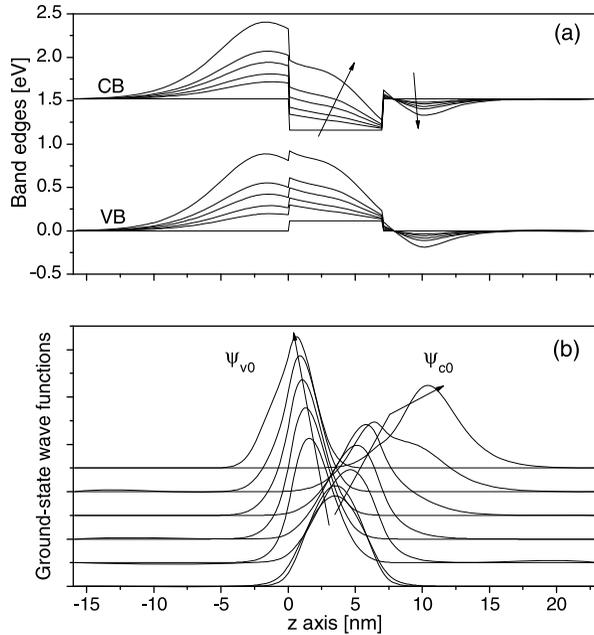
$$N(z, t) = \frac{N_{D,A}}{2\sqrt{\pi}L_D} \exp\left[-\frac{(z - z_{D,A})^2}{4L_D^2}\right] \quad (5.181)$$

where N_D and N_A are sheet densities of dopants, $z_A = -3$ and $z_D = 10$ nm. The diffusion coefficient is

$$D = D_0 \exp\left(-\frac{E_a}{k_B T}\right) \quad (5.182)$$

For Be, $E_a = 1.95$ eV and $D_0 = 2 \times 10^{-5}$ cm^2/s , and for Si, $E_a = 2.45$ eV, $D_0 = 4 \times 10^{-4}$ cm^2/s [49]. Since the growth temperature of the DDD layers in the samples is $T = 580$ $^\circ\text{C}$, it is easy to obtain that $L_D = 2.50$ nm for Be and 0.38 nm for Si. Because of the thin layers between the δ -doped layers and the InGaAs quantum well, it is expected that the carriers from dopants are transferred to the quantum well region. The electric fields and Coulombic potential energies in the region of $z > 10$ and $z < -3$ nm are zero, which are the boundary conditions for the Poisson equation to be self-consistently calculated with Schrödinger equation

Fig. 5.44 Conduction-band (CB) and valence-band (VB) structures and wave functions of electron and hole ground states. The arrows indicate the increase of N_D (N_A) from 0.0 to 2.0, 3.0, 4.5 and 6.0×10^{12} (0.0 to 10.0, 15.0, 22.5, and $30.0 \times 10^{12} \text{ cm}^{-2}$). The distribution profiles of the dopants are approximated as rectangular having a width of only 5 Å (Reprinted with permission from Y. Fu, S.-M. Wang, X.-D. Wang, and A. Larsson, Redshift of the light emission from highly strained $\text{In}_{0.3}\text{Ga}_{0.7}\text{As}/\text{GaAs}$ quantum wells by dipole d-doping, *J. Appl. Phys.*, vol. 98, p. 043501(5), 2005. Copyright 2005, American Institute of Physics)



(see Sect. 1.6). It can be expected that the spatial distributions of the dopants at $z_{D,A}$ are unimportant to the numerical solution of the Poisson equation as long as the spreading of the dopants is small compared with the distance of 3 nm between the doping layer and the InGaAs quantum well. Figure 5.44 shows that due to the high Be concentration, the quantum well is effectively *p*-type. By increasing the doping level, an electric field is induced in the InGaAs quantum well region, so that the energy separation between the electron and the hole ground states decreases. At the same time, the overlap between the wave functions of the two sublevels decreases, as can be expected in the Stark effect.

We now modify the sample into laser structures by placing the DDD active layers between *p*- and *n*-type cladding layers, and applying an external bias across the whole system. The active $\text{In}_{0.36}\text{Ga}_{0.64}\text{As}$ quantum well was placed between two 91.5-nm GaAs confined layers, which are positioned between $z = -3$ nm and the *p*-type cladding layer, and the other from $z = 10$ nm to the *n*-type cladding layer. The two *p*- and *n*-cladding layers were 1.2- μm -thick $\text{Al}_{0.5}\text{Ga}_{0.5}\text{As}$. Finally, 100-nm-thick highly Be-doped GaAs was grown as a contact layer. The doping levels in the cladding layers are $5.0 \times 10^{17} \text{ cm}^{-3}$. The laser structure is schematically represented in Fig. 5.43(b).

Figure 5.45 shows that the electric field induced by the external bias across the InGaAs quantum well partially cancels the electric field induced by the DDD. The reduced total electric field in the quantum well decreases the transition wavelength redshift. At a bias of 0.67 V, the transition energy shift is 0.298 eV. At a bias of 0.92 V, it becomes 0.201 eV. Figure 5.45(c) presents the optical transition energy shift and wave function overlaps as functions of the external bias. The redshift is

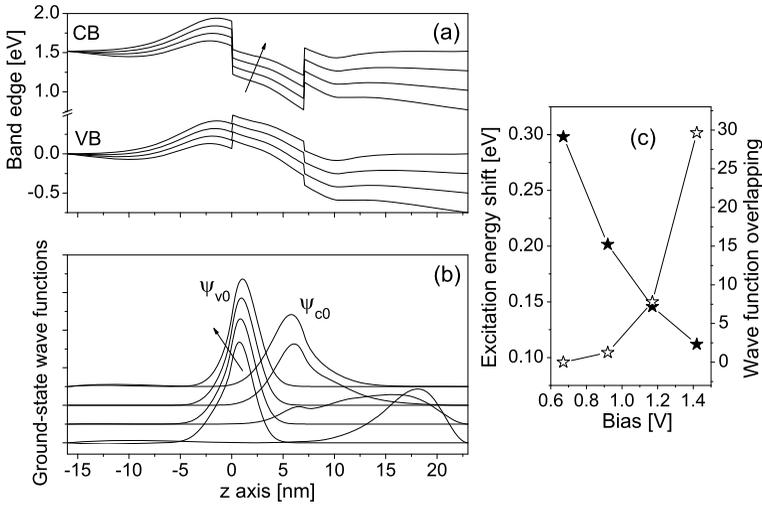


Fig. 5.45 Bandedge profiles and ground-state wave functions at different biases. The *arrows* indicate the increase of the bias V_D from 0.67 to 0.92, 1.17 and 1.42 V. (c) Optical transition energy shift and wave-function overlaps as functions of the external bias. $N_D = 4.5 \times 10^{12} \text{ cm}^{-2}$ and $N_A = 5N_D$ (Reprinted with permission from Y. Fu, S.-M. Wang, X.-D. Wang, and A. Larson, Redshift of the light emission from highly strained $\text{In}_{0.3}\text{Ga}_{0.7}\text{As}/\text{GaAs}$ quantum wells by dipole d-doping, *J. Appl. Phys.*, vol. 98, p. 043501(5), 2005. Copyright 2005, American Institute of Physics)

much reduced by the external bias. However, the overlap between the ground-state wave functions, i.e., the optical transition intensity, is enhanced at the same time.

Figure 5.46 shows the below-threshold electroluminescence signals measured from the broad area laser. Together with the electroluminescence spectrum from a similar laser structure which is not dipole δ -doped (dotted curve in Fig. 5.46), the electroluminescence spectra shows a much reduced redshift in the wavelength, as compared with the interband transition energy modification shown in Fig. 5.45(c).

5.11 Quantum Dot Biomarker

The fundamental cause of the astonishing properties of nanostructures we have studied thus far is the quantum confinement. In a three-dimensionally quantum dot where the quantum dot radius is smaller than the exciton Bohr radius, the electron and hole are forced to stay close to each other. Such an electron-hole pair, i.e., the exciton in a quantum dot radiatively recombines efficiently at room temperature, which is only observable in bulk materials at low temperature.

As we learned in Sect. 3.8, exciton in quantum dot can be optically excited by light with basically any photon energy $\hbar\omega$, while the radiative recombination of the exciton is characterized by the transition energy $\hbar\omega_{\text{QD}}$ between the ground conduction- and valence-band states in the quantum dot. Furthermore, quantum dot

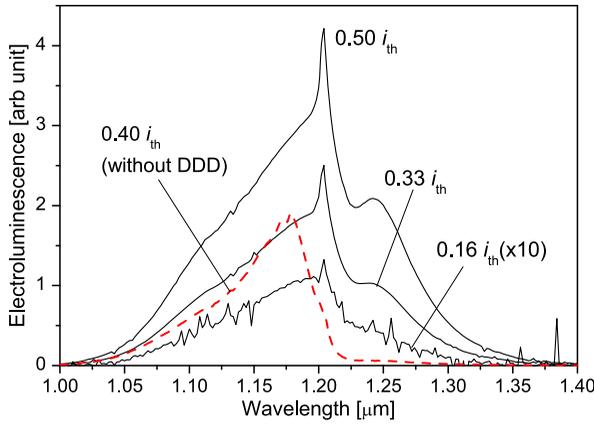
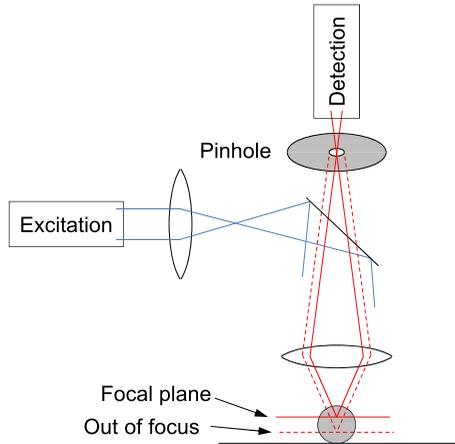


Fig. 5.46 Experimental electroluminescence spectra. $N_D = 7.5 \times 10^{12} \text{ cm}^{-2}$ ($N_A = 5N_D$). The dotted electroluminescence spectrum comes from a similar laser structure which is not double δ -doped (Reprinted with permission from Y. Fu, S.-M. Wang, X.-D. Wang, and A. Larsson, Red-shift of the light emission from highly strained $\text{In}_{0.3}\text{Ga}_{0.7}\text{As}/\text{GaAs}$ quantum wells by dipole ddoping, *J. Appl. Phys.*, vol. 98, p. 043501(5), 2005. Copyright 2005, American Institute of Physics)

Fig. 5.47 Schematic diagram of the confocal principle



is normally about tens nm in diameter so that even the wavelength of the emitted photon due to the exciton radiative recombination is in the visible range in many applications, i.e., 400 ~ 700 nm, and the spatial size of an optical imaging of the quantum dot due to the exciton radiative recombination is in the similar range, we can theoretically identify the exact location of the quantum dot to the center of its optical imaging, i.e., down to tens nm, reaching a super imaging resolution. Quantum dot has therefore been studied and developed vastly since late 1990’s as an optical contrast element.

The most common used imaging technique is the confocal microscope, see Fig. 5.47, which eliminates out-of-focus information by means of a confocal “pin-

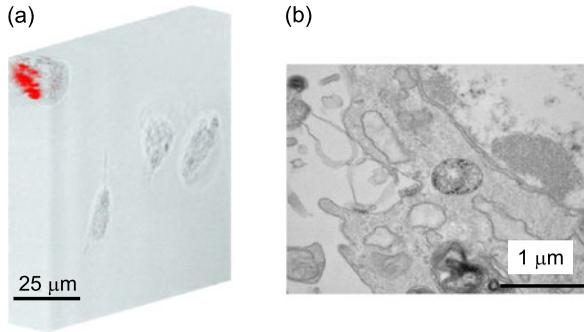


Fig. 5.48 (a) Reconstructed three-dimensional confocal imaging. A cross section of an endothelial progenitor cell is shown in the *upper left corner*. Quantum dots (*red*) are located in the middle of the EPC surrounded by the cell membrane. (b) Transmission-electron-microscopy image of quantum dots inside of membrane-boarded compartments of the cytoplasmic region

hole” situated in front of the detection which acts as a spatial filter and allows only the in-focus portion of the light to be detected and imaged. Light from above and below the plane of focus, i.e., the confocal plane, of the object is eliminated from the final image. A three-dimensional image can thereafter be easily re-constructed from a series of confocal images.

Figure 5.48(a) shows a reconstructed three-dimensional confocal imaging of colloidal CdSe-based biomarkers in endothelial progenitor cells (EPCs). Each CdSe quantum dot was coated with one monolayer CdS shell. The CdSe-CdS core-shell quantum dot has a nominal diameter 5.5 nm and an emission wavelength of 625 nm. EPCs are heterogeneous groups of endothelial cell precursors which are circulating in the blood vessel. These cells play an important role in atherogenesis and cardiovascular regeneration [51, 52]. The EPC cells were incubated with quantum dots (for more details about the bio conjugation between quantum dots and EPC cells, see Ref. [53]). Figure 5.48(a) shows that the red quantum dots (emission wavelength 625 nm) are up-taken by the EPC cells, they locate in the middle of the cell surrounded by the cell membrane. Figure 5.48(b) is a transmission-electron-microscopy image of quantum dots inside of membrane-boarded compartments of the cytoplasmic region.

References

1. Lee G, Chun SK, Wang KL (1993). In: 51th annual device research conference, Santa Barbara, CA
2. Fu Y, Chao KA (1989) Subband structures of GaAs/AlGaAs multiple quantum wells. *Phys Rev B* 40:8349–8356
3. Andersson JY, Lundqvist L (1992) Grating coupled quantum well infrared detectors: theory and performance. *J Appl Phys* 71:3600–3610
4. Pan D, Li JM, Li YP, Kong MY (1996) Long period two-dimensional gratings for 8–12 μm quantum well infrared photodetectors. *J Appl Phys* 80:7069–7071

5. Andersson JY, Lundqvist L (1991) Near-unity quantum efficiency of AlGaAs/GaAs quantum well infrared detectors using a waveguide with a doubly periodic grating coupler. *Appl Phys Lett* 59:857–859
6. Andersson JY, Lundqvist L, Paska ZF (1991) Quantum efficiency enhancement of AlGaAs/GaAs quantum well infrared detectors using a waveguide with a grating coupler. *Appl Phys Lett* 58:2264–2266
7. Andersson JY, Lundqvist L, Paska ZF, Borglind J, Haga D (1993) Efficiency of grating coupled AlGaAs/GaAs quantum well infrared detectors. *Appl Phys Lett* 63:3361–3363
8. Goossen KW, Lyon SA, Alavi K (1988) Grating enhancement of quantum well detector response. *Appl Phys Lett* 53:1027–1029
9. Goossen KW, Lyon SA (1988) Performance aspects of a quantum well detector. *J Appl Phys* 63:5149–5153
10. Yu LS, Li SS, Wang YH, Kao YC (1992) A study of the coupled efficiency versus grating periodicity in a normal incident GaAs/AlGaAs multi-quantum-well infrared detector. *J Appl Phys* 72:2105–2109
11. Levine BF, Choi KK, Bethea CG, Walker J, Malik RJ (1987) New 10 μm infrared detector using intersubband absorption in resonant tunneling GaAlAs superlattices. *Appl Phys Lett* 50:1092–1094
12. Levine BF (1993) Quantum-well infrared photodetectors. *J Appl Phys* 74:R1–R81
13. Kishino K, Arai S (1994) Integrated lasers. In: *Handbook of semiconductor lasers and photonic integrated circuits*. Chapman & Hall, London, p 350, Chap. 11
14. Stover JC (1990) *Optical scattering: measurement and analysis*. McGraw-Hill, New York, p 51
15. Cowley J (1995) *Diffraction physics*. Elsevier, Amsterdam, p 11
16. Choi K-K, Forrai DP, Endres DW, Sun J (2009) *IEEE J Quantum Electron* 45:1255
17. Yee KS (1966) Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media. *IEEE Trans Antennas Propag* 14:302–307
18. Taflov A, Hagness SC (2000) *Computational electrodynamics: the finite-difference time-domain method*, 2nd edn. Artech House, Boston
19. Chen Z-H, Hellström S, Yu Z-Y, Qiu M, Fu Y (2012) Time-resolved photocurrents in quantum well/dot infrared photodetectors with different optical coupling structures. *Appl Phys Lett* 100 043502
20. Fu Y, Willander M, Liu X-Q, Lu W, Shen SC, Tan HH, Jagadish C, Zou J, Cockayne DJH (2001) Optical transition in infrared photodetector based on V-groove $\text{Al}_{0.5}\text{Ga}_{0.5}\text{As}/\text{GaAs}$ multiple quantum wire. *J Appl Phys* 89:2351–2356
21. Ebbesen TW, Lezec HJ, Ghaemi HF, Thio T, Wolff PA (1998) Extraordinary optical transmission through sub-wavelength hole arrays. *Nature* 391:667
22. Bethe HA (1944) Theory of diffraction by small holes. *Phys Rev* 66:163
23. Posani KT, Tripathi V, Annamalai S, Weisse-Bernstein NR, Krishna S, Perahia R, Crisafulli O, Painter OJ (2006) Nanoscale quantum dot infrared sensors with photonic crystal cavity. *Appl Phys Lett* 88 151104
24. Shenoia RV, Ramirez DA, Sharma Y, Attaluria RS, Rosenbergb J, Painterb OJ, Krishna S (2007) Plasmon assisted photonic crystal quantum dot sensors. *Proc SPIE* 6713, 67130
25. Chang C-Y, Chang H-Y, Chen C-Y, Tsai M-W, Chang Y-T, Lee S-C, Tang S-F (2007) Wavelength selective quantum dot infrared photodetector with periodic metal hole arrays. *Appl Phys Lett* 91:163107
26. Lee SC, Krishna S, Brueck SRJ (2009) Quantum dot infrared photodetector enhanced by surface plasma wave excitation. *Opt Express* 17:23160–23168
27. Tsaor BY, Chen CK, Marino SA (1991) Long wavelength $\text{Ge}_x\text{Si}_{1-x}/\text{Si}$ heterojunction infrared detectors and 400×400 element image arrays. *IEEE Electron Device Lett* 12:293–296
28. Lin TL, Ksendzov A, Dejewski SM, Jones EW, Fathauer RW, Krabach TN, Maserjian J (1991) SiGe/Si heterojunction internal photoemission long-wavelength infrared detectors fabricated by molecular beam epitaxy. *IEEE Trans Electron Devices* 38:1141–1144

29. Van de Walle CG, Martin RM (1986) Theoretical calculations of heterojunction discontinuities in the Si/Ge system. *Phys Rev B* 34:5621–5634
30. Jain SC, Hayes W (1991) Structure, properties and applications of $\text{Ge}_x\text{Si}_{1-x}$ strained layers and superlattices. *Semicond Sci Technol* 6:547–576
31. Bean JC (1992) Silicon-based semiconductor heterostructures: column IV bandgap engineering. *Proc IEEE* 80:571–581
32. Jain SC, Poortmans J, Nijs J, Van Mieghem P, Mertens RP, Van Overstraeten R (1992) Band offsets in heavily doped *p*-type GeSi/Si(100) strained layers: applications to design of long-wavelength infrared (LWIR) detectors. *Microelectron Eng* 19:439–442
33. Jain SC, Roulston DJ (1991) A simple expression for bandgap narrowing (BGN) in heavily doped Si, Ge, GaAs and $\text{Ge}_x\text{Si}_{1-x}$ strained layers. *Solid-State Electron* 34:453–465
34. Shockley W, Queisser HJ (1961) *J Appl Phys* 32:510
35. Maxwell-Garnett JC (1906) Colours in metal glasses, in metallic films, and in metallic solutions. II. *Philos Trans R Soc Lond* 205:237–288
36. Cohen RW, Cody GD, Coutts MD, Abeles B (1973) Optical properties of granular silver and gold films. *Phys Rev B* 8:3689–3701
37. Gittleman JI, Abeles B (1977) Comparison of the effective medium and the Maxwell-Garnett predictions for the dielectric constants of granular metals. *Phys Rev B* 15:3273–3275
38. Stockman MI (2007) *Phys Rev Lett* 98:177404
39. Sun JP, Mains RK, Chen WL, East JR, Haddad GI (1992) C–V and I–V characteristics of quantum well varactors. *J Appl Phys* 76:2340–2346
40. Capasso F, Sen S, Beltram F, Lunardi LM, Vengurlekar AS, Smith PR, Shah NJ, Malik RJ, Cho AY (1989) Quantum functional devices: resonant-tunneling transistors, circuits with reduced complexity, and multiple valued logic. *IEEE Trans Electron Devices* 36:2065–2082
41. England PE, Golub JE, Florez LT, Harbison JP (1991) Optical switching in a resonant tunneling structure. *Appl Phys Lett* 58:887–889
42. White CRH, Skolnick MS, Eaves L, Leadbeater ML (1991) Electroluminescence and impact ionization phenomena in a double-barrier resonant tunneling structure. *Appl Phys Lett* 58:1164–1166
43. Van Hoof C, Borghs G, Goovaerts E (1991) Optical detection of light- and heavy-hole resonant tunneling in *p*-type resonant tunneling structures. *Appl Phys Lett* 59:2139–2141
44. Van Hoof C, Genoe J, Mertens R, Borghs G, Goovaerts E (1992) Electroluminescence from bipolar resonant tunneling diodes. *Appl Phys Lett* 60:77–79
45. Tolstikhin VI, Willander M (1995) Resonant tunneling injection hot electron laser: an approach to picosecond gain-switching and pulse generation. *Appl Phys Lett* 67:2684–2686
46. Faist J, Capasso F, Sirtori C, Sivco DL, Hutchinson AL, Cho AY (1994) Quantum cascade laser. *Science* 264:553–556
47. Sirtori C, Faist J, Capasso F, Sivco DL, Hutchinson AL, Chu SN, Cho AY (1996) Continuous wave operation of midinfrared (7.4–8.6 μm) quantum cascade lasers up to 110 K temperature. *Appl Phys Lett* 68:1745–1747
48. Faist J, Capasso F, Sirtori C, Sivco DL, Baillargeon JN, Hutchinson AL, Chu SN, Cho AY (1996) High power midinfrared ($\lambda \sim 5 \mu\text{m}$) quantum cascade lasers operating above room temperature. *Appl Phys Lett* 68:3680–3682
49. Schubert EF (1996) *Delta-doping of semiconductors*. Cambridge University Press, Cambridge
50. Wang SM, Zhao QX, Wang XD, Wei YQ, Sadeghi M, Larsson A (2004) 1.3 to 1.5 μm light emission from InGaAs/GaAs quantum wells. *Appl Phys Lett* 85:875–877
51. Hirschi KK, Ingram DA, Yoder MC (2008) Assessing identity, phenotype, and fate of endothelial progenitor cells. *Arterioscler Thromb Vasc Biol* 28:1584–1595
52. Fadini G, Baesso I, Albiero M, Sartore S, Agostini C, Avogaro A (2008) Technical notes on endothelial progenitor cells: ways to escape from the knowledge plateau. *Atherosclerosis* 197:496–503
53. Molnár M, Fu Y, Friberg P, Chen Y (2010) Optical characterization of colloidal CdSe quantum dots in endothelial progenitor cells. *J Nanobiotechnol* 8:2. doi:[10.1186/1477-3155-8-2](https://doi.org/10.1186/1477-3155-8-2)

Chapter 6

Numerical Recipes

Abstract In this chapter we present key FORTRAN codes that calculate and simulate fundamental physical properties of electrons in nanostructures based on Bloch theorem and Schrödinger equation for the envelope function of electrons presented in Chap. 1. We first calculate the Fermi level as a function of the doping concentration, then the transmission of an electron wave through a potential barrier. Numerical calculations of localized states in one-, two-, and three-dimensionally confined nanostructures are presented. Finally we present the code to simulate time-dependent wave packet transmission through nanostructures. All these codes are used extensively through the book to study and discuss electron transport and light-matter interactions in nanostructures.

6.1 Fermi-Dirac Integral

The occupation of conduction-band state \mathbf{k} by a conduction-band electron is given by the Fermi-Dirac distribution, and the concentration of free carriers in the conduction band can be calculated by

$$n = \int \frac{1}{e^{[E(\mathbf{k})-E_f]/k_B T} + 1} \frac{2d\mathbf{k}}{(2\pi)^3} \quad (6.1)$$

where E_f is the Fermi energy. The integration is limited to the first Brillouin zone.

We consider electrons in a single parabolic conduction band

$$E(\mathbf{k}) = E_c + \frac{\hbar^2 k^2}{2m_c^*} \quad (6.2)$$

where E_c is the conduction bandedge and m_c^* is the effective mass of the electron in the conduction band. By using the above energy dispersion relationship, denoting $x = E/k_B T$ as the electron energy in unit of $k_B T$, $\eta = E_f/k_B T$ as the Fermi level in unit of $k_B T$, E_c as the energy reference, it is easy to obtain

$$n = 4\pi \left(\frac{2m_c^* k_B T}{h^2} \right)^{3/2} \int_0^\infty \frac{x^{1/2} dx}{e^{x-\eta} + 1} = \frac{2N_{3D}}{\sqrt{\pi}} F_{1/2}(\eta) \quad (6.3)$$

$$N_{3D} = 2 \left(\frac{2\pi m_c^* k_B T}{h^2} \right)^{3/2}, \quad F_{1/2}(\eta) = \int_0^\infty \frac{x^{1/2} dx}{e^{x-\eta} + 1}$$

where N_{3D} is the three-dimensional density of states in the conduction band. $F_{1/2}(\eta)$ is the Fermi-Dirac integral of order of 1/2.

It is known that the Fermi-Dirac integral cannot be integrated analytically. It is only for the two extreme cases, namely, for $\eta \ll -1$ and $\eta \gg 1$ that the integral can be expressed in the form of a rapidly converging series:

$$F_{1/2}(\eta) = \begin{cases} \frac{\sqrt{\pi}}{2} \exp[\eta(1 - 2^{-3/2}e^\eta + \dots)] & \text{if } \eta \ll -1 \\ \frac{2}{3}\eta^{3/2}(1 + 0.125\pi^2\eta^{-2} + 0.267\eta^{-4} + \dots) & \text{if } \eta \gg 1 \end{cases} \quad (6.4)$$

It was shown that the function [1]

$$F_{1/2}(\eta) \approx \frac{2\sqrt{\pi}}{3\sqrt{\pi}a^{-3/8} + 4e^{-\eta}} \quad (6.5)$$

approximates the Fermi-Dirac integral of order of 1/2 with a relative error not exceeding 0.4% for the value of η included in the range from $-\infty$ to $+\infty$, where

$$a = \eta^4 + 33.6\eta[1 - 0.68e^{-0.17(\eta+1)^2}] + 50 \quad (6.6)$$

The following Fortran program is designed to use the above expressions for calculating the Fermi level F in unit of eV for an n -type material, with a donor concentration DN in unit of cm^{-3} , with respect to the conduction bandedge, at temperature T in unit of K. The conduction band of the material is described by a conduction-band effective mass DM in unit of free electron mass. EB is the excitation energy of the donor and D is the donor concentration at Mott transition.

```
PROGRAM TEST
DOUBLE PRECISION DM, DN, D, EB, T, F
T=300.0
DM=0.067
DN=1.0D14
EB=8.0D-3
D=1.0D17
F=-0.5
T=T/1.1604D4
CALL FERMI (DN, DM, D, EB, T, F)
PRINT *, DN, DM, T, F
STOP
END
```

C

```
SUBROUTINE FERMI (DN, DM, D, EB, T, F)
DOUBLE PRECISION DM, DN, D, EB, F, T, S, X, Y, Z
```

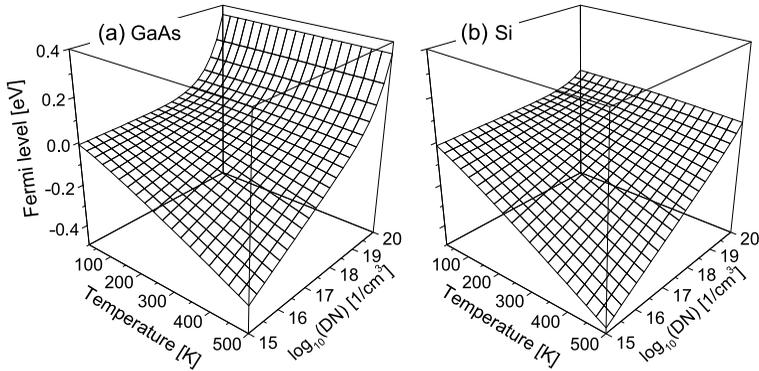


Fig. 6.1 Fermi level F [eV] as a function of temperature T [K] and doping concentration DN in the unit of cm^{-3} . (a) n -type GaAs; (b) n -type Si

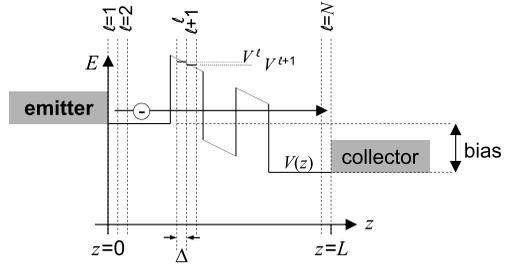
```

X=DN/(1.0+DEXP(EB/T))/((T*DM)**1.5*6.037E21)
IF (DN.GE.D) X=DN/((T*DM)**1.5*6.037E21)
S=1.0E-3/T
F=F/T
DO 100 I=1,10
S=S*0.5
DO 110 J=1,5000
F=F+S
Y=0.0
IF (-F.GT.90.0) GOTO 110
Y=F**4+33.6*F*(1.0-0.68*DEXP(-0.17*(1.0+F)**2))
Y=1.0/(DEXP(-F)+1.32934038675/(50.0+Y)**0.375)-X
IF (I+J.EQ.2) Z=Y
110 IF (Y*Z.LE.0.0) GOTO 100
100 F=F-S
F=(F+0.5*S)*T
RETURN
END

```

In Fig. 6.1 we plot the Fermi level as a function of the sample temperature and doping concentration. We have assumed here that the impurities are completely ionized (we neglect the Mott transition effect). For n -type GaAs we use the effective mass of 0.067. The results of Fermi level higher than 0.3 eV are artificial because we have neglected other conduction band valleys, e.g., the nearby X -valleys. The n -type Si is described by six ellipsoids which are grouped into: (1) the longitudinal effective mass is $m_{\ell}^* = 0.98$, the transverse effective mass is $m_t^* = 0.19$ and the degeneracy $g = 4$; (2) $m_{\ell}^* = 0.19$, $m_t^* = 0.98$, $g = 2$. The computer code listed above needs some modifications to take into account the occupations of electrons in the six ellipsoids.

Fig. 6.2 A schematic one-dimensional structure defined by potential $V(z)$



The difference between the Fermi levels in n -type GaAs and n -type Si is due to the difference in the corresponding effective masses of the two types of materials.

6.2 Amplitude of Transmitted Wave

Refer to Fig. 6.2, we now discuss a wave transmission in a one-dimensional structure $V(z)$ along the z axis, $z \in (0, L)$, which is meshed into sublayers marked with index $\ell \in (1, N)$, and the mesh size is Δ . We call the left side of the system ($z < 0$, $\ell \leq 0$) the emitter and the right side ($z > L$, $\ell > N$) the collector. The potential energy is constant in each sublayer, such as V^ℓ at sublayer ℓ in Fig. 6.2. A wave is injected from the left with an energy E along the z axis and wave vector (k_x, k_y) in the xy plane. E , k_x and k_y are conserved during the quantum wave transport.

At sublayer ℓ , the eigen states are obtained from the $\mathbf{k} \cdot \mathbf{p}$ theory including the spin-orbital interactions and strain [2–4]. The eigen states are expressed as

$$\theta_j^\ell = \sum_n C_{jn}^\ell |n\rangle \quad (6.7)$$

where n is the index of orbitals, $|x \uparrow\rangle$, $|y \uparrow\rangle$, $|z \uparrow\rangle$, $|x \downarrow\rangle$, $|y \downarrow\rangle$ and $|z \downarrow\rangle$ used in the $\mathbf{k} \cdot \mathbf{p}$ theory.

The total wave function and its first-order derivative at sublayer ℓ are

$$\begin{aligned} \Psi^\ell &= e^{i(k_x x + k_y y)} \sum_j \psi_j^\ell \theta_j^\ell \\ \frac{d\Psi^\ell}{dz} &= e^{i(k_x x + k_y y)} \sum_j \frac{d\psi_j^\ell}{dz} \theta_j^\ell \end{aligned} \quad (6.8)$$

where

$$\psi_j^\ell = A_j^\ell e^{ik_j^\ell z} + B_j^\ell e^{-ik_j^\ell z} \quad (6.9)$$

Since the phase factor $e^{i(k_x x + k_y y)}$ is the same for any wave functions (momentum conservation in the xy plane), it is neglected. At the interface between sublayer ℓ

and $\ell + 1$, the wave function and its first-order derivative are continuous

$$\begin{aligned}\sum_j (A_j^\ell + B_j^\ell) \theta_j^\ell &= \sum_m (A_m^{\ell+1} + B_m^{\ell+1}) \theta_m^{\ell+1} \\ \sum_j k_j^\ell (A_j^\ell - B_j^\ell) \theta_j^\ell &= \sum_m k_m^{\ell+1} (A_m^{\ell+1} - B_m^{\ell+1}) \theta_m^{\ell+1}\end{aligned}\quad (6.10)$$

from which we obtain

$$\begin{aligned}A_m^{\ell+1} + B_m^{\ell+1} &= \sum_j X_{mj}^{\ell+1, \ell} (A_j^\ell + B_j^\ell) \\ k_m^{\ell+1} (A_m^{\ell+1} - B_m^{\ell+1}) &= \sum_j X_{mj}^{\ell+1, \ell} k_j^\ell (A_j^\ell - B_j^\ell)\end{aligned}\quad (6.11)$$

where

$$X_{mj}^{\ell+1, \ell} = \langle \theta_m^{\ell+1} | \theta_j^\ell \rangle \quad (6.12)$$

By careful mathematical manipulations,

$$\begin{aligned}\Psi^{\ell+1} &= \sum_{mj} X_{mj}^{\ell+1, \ell} \left(\alpha_m^{\ell+1} \psi_j^\ell + \frac{\beta_m^{\ell+1}}{k_m^{\ell+1}} \frac{d\psi_j^\ell}{dz} \right) \theta_m^{\ell+1} \\ \frac{d\Psi^{\ell+1}}{dz} &= \sum_{mj} X_{mj}^{\ell+1, \ell} \left(-k_m^{\ell+1} \beta_m^{\ell+1} \psi_j^\ell + \alpha_m^{\ell+1} \frac{d\psi_j^\ell}{dz} \right) \theta_m^{\ell+1}\end{aligned}\quad (6.13)$$

where

$$\alpha_m^{\ell+1} = \frac{e^{ik_m^{\ell+1}\Delta} + e^{-ik_m^{\ell+1}\Delta}}{2}, \quad \beta_m^{\ell+1} = \frac{e^{ik_m^{\ell+1}\Delta} - e^{-ik_m^{\ell+1}\Delta}}{2i} \quad (6.14)$$

Δ is the thickness of sublayers. Here we have to emphasize that ψ_j^ℓ and $d\psi_j^\ell/dz$ are envelope function and its first-order derivative at the right edge of sublayer ℓ .

We can reorganize the above equation in the following matrix form

$$\begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi'_1 \\ \psi'_2 \\ \vdots \end{pmatrix}^{\ell+1} = \mathbf{T}^{\ell+1, \ell} \begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi'_1 \\ \psi'_2 \\ \vdots \end{pmatrix}^\ell \quad (6.15)$$

By the above equations, we can obtain the wave function and its first-order derivative at sublayer $\ell + 1$ from the ones at sublayer ℓ .

Consider now a problem of carrier transport. The carrier is injected into the system at $\ell = 1$ ($z = 0$) in the form of an initial state

$$\sum_j a_j e^{ik_j^1 z} \theta_j^1$$

By including the reflection wave

$$\sum_j r_j e^{-ik_j^1 z} \theta_j^1$$

the total wave function and its first-order differentiation are

$$\begin{aligned} \psi^1 &= \sum_j (a_j e^{ik_j^1 z} + r_j e^{-ik_j^1 z}) \theta_j^1 = \sum_j (a_j + r_j) \theta_j^1 \\ \frac{d\psi^1}{dz} &= \sum_j ik_j^1 (a_j e^{ik_j^1 z} - r_j e^{-ik_j^1 z}) \theta_j^1 = \sum_j ik_j^1 (a_j - r_j) \theta_j^1 \end{aligned} \quad (6.16)$$

The transmitted wave at $\ell = N$ ($z = L$) is expressed as

$$\begin{aligned} \psi^N &= \sum_j t_j e^{ik_j^N L} \theta_j^N \\ \frac{d\psi^N}{dz} &= \sum_j ik_j^N t_j e^{ik_j^N L} \theta_j^N \end{aligned} \quad (6.17)$$

The above two wave functions are related by

$$\begin{pmatrix} a_1 + r_1 \\ a_2 + r_2 \\ \vdots \\ ik_1^1 (a_1 - r_1) \\ ik_2^1 (a_2 - r_2) \\ \vdots \end{pmatrix} = \mathbf{T} \begin{pmatrix} t_1 e^{ik_1^N L} \\ t_2 e^{ik_2^N L} \\ \vdots \\ ik_1^N t_1 e^{ik_1^N L} \\ ik_2^N t_2 e^{ik_2^N L} \\ \vdots \end{pmatrix} \quad (6.18)$$

where

$$\mathbf{T} = \Pi_\ell \mathbf{T}^{\ell+1, \ell}$$

Reflection and transmission are easily obtained from the above relation. Finally, the transmitted current density at E , k_x , and k_y is obtained by

$$i_e(E, k_x, k_y) = \frac{e\hbar}{2im} \left(\left\langle \psi^N \left| \frac{d}{dz} \right| \psi^N \right\rangle - \text{c.c.} \right) = \frac{e\hbar}{m} \sum_j k_j^N |t_j|^2 \quad (6.19)$$

And the total current density from the emitter to the collector is

$$I_e = \int f(E, E_{fe}) i_e(E, k_x, k_y) N(E) dE \quad (6.20)$$

where E_{fe} is the local Fermi level in the emitter where the electron waves are injected from.

Similar expressions can be written down for electron wave transmission from the collector to the emitter. The total current density of the carrier transport through the one dimensional system is

$$I = I_e - I_c \quad (6.21)$$

The following Fortran program calculates the amplitude of the transmitted wave, $A = |t|^2$, at $z = L$ when a wave $\exp(ik^1 z)$ is injected from $z = 0$, as a function of the kinetic energy of the incoming wave, $E = \hbar^2(k^1)^2/2m^*$. The mesh size is 1 \AA . The system is an n -type AlGaAs/GaAs double barrier resonant tunneling diode. The structure is represented by the conduction band offset $DE(N)$, the energy potential is $V(N)$, and the electron effective mass m^* is denoted as AM . The wave function $W(N)$ is also an output. Because of the much simplified single-band structure, the envelope function and its first-order differentiation are

$$\begin{bmatrix} a + r \\ ik^1(a - r) \end{bmatrix} = T \begin{bmatrix} t e^{ik^N L} \\ ik^N t e^{ik^N L} \end{bmatrix} \quad (6.22)$$

Re-define $t \equiv t e^{ik^N L}$. And because of the wave function normalization, we can put $a = 1$, thus the above equation becomes

$$\begin{bmatrix} 1 + r \\ ik^1(1 - r) \end{bmatrix} = \begin{pmatrix} P1 & P2 \\ P3 & P4 \end{pmatrix} \begin{pmatrix} t \\ ik^N t \end{pmatrix} \quad (6.23)$$

from which it is easy to obtain

$$t = \frac{2}{P1 + ik^N \cdot P2 + \frac{P3}{ik^1} + \frac{k^N}{k^1} P4} \quad (6.24)$$

Note that in the above equations, we have expressed the wave function and its first-order differentiation at $z = 0$ in terms of the ones at $z = L$. We can move the transfer matrix T from the right side to the left side in Eq. (6.22) so that the wave function and its first-order differentiation at $z = L$ are expressed in terms of the ones at $z = 0$. The expression for t becomes a little bit more complicated than Eq. (6.24).

```
PROGRAM TEST
DOUBLE PRECISION DE(8500), V(8500), G, A, AM, E
DOUBLE COMPLEX W(8500)
G=0.1
```

```

AM=0.065
N=1000
DO 106 I=1,85
N=N+1
106 DE(N)=0.6*1.247*0.4
DO 102 I=1,56
N=N+1
102 DE(N)=0.0
DO 107 I=1,85
N=N+1
107 DE(N)=0.6*1.247*0.4
N=N+1000
DO 300 I=1,1000
300 V(I)=0.0
A=-G/DFLOAT(N-2000)
DO 301 I=1000,N-1000
301 V(I)=A*DFLOAT(I-1000)
DO 302 I=N-1000,N
302 V(I)=-G
DO 200 I=1,N
200 V(I)=V(I)+DE(I)
E=0.0
DO 210 I=1,1000
E=E+1.0D-3
CALL IW(AM,V,N,E,A,W)
210 PRINT *,E,A
STOP
END

C
SUBROUTINE IW(AM,V,N,E,A,W)
DOUBLE PRECISION A,E,AM,R,S,V(8500),Y,Z,P1,P2,
1 P3,P4,P5,P6
DOUBLE COMPLEX G1,G2,W(8500)
P1=1.0
P2=0.0
P3=0.0
P4=1.0
DO 210 L=1,N
S=DSQRT(0.262464913312*AM*DABS(E-V(L)))
IF (S.NE.0.0) GOTO 30
P1=P1+P3
P2=P2+P4
P5=P3
P6=P4
GOTO 60

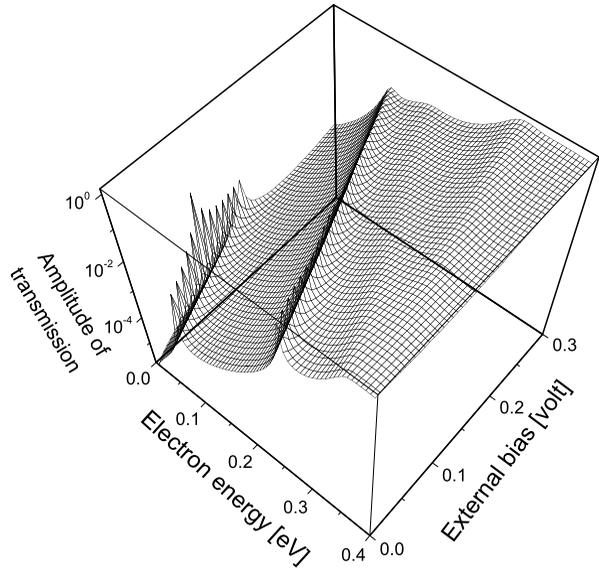
```

```

30      IF (E.GT.V(L)) GOTO 40
        Y=DSINH(S)
        Z=DCOSH(S)
        R=S
        GOTO 50
40      Y=DSIN(S)
        Z=DCOS(S)
        R=-S
50      P5=P1*R*Y+P3*Z
        P6=P2*R*Y+P4*Z
        P1=P1*Z+P3*Y/S
        P2=P2*Z+P4*Y/S
60      P3=P5
210     P4=P6
        R=DSQRT(0.262464913312*AM*DABS(E-V(N)))
        S=DSQRT(0.262464913312*AM*DABS(E-V(1)))
        G1=2.0/DCMPLX(P1+P4*S/R,P2*S-P3/R)
        G2=DCMPLX(0.0,S)*G1
        A=CDABS(G1)**2
        P1=1.0
        P2=0.0
        P3=0.0
        P4=1.0
        DO 211 L=1,N
          S=DSQRT(0.262464913312*AM*DABS(E-V(L)))
          IF (S.NE.0.0) GOTO 31
          P1=P1+P3
          P2=P2+P4
          P5=P3
          P6=P4
          GOTO 61
31      IF (E.GT.V(L)) GOTO 41
        Y=DSINH(S)
        Z=DCOSH(S)
        R=S
        GOTO 51
41      Y=DSIN(S)
        Z=DCOS(S)
        R=-S
51      P5=P1*R*Y+P3*Z
        P6=P2*R*Y+P4*Z
        P1=P1*Z+P3*Y/S
        P2=P2*Z+P4*Y/S
61      P3=P5
        P4=P6

```

Fig. 6.3 Amplitude of the transmitted wave through an n -type $\text{Al}_{0.4}\text{Ga}_{0.6}\text{As}/\text{GaAs}$ double barrier resonant tunneling structure



```

211   W(L) = P1 * G1 + P2 * G2
      RETURN
      END

```

We now apply the above computer code to calculate the amplitude of the transmitted wave through an n -type $\text{Al}_{0.4}\text{Ga}_{0.6}\text{As}/\text{GaAs}$ double-barrier resonant-tunneling structure. The barrier is 85 \AA thick and the thickness of the central well is 56 \AA . We use a uniform effective mass of $m^* = 0.067$ in the unit of free electron mass. The conduction-band offset between $\text{Al}_x\text{Ga}_{1-x}\text{As}$ and GaAs is $0.6 \times 1.247x \text{ eV}$, where 0.6 is the ratio between conduction-band offset and valence-band offset, $1.247x$ is the energy-bandgap difference. The result is presented in Fig. 6.3 which clearly demonstrates the resonance of electron wave transmission through the double-barrier structure.

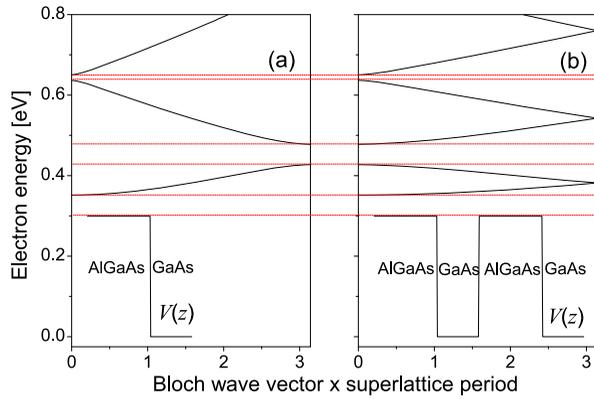
Another important boundary condition is the so-called Bloch type for multiple quantum well structures, such as an $\text{AlGaAs}/\text{GaAs}$ superlattice, see Fig. 6.4. For E above the AlGaAs bandedge, we write the envelope function and its first-order differentiation

$$\begin{aligned} \psi(z) &= u(z)e^{iqz} \\ \frac{d\psi(z)}{dz} &= [u'(z) + iqu(z)]e^{iqz} \end{aligned} \quad (6.25)$$

which are to fulfill the following Bloch boundary conditions

$$\begin{aligned} u(0) &= u(L), & u'(0) &= u'(L) \\ \psi(0) &= \psi(L), & \left. \frac{d\psi(z)}{dz} \right|_{z=0} &= \left. \frac{d\psi(z)}{dz} \right|_{z=L} \end{aligned} \quad (6.26)$$

Fig. 6.4 Dispersion relationship in an *n*-type Al_{0.4}Ga_{0.6}As/GaAs superlattice. (a) The superlattice period $L = W + B$; (b) The superlattice period $L = 2(W + B)$. Formation of the minibands and minibandgaps is clearly shown, as well as the zone folding effect when the period becomes enlarged



The relationship between E and q , i.e., $E(q)$, is called the dispersion relation.

Apply the above transfer matrix method,

$$\begin{pmatrix} P1 & P2 \\ P3 & P4 \end{pmatrix} \begin{bmatrix} u(0) \\ u'(0) + iqu(0) \end{bmatrix} = \left\{ \begin{array}{l} u(L)e^{iqL} \\ [u'(L) + iqu(L)]e^{iqL} \end{array} \right\} \quad (6.27)$$

By Eq. (6.26), and denote $u(0) = u(L) \equiv u$ and $u'(0) = u'(L) \equiv u'$, the above equation becomes

$$\begin{pmatrix} P1 & P2 \\ P3 & P4 \end{pmatrix} \begin{pmatrix} u \\ u' + iqu \end{pmatrix} = \begin{bmatrix} ue^{iqL} \\ (u' + iqu)e^{iqL} \end{bmatrix} \quad (6.28)$$

There are three unknown parameters in the above equation, q , u and u' . However, we can simply set $u = 1$ because of the wave function normalization condition. Thus, we obtain the following equations

$$\begin{aligned} G &= P1 + P2(u' + iq) \\ (u' + iq)G &= P3 + P4(u' + iq) \end{aligned} \quad (6.29)$$

where $G = e^{iqL}$. By the first equation, $u' + iq = (G - P1)/P2$. Insert it into the second equation

$$\frac{G - P1}{P2}G = P3 + P4 \frac{G - P1}{P2} \quad (6.30)$$

from which we obtain

$$G = \frac{P1 + P4 \pm \sqrt{(P1 - P4)^2 + 4P2 \cdot P3}}{2} \quad (6.31)$$

Knowing G , the wave function is easily obtained. Note the boundary condition requires that the Bloch wave vector q must be real. The following Fortran subroutine calculate the Bloch wave function.

```

SUBROUTINE IIW1(E,AM,V,W,N)
C Bloch wave
DOUBLE PRECISION T,E,AM,R,S,V(1050),Y,Z,P1,P2,
1 P3,P4,P5,P6
DOUBLE COMPLEX A,G,W(1050)
A=DCMPLX(0.0,1.0)
P1=1.0
P2=0.0
P3=0.0
P4=1.0
DO 210 L=1,N
W(L)=0.0D0
S=DSQRT(0.262464913312*AM*DABS(E-V(L)))
IF (S.NE.0.0) GOTO 30
P1=P1+P3
P2=P2+P4
P5=P3
P6=P4
GOTO 60
30 IF (E.GT.V(L)) GOTO 40
Y=DSINH(S)
Z=DCOSH(S)
R=S
GOTO 50
40 Y=DSIN(S)
Z=DCOS(S)
R=-S
50 P5=P1*R*Y+P3*Z
P6=P2*R*Y+P4*Z
P1=P1*Z+P3*Y/S
P2=P2*Z+P4*Y/S
60 P3=P5
210 P4=P6
R=(P1-P4)**2+4.0*P2*P3
IF (R.GE.0.0) THEN
G=0.5*(P1+P4+DSQRT(R))
ELSE
G=0.5*(P1+P4+A*DSQRT(DABS(R)))
END IF
IF (DABS(DIMAG(CDLOG(G)/DFLOAT(N)*A)).GT.1.0D-5)
1 RETURN
G=(G-P1)/P2
P1=1.0
P2=0.0
P3=0.0

```

```

P4=1.0
T=0.0
DO 211 L=1,N
S=DSQRT(0.262464913312*AM*DABS(E-V(L)))
IF (S.NE.0.0) GOTO 31
P1=P1+P3
P2=P2+P4
P5=P3
P6=P4
GOTO 61
31 IF (E.GT.V(L)) GOTO 41
Y=DSINH(S)
Z=DCOSH(S)
R=S
GOTO 51
41 Y=DSIN(S)
Z=DCOS(S)
R=-S
51 P5=P1*R+Y+P3*Z
P6=P2*R+Y+P4*Z
P1=P1*Z+P3*Y/S
P2=P2*Z+P4*Y/S
61 P3=P5
P4=P6
W(L)=P1+P2*G
211 T=T+(CDABS(W(L)))**2
T=2.0/DSQRT(T)
DO 212 L=1,N
212 W(L)=W(L)*T
RETURN
END

```

We apply the above computer code to calculate the dispersion relationship of $E(q)$ of an n -type $\text{Al}_{0.4}\text{Ga}_{0.6}\text{As}/\text{GaAs}$ superlattice. The barrier is $B = 85 \text{ \AA}$ thick and the thickness of the GaAs well is $W = 56 \text{ \AA}$. We use a uniform effective mass of $\text{AM} = 0.067$ in the unit of free electron mass. The conduction-band offset between $\text{Al}_x\text{Ga}_{1-x}\text{As}$ and GaAs is $0.6 \times 1.247x \text{ eV}$, where 0.6 is the ratio between the conduction-band offset and the energy-bandgap difference, $1.247x$ is the energy-bandgap difference. The result is presented in Fig. 6.4 which clearly demonstrates the formation of minibands in the energy above the $\text{Al}_{0.4}\text{Ga}_{0.6}\text{As}$ bandedge.

6.3 Localized State

In this section we present the Fortran program to calculate localized state in an arbitrary one-dimensional confinement potential $V(N)$. The output of the program is the eigen energy $E(I)$ and the corresponding wave function $W(I, N)$ of state I . The wave functions of the localized state at two ends, i.e., $z = 0$ and $z = L$, are zero, i.e., $\psi(0) = \psi(L) = 0$. The first-order differentiation of the wave function at $z = 0$ can be set to be 1 due to the wave function normalization, i.e., $\psi'(0) = 1$. Thus, we have the following relation

$$\begin{bmatrix} \psi(L) \\ \psi'(L) \end{bmatrix} = \begin{pmatrix} P1 & P2 \\ P3 & P4 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (6.32)$$

i.e., $\psi(L) = P4$. It is found that when E scans across one eigen energy, $P4$ changes its signs through 0. This gives us the criterion to determine the eigen energy.

```

PROGRAM TEST
DOUBLE PRECISION V(2000), W(10, 2000), E(10), AM
AM=0.065
N=0
DO 106 I=1, 500
N=N+1
106 V(N)=0.6*1.247*0.4
DO 102 I=1, 100
N=N+1
102 V(N)=0.0
DO 107 I=1, 500
N=N+1
107 V(N)=0.6*1.247*0.4
CALL WAVE(AM, V, N, E, W)
STOP
END

C
SUBROUTINE WAVE(AM, V, N, Q, W)
DOUBLE PRECISION V(2000), W(10, 2000), Q(10),
1 B, C, D, E, AM, R, S, X, Y, Z
DO 99 J=1, 10
DO 99 L=1, N
99 W(J, L)=0.0
E=0.0
DO 100 L=1, N
100 IF (E.GT.V(L)) E=V(L)
C=1.0
DO 556 J=1, 10
DE=1.0D-4
DO 666 M=1, 6000

```

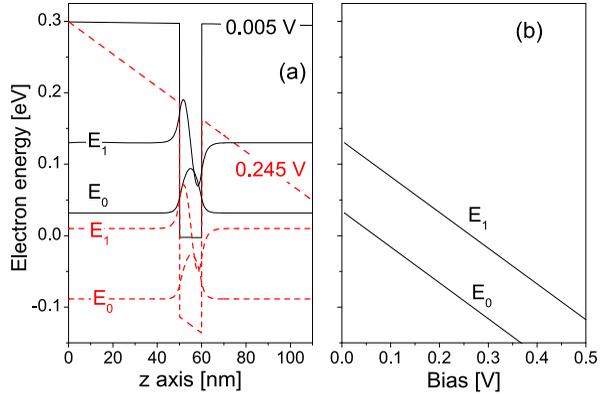
```

      E=E+DE
      IF (E.GE.V(1)) GOTO 88
      D=0.0
      B=1.0D-10
      DO 200 L=1,N
      S=SQRT(0.262464913312*AM*DABS(E-V(L)))
      IF (S.NE.0.0) GOTO 30
      D=D+B
      GOTO 200
30    IF (E.GT.V(L)) GOTO 40
      Y=DSINH(S)
      Z=DCOSH(S)
      R=S
      GOTO 50
40    Y=DSIN(S)
      Z=DCOS(S)
      R=-S
      GOTO 50
50    X=D*Z+B*Y/S
      B=D*R*Y+B*Z
      D=X
      GOTO 200
200   W(J,L)=D
      IF (C*D.GT.0.0) GOTO 666
      E=E-DE
      DE=DE*0.1
      IF (DE.LE.1.0D-8) GOTO 77
666   CONTINUE
      GOTO 88
77    C=-C
      E=E+1.0D1*DE
      Q(J)=E
      DO 211 L=N,1,-1
      IF (C*W(J,L).LT.0.0) GOTO 123
211   W(J,L)=0.0
123   B=0.0
      DO 210 L=1,N
210   B=B+W(J,L)*W(J,L)
      B=1.0/DSQRT(B)
      DO 240 L=1,N
240   W(J,L)=W(J,L)*B
556   CONTINUE
88    RETURN
      END

```

We apply the above computer program to study the Stark effect in an n -type GaAs quantum well embedded in the n -type $\text{Al}_{0.4}\text{Ga}_{0.6}\text{As}$ material. The quantum

Fig. 6.5 (a) Potentials and wave functions of ground sublevel E_0 and the first excited sublevel E_1 at two different biases (*black solid lines*: $V_D = 0.005$ V; *red dash lines*: $V_D = 0.245$ V). (b) E_0 and E_1 as functions of the external bias V_D



well is 100 Å thick. The left and right AlGaAs barriers are 500 Å thick. We assume that an external bias V_D applies linearly across the structure. In Fig. 6.5 we plot the energy sublevels, E_0 the ground state, E_1 the first excited state, and so on, localized in the quantum well as functions of the bias V_D (it is easy to convert the bias into the electric field strength in the quantum well). The electrical potential is approached as

$$V(z) = -V_D \frac{z}{L} \quad (6.33)$$

where L is the total thickness of the system. Numerical results are shown in Fig. 6.5 where we observe here clearly the Stark effect, and the peaks of the wave functions move toward the right boundary of the quantum well.

The program is readily applicable to any one-dimensional system, e.g., the inversion layer in a metal-oxide-semiconductor field-effect transistor, when the Poisson equation is taken into account.

6.4 Local Density of States: Recursion Method

When we study the electronic properties of systems which do not have the translational symmetry in more than one direction, e.g., semiconductor surface, quantum wire, quantum dot, nano-scale MOSFETs, we have to solve the two- and three-dimensional Schrödinger equations, which is proven to be a very difficult task. In many cases however we do not need to know all the details about the eigen solutions of the multiple-dimensional Schrödinger equation. We only need to know the amplitude of the wave function, i.e., $|\psi|$ when we calculate, self-consistently the potential energy from Poisson's equation.

In this case, the recursion method provides us precisely the information we need. We calculate the local Green's function, $g(\mathbf{r}, E)$, as we determined in the threshold voltage in a nano-scale MOSFET. Let us mark the spatial position by ℓ ($\ell = 1, 2, \dots$) which denotes a spatial position $\mathbf{r} = (x, y, z)$, $\ell + 1$ as $(x + \Delta, y, z)$ and so on.

Denoting the Hamiltonian of the system as H , we construct a base $|1\rangle, |2\rangle, \dots$ of

$$|1\rangle = \begin{pmatrix} \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix} \dots \dots \begin{matrix} \ell - 1 \\ \ell \\ \ell + 1 \end{matrix} \tag{6.34}$$

where lattice site ℓ corresponds to a spatial coordinate of \mathbf{r} . $|2\rangle$ and higher $|n\rangle$ are defined by

$$\begin{aligned} |2\rangle &= H|1\rangle - a_1|1\rangle \\ \dots & \\ |n+1\rangle &= H|n\rangle - a_n|n\rangle - \sqrt{b_{n-1}}|n-1\rangle \\ \dots & \end{aligned} \tag{6.35}$$

In this base H becomes tri-diagonalized with diagonal $H_{nn} = a_n$ and off-diagonal elements $H_{n-1,n} = H_{n,n-1} = b_{n-1}$,

$$a_n = \frac{\langle n|H|n\rangle}{\langle n|n\rangle}, \quad b_{n-1} = \frac{\langle n|n\rangle}{\langle n-1|n-1\rangle} \tag{6.36}$$

The local density of states is obtained from local Green's function

$$g(\mathbf{r}, E) = \frac{1}{E - a_1 - \frac{b_1}{E - a_2 - \frac{b_2}{E - a_3 - \dots}}} \tag{6.37}$$

The local density of states is then calculated

$$v(\mathbf{r}, E) = -\frac{1}{\pi} \lim_{\beta \rightarrow 0} \{ \text{Im}[g(\mathbf{r}, E + i\beta)] \} \tag{6.38}$$

From the theory of Green's function [5], we know that

$$v(\mathbf{r}, E) = \sum_j |\psi_{E_i,j}(\mathbf{r})|^2 \delta_{E,E_i} \tag{6.39}$$

where E_i is the i th energy state and j is the index of degenerate states at E_i .

We take a two-dimensional (2D) quantum well (three-dimensional quantum wire) as an example. The width of the 2D quantum well is defined as L_X and L_Y in x - and y -directions, respectively. We take GaAs/AlGaAs as the hetero-material so that the effective mass $m^* = 0.067$ (in unit of free electron mass m_0). The barrier height is defined as B . The following computer program calculates the local density of states W at (IX, IY) in the energy range $(E1, E2)$, the GaAs quantum well is

located within $(-LX/2 : LX/2, -LY/2 : LY/2)$. In the program, the unit of length is Å and eV for energy.

We discretize the xy plane into unit square cells of $\Delta \times \Delta$ (in the computer code, $\Delta = 1$ Å). The two-dimensional effective-mass Schrödinger equation

$$-\frac{\hbar^2}{2m^*} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) \psi(x, y) + V(x, y) \psi(x, y) = E \psi(x, y) \quad (6.40)$$

$$-\frac{\hbar^2}{2m^*} \left[\frac{\psi(x + \Delta, y) - 2\psi(x, y) + \psi(x - \Delta, y)}{\Delta^2} + \frac{\psi(x, y + \Delta) - 2\psi(x, y) + \psi(x, y - \Delta)}{\Delta^2} \right] + V(x, y) \psi(x, y) = E \psi(x, y) \quad (6.41)$$

Let

$$D = \frac{\hbar^2}{2m^* \Delta^2} \quad (6.42)$$

Equation (6.41) becomes

$$(4D + V^\ell) \psi^\ell - D \sum_{n=1}^4 \psi^{E(\ell, n)} = E \psi^\ell \quad (6.43)$$

Here $E(\ell, n)$ is the correlation function among discretized lattice sites. For a two-dimensional square lattice, each lattice $\ell = (x, y)$ site has four nearest neighbors. Neighbor 1 is $(x + \Delta, y)$, noted as $n = 1$ and its lattice index is $E(\ell, 1)$. Neighbor 2 is $(x - \Delta, y)$, noted as $n = 2$ and its lattice index is $E(\ell, 2)$. Neighbor 3 is $(x, y + \Delta)$, noted as $n = 3$ and its lattice index is $E(\ell, 3)$. Finally, neighbor 4 is $(x, y - \Delta)$, noted as $n = 4$ and its lattice index is $E(\ell, 4)$.

```

PROGRAM TEST
DOUBLE PRECISION AM, B, E1, E2, W(1000), V(319201)
INTEGER A(4, 2), E(319201, 4), I, J, JJ(400), K, M, M1,
1 M2, M3, N, P(319201, 2), X, Y
COMMON E, JJ, AM, V
OPEN (UNIT=21, STATUS='UNKNOWN', FILE='a10.dat')
AM=0.067
B=0.3
E1=0.0
E2=0.3
LX=100
LY=100
IX=20
IY=20
A(1, 1)=1

```

```

A(1,2)=0
A(2,1)=-1
A(2,2)=0
A(3,1)=0
A(3,2)=1
A(4,1)=0
A(4,2)=-1
M1=1
M2=1
M3=1
J=1
JJ(1)=J
P(1,1)=0
P(1,2)=0
DO 100 K=2,400
DO 110 M=M1,M2
DO 200 N=1,4
X=P(M,1)+A(N,1)
Y=P(M,2)+A(N,2)
DO 210 I=M3,J
IF (P(I,1).NE.X) GOTO 210
IF (P(I,2).EQ.Y) GOTO 220
210 CONTINUE
J=J+1
IF (J.GE.319201) GOTO 777
P(J,1)=X
P(J,2)=Y
E(M,N)=J
GOTO 200
220 E(M,N)=I
200 CONTINUE
110 CONTINUE
M3=M1
JJ(K)=J
M1=M2+1
M2=J
100 CONTINUE
777 DO 400 I=1,J
V(I)=B
400 IF ((P(I,1)+IX+LX/2)*(LX/2-IX-P(I,1)).GE.0.AND.
1 (P(I,2)+IY+LY/2)*(LY/2-IY-P(I,2)).GE.0) V(I)=0.0
CALL CARRIER(E1,E2,W)
DO 500 I=1,1000
500 WRITE (21,77) (E2-E1)*1.0D-3*DFLOAT(I),W(I)
77 FORMAT (2(E10.4,1X))

```

```
STOP
END
```

C

```

SUBROUTINE CARRIER(E1,E2,W)
DOUBLE PRECISION W(1000),A(400),B(400),E1,E2,
1 DE,E
DOUBLE COMPLEX C,D
II=395
CALL aANDb(II,A,B)
DO 210 I=1,II
210 B(I)=B(I)*B(I)
E=E1
DE=(E2-E1)*1.0D-3
DO 100 I=1,1000
E=E+DE
D=DCMPLX(E,DE*1.0D-2)
C=0.5*(D-A(II)-CDSQRT((D-A(II))*2-4.0*B(II)))/
1 B(II)
DO 200 J=II,1,-1
200 C=1.0D0/(D-A(J)-B(J)*C)
100 W(I)=DABS(DIMAG(C))/3.14159265
111 CONTINUE
77 RETURN
END
```

C

```

SUBROUTINE aANDb(II,A,B)
DOUBLE PRECISION A(400),B(400),C,D,
1 U(319201),V(319201),W(319201),AM,P(319201)
INTEGER E(319201,4),JJ(400)
COMMON E,JJ,AM,P
DO 99 I=1,319201
V(I)=0.0
W(I)=0.0
99 U(I)=0.0
U(1)=1.0
D=3.81/AM
DO 100 K=1,II
A(K)=0.0
B(K)=0.0
DO 110 I=1,JJ(K+2)
C=4.0*D+P(I)
W(I)=C*U(I)-D*(U(E(I,1))+U(E(I,2))+U(E(I,3))
1 +U(E(I,4)))
110 A(K)=A(K)+U(I)*W(I)
DO 140 I=1,JJ(K+2)
```

```

      W(I)=W(I)-A(K)*U(I)
      IF (K.NE.1) W(I)=W(I)-B(K-1)*V(I)
      V(I)=U(I)
      U(I)=W(I)
140   B(K)=B(K)+U(I)*U(I)
      B(K)=DSQRT(B(K))
      DO 150 I=1, JJ(K+2)
150   U(I)=U(I)/B(K)
100   CONTINUE
      RETURN
      END

```

When $L_x=L_y=100 \text{ \AA}$ and $B=0.3 \text{ eV}$, we find three localized energy sublevels confined in the quantum well, namely, $E_0 = 67.2$, $E_1 = 162.9$ and $E_2 = 255.0 \text{ meV}$. In the infinitely high barrier approximation, we know that for a square quantum well,

$$E_i = (n_x^2 + n_y^2)e_0 \quad (6.44)$$

where n_x and n_y are quantum numbers in the x - and y -directions.

$$e_0 = \frac{\pi^2 \hbar^2}{2L_x^2 m^*} = 56.1 \text{ meV} \quad (6.45)$$

when $L_x = L_y = 100 \text{ \AA}$. It is then expected that $E_0 = 2e_0 = 112.2$, $E_1 = 5e_0 = 281.0$ and $E_2 = 8e_0 = 448.8 \text{ meV}$ when the barrier is infinitely high. These results are higher than what we have obtained numerically when the barrier height is finite. When we increase the barrier height to $B=0.93 \text{ eV}$, $E_0 = 82.2 \text{ meV}$, $E_1 = 204.6 \text{ meV}$ and $E_2 = 326.2 \text{ meV}$. For $B=1.93 \text{ eV}$, $E_0 = 102.8 \text{ meV}$, $E_1 = 256.5 \text{ meV}$, and $E_2 = 410.5 \text{ meV}$, which are much closer to the ones of infinitely high barrier approximation.

By slightly modifying the above Fortran program we can calculate the local density of states in an xy plane at certain energy E_1 . For the following program we calculate the local density of states (LDOS) in a plane of $(-8 : 8, -8 : 8)$ in the unit of nm. Note that the quantum well is located within $(-5 : 5, -5 : 5) \text{ nm}$. Subroutine `aANDb` is not changed so is not repeated.

```

PROGRAM TEST
DOUBLE PRECISION AM,B,E1,W(0:20),V(319201)
INTEGER A(4,2),E(319201,4),I,J,JJ(400),K,M,M1,
1 M2,M3,N,P(319201,2),X,Y
COMMON E,JJ,AM,V
OPEN (UNIT=21,STATUS='UNKNOWN',FILE='a10.dat')
AM=0.067
B=0.3
LX=100
LY=100

```

```

E1=67.2D-3
A(1,1)=1
A(1,2)=0
A(2,1)=-1
A(2,2)=0
A(3,1)=0
A(3,2)=1
A(4,1)=0
A(4,2)=-1
M1=1
M2=1
M3=1
J=1
JJ(1)=J
P(1,1)=0
P(1,2)=0
DO 100 K=2,400
DO 110 M=M1,M2
DO 200 N=1,4
X=P(M,1)+A(N,1)
Y=P(M,2)+A(N,2)
DO 210 I=M3,J
IF (P(I,1).NE.X) GOTO 210
IF (P(I,2).EQ.Y) GOTO 220
210 CONTINUE
J=J+1
IF (J.GE.319201) GOTO 777
P(J,1)=X
P(J,2)=Y
E(M,N)=J
GOTO 200
220 E(M,N)=I
200 CONTINUE
110 CONTINUE
M3=M1
JJ(K)=J
M1=M2+1
M2=J
100 CONTINUE
777 DO 300 IX=0,80,4
DO 310 L=0,20
IY=4*L
DO 400 I=1,J
V(I)=B
400 IF ((P(I,1)+IX+LX/2)*(LX/2-IX-P(I,1)).GE.0.AND.

```

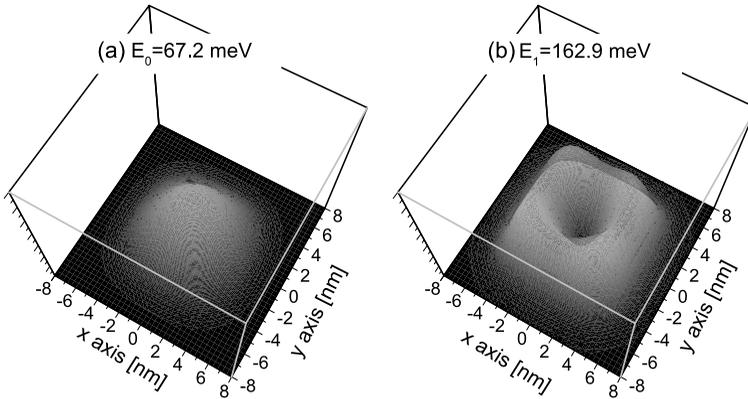


Fig. 6.6 Local densities of states in the region of $(-8 : 8, -8 : 8) \text{ nm}^2$ at (a) $E_0 = 67.2$ and (b) $E_1 = 162.9$ meV. $B = 0.3$ eV. Note that the quantum wire size is $(-5 : 5, -5 : 5) \text{ nm}^2$

```

1 (P(I,2)+IY+LY/2)*(LY/2-IY-P(I,2)).GE.0)V(I)=0.0
CALL CARRIER(E1,WW)
310 W(L)=WW
WRITE(21,77)(W(I),I=0,20)
77 FORMAT(21(E10.4,1X))
300 CONTINUE
STOP
END
C
SUBROUTINE CARRIER(E,W)
DOUBLE PRECISION W,A(400),B(400),E
DOUBLE COMPLEX C,D
II=395
CALL aANDb(II,A,B)
DO 210 I=1,II
210 B(I)=B(I)*B(I)
D=DCMPLX(E,1.0D-4)
C=0.5*(D-A(II)-CDSQRT((D-A(II))*2-4.0*B(II)))/
1 B(II)
DO 200 J=II,1,-1
200 C=1.0D0/(D-A(J)-B(J)*C)
100 W=DABS(DIMAG(C))/3.14159265
77 RETURN
END

```

Figure 6.6 shows the local densities of states in the region of $(-8 : 8, -8 : 8) \text{ nm}^2$ when $B = 0.3$ eV. The ground state is what we can expect:

$$\psi_{E_0} = \sin(\alpha x) \sin(\alpha y) \quad (6.46)$$

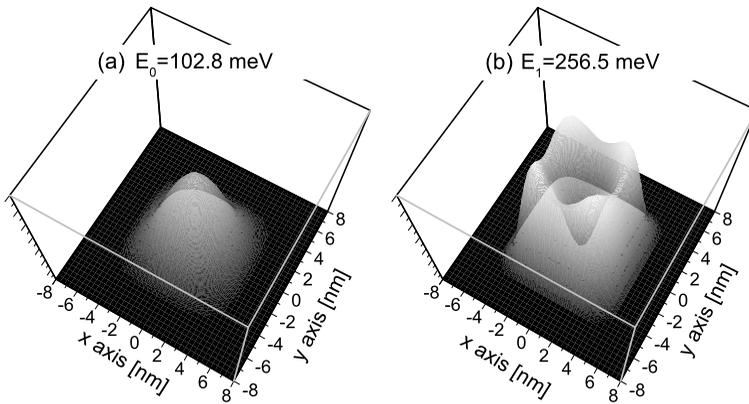


Fig. 6.7 Same as Fig. 6.6 but $B = 19.3$ eV

where $\alpha = \pi/L_x$. The first excited state is two-fold degenerate

$$\begin{aligned} \psi_{E_{1,1}} &= \sin(\alpha x) \sin(2\alpha y) \\ \psi_{E_{1,2}} &= \sin(2\alpha x) \sin(\alpha y) \end{aligned} \tag{6.47}$$

and the local densities of states provide us with the information of

$$|\psi_{E_{1,1}}|^2 + |\psi_{E_{1,2}}|^2 \tag{6.48}$$

As mentioned before, when B is relatively small, the confined energy levels in the quantum wire is low. This is reflected in the penetration of the wave function into the barrier region. Figure 6.7 shows the local densities of the ground state and the first excited state in a quantum wire of $(-5 : 5, -5 : 5)$ nm², just as in Fig. 6.6. However, the quantum wire has a much higher barrier potential of $B = 19.3$ eV so that the penetrations of the wave functions into the barrier regions are almost zero.

Information about the phase factors in the wave functions can also be obtained from the local Green's function. For example, we can construct the base for tri-diagonalized from

$$|1\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} \vdots \\ 0 \\ 1 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix} \begin{matrix} \dots\dots \\ \dots\dots \end{matrix} \begin{matrix} \ell - 1 \\ \ell \\ \ell + 1 \\ \ell' - 1 \\ \ell' \\ \ell' + 1 \end{matrix} \tag{6.49}$$

where lattice site ℓ' corresponds to a spatial coordinate of \mathbf{r}' . We then have the following local Green's function [5, 6]

$$v(\mathbf{r}\mathbf{r}', E) = \sum_j \psi_{E_i, j}^*(\mathbf{r}') \psi_{E_i, j}(\mathbf{r}) \delta_{E, E_i} \quad (6.50)$$

where E_i is the i th energy state and j is the index of degenerate states at E_i . The phase factor can be easily deduced from the comparison between $v(\mathbf{r}\mathbf{r}', E)$ and $v(\mathbf{r}, E)$.

The extension of the computer program to three-dimensional structures is rather straightforward. The method is well used in the theoretical frame of tight-binding approximation [7].

6.5 Time-Dependent Wave Packet Transmission

As we learned in Sect. 2.1, the motion of the electron state is governed by the time-dependent Schrödinger equation

$$i\hbar \frac{\partial \Psi(\mathbf{r}, t)}{\partial t} = H\Psi(\mathbf{r}, t) \quad (6.51)$$

Discretizing the time t by step δt so that $t = n\delta t$, where n is a positive integer. The wave function of the electron becomes $\Psi(\mathbf{r}, n\delta t)$ which is further denoted as $\Psi^n(\mathbf{r})$, and the time-dependent Schrödinger equation employs the Cayley form

$$\left(1 + \frac{i\delta t}{2\hbar} H\right) \Psi^{n+1}(\mathbf{r}) = \left(1 - \frac{i\delta t}{2\hbar} H\right) \Psi^n(\mathbf{r}) \quad (6.52)$$

with which we have studied time-dependent electron transport through nanostructures in Chap. 2.

The following FORTRAN codes calculate an electron wave packet transmitting through a square quantum wire, $y \in (-LY : LY)$ and $z \in (-LZ : LZ)$, along the x axis, $x \in (1 : LX)$. Mesh size is 2 \AA (2 \AA denoted in the code). The effective mass of the electron can be anisotropic, denoted as AMX , AMY , AMZ . Actually the codes were used to calculate the motion pictures, Fig. 4.43, of a wave packet transporting through a quantum wire having an ionic scattering center in Sect. 4.6.5.

```

PROGRAM wavepacket
PARAMETER (LX=5000,LY=13,LZ=13)
DOUBLE PRECISION AMX,AMY,AMZ,C,F,Q,G,T,E,
1 V(LX,-LY:LY,-LZ:LZ)
DOUBLE COMPLEX B,D,S0(LX,-LY:LY,-LZ:LZ),
1 S(LX,-LY:LY,-LZ:LZ),
1 AX,AY,AZ
INTEGER X,Y,Z

```

```

OPEN (UNIT=22,STATUS='UNKNOWN',FILE='x600ep.dat')
OPEN (UNIT=23,STATUS='UNKNOWN',FILE='x600qp.dat')
D=DCMPLX(0.0D0,1.0D0)
AMX=0.067D0
AMY=0.067D0
AMZ=0.067D0

```

C initial Gaussian wave packet E eV at (1050,0,0) along
C x axis

```

E=0.12D0
C=2.0D0*DSQRT(E*AMX/3.81D0)
DO 153 X=2,2100
AX=CDEXP(D*(X-1050)*C)
DO 153 Y=-LY,LY
DO 153 Z=-LZ,LZ
153 S(X,Y,Z)=AX
Q=0.0D0
DO 154 X=2,2100
DO 154 Y=-LY,LY
DO 154 Z=-LZ,LZ
S(X,Y,Z)=S(X,Y,Z)*S(1,Y,Z)
154 Q=Q+(CDABS(S(X,Y,Z)))**2
Q=1.0D0/DSQRT(Q)
DO 155 X=1,LX
DO 155 Y=-LY,LY
DO 155 Z=-LZ,LZ
S(X,Y,Z)=S(X,Y,Z)*Q
155 IF ((X-1)*(X-LX)*(Y-LY)*(Y+LY)*(Z-LZ)*(Z+LZ).
1 EQ.0) S(X,Y,Z)=0.0D0

```

C temporal evolution (step 2.0D-2 fs, mesh 2 AA) of 3D
C wave packet

```

AX=2.0D-2*D*2.8941722D0/4.0D0/AMX
AY=2.0D-2*D*2.8941722D0/4.0D0/AMY
AZ=2.0D-2*D*2.8941722D0/4.0D0/AMZ
B=2.0D-2*D*0.75962696D0

```

C potential energy definition, ionic scattering center
C at (2250,MY,0)

```

MY=13
DO 160 X=1,LX
DO 160 Z=-LZ,LZ
DO 160 Y=-LY,LY
C=0.0D0
IF ((Y-LY)*(Y+LY)*(Z-LZ)*(Z+LZ).EQ.0) C=3.5D0

```

```

      Q=DSQRT(1.0D0*(X-2250)**2+(Y-MY)**2+Z*Z)
160   V(X,Y,Z)=C-0.6D0*DEXP(-Q/1.0D3)/(1.0D-2+Q)

      DO 333 I=1,6000
      DO 200 Z=-LZ+1,LZ-1
      DO 200 Y=-LY+1,LY-1
      DO 200 X=2,LX-1
200   S0(X,Y,Z)=(1.0D0-B*V(X,Y,Z)-2.0D0*(AX+AY+AZ))*
1     S(X,Y,Z)+AX*(S(X-1,Y,Z)+S(X+1,Y,Z))+AY*
1     (S(X,Y-1,Z)+S(X,Y+1,Z))+AZ*(S(X,Y,Z-1)+
1     S(X,Y,Z+1))
      DO 201 L=1,100
      DO 201 Z=-LZ+1,LZ-1
      DO 201 Y=-LY+1,LY-1
      DO 201 X=2,LX-1
      D=(S0(X,Y,Z)+AX*(S(X-1,Y,Z)+S(X+1,Y,Z))+
1     AY*(S(X,Y-1,Z)+S(X,Y+1,Z))+AZ*(S(X,Y,Z-1)+
1     S(X,Y,Z+1)))/(1.0D0+B*V(X,Y,Z)+2.0D0*(AX+AY+AZ))
201   S(X,Y,Z)=0.5D0*(D+S(X,Y,Z))
      CONTINUE
      Q=0.0D0
      F=0.0D0
      C=0.0D0
      DO 203 Z=-LZ+1,LZ-1
      DO 203 Y=-LY+1,LY-1
      DO 203 X=2,LX-1
      D=(S0(X,Y,Z)+AX*(S(X-1,Y,Z)+S(X+1,Y,Z))+
1     AY*(S(X,Y-1,Z)+
1     S(X,Y+1,Z))+AZ*(S(X,Y,Z-1)+S(X,Y,Z+1)))/
1     (1.0D0+2.0D0*(AX+AY+AZ)+B*V(X,Y,Z))
      IF (CDABS(D-S(X,Y,Z)).GT.Q) Q=CDABS(D-S(X,Y,Z))
      S(X,Y,Z)=0.5D0*D+0.5D0*S(X,Y,Z)
      T=CDABS(S(X,Y,Z))
      C=C+T*T
203   IF (T.GT.F) F=T
      C=1.0D0/DSQRT(C)
      DO 206 Z=-LZ+1,LZ-1
      DO 206 Y=-LY+1,LY-1
      DO 206 X=2,LX-1
206   S(X,Y,Z)=S(X,Y,Z)*C
      IF (I.NE.10*(I/10)) GOTO 332

```

C T/C are currents in x direction

```

      T=0.0D0
      C=0.0D0

```

```

DO 211 Z=-LZ+1,LZ-1
DO 211 Y=-LY+1,LY-1
DO 211 X=2,LX/2
IQ=X+LX/2
C=C+DIMAG(DCONJG(S(IQ,Y,Z))*(S(IQ,Y,Z)-
1 S(IQ-1,Y,Z)))
211 T=T+DIMAG(DCONJG(S(X,Y,Z))*(S(X,Y,Z)-
1 S(X-1,Y,Z)))
PRINT *,I,T,C,Q/F
WRITE(22,*)I,T,C
332 IF(I.NE.1.AND.I.NE.1000*(I/1000))GOTO 333
DO 4000 X=1,LX
Q=0.0D0
DO 4001 Z=-LZ,LZ
DO 4001 Y=-LY,LY
4001 Q=Q+(CDABS(S(X,Y,Z)))**2
Q=Q+1.0D-15
4000 WRITE(23,*)Q
333 CONTINUE
STOP
END

```

References

1. Bednarczyk D, Bednarczyk J (1978) The approximation of the Fermi-Dirac integral $F_{1/2}(\eta)$. Phys Lett A 64:409-410
2. Dresselhaus G, Kip AF, Kittel C (1955) Cyclotron resonance of electrons and holes in silicon and germanium crystals. Phys Rev 98:368-384
3. Tiersten M (1961) Acoustic-mode scattering of holes. IBM J Res Dev 5:122-131
4. Pikus GE, Bir GL (1959) Effect of deformation on the hole energy spectrum of germanium and silicon. Sov Phys, Solid State 1:1502-1517
5. Economons EN (1979) Green's function in quantum physics, Cardona M (ed). Springer, Berlin
6. Fu Y (1987) Correlations of lattice vibrations. Chin Phys Lett 4:309-312
7. Fu Y, Willander M, Han P, Matsuura T, Murota J (1998) Local environment effect on electronic band structure of cubic $\text{Si}_{1-y}\text{C}_y$ alloy. Phys Rev B 58:7717-7722

Appendix A

Fundamental Constants

Table A.1 Fundamental constants

Quantity & symbol		CGS	MKS (SI)
Speed of light in vacuum c	2.997925	10^{10} cm sec ⁻¹	10^8 m s ⁻¹
Magnetic constant $\mu_0 = 4\pi \times 10^{-7}$ (permeability of free space)	1.25664		10^{-6} H m ⁻¹
Electric constant $\epsilon_0 = 1/\mu_0 c^2$ (permittivity of free space)	8.8541853		10^{-12} F m ⁻¹
Electron charge e	1.60219		10^{-19} C
	4.80324	10^{-10} esu	
Electron volt eV	1.60219	10^{-12} erg eV ⁻¹	10^{-19} J eV ⁻¹
Electron rest mass m_0	9.10956	10^{-28} gm	10^{-31} kg
Planck's constant h	6.626196	10^{-27} erg sec	10^{-34} J s
Planck's constant $\hbar = h/2\pi$	1.05459	10^{-27} erg sec	10^{-34} J s
Bohr radius $a_0 = 4\pi\epsilon_0\hbar^2/m_0e^2$	0.529177	10^{-8} cm	10^{-10} m
Rydberg constant $R_y = \hbar^2/2m_0a_0^2$	13.6058	eV	eV
Fine structure constant $\alpha = e^2/2\epsilon_0hc$	7.297351	10^{-3}	10^{-3}
Avogadro's constant L, N_A	6.022169	10^{23} mol ⁻¹	10^{23} mol ⁻¹
Loschmidt's constant N_L	2.68719		10^{25} m ⁻³
Boltzmann's constant $k_B = R/L$	1.380622	10^{-16} erg K ⁻¹	10^{-23} J K ⁻¹
Faraday constant $F = Le$	9.64867		10^4 C mol ⁻¹
Universal gas constant $R = Lk_B$	8.31435	10^7 erg K ⁻¹ mol ⁻¹	J K ⁻¹ mol ⁻¹
Mechanical equivalent of heat	4.184	10^7 erg cal ⁻¹	J cal ⁻¹
Bohr magneton $\mu_B = e\hbar/2m_0c$	9.2741	10^{-21} erg G ⁻¹	10^{-24} J T ⁻¹
Proton rest mass m_p	1.67251	10^{-24} gm	10^{-27} kg
Nuclear magneton $\mu_N = e\hbar/2m_p c$	5.0508	10^{-24} erg G ⁻¹	10^{-27} J T ⁻¹
Neutron rest mass m_n	1.67492	10^{-24} gm	10^{-27} kg
Stefan-Boltzmann constant $\sigma = 2\pi^5 k_B^4/15h^3c^2$	5.6697		10^{-8} W m ⁻² K ⁻⁴

Table A.1 (Continued)

Quantity & symbol	CGS	MKS (SI)
Gravitational constant G	6.6732	$10^{-11} \text{ N m}^2 \text{ kg}^{-2}$
Acceleration of free fall g_n	9.80665	m s^{-2}

$1.0 \text{ [eV]} = 2.41796 \times 10^{14} \text{ [Hz]} = 8.0655 \times 10^3 \text{ wavenumber [cm}^{-1}] =$
 $1.1604 \times 10^4 \text{ [K]} = 1.239855 \text{ wavelength } [\mu\text{m}]$

W: watt, G: gauss, T: tesla, N: newton, C: coulomb

Appendix B

Quantum Physics

B.1 Black Body Radiation

Planck's law for the energy density distribution for the radiation from a black body at temperature T is

$$w(f, T) = \frac{8\pi hf^3}{c^3} \frac{1}{e^{hf/k_B T} - 1} \tag{B.1}$$

where h is the Planck's constant, and f is the frequency. Note that $\omega = 2\pi f$ is the angular frequency, and $\hbar = h/2\pi$. The low-frequency Rayleigh-Jens law

$$w(f, T) = \frac{8\pi f^2 k_B T}{c^3}$$

is obtained when $hf \ll k_B T$. Stefan-Boltzmann's law for the total radiation energy per unit volume can be derived

$$W(T) = \int_0^\infty w(f, T) df = \frac{8\pi^5 k_B^4}{15h^3 c^3} T^4 = \frac{4\sigma}{c} T^4 \tag{B.2}$$

where $\sigma = 2\pi^5 k_B^4 / 15h^3 c^2$ is the Stefan-Boltzmann constant. Wien's law for the wavelength λ_{\max} at which the energy density has its maximum value can be derived from Eq. (B.1)

$$\lambda_{\max} = \frac{b}{T} \tag{B.3}$$

where $b = 0.2898 \text{ cm}\cdot\text{K}$ is a universal constant. Note $\lambda = c/f$.

B.2 The Compton Effect

A photon with initial wavelength λ incident upon an electron at rest with a rest mass m_0 . After the collision, the photon has a wavelength of λ' scattered into a direction

at an angle θ with respect to its initial propagation direction, which is also referred to be the photon scattering angle.

$$\lambda' = \lambda + \frac{h}{m_0c}(1 - \cos\theta) = \lambda + \lambda_C(1 - \cos\theta) \quad (\text{B.4})$$

where $\lambda_C = h/m_0c$ is called the Compton wavelength of the electron.

B.3 Electron Diffraction

Diffraction pattern of electron from a crystal can be explained by the dual particle-wave nature of matter. A particle having a momentum p is associated the so-called de Broglie wavelength

$$\lambda = \frac{h}{p} \quad (\text{B.5})$$

B.4 Operators in Quantum Physics

In quantum mechanics, classical physical quantities are translated into quantum operators. The measurement of a physical quantity $\langle A \rangle$ is described quantum mechanically by

$$\langle A \rangle = \int \psi_1^* A \psi_2 d\mathbf{r} = \int (A \psi_1)^* \psi_2 d\mathbf{r} \quad (\text{B.6})$$

where ψ is the wave function that describes the quantum mechanical state of the system under measurement. The only possible values that can be obtained when measuring a physical quantity $\langle A \rangle$ are the eigen values of the quantum operator A . Thus, the operator is hermitian and its eigen values are real, i.e., $Au_n = a_n u_n$, where u_n is the eigen function, a_n is the eigen value which is real. Subscript n denotes the index of the eigen functions, i.e., operator A may have multiple eigen functions and eigen values.

B.5 The Schrödinger Equation

The momentum operator is given by $\mathbf{p} = -i\hbar\nabla$, and the position operator is $\mathbf{r} = i\hbar\nabla_p$. The energy operator is

$$E = i\hbar\frac{\partial}{\partial t}$$

The Hamiltonian of a particle with mass m and potential energy V is given by

$$H = \frac{p^2}{2m} + V$$

From the eigen equation $H\psi = E\psi$ follows the time-dependent Schrödinger equation

$$-\frac{\hbar^2}{2m}\nabla^2\psi(\mathbf{r}, t) + V\psi(\mathbf{r}, t) = i\hbar\frac{\partial\psi(\mathbf{r}, t)}{\partial t} \quad (\text{B.7})$$

where $\psi(\mathbf{r}, t)$, which is a function of space \mathbf{r} and time t , is the wave function that describes the quantum mechanical state of the particle. The wave character of a particle is described by $\psi(\mathbf{r}, t)$, and can be interpreted as a measure for the probability $\rho(\mathbf{r}, t)$ to find the particle at \mathbf{r} and t

$$\rho(\mathbf{r}, t) = |\psi(\mathbf{r}, t)|^2 \quad (\text{B.8})$$

The normalizing condition for the wave function is

$$\langle\psi|\psi\rangle = \int \psi^*\psi d\mathbf{r} = 1$$

At steady state,

$$i\hbar\frac{\partial\psi(\mathbf{r}, t)}{\partial t} = E\psi(\mathbf{r}, t) \quad (\text{B.9})$$

so that

$$\psi(\mathbf{r}, t) = \psi(\mathbf{r})e^{-iEt/\hbar} \quad (\text{B.10})$$

we have the time-independent Schrödinger equation

$$-\frac{\hbar^2}{2m}\nabla^2\psi(\mathbf{r}) + V\psi(\mathbf{r}, t) = E\psi(\mathbf{r}) \quad (\text{B.11})$$

for $\psi(\mathbf{r})$.

The probability flux \mathbf{s} is given by

$$\mathbf{s}(\mathbf{r}, t) = \frac{\hbar}{2im}[\psi^*(\mathbf{r}, t)\nabla\psi(\mathbf{r}, t) - \psi(\mathbf{r}, t)\nabla\psi^*(\mathbf{r}, t)] \quad (\text{B.12})$$

which is related to the current density of the particle $\mathbf{j} = q\mathbf{s}$. Here q is the electric charge of the particle (for electron $q = -e$).

The following conservation law holds

$$\frac{\partial\rho(\mathbf{r}, t)}{\partial t} = -\nabla \cdot \mathbf{s}(\mathbf{r}, t) \quad (\text{B.13})$$

The time dependence of an operator A is given by (Heisenberg):

$$\frac{dA}{dt} = \frac{\partial A}{\partial t} + \frac{[A, H]}{i\hbar} \quad (\text{B.14})$$

where H is the Hamiltonian and $[A, B] \equiv AB - BA$ is the commutator of A and B . For hermitian operators the commutator is always complex. If $[A, B] = 0$, the operators A and B have a common set of eigen functions. By applying this to \mathbf{p} and \mathbf{r} it follows (Ehrenfest)

$$m \frac{d^2 \langle \mathbf{r} \rangle_t}{dt^2} = -\langle \nabla V \rangle \quad (\text{B.15})$$

which is the classical Newton's second law of motion.

A classical product AB becomes $\frac{1}{2}(AB + BA)$ in quantum mechanics.

B.6 The Uncertainty Principle

The uncertainty ΔA in A is defined as

$$(\Delta A)^2 = \langle \psi | A - \langle A \rangle |^2 \psi \rangle = \langle A^2 \rangle - \langle A \rangle^2$$

it follows

$$\Delta A \cdot \Delta B \geq \frac{1}{2} |\langle \psi | [A, B] | \psi \rangle| \quad (\text{B.16})$$

from which it follows:

$$\Delta E \cdot \Delta t \geq \frac{\hbar}{2}, \quad \Delta p_x \cdot \Delta x \geq \frac{\hbar}{2} \quad (\text{B.17})$$

B.7 Parity

The parity operator in one dimension is given by $\mathcal{P}\psi(x) = \psi(-x)$. If the wave function is split into even and odd functions, it can be expanded into eigen functions of \mathcal{P} :

$$\psi(x) = \underbrace{\frac{1}{2}[\psi(x) + \psi(-x)]}_{\text{even: } \psi^+} + \underbrace{\frac{1}{2}[\psi(x) - \psi(-x)]}_{\text{odd: } \psi^-} \quad (\text{B.18})$$

with

$$\psi^+ = \frac{1}{2}(1 + \mathcal{P})\psi(x), \quad \psi^- = \frac{1}{2}(1 - \mathcal{P})\psi(x)$$

both of which satisfy the Schrödinger equation. Hence, parity is a conserved quantity. Moreover, $[\mathcal{P}, H] = 0$.

B.8 The Tunneling Effect

The wave function of a particle in a one-dimensional infinitely high potential step from $x = 0$ to $x = a$ is given by

$$\psi_n(x) = \sqrt{\frac{2}{a}} \sin(k_n x) \quad (\text{B.19})$$

where $k_n = n\pi/a$, n is an integer. The energy levels are given by

$$E_n = \frac{n^2 \hbar^2 \pi^2}{2ma^2}$$

where m is the mass of the particle.

If the wave function with energy E meets a potential barrier of W , while $E < W$, the wave function will, unlike the classical case, be non-zero within the potential barrier. Consider a one-dimensional system along the x axis. Denote the potential barrier as

$$V(x) = \begin{cases} 0 & \text{for } x < 0 \\ W & \text{for } 0 \leq x < a \\ 0 & \text{for } x \geq a \end{cases} \quad (\text{B.20})$$

For an wave Ae^{ikx} running into the potential barrier from $x = -\infty$, we express the total wave function as

$$\psi(x) = \begin{cases} Ae^{ikx} + Be^{-ikx} & \text{for } x < 0 \\ Ce^{iqx} + De^{-iqx} & \text{for } 0 \leq x < a \\ A'e^{ikx} & \text{for } x \geq a \end{cases} \quad (\text{B.21})$$

with $q^2 = 2m(E - W)/\hbar^2$ and $k^2 = 2mE/\hbar^2$. Boundary conditions require that $\psi(x)$ and $\partial\psi(x)/\partial x$ continuous at $x = 0$ and $x = a$, which give B , C and D and A' in terms of A . The transmission coefficient T of the incident wave Ae^{ikx} through the potential barrier is defined by

$$T = \frac{|A'|^2}{|A|^2} = \left| \frac{(1 + k/q)^2 - (1 - k/q)^2}{(1 + k/q)^2 e^{-iqa} - (1 - k/q)^2 e^{iqa}} \right|^2 \quad (\text{B.22})$$

which is not zero even when $E < W$. The nonzero transmission coefficient is the tunneling effect. Moreover, $T = 1$ is not always true when the energy of the incident wave is higher than the potential barrier (i.e., $E \geq W$). $T = 1$ occurs only when $e^{2iqa} = 1$, i.e., $2qa = 2n\pi$, n is an integer.

B.9 Harmonic Oscillator

For the one-dimensional potential energy

$$V(x) = \frac{1}{2}bx^2 \quad (\text{B.23})$$

Let $\omega^2 = b/m$, the Hamiltonian H is then given by:

$$H = \frac{p^2}{2m} + \frac{1}{2}m\omega^2x^2 = \frac{1}{2}\hbar\omega + \omega A^\dagger A \quad (\text{B.24})$$

with

$$A = \sqrt{\frac{m}{2\omega}}x + \frac{ip}{\sqrt{2m\omega}}, \quad A^\dagger = \sqrt{\frac{m}{2\omega}}x - \frac{ip}{\sqrt{2m\omega}} \quad (\text{B.25})$$

$A \neq A^\dagger$ is non hermitian. $[A, A^\dagger] = \hbar$ and $[A, H] = \hbar\omega A$. A is a so called creation operator, and A^\dagger an annihilation operator. $HAu_E = (E - \hbar\omega)Au_E$. There is a ground-state u_0 such $Au_0 = 0$. The energy in this ground state is $\frac{1}{2}\hbar\omega$, which is normally known as the zero-point energy. Let n be a positive integer, normalized eigen functions and corresponding eigen values are

$$u_n = \frac{1}{\sqrt{n!}} \left(\frac{A^\dagger}{\sqrt{\hbar}} \right)^n u_0, \quad u_0 = \sqrt[4]{\frac{m\omega}{\pi\hbar}} \exp\left(-\frac{m\omega x^2}{2\hbar}\right), \quad E_n = \left(\frac{1}{2} + n\right)\hbar\omega \quad (\text{B.26})$$

B.10 Angular Momentum and Spin

The orbital angular momentum operator is defined as

$$\mathbf{L} = -i\hbar\mathbf{r} \times \nabla \quad (\text{B.27})$$

$[L_z, L^2] = [L_z, H] = [L^2, H] = 0$, $[L_x, L_y] = i\hbar L_z$, $[L_y, L_z] = i\hbar L_x$, $[L_z, L_x] = i\hbar L_y$. Not all components of \mathbf{L} can be known at the same time with arbitrary accuracy

$$\Delta L_x \Delta L_y \geq \frac{\hbar}{2} L_z \quad (\text{B.28})$$

L_z in spherical polar and Cartesian coordinates is

$$L_z = x\mathbf{p}_y - y\mathbf{p}_x - i\hbar \frac{\partial}{\partial \varphi} = -i\hbar \left(x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right) \quad (\text{B.29})$$

The creation and annihilation operators L_\pm are defined by: $L_\pm = L_x \pm iL_y$. $L^2 = L_+L_- + L_z^2 - \hbar L_z$.

Eigen value relations of angular momentum \mathbf{L} with eigen function $Y_{\ell m}$

$$\begin{aligned} \mathbf{L}^2 Y_{\ell m} &= \ell(\ell + 1)\hbar^2 Y_{\ell m} \\ L_z Y_{\ell m} &= m\hbar Y_{\ell m} \end{aligned} \quad (\text{B.30})$$

where $\ell \geq 0$ and $2\ell + 1$ is an integer, $-\ell \leq m \leq \ell$ and can take on the values

$$m = -\ell, -\ell + 1, -\ell + 2, \dots, \ell - 1, \ell$$

$Y_{\ell m}$ is the spherical harmonics

$$Y_{\ell m}(\theta, \psi) = N_{\ell m} P_{\ell}^{|m|}(\cos \theta) e^{im\psi} \quad (\text{B.31})$$

where $P_{\ell}^{|m|}(\cos \theta)$ is the associated Legendre polynomials.

For integral values of ℓ , we discuss orbit angular momentum,

$$\begin{aligned} L_+ Y_{\ell m} &= \sqrt{\ell(\ell + 1) - m(m + 1)}\hbar Y_{\ell m+1} \\ L_- Y_{\ell m} &= \sqrt{\ell(\ell + 1) - m(m - 1)}\hbar Y_{\ell m-1} \end{aligned} \quad (\text{B.32})$$

Addition theorem for angular momentum:

$$\mathbf{L} = \mathbf{L}_1 + \mathbf{L}_2 \quad (\text{B.33})$$

\mathbf{L} has eigen functions $Y_{\ell m}$ with

$$\ell = |\ell_1 - \ell_2|, |\ell_1 - \ell_2| + 1, \dots, \ell_1 + \ell_2, \quad m = -\ell, -\ell + 1, \dots, \ell \quad (\text{B.34})$$

When ℓ takes on the values of half-odd integral, i.e., $\ell = 1/2, 3/2, \dots$, we discuss spin. Spin operators are defined by their commutation relations: $[S_x, S_y] = i\hbar S_z$, they do not act in the physical space (x, y, z) . Furthermore, $[\mathbf{L}, \mathbf{S}] = 0$ so that spin and angular momentum operators do not have a common set of eigen functions. The spin operators are given by $\mathbf{S} = \frac{1}{2}\hbar\boldsymbol{\sigma}$, where

$$\boldsymbol{\sigma}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \boldsymbol{\sigma}_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \boldsymbol{\sigma}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (\text{B.35})$$

are Pauli spin matrices. Denote the eigen function of spin as χ ,

$$\mathbf{S}^2 \chi_{s, m_s} = s(s + 1)\hbar^2 \chi_{s, m_s}, \quad s = \frac{1}{2} \quad (\text{B.36})$$

We normally denote $\chi_{\frac{1}{2}, \frac{1}{2}} = \alpha$ and $\chi_{\frac{1}{2}, -\frac{1}{2}} = \beta$.

Addition of two spins $\mathbf{S} = \mathbf{S}_1 + \mathbf{S}_2$,

$$\left\{ \begin{array}{l} \text{triplet, parity} = 1 \\ \text{singlet, parity} = -1 \end{array} \right. \left\{ \begin{array}{l} \chi_{1,1} = \alpha_1 \alpha_2 \\ \chi_{1,0} = \frac{1}{\sqrt{2}}(\alpha_1 \beta_2 + \beta_1 \alpha_2) \\ \chi_{1,-1} = \beta_1 \beta_2 \\ \chi_{0,0} = \frac{1}{\sqrt{2}}(\alpha_1 \beta_2 - \beta_1 \alpha_2) \end{array} \right. \quad (\text{B.37})$$

The electron has an intrinsic magnetic dipole moment \mathbf{M} due to its spin, $\mathbf{M} = -eg_s\mathbf{S}/2m$, with $g_s = 2(1 + \alpha/2\pi + \dots)$ is the gyromagnetic ratio. In the presence of an external magnetic field this gives a potential energy $V = -\mathbf{M} \cdot \mathbf{B}$. The Schrödinger equation then becomes (because $\partial\chi/\partial x_i \equiv 0$):

$$i\hbar \frac{\partial\chi(t)}{\partial t} = \frac{eg_s\hbar}{4m} \boldsymbol{\sigma} \cdot \mathbf{B}\chi(t) \quad (\text{B.38})$$

with $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$. If $\mathbf{B} = B\mathbf{e}_z$ there are two eigenvalues for this equation, $\pm eg_s\hbar B/4m = \pm\hbar\omega$, and the general solution is given by $\chi(t) = Ae^{-i\omega t} + Be^{i\omega t}$. From these,

$$\langle S_x \rangle = \frac{1}{2}\hbar \cos(2\omega t), \quad \langle S_y \rangle = \frac{1}{2}\hbar \sin(2\omega t) \quad (\text{B.39})$$

Thus the spin precesses about the z axis with frequency 2ω . This causes the Zeeman splitting of spectral lines.

B.11 Hydrogen Atom

The hydrogen atom contains one proton and one electron. The Schrödinger equation becomes a one-particle equation after the center-of-mass motion is separated out. In spherical coordinate, the potential energy is

$$V(r) = -\frac{e^2}{4\pi\epsilon r} \quad (\text{B.40})$$

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + V(r) \right] \psi(\mathbf{r}) = E\psi(\mathbf{r}) \quad (\text{B.41})$$

where m is the reduced mass which is approximately the same as the electron mass m_0 .

The solutions of the above Schrödinger equation in spherical coordinates if the potential energy is a function of r can be written as

$$\psi(r, \theta, \varphi) = R_{n\ell}(r) Y_{\ell m}(\theta, \varphi)$$

$Y_{\ell m}$ is the spherical harmonics in Eq. (B.31). Let $u_{n\ell}(r) = rR_{n\ell}(r)$, the equation that determines $u_{n\ell}(r)$ is

$$\frac{d^2 u_{n\ell}(r)}{dr^2} + \frac{2m}{\hbar^2} \left[E_n - V(r) - \frac{\ell(\ell+1)\hbar^2}{2mr^2} \right] u_{n\ell}(r) = 0 \quad (\text{B.42})$$

Solution of the above equation for $V(r)$ in Eq. (B.40) when $\epsilon = \epsilon_0$ and $m = m_0$

$$E_n = -\frac{R_y}{n^2} \quad (\text{B.43})$$

where $R_y = \hbar^2/2m_0a_0^2 = 13.6058$ eV is Rydberg, $a_0 = 4\pi\epsilon_0\hbar^2/m_0e^2 = 0.529$ Å is Bohr radius. The parity of these solutions is $(-1)^\ell$, and the functions are

$$2 \sum_{\ell=0}^{n-1} (2\ell + 1) = 2n^2$$

fold degenerated.

B.12 Interaction with Electromagnetic Fields

The Hamiltonian of an electron in an electromagnetic field is given by:

$$H = \frac{1}{2\mu}(\mathbf{p} + e\mathbf{A})^2 - e\phi = -\frac{\hbar^2}{2\mu}\nabla^2 + \frac{i\hbar e}{2\mu}(\mathbf{A} \cdot \nabla + \nabla \cdot \mathbf{A}) + \frac{e^2}{2\mu}A^2 - e\phi \quad (\text{B.44})$$

where μ is the reduced mass of the system, \mathbf{A} is the vector field and ϕ is the scalar field of the electromagnetic field. The term $\sim A^2$ can usually be neglected, except for very strong fields or macroscopic motions.

B.13 Time-Independent Perturbation Theory

Consider a time-independent perturbation V' so that the Schrödinger equation becomes $(H_0 + \lambda V')\Psi_n = E_n\Psi_n$. Let ψ_n be the complete set of eigen functions of the non-perturbed Hamiltonian H_0 , i.e., $H_0\psi_n = E_n^0\psi_n$. We write

$$\Psi_n = \psi_n + \sum_{k \neq n} c_{nk}(\lambda)\psi_k \quad (\text{B.45})$$

Expanding c_{nk} and E_n into λ

$$\begin{aligned} c_{nk} &= \lambda c_{nk}^{(1)} + \lambda^2 c_{nk}^{(2)} + \dots \\ E_n &= E_n^0 + \lambda E_n^{(1)} + \lambda^2 E_n^{(2)} + \dots \end{aligned} \quad (\text{B.46})$$

inserting them into the Schrödinger equation result in the first-order correction

$$\begin{aligned} E_n^{(1)} &= \langle \phi_n | V' | \phi_n \rangle, & c_{nk, k \neq n}^{(1)} &= \frac{\langle \psi_k | V' | \psi_n \rangle}{E_n^0 - E_k^0} \\ \Psi_n &= \psi_n + \sum_{k \neq n} \frac{\langle \psi_k | V' | \psi_n \rangle}{E_n^0 - E_k^0} \psi_k \end{aligned} \quad (\text{B.47})$$

where $m \neq n$, and the second-order correction of the energy

$$E_n^{(2)} = \sum_{k \neq n} \frac{|\langle \psi_k | V' | \psi_n \rangle|^2}{E_n^0 - E_k^0} \quad (\text{B.48})$$

B.14 Time-Dependent Perturbation Theory

When the perturbation is time-dependent, i.e., $V'(t)$, the Schrödinger equation is

$$i\hbar \frac{\partial \Psi(t)}{\partial t} = [H_0 + \lambda V'(t)] \Psi(t) \quad (\text{B.49})$$

and

$$\Psi(t) = \sum_n c_n(t) \exp\left(\frac{-iE_n^0 t}{\hbar}\right) \psi_n \quad (\text{B.50})$$

with $c_n(t) = \delta_{nk} + \lambda c_n^{(1)}(t) + \dots$. The first-order correction follows

$$c_n^{(1)}(t) = \frac{1}{i\hbar} \int_0^t \langle \psi_n | V'(\tau) | \psi_k \rangle \exp\left[\frac{i(E_n^0 - E_k^0)\tau}{\hbar}\right] d\tau \quad (\text{B.51})$$

B.15 N -Particle System

Identical particles are indistinguishable. For the total wave function of a system of identical particles,

1. Particles with a half-odd integer spin (Fermions): Ψ_{total} must be anti-symmetric with respect to interchanges of the coordinates (spatial and spin) of each pair of particles. The Pauli principle results from this: two Fermions cannot exist in an identical state because then $\Psi_{\text{total}} = 0$.
2. Particles with an integer spin (Bosons): Ψ_{total} must be symmetric with respect to interchange of the coordinates (spatial and spin) of each pair of particles.

For a system of two electrons there are 2 possibilities for the spatial wave function. When a and b are the quantum numbers of electron 1 and 2,

$$\begin{aligned} \psi_S(1, 2) &= \psi_a(1)\psi_b(2) + \psi_a(2)\psi_b(1) \\ \psi_A(1, 2) &= \psi_a(1)\psi_b(2) - \psi_a(2)\psi_b(1) \end{aligned} \quad (\text{B.52})$$

Following spin wave functions are possible:

$$\chi_A = \frac{1}{\sqrt{2}} [\chi_+(1)\chi_-(2) - \chi_+(2)\chi_-(1)] \quad (\text{B.53})$$

$$\chi_S = \begin{cases} \chi_+(1)\chi_+(2) \\ \frac{1}{\sqrt{2}}[\chi_+(1)\chi_-(2) + \chi_+(2)\chi_-(1)] \\ \chi_-(1)\chi_-(2) \end{cases}$$

Because the total wave function must be anti-symmetric, $\Psi_{\text{total}} = \psi_S \chi_A$, or $\Psi_{\text{total}} = \psi_A \chi_S$.

For N particles the symmetric spatial function is given by:

$$\psi_S(1, \dots, N) = \sum \psi(\text{all permutations of } 1, \dots, N) \quad (\text{B.54})$$

The anti-symmetric wave function is given by the Slater determinant

$$\psi_A(1, \dots, N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_{E_1}(1) & \psi_{E_1}(2) & \cdots & \psi_{E_1}(N) \\ \psi_{E_2}(1) & \psi_{E_2}(2) & \cdots & \psi_{E_2}(N) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{E_N}(1) & \psi_{E_N}(2) & \cdots & \psi_{E_N}(N) \end{vmatrix} \quad (\text{B.55})$$

B.16 Quantum Statistics

If a system exists in a state in which one has not the disposal of the maximal amount of information about the system, it can be described by a *density matrix* ρ . If the probability that the system is in state Ψ_i is given by c_i , one can write for the expectation value a of A

$$a = \langle A \rangle = \sum_i c_i \langle \Psi_i | A | \Psi_i \rangle \quad (\text{B.56})$$

If Ψ is expanded into an orthonormal basis $\{\psi_k\}$ as $\psi^{(i)} = \sum_k c_k^{(i)} \phi_k$,

$$\langle A \rangle = \sum_k (A\rho)_{kk} = \text{Tr}(A\rho) \quad (\text{B.57})$$

where $\rho_{\ell k} = c_k^* c_\ell$. ρ is hermitian, with $\text{Tr}(\rho) = 1$. Further holds

$$\rho = \sum r_i |\psi_i\rangle \langle \psi_i|$$

The probability to find eigenvalue a_n when measuring A is given by ρ_{nn} if one uses a basis of eigen vectors of A for $\{\phi_k\}$. For the time-dependence holds (in the Schrödinger image operators are not explicitly time-dependent):

$$i\hbar \frac{d\rho}{dt} = [H, \rho] \quad (\text{B.58})$$

For a macroscopic system in equilibrium holds $[H, \rho] = 0$. If all quantum states with the same energy are equally probable: $P_i = P(E_i)$, one can obtain the distribution:

$$P_n(E) = \rho_{nn} = \frac{e^{-E_n/k_B T}}{Z} \quad \text{with the state sum } Z = \sum_n e^{-E_n/k_B T} \quad (\text{B.59})$$

The thermodynamic quantities are related to these definitions as follows:

$$\begin{aligned} F &= -k_B T \ln(Z) \\ U = \langle H \rangle &= \sum_n p_n E_n = -\frac{\partial}{\partial k_B T} \ln(Z) \\ S &= -k_B \sum_n P_n \ln(P_n) \end{aligned} \quad (\text{B.60})$$

For a mixed state of M orthonormal quantum states with probability $1/M$ follows: $S = k_B \ln(M)$.

The distribution function for the internal states for a system in thermal equilibrium is the most probable function. This function can be found by taking the maximum of the function which gives the number of states with Stirling's equation:

$$\ln(n!) \approx n \ln(n) - n$$

and the conditions $\sum_k n_k = N$ and $\sum_k n_k W_k = W$. For identical particles which obey the Pauli exclusion principle the possible number of states is given by:

$$P = \prod_k \frac{g_k!}{n_k!(g_k - n_k)!} \quad (\text{B.61})$$

This results in the *Fermi-Dirac statistics*. For indistinguishable particles which *do not* obey the exclusion principle the possible number of states is given by:

$$P = N! \prod_k \frac{g_k^{n_k}}{n_k!} \quad (\text{B.62})$$

This results in the *Bose-Einstein statistics*.

The distribution functions which explain how particles are distributed over the different one-particle states k which are each g_k -fold degenerate depend on the spin of the particles. They are given by:

1. Fermi-Dirac statistics: $n_k = \frac{g_k}{\exp[(E_k - E_f)/k_B T] + 1}$
2. Bose-Einstein statistics: $n_k = \frac{g_k}{\exp[(E_k - E_f)/k_B T] - 1}$

where E_f is the Fermi-energy.

Appendix C

Electricity & Magnetism

C.1 The Maxwell Equations

The classical electromagnetic field is described by electric displacement \mathbf{D} , polarization \mathbf{P} , electric field strength \mathbf{E} , magnetic field strength \mathbf{H} , the magnetization \mathbf{M} and the magnetic flux density \mathbf{B} , via the Maxwell equations, in integral and differential forms:

$$\begin{aligned}
 \oiint_S \mathbf{D} \cdot d\mathbf{S} &= \iiint_{\Omega} \rho d\Omega & \nabla \cdot \mathbf{D} &= \rho \\
 \oiint_S \mathbf{B} \cdot d\mathbf{S} &= 0 & \nabla \cdot \mathbf{B} &= 0 \\
 \oint_{\ell} \mathbf{E} \cdot d\boldsymbol{\ell} &= -\frac{\partial}{\partial t} \iint_S \mathbf{B} \cdot d\mathbf{S} & \nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t} \\
 \oint_{\ell} \mathbf{H} \cdot d\boldsymbol{\ell} &= \iint_S \left(\mathbf{j} + \frac{\partial \mathbf{D}}{\partial t} \right) \cdot d\mathbf{S} & \nabla \times \mathbf{H} &= \mathbf{j} + \frac{\partial \mathbf{D}}{\partial t}
 \end{aligned} \tag{C.1}$$

In the first two integral equations, Ω is the volume totally enclosed by surface S ; In the last two integral equations, S is the surface totally enclosed by line ℓ . In the above equations, ρ is the density of free electric charge and \mathbf{j} is the electric current density. \mathbf{D} , \mathbf{P} and \mathbf{E} depend on each other according to $\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} = \epsilon \mathbf{E}$, while \mathbf{H} , \mathbf{M} and \mathbf{B} depend on each other according to $\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M}) = \mu \mathbf{H}$.

There is the law of charge conservation which is implicit in the above Maxwell equations (also in its integral and differential forms)

$$\oiint_S \mathbf{j} \cdot d\mathbf{S} + \frac{\partial}{\partial t} \oiint_S \mathbf{D} \cdot d\mathbf{S} = 0, \quad \nabla \cdot \mathbf{j} + \frac{\partial \rho}{\partial t} = 0 \tag{C.2}$$

C.2 Force and Potential

The vector fields from a particle of charge Q_1 and velocity \mathbf{v} are

$$\mathbf{E}_1 = \frac{Q_1}{4\pi\epsilon_0} \frac{\mathbf{r}}{r^3} \left(1 - \frac{v^2}{c_0^2}\right), \quad \mathbf{B}_1 = \frac{\mu_0 Q_1}{4\pi} \frac{\mathbf{v} \times \mathbf{r}}{r^3} \left(1 - \frac{v^2}{c_0^2}\right) \quad (\text{C.3})$$

The force between two stationary (or quasi stationary, i.e., their velocities are far less than the speed of light) point charges Q_1 and Q_2 in a homogeneous space with a dielectric constant ϵ are given by:

$$\mathbf{F} = Q_2 \mathbf{E}_1 = \frac{Q_1 Q_2}{4\pi\epsilon_0} \frac{\mathbf{r}}{r^3} \quad (\text{C.4})$$

A particle with charge Q_1 experiences the so-called Lorentz force when moving at an instantaneous velocity \mathbf{v} in a \mathbf{B} field

$$\mathbf{F} = Q_1 \mathbf{v} \times \mathbf{B} \quad (\text{C.5})$$

The electromagnetic field can be described by a vector potential \mathbf{A} and a scalar potential ϕ

$$\mathbf{E} = -\nabla\phi - \frac{\partial\mathbf{A}}{\partial t}, \quad \mathbf{B} = \nabla \times \mathbf{A} \quad (\text{C.6})$$

The magnetic field, resulting from an electric current $\mathbf{j} = j d\ell$ is given by the law of Biot-Savart, also known as the law of Laplace.

$$\mathbf{B} = \int \frac{\mu_0 j}{4\pi} \frac{d\ell \times \mathbf{r}}{r^3} \quad (\text{C.7})$$

where \mathbf{r} is the spatial vector from the current $j d\ell$ to the location at which \mathbf{B} is being calculated. If the current is time-dependent one has to take retardation into account that the substitution $j(t) \rightarrow j(t - r/c)$ has to be applied.

C.3 Electromagnetic Waves

By introducing a vector potential, \mathbf{A} , and a scalar potential, ϕ , the electric and magnetic fields can be obtained from the relations of

$$\begin{aligned} \mathbf{E} &= -\nabla\phi - \frac{\partial\mathbf{A}}{\partial t} \\ \mathbf{B} &= \nabla \times \mathbf{A} \end{aligned} \quad (\text{C.8})$$

The first and last Maxwell equations are automatically satisfied by the definitions. By the relation of

$$\nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}$$

and in the Lorentz gauge of

$$\frac{1}{\mu} \nabla \cdot \mathbf{A} + \epsilon \frac{\partial \phi}{\partial t} = 0 \quad (\text{C.9})$$

we have the following equations for the vector and scalar potentials

$$\nabla^2 \mathbf{A} - \epsilon \mu \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mu \mathbf{J} \quad (\text{C.10})$$

$$\nabla^2 \phi - \epsilon \mu \frac{\partial^2 \phi}{\partial t^2} = -\frac{\rho}{\epsilon} \quad (\text{C.11})$$

The wave equation $\square \Psi(\mathbf{r}, t) = -f(\mathbf{r}, t)$ has the following general solution

$$\Psi(\mathbf{r}, t) = \int \frac{f(\mathbf{r}', t - |\mathbf{r} - \mathbf{r}'|/c)}{4\pi |\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \quad (\text{C.12})$$

where $c = 1/\sqrt{\epsilon\mu}$. When $\mathbf{J}(\mathbf{r}, t)$ and $\rho(\mathbf{r}, t)$ can be expressed as $\mathbf{J}(\mathbf{r}) \exp(-i\omega t)$ and $\rho(\mathbf{r}) \exp(-i\omega t)$, respectively, $\mathbf{A}(\mathbf{r}, t)$ and $\phi(\mathbf{r}, t)$ have the similar forms of $\mathbf{A}(\mathbf{r}) \exp(-i\omega t)$ and $\phi(\mathbf{r}) \exp(-i\omega t)$ with:

$$\begin{aligned} \mathbf{A}(\mathbf{r}) &= \frac{\mu}{4\pi} \int \mathbf{J}(\mathbf{r}') \frac{\exp[i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}')] }{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \\ \phi(\mathbf{r}) &= \frac{1}{4\pi\epsilon} \int \rho(\mathbf{r}') \frac{\exp[i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}')] }{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \end{aligned} \quad (\text{C.13})$$

An ideal dipole that oscillates in time

$$\mathbf{p}(t) = \mathbf{p}_0 \cos(\omega t) \quad (\text{C.14})$$

The electric and magnetic fields, the Poynting flux, and angular distribution of this oscillating dipole are

$$\begin{aligned} \mathbf{E} &= -\theta_0 \frac{p_0}{4\pi\epsilon_0 r} \frac{\omega^2}{c_0^2} \sin\theta \cos(\omega t - \mathbf{k} \cdot \mathbf{r}) \\ \mathbf{B} &= -\phi_0 \frac{\mu_0 p_0}{4\pi r} \frac{\omega^2}{c_0} \sin\theta \cos(\omega t - \mathbf{k} \cdot \mathbf{r}) \\ \mathbf{S} &= \mathbf{r}_0 \frac{p_0^2}{16\pi^2 \epsilon_0 r^2} \frac{\omega^4}{c^3} \sin^2\theta \cos^2(\omega t - \mathbf{k} \cdot \mathbf{r}) \\ \langle \mathbf{S} \rangle_t &= \mathbf{r}_0 \frac{p_0^2 \sin^2\theta}{32\pi^2 \epsilon_0 r^2} \frac{\omega^4}{c^3} \\ \frac{dP}{d\Omega} &= r^2 \mathbf{r}_0 \cdot \langle \mathbf{S} \rangle_t = \frac{p_0^2 \sin^2\theta}{32\pi^2 \epsilon_0 r^2} \frac{\omega^4}{c^3} \end{aligned} \quad (\text{C.15})$$

in spherical coordinate, where \mathbf{r}_0 , $\boldsymbol{\theta}_0$ and $\boldsymbol{\phi}_0$ are the three unit vectors. Note that the above expressions are valid for the following conditions: the dipole dimension is much less than the distance to the dipole, it is also very small compared with the wavelength of radiation. The distance to the dipole is much longer than the wavelength.

The wave equations in matter, with $c_{\text{mat}} = (\varepsilon\mu)^{-1/2}$ the light speed in matter, are:

$$\begin{aligned} \left(\nabla^2 - \varepsilon\mu \frac{\partial^2}{\partial t^2} - \frac{\mu}{\rho} \frac{\partial}{\partial t} \right) \mathbf{E} &= 0 \\ \left(\nabla^2 - \varepsilon\mu \frac{\partial^2}{\partial t^2} - \frac{\mu}{\rho} \frac{\partial}{\partial t} \right) \mathbf{B} &= 0 \end{aligned} \quad (\text{C.16})$$

give, after substitution of monochromatic plane waves:

$$\mathbf{E} = \mathbf{E}_0 \exp[i(\mathbf{k} \cdot \mathbf{r} - \omega t)] \quad \text{and} \quad \mathbf{B} = \mathbf{B}_0 \exp[i(\mathbf{k} \cdot \mathbf{r} - \omega t)]$$

the dispersion relation:

$$k^2 = \varepsilon\mu\omega^2 + \frac{i\mu\omega}{\rho} \quad (\text{C.17})$$

The first term arises from the displacement current, the second from the conductance current. If k is written in the form $k = k' + ik''$ it follows that:

$$k' = \omega \sqrt{\frac{1}{2}\varepsilon\mu} \sqrt{1 + \sqrt{1 + \frac{1}{(\rho\varepsilon\omega)^2}}} \quad \text{and} \quad k'' = \omega \sqrt{\frac{1}{2}\varepsilon\mu} \sqrt{-1 + \sqrt{1 + \frac{1}{(\rho\varepsilon\omega)^2}}} \quad (\text{C.18})$$

This results in a damped wave:

$$\mathbf{E} = \mathbf{E}_0 \exp(-k'' \mathbf{n} \cdot \mathbf{r}) \exp[i(k' \mathbf{n} \cdot \mathbf{r} - \omega t)] \quad (\text{C.19})$$

Appendix D

Solid State Physics

D.1 Crystal Structure

A lattice is defined by the 3 translation vectors \mathbf{a}_i , so that the atomic composition looks the same from each point \mathbf{r} and $\mathbf{r}' = \mathbf{r} + \mathbf{R}$, where \mathbf{R} is a translation vector given by: $\mathbf{R} = u_1\mathbf{a}_1 + u_2\mathbf{a}_2 + u_3\mathbf{a}_3$ with u_i are integers. A lattice can be constructed from primitive cells. As a primitive cell one can take a parallelepiped, with volume

$$\Omega_{\text{cell}} = \mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3) \tag{D.1}$$

Because a lattice has a periodical structure the physical properties n which are connected with the lattice have the same periodicity (neglecting boundary effects):

$$n(\mathbf{r} + \mathbf{R}) = n(\mathbf{r}) \tag{D.2}$$

This periodicity is suitable to use Fourier analysis: $n(\mathbf{r})$ is expanded as:

$$n(\mathbf{r}) = \sum_{\mathbf{G}} n_{\mathbf{G}} \exp(i\mathbf{G} \cdot \mathbf{r}) \tag{D.3}$$

with

$$n_{\mathbf{G}} = \frac{1}{\Omega_{\text{cell}}} \int_{\text{cell}} n(\mathbf{r}) \exp(-i\mathbf{G} \cdot \mathbf{r}) d\mathbf{r} \tag{D.4}$$

\mathbf{G} is the *reciprocal lattice vector*. If \mathbf{G} is written as $\mathbf{G} = v_1\mathbf{b}_1 + v_2\mathbf{b}_2 + v_3\mathbf{b}_3$ with v_i as integers, it follows for the vectors \mathbf{b}_i , cyclically:

$$\mathbf{b}_i = 2\pi \frac{\mathbf{a}_{i+1} \times \mathbf{a}_{i+2}}{\mathbf{a}_i \cdot (\mathbf{a}_{i+1} \times \mathbf{a}_{i+2})} \tag{D.5}$$

The set of \mathbf{G} -vectors determines the Röntgen diffractions: a maximum in the reflected radiation occurs if: $\Delta\mathbf{k} = \mathbf{G}$ with $\Delta\mathbf{k} = \mathbf{k} - \mathbf{k}'$. So: $2\mathbf{k} \cdot \mathbf{G} = G^2$. From this follows for parallel lattice planes (Bragg reflection) that for the maxima holds: $2d \sin(\theta) = n\lambda$.

The Brillouin zone is defined as a Wigner-Seitz cell in the reciprocal lattice.

D.2 Crystal Binding

A distinction can be made between 4 binding types:

1. van der Waals bond
2. Ion bond
3. Covalent or homopolar bond
4. Metallic bond

The interaction in a covalent bond depends on the relative spin orientations of the electrons constituting the bond. The potential energy for two parallel spins is higher than the potential energy for two antiparallel spins. Furthermore the potential energy for two parallel spins has sometimes no minimum. In that case binding is not possible.

D.3 Crystal Vibrations

For a lattice with one type of atoms and only nearest-neighbor interactions are taken into account, the force on atom s with mass M can then be written as:

$$F_s = M \frac{d^2 u_s}{dt^2} = C(u_{s+1} - u_s) + C(u_{s-1} - u_s) \quad (\text{D.6})$$

Assuming that all solutions have the same time-dependence $\exp(-i\omega t)$ this results in:

$$-M\omega^2 u_s = C(u_{s+1} + u_{s-1} - 2u_s) \quad (\text{D.7})$$

Further it is postulated that: $u_{s\pm 1} = u \exp(isKa) \exp(\pm iKa)$. This gives: $u_s = \exp(iKsa)$. Substituting the latter two equations in the first results in a system of linear equations, which has only one solution if their determinant is 0. This gives:

$$\omega^2 = \frac{4C}{M} \sin^2\left(\frac{1}{2}Ka\right) \quad (\text{D.8})$$

Only vibrations with a wavelength within the first Brillouin Zone have a physical significance. This requires that $-\pi < Ka \leq \pi$. The group velocity of these vibrations is given by:

$$v_g = \frac{d\omega}{dK} = \sqrt{\frac{Ca^2}{M}} \cos\left(\frac{1}{2}Ka\right) \quad (\text{D.9})$$

and is 0 on the edge of a Brillouin Zone. Here, there is a standing wave.

For a lattice with two types of atoms, the solutions are:

$$\omega^2 = C \left(\frac{1}{M_1} + \frac{1}{M_2} \right) \pm C \sqrt{\left(\frac{1}{M_1} + \frac{1}{M_2} \right)^2 - \frac{4 \sin^2(Ka)}{M_1 M_2}} \quad (\text{D.10})$$

Connected with each value of K are two values of ω . The larger value describes the optical vibrational mode, the lower value the acoustical mode. In the optical mode, both types of ions oscillate in opposite phases, in the acoustical mode they oscillate in the same phase. This results in a much larger induced dipole moment for optical oscillations, and also a stronger emission and absorption of radiation. Furthermore each vibrational mode has three polarization directions, one longitudinal and two transversal.

D.4 Free Electron Fermi Gas

D.4.1 Thermal Heat Capacity

The solution with period L of the one-dimensional Schrödinger equation is:

$$\psi_n(x) = A \sin\left(\frac{2\pi x}{\lambda_n}\right)$$

with $n\lambda_n = 2L$. From this follows

$$E = \frac{\hbar^2}{2m} \left(\frac{n\pi}{L}\right)^2 \quad (\text{D.11})$$

In a linear lattice the only important quantum numbers are n and m . The *Fermi level* is the uppermost filled level, which has the *Fermi-energy* E_f . If n_f is the quantum number of the Fermi level, it can be expressed as: $2n_f = N$ so $E_f = \hbar^2\pi^2 N^2/8mL$. Here N is the total number of electrons. In three dimensions holds:

$$k_f = \left(\frac{3\pi^2 N}{\Omega}\right)^{1/3}, \quad E_f = \frac{\hbar^2}{2m} \left(\frac{3\pi^2 N}{\Omega}\right)^{2/3} \quad (\text{D.12})$$

The number of states with energy $\leq E$ is then: $N = \frac{\Omega}{3\pi^2} \left(\frac{2mE}{\hbar^2}\right)^{3/2}$, and the density of states becomes

$$D(E) = \frac{dN}{dE} = \frac{\Omega}{2\pi^2} \left(\frac{2m}{\hbar^2}\right)^{3/2} \sqrt{E} = \frac{3N}{2E} \quad (\text{D.13})$$

The heat capacity of the electrons is approximately 0.01 times the classical expected value $\frac{3}{2}Nk_B$. This is caused by the Pauli exclusion principle and the Fermi-Dirac distribution: only electrons within an energy range $\sim k_B T$ of the Fermi level are excited thermally. There is a fraction $\approx T/T_f$ excited thermally. The internal energy then becomes:

$$U \approx Nk_B T \frac{T}{T_f}, \quad C = \frac{\partial U}{\partial T} \approx Nk_B \frac{T}{T_f} \quad (\text{D.14})$$

A more accurate analysis gives:

$$C_{\text{electrons}} = \frac{1}{2} \pi^2 N k_B T / T_f \sim T$$

Together with the T^3 dependence of the thermal heat capacity of the phonons the total thermal heat capacity of metals is described by

$$C = \gamma T + AT^3$$

D.4.2 Electric Conductance

The equation of motion for the charge carriers is:

$$F = m \frac{d\mathbf{v}}{dt} = \hbar \frac{d\mathbf{k}}{dt} \quad (\text{D.15})$$

The variation of \mathbf{k} is given by

$$\delta\mathbf{k} = \mathbf{k}(t) - \mathbf{k}(0) = \frac{-e\mathbf{E}t}{\hbar}$$

If τ is the characteristic collision time of the electrons, $\delta\mathbf{k}$ remains stable if $t = \tau$.

$$\langle \mathbf{v} \rangle = \mu \mathbf{E} \quad (\text{D.16})$$

with $\mu = e\tau/m$ the mobility of the electrons. The current in a conductor is given by:

$$\mathbf{J} = nq\mathbf{v} = \sigma \mathbf{E} = \frac{\mathbf{E}}{\rho} = ne\mu \mathbf{E} \quad (\text{D.17})$$

D.5 Energy Bands

In the tight-bond approximation it is assumed that

$$\psi = e^{ikna} \phi(x - na)$$

from this follows for the energy:

$$\langle E \rangle = \langle \psi | H | \psi \rangle = E_{\text{at}} - \alpha - 2\beta \cos(ka)$$

This gives a cosine superimposed on the atomic energy, which can often be approximated by a harmonic oscillator. If it is assumed that the electron is nearly free one can postulate

$$\psi = e^{i\mathbf{k}\cdot\mathbf{r}}$$

i.e., a traveling wave. This wave can be decomposed into two standing waves:

$$\psi(+) = \exp(i\pi x/a) + \exp(-i\pi x/a) = 2 \cos(\pi x/a)$$

$$\psi(-) = \exp(i\pi x/a) - \exp(-i\pi x/a) = 2i \sin(\pi x/a)$$

The probability density $|\psi(+)|^2$ is high near the atoms of the lattice and low in between. The probability density $|\psi(-)|^2$ is low near the atoms of the lattice and high in between. Hence the energy of $\psi(+)$ is also lower than the energy of $\psi(-)$. Suppose that $V(x) = V \cos(2\pi x/a)$, then the bandgap is given by:

$$E_g = \int_0^1 V(x)[|\psi(+)|^2 - |\psi(-)|^2]dx = V \quad (\text{D.18})$$

Index

A

Acceleration effective mass, 31
Angular momentum, 4–5, 7, 142, 149,
388–389

B

Biaxially strain, 42
Bloch theorem, 1, 19–21, 50, 59, 75, 84, 130,
226, 353
Bloch wave function, 69, 363
Body-centered cubic lattics, 11
Bohr radius, 4, 6, 96–97, 141, 143–144, 160,
178, 324, 330, 348, 391
Boltzmann (electron wave) transport, 97, 100,
239
Bose-Einstein distribution, 394
Bowling parameter, 38
Bravais crystal, 10, 13, 44, 90, 137
Bravais lattice, 10, 13
Brillouin zone, 14, 32, 130, 135, 138, 288,
353, 399–400

C

Carrier-concentration effective mass, 32,
315–316
Cauchy stress tensor, 42
Cauchy's infinitesimal strain tensor, 42
Cayley form, 71, 377
Cladding layer, 347
Complementary metal-oxide-semiconductor,
CMOS, 190–191, 233, 253, 257–258
Complex conjugate, c.c., 114–116, 158, 211,
358
Compound semiconductor, 103
Conduction channel, 190, 237, 239–242,
244–247, 249–250, 252–254, 258
Conductivity effective mass, 31, 33

Core electron, 9–10, 302, 345–346
Coulomb blockade, 245, 258, 261
Crystal lattice and lattice basis, 10, 14, 17, 44,
89–91, 137
Crystal momentum, 118
Czochralski (CZ) technique, 15

D

Deformation potential, 34–35, 103, 170
 δ doping, 230–231, 233, 302, 345–346
Density of states, 32–33, 59–62, 82, 93, 98,
170, 180, 187, 200, 216, 233–236,
285–286, 338–339, 369, 373
Density-of-states effective mass, 32
Depletion layer, 219
Diamond crystal structure, 10
Dielectric coefficient, 323
Dielectric polarization, 158, 160, 162, 165,
324, 329
Diffusion coefficient, 90, 100–101, 192, 346
Diffusion length, 56
Distributed Bragg reflector, DBR, 399
Dynamic random access memory, DRAM, 2

E

Edge emission, 222
Effective mass, 29–33, 38, 53–54, 56–57, 93,
118, 136–137, 142–143, 147–148,
249–250, 276–280, 282–286, 306–307,
314–316, 353–355
Effective mass approximation, 53–54, 118,
137, 142, 147–148, 177, 307
Effective medium approximation, 25–26, 48
Einstein ratio, 101, 191
Einstein relation, 101, 191
Electron affinity, 24, 147, 188, 190
Electron-beam direct-writing technique, 3,
257–258

Electron-phonon interaction, 104–105,
107–108, 207, 209–213

Electron wavelength, 2

Elemental semiconductor, 103

Ellipsoidal band, 31

Energy band offset, 24

Energy dispersion relation, 85, 114

Envelope function, 1, 27–28, 52–53, 56–59,
92, 134, 140, 143–144, 160, 226, 275,
277, 311, 340, 353

Evanescent states, 108, 195, 209, 211–212

Exciton, *x*, 79, 111, 117, 134–135, 137–146,
157–163, 165–168, 172–173, 176, 298,
317, 323–325, 327–335, 348–349

Exciton binding energy, 141, 144, 146

Exciton Bohr radius, 141, 143, 324, 330, 348

Exciton polariton, 117, 162, 298, 323, 332

F

Fabry-Perot microcavity, 342

Face-centered cubic lattice, 11–12

Fermi level, 23, 89, 95–96, 102–103, 106–107,
187–188, 190, 208, 219, 228, 230–231,
315–316, 336–337, 353–355, 401

Field-effect transistor, *v*, 2, 185, 188, 230–231,
233, 239, 245, 248–249, 253, 255, 368

First-order perturbations, 28, 74, 78–79, 83,
85, 157–158, 167, 204

Fourier transform, 53, 138–139, 332

Frequency multiplier, 213–214

G

Gate length, 2, 236, 252

Gaussian wave packet, 71–72, 378

Gunn effect, 213

H

Harmonics, 5, 153, 179, 213, 295, 326,
389–390

Heterostructure, *vi*, 1, 3, 17–18, 39, 50, 53–56,
104, 185, 213–216, 218, 224, 249,
313–314, 343–344

Heterostructure barrier varactor, HBV, 185,
213–214, 216, 249

Heterostructure barrier varactor diode, HBV,
213

High-electron-mobility transistor, HEMT, 185,
224–226, 229–230, 243–245, 247, 263

I

Impurity state, 93–94, 96

Index guiding, 343

Infinitely deep quantum well, 62, 144–145,
272–274, 278, 305, 373

Interface diffusion, 18, 51

Inverse effective mass, 276–277, 279–280,
282–283

K

Kane parameter, 36

Kramers' theorem, 21

L

Large scale integration, LSI, 112, 262

Localized state, 107, 223, 366

Longitudinal effective mass, 31, 38, 276, 355

Luttinger parameters, 36

M

Maxwell equations, 113–114, 324, 395–396

Metal-organic chemical vapor deposition,
MOCVD, 3, 15, 18–19, 43, 310

Metal-oxide-semiconductor, MOSFET, *v*, 2,
188–191, 230–231, 252–255, 257, 262,
368

Misfit dislocation, 24, 40

Mobility, *vi*, 33, 67, 100–101, 185, 192,
224–226, 228–230, 233, 238, 243, 247,
263, 287, 320

Mole fraction, 26–27, 36, 47–48, 55

Molecular beam epitaxy, MBE, 3, 15, 17–19,
40, 200, 224, 230, 246, 310

Momentum operator, 27–28, 118, 141, 159,
384, 388

Monte Carlo scheme, 45, 170–172

Mott phase transition, 94, 354–355

N

Nanofabrication technique, 3

O

Optical absorption coefficient, 279, 302,
306–307

Optical grating, 271, 289, 294, 298, 307–308,
310, 312

Optical transition matrix, 35, 128, 131, 142,
163, 275–276, 302, 306, 322

Orbital angular momentum, 7, 388

Oscillator strength, 402

P

Pauli exclusion principle, 8, 68, 89, 93, 98,
117, 127, 193, 227, 296, 317, 394, 401

Photocurrent, 302, 305, 307–308, 310,
312–313, 316, 319–323
 Photonic bandgap, 328
 Photonic bandgap crystal, PBC, 323
 Photonic crystal, 309, 323
 Poisson effect, 234, 315, 368
 Population inversion, 128, 133, 167, 336–337,
342–343, 345
 Poynting vector, 114, 116–117, 165–166
 Primitive vector, 10–13, 20

Q

Quantum cascade laser, 345
 Quantum confinement, 57–58, 61, 133, 146,
160, 178, 243, 348
 Quantum dot, QD, 43, 45–47, 61–62,
133–134, 146–148, 150, 160–163,
167–174, 257–261, 312, 318, 322–324,
328–330, 332–333, 348–350
 Quantum dot cellular automata, QCA,
258–259, 261
 Quantum effect device, 3
 Quantum efficiency, 317
 Quantum number, 4, 7, 20, 57, 138, 142, 401
 Quantum well, QW, 46–48, 53–55, 59,
130–131, 133–134, 144–146, 223,
229–230, 271–279, 281–291, 299–308,
341, 346–347, 367–369, 373
 Quantum well infrared photodetector, QWIP,
290–291, 294, 299–300, 302–303,
305–308
 Quantum wire, QWR, 57, 59, 61–62, 71–73,
108, 132–134, 154–155, 238–239, 253,
255–257, 262–263, 307–308, 310–311,
368–369, 375–377

R

Radial Schrödinger equation, 6, 149
 Random walking, 45
 Reciprocal lattice vector, 13, 130, 399
 Reciprocal space, 12–14
 Recursion method, 234, 368
 Reflection high energy electron diffraction,
RHEED, 18
 Resonant cavity, 342
 Resonant tunneling diode, RTD, 106, 185,
195–196, 198, 200, 208, 218, 339–340,
359
 Resonant-tunneling light-emitting diode,
RTLED, 339
 Rydberg constant, 4, 381

S

Schottky barrier, 213, 313
 Schottky diode, 213
 Schrödinger equation, 5–6, 19–20, 27–28,
68–71, 77, 92–93, 143, 168–169, 198,
234, 239–240, 259–260, 368, 385–386,
390–392
 Second-order perturbations, 28
 Secular determinant, 22
 Secular equation, 22
 Selection rule, 153, 276, 282, 307–308
 Semiclassical picture (theory), 192
 Semiconductor microelectronics, 2
 Semiconductor optoelectronics, 2
 Sequential tunneling, 199, 202
 Shockley matrix, 35
 Si-on-insulator, 253, 257
 Silicon on insulator, SOI, 253, 257
 Simple cubic lattice, 11, 14
 Single electron transistor, SET, 257–258,
261–263
 Slater determinant, 135, 177, 393
 Soft lithography, 3
 Spherical band, 31, 276
 Spherical harmonics, 5, 153, 179, 389–390
 Spherical symmetry, 5
 Spin angular momentum, 7
 Spin split off, 30, 33
 Stark effect, 347, 367–368
 Stranski-Krastanov growth, 40, 42–43
 Superlattice, 18, 51, 195, 271, 287–289, 345,
362–363, 365
 Surface emission, 51

T

Tight-binding model, 21–23, 55, 136, 148
 Transfer-matrix method, 288, 363
 Transition matrix, 35, 76, 87, 109, 121, 128,
131, 142, 163–164, 275–276, 302, 306,
321–322, 340
 Transverse effective mass, 31, 38, 276, 355
 Tunneling probability, 200–201, 206, 209–211,
314
 Two-dimensional electron gas, 2DEG,
225–226, 228–230

U

Uniaxially strain, 43
 Unit cells, 20–25, 44, 50–51, 57, 70, 86–87,
89–90, 104, 129–131, 137, 211

V

Valence-band offset, VBO, 36, 362
 Valence electron, 9–10, 21–23, 91–92, 185

Valence-force-field (VFF) approach, 43
Varactor, 185, 213–218, 221, 223–224, 249
Varshni approximation, 36
Varshni parameters, 36
Vertical cavity surface emitting laser, VCSEL,
345
Virtual crystal approximation, 37, 56

W

Wannier function, 50–52, 92
Wave vector space, 12
Wavelength division multiplexing, 116, 165
Work function, 188, 190

Z

Zincblende crystal structure, 12, 22, 44