From: Fundamentals of Carrier Transport
by Mark Lundstrom

1 The quantum foundation

1.1 Electrons in a nonuniform potential, $E_{C0}(\mathbf{r})$
1.2 Electrons in a periodic potential, $U_C(\mathbf{r})$
1.3 Semiconductor heterostructures
1.4 Counting electron states
1.5 Electron wave propagation in devices
1.6 Semiclassical electron dynamics
1.7 Scattering of electrons by the random potential, $U_s(\mathbf{r}, t)$
1.8 Lattice vibrations (phonons)
1.9 Summary

Conventional device analysis begins by assuming that carriers behave as classical particles which obey Newton's laws. A more fundamental treatment describes the electron by its wave function, $\Psi(\mathbf{r}, t)$, which is obtained by solving the Schrödinger equation,

$$i\hbar\frac{\partial\Psi}{\partial t} = -\frac{\hbar^2}{2m_0}\nabla^2\Psi + [E_{C0}(\mathbf{r}) + U_C(\mathbf{r}) + U_S(\mathbf{r}, t)]\Psi(\mathbf{r}, t). \tag{1.1}$$

The quantity $\Psi^*(\mathbf{r}, t)\Psi(\mathbf{r}, t)d\mathbf{r}$ is the probability of finding the electron between $\mathbf{r}$ and $\mathbf{r} + d\mathbf{r}$. Three different potential energies appear in the wave equation; the first, $E_{C0}(\mathbf{r})$, describes potentials that are built-in or applied to the device. (The energy band diagram of a semiconductor device is just a plot of this potential versus position. Device engineers usually refer to this potential as $E_C(\mathbf{r})$, but in this text E_C will refer to the position and momentum-dependent conduction band potential; it contains a potential energy component, $E_{C0}(\mathbf{r})$, and a kinetic energy component.) The second potential is the *crystal potential*, $U_C(\mathbf{r})$, which describes the electrostatic potential due to the atoms. (Since eq. (1.1) is a wave equation for a single electron, $U_C(\mathbf{r})$ also includes the average potential due to the other electrons in the solid.) Finally, U_S is a scattering potential due to random deviations in potential caused by ionized impurities or by lattice vibrations. Device analysis is usually based on an approximate solution to eq. (1.1) known as the *semiclassical* treatment which describes carrier dynamics in the applied and built-

in potentials by Newton's laws without explicitly treating the crystal potential. The influence of the crystal potential is treated indirectly by the use of an effective mass or an energy band structure. Carrier scattering is treated quantum mechanically.

This chapter reviews techniques for treating the three different potentials in the wave equation. The emphasis is on justifying the semiclassical approach to carrier transport because it serves as the basis for most of conventional device analysis and for most of this text. It is important to understand the underlying approximations because they can be violated in advanced, ultra-small devices. For the most part, this chapter should be a review of introductory quantum mechanics and solid-state physics; results are stated, not derived, and their significance and relevance to device analysis is noted. For a thorough treatment of the fundamentals surveyed in this introductory chapter, the reader is referred to *Quantum Phenomena*, by Supriyo Datta [1.1].

1.1 Electrons in a nonuniform potential, $E_{C0}(\mathbf{r})$

Let us first review the nature of solutions to the wave equation in the absence of the crystal and scattering potentials; we further simplify the problem by reducing it to one spatial dimension. Application of the technique of separation of variables to the wave equation then shows that the solutions are of the form

$$\psi(z,t) = \psi(z)e^{-iEt/\hbar} = \psi(z)e^{-i\omega t}, \tag{1.2}$$

which oscillate in time with a frequency of $\omega = E/\hbar$. When eq. (1.2) is inserted in the wave equation, we obtain the time-independent wave equation,

$$\frac{d^2\psi}{dz^2} + k^2\psi = 0, \tag{1.3a}$$

where

$$k^2 = \frac{2m_0}{\hbar^2}[E - E_{C0}(z)]. \tag{1.3b}$$

The nature of the solutions is determined by whether k^2 is greater or less than zero.

General features of the solutions can be illustrated by a few very simple examples (see Fig. 1.1). First, we let $E_{C0}(z)$ be constant and set the constant to zero. Since $k^2 \geq 0$, the solutions to eq. (1.3) are of the form

$$\psi(z) = a_k e^{\pm ikz} \tag{1.4}$$

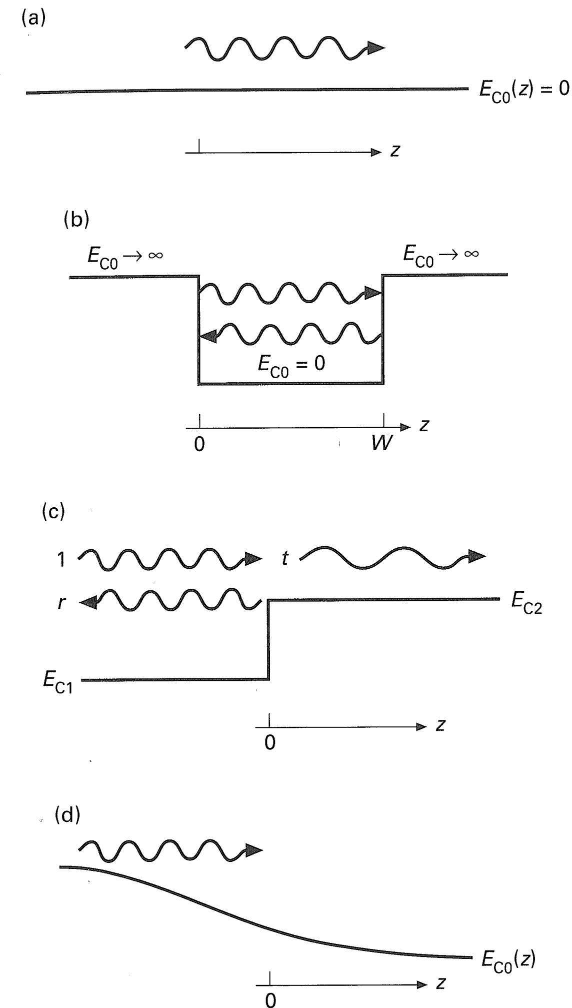


Fig. 1.1 Example potential profiles. (a) free electron, (b) infinite potential well, (c) finite potential step, and (d) slowly varying potential.

(or, equivalently, sin kz or cos kz), where

$$\hbar k = \sqrt{2m_0 E}, \tag{1.5}$$

and a_k is an arbitrary constant. According to eq. (1.5),

$$E(k) = \frac{\hbar^2 k^2}{2m_0} \tag{1.6}$$

is the relation between the electron's energy and its wave vector. Since $\hbar k$ can be shown to be the electron's momentum [1.2], the energy and momentum of a free

electron are related exactly as they are in classical physics. The time-dependent solution,

$$\Psi(z,t) = a_k e^{i(\pm kz - \omega t)}, \tag{1.7}$$

represents a wave traveling in the $\pm\hat{z}$ direction.

The probability of finding the electron, $P(z) = \psi(z)^*\psi(z)$, is simply $|a_k|^2$; the electron has an equal probability of being anywhere. To describe a particle located near z_0 with a momentum of about $\hbar k_0$, we form a linear combination of the solutions, eq. (1.4). Such a solution,

$$\Psi(z,t) = \int_{-\infty}^{+\infty} a(k-k_0)e^{ik(z-z_0)}e^{-i\omega(k)t}dk, \tag{1.8}$$

is known as a *wave packet*. The weighting function, $a(k-k_0)$, is large only near k_0 as shown in Fig. 1.2. At $t=0$ and $z=z_0$ the phase is zero so the contributions for all wave vectors add in phase and the result is a large amplitude. But for $|z| \gg |z_0|$, the exponential, $e^{ik(z-z_0)}$, oscillates rapidly with k, and the contributions from different k add destructively. The result is that eq. (1.8) describes an electron that is located with high probability near $z=z_0$. The plane wave solution, eq. (1.7), had a well-defined momentum ($\hbar k$) but the particle's location was undefined. Equation (1.8) localizes the particle, but since we had to add waves with different momentum, an uncertainty, $\hbar \Delta k$, has been introduced in the electron's momentum.

The uncertainty in the particle's position is related to the spread in wave vectors by [1.2]

$$\Delta z \Delta k \simeq 1, \tag{1.9}$$

which states that many Fourier components are needed to describe a small particle. Similarly, a particle can be localized in time by adding contributions with different frequencies such that

$$\Delta\omega\Delta t \simeq 1. \tag{1.10}$$

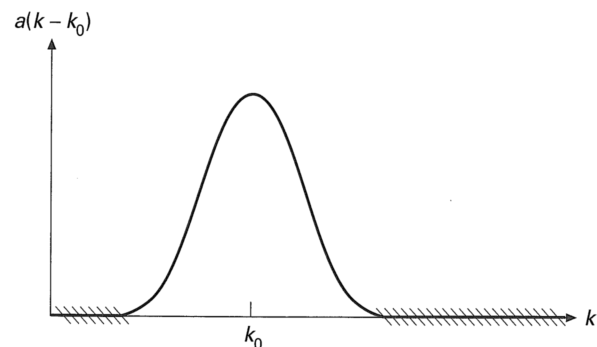


Fig. 1.2 Weighting function $a(k-k_0)$ used to localize an electron.

When a change of variables to momentum and energy is made in eq. (1.9) and eq. (1.10), we find the *uncertainty relations*:

$$\Delta z \Delta p \geq \hbar \tag{1.11}$$

and

$$\Delta E \Delta t \geq \hbar \tag{1.12}$$

which state that we cannot know both a carrier's position and momentum exactly and that we cannot determine its exact energy in a finite time.

Each of the individual components of the wave packet, eq. (1.8), travels at its own *phase velocity*

$$v_p = \frac{\omega(k)}{k}. \tag{1.13}$$

Since each of the components travels at a different velocity, both the center of the wave packet and its shape change with time. The center, which occurs where the components add constructively, moves at the *group velocity*, v_g, where

$$\boxed{v_g(k_0) = \left.\frac{d\omega(k)}{dk}\right|_{k=k_0} = \left.\frac{1}{\hbar}\frac{dE(k)}{dk}\right|_{k=k_0}}. \tag{1.14}$$

According to eq. (1.14) and eq. (1.6), for free electrons

$$v_g = \frac{\hbar k_0}{m_0} = \frac{\langle p \rangle}{m_0},$$

which shows that the group velocity of the electron wave packet is simply its average momentum divided by its mass – just what classical physics would give.

For electrons constrained within a potential well, the solutions are much different. Consider a second example (Fig. 1.1b) for which $E_{C0} = 0$ between 0 and W, but assume now that W is small and that $E_{C0} \to \infty$ at $x=0$ and W so that the electron is *bound* – it must remain between 0 and W. The solutions are still given by eq. (1.4) but it is more convenient to use linear combinations of these, or $\sin(kz)$ and $\cos(kz)$. Because $\psi(z)=0$ at $z=0$, we find

$$\psi(z) = a_k \sin kz. \tag{1.15}$$

But $\psi(z)$ must also be zero at $z=W$, so the wave vector must be restricted to

$$k_n = \frac{n\pi}{W} \quad n = 1, 2, \ldots \tag{1.16}$$

In the first example, the electron was free, and we found a continuous distribution of wave vectors given by eq. (1.6), but in this example, the electron is bound

and only certain k's, specified by eq. (1.16) are permitted. From eq. (1.6) we find that the energy is also quantized according to

$$\varepsilon_n = E(k_n) = \frac{\hbar^2 k_n^2}{2m_0} = \frac{\hbar^2 n^2 \pi^2}{2m_0 W^2} \quad n = 1, 2, \ldots \tag{1.17}$$

With modern microfabrication technology, potential wells can be engineered into devices by appropriate variations of doping or composition. In a silicon metal-oxide-semiconductor field effect transistor (MOSFET), inversion layer carriers are confined in a potential well at the oxide–silicon interface. For such structures, the carriers are confined in only one direction; in the orthogonal plane, the potential is constant and the wave functions have the plane wave character of the first example. Electrons confined in the so-called *quantum wells* are discussed in Section 1.3.

As a third example, we consider an electron, which may be propagating through a device, when it encounters the potential step sketched in Fig. 1.1c. For the conditions shown, $E > E_{C0}(z)$ everywhere so the solutions are traveling waves of the form

$$\psi(z) = e^{ik_1 z} + re^{-ik_1 z} \quad z < 0 \tag{1.18a}$$

and

$$\psi(z) = te^{ik_2 z} \quad z > 0 \tag{1.18b}$$

where

$$k_1 = \frac{\sqrt{2m_0(E - E_{C1})}}{\hbar} \tag{1.19a}$$

and

$$k_2 = \frac{\sqrt{2m_0(E - E_{C2})}}{\hbar}. \tag{1.19b}$$

Both $\psi(z)$ and $d\psi/dz$ must be continuous everywhere otherwise $d^2\psi/dz^2$ would be infinite and eq. (1.3a) could not be satisfied (unless $E_{C0}(z)$ goes to infinity as it did for the second example). Applying these continuity conditions at $z = 0$, we find

$$r = \frac{k_1 - k_2}{k_1 + k_2} \tag{1.20a}$$

and

$$t = \frac{2k_1}{k_1 + k_2}. \tag{1.20b}$$

From a classical perspective, we expect that when the electron's energy is greater than the top of the step, it will simply transmit across. The finite probability of reflection at a step is a quantum mechanical effect due to the wave nature of carriers. Such reflections occur when the potential changes rapidly (meaning that it varies significantly on a distance comparable to the electron's wavelength).

The final example, illustrated in Fig. 1.1d, shows an electron moving through a slowly varying potential (the electron's energy is assumed to be greater than the potential energy). For such potentials, reflections do not occur and the solution can be obtained using the Wentzel–Kramers–Brillouin approximation as in [1.2]

$$\psi(z) = \frac{1}{\sqrt{k}} e^{i\int^z k(\eta)d\eta} \tag{1.21}$$

where $k(\eta)$ is given by eq. (1.3b). The absence of reflections can be understood if the slowly varying potential is approximated by a large number of small potential steps. The small reflections that occur at each interface add destructively so that no net reflected wave occurs. Electron motion in a slowly varying potential can be described classically because reflections do not occur.

1.1.1 Probability current

The goal of device analysis is often to compute the current through the device. Since quantum mechanics is based on probability, we evaluate the flow of probability,

$$\frac{\partial P(z)}{\partial t} = \frac{\partial}{\partial t}(\Psi^* \Psi) = \frac{\partial \Psi^*}{\partial t}\Psi + \Psi^* \frac{\partial \Psi}{\partial t}. \tag{1.22}$$

The derivatives can be evaluated from the wave equation, eq. (1.1), and its complex conjugate to find

$$\frac{\partial P}{\partial t} + \frac{\partial J}{\partial z} = 0 \tag{1.23}$$

where

$$J = \frac{\hbar}{2m_0 i}\left[\Psi^* \frac{\partial \Psi}{\partial z} - \Psi \frac{\partial \Psi^*}{\partial z}\right] \tag{1.24}$$

Equation (1.23) is a continuity equation – the first term is the rate of increase of probability and the second term, the divergence of a vector J, represents the flow of probability away from z. The vector J, which describes the flow of probability, is termed the *probability current*. For an ensemble of electrons, we interpret $\Psi(z)^* \Psi(z)$ as $n(z)$, the electron density and $J(z)$ as the electron flux.

For the plane wave solutions found in the first example, the probability current works out to be

$$J = \Psi^* \Psi \frac{\hbar k}{m_0}. \tag{1.25}$$

Since $\Psi^* \Psi$ is the electron density and $\hbar k$ the electron momentum, eq. (1.25) simply states that $J = nv$. For the second example, the states were bound, and $\psi(z)$ was proportional to $\sin kz$. For such states, eq. (1.24) shows that $J = 0$ which is consistent with the fact that the electron is localized and not going anywhere. For the potential step considered in the third example, the wave functions, eq. (1.18), produce a current

$$J = \frac{\hbar k_1}{m_0}(1 - |r|^2) \quad z < 0 \tag{1.26a}$$

and

$$J = \frac{\hbar k_2}{m_0}|t|^2 \quad z > 0, \tag{1.26b}$$

from which the incident, reflected, and transmitted currents J_i, J_r, J_t, are apparent. Transmission and reflection coefficients for the currents can be defined as

$$T \equiv \frac{J_t}{J_i} = \frac{k_2}{k_1}|t|^2 \tag{1.27}$$

and

$$R \equiv \frac{J_r}{J_i} = |r|^2. \tag{1.28}$$

Notice that T and R are real numbers unlike t and r and that $T + R = 1$ as expected.

For electrons moving in a slowly varying potential, the wave function is given by eq. (1.21). For electrons in such a potential,

$$J = \psi(z)^* \psi(z) \frac{\hbar k(z)}{m_0}. \tag{1.29}$$

Since $\psi^* \psi \sim 1/k(z)$, the current is constant – as it should be. When $\psi^* \psi$ is interpreted as the electron density, eq. (1.29) is seen to correspond to the classical current density, $J = n(z)v(z)$. The electron's momentum is $\hbar k(z)$, so from eq. (1.29) its velocity is

$$v(z) = \frac{\sqrt{2m_0(E - E_{C0}(z))}}{m_0}. \tag{1.30}$$

E is the electron's total energy and $E_{C0}(z)$ its potential energy, so eq. (1.30) is just the expression for the velocity of a classical particle.

These examples show that the wave nature of carriers is important when the potential changes rapidly, but when the potential varies gradually, reflections don't occur and electrons can be treated as classical particles which obey Newton's Laws. The electrons in the conduction band of a semiconductor always see at least one rapidly varying potential – it is the crystal potential due to the nucleus and core electrons.

1.2 Electrons in a periodic potential, $U_C(\mathbf{r})$

Electrons in a semiconductor crystal respond when fields are applied, but a crystal potential due to the lattice atoms and other electrons is always present. As sketched in Fig. 1.3a, the crystal potential, $U_C(z)$, displays the periodicity of the lattice. To find the wave functions we solve the one-electron wave equation

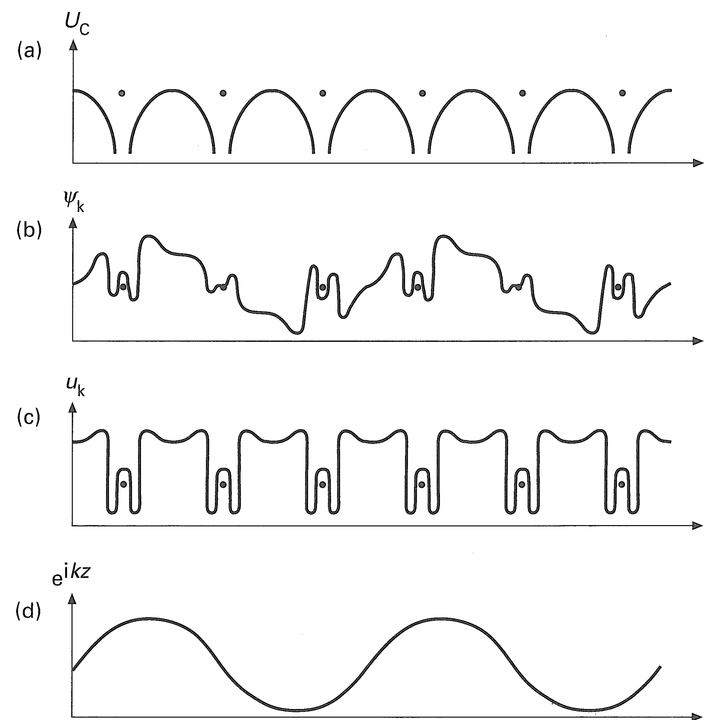


Fig. 1.3 Illustration of wave functions in a periodic potential. (a) $U_C(z)$, the crystal potential, (b) ψ_k, the eigenfunction, (c) u_k, the Bloch function, (d) $e^{ik_t \cdot z}$, a plane wave. From Harrison [1.4]. (Reproduced with permission from Dover, New York.)

$$\left[-\frac{\hbar^2}{2m_0}\frac{d^2}{dz^2} + U_C(z)\right]\psi(z) = E\psi(z). \tag{1.31}$$

The solutions sketched in Fig. 1.3b also reflect the periodicity of the lattice. The solutions for a periodic are called *Bloch waves* and consist of a function with the periodicity of the lattice multiplied by a plane wave. Figures 1.3c and 1.3d display the two components of a Bloch wave, which is mathematically defined by

$$\psi_k = u_k e^{ikz}, \tag{1.32a}$$

where

$$u_k(z + a) = u_k(z). \tag{1.32b}$$

The electron's momentum varies with position because the crystal potential alternately speeds up and slows down electrons. Nevertheless, the quantity $\hbar k$, termed the *crystal momentum*, often acts like the carrier's momentum.

To find $u_k(z)$ we insert eq. (1.32a) in eq. (1.31) and find

$$\left[\frac{1}{2m_0}\left(\frac{\hbar}{i}\frac{\partial}{\partial z} + \hbar k\right)^2 + U_C(z)\right]u_k = E(k)u_k, \tag{1.33}$$

which must be solved with the boundary condition eq. (1.32b). For any k we select, eq. (1.33) can be solved for the energy eigenvalue, $E(k)$, and the eigenfunction, u_k. Since there are an infinite number of eigenvalues, we should label them as $E_n(k)$, where $n = 1, 2, 3, \ldots$ labels the particular eigenvalue. Choosing another k results in another infinite set of eigenvalues. Each eigenvalue is associated with a *band* because as k varies a band of energies is traversed.

The general features of energy bands found by solving eq. (1.33) are summarized in Fig. 1.4. Only four eigenvalues for each k are shown. The eigenvalues $E_1(k_1)$, $E_2(k_1)$, $E_3(k_1)$, $E_4(k_1)$, etc., are found by solving eq. (1.33) for $k = k_1$. By repeating the procedure for other choices of k, one continuous curve, $E_n(k)$, is mapped out for each of the various eigenvalues. The dashed lines show that $E_n(k)$ is periodic in **k**-space. Note also that there are certain energy gaps – forbidden regions on the energy axis that cannot be reached by any real k in any band.

The periodicity of $E(k)$ is a consequence of the spatial periodic crystal potential. In general,

$$E_n(k) = E_n(k + K_j), \tag{1.34}$$

where

$$K_j = j\frac{2\pi}{a} \quad j = 1, 2, \ldots \tag{1.35}$$

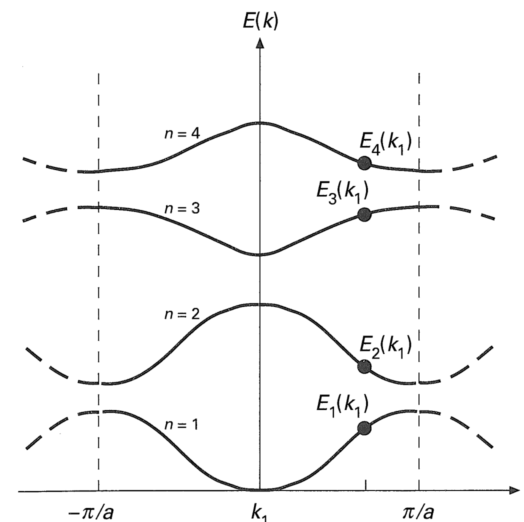


Fig. 1.4 Electron energy versus wave vector, $E(k)$.

is a *reciprocal lattice* vector, and a is the lattice constant. Because $E_n(k)$ is periodic, all information is available in one period or *Brillouin zone*. It is convenient to use a period centered about the origin (the so-called *reduced zone* representation).

The band structures of common semiconductors are well known from various experiments and from numerical solutions to the wave equation. For semiconductor work, approximate solutions to eq. (1.33), accurate near a band minimum, are often adequate. The so-called $\mathbf{k}\cdot\mathbf{p}$ method for obtaining such approximate solutions is discussed in the texts by Datta [1.1] and by Singh [1.6]. To treat very energetic carriers, however, a full, numerical tabulation of $E(\mathbf{k})$ throughout the Brillouin zone is essential [1.7].

If the band structure is known, $E(k)$ can always be expanded in a Taylor series as

$$E(k) = E(0) + \left.\frac{\partial E(k)}{\partial k}\right|_{k=0} k + \frac{1}{2}\left.\frac{\partial^2 E(k)}{\partial k^2}\right|_{k=0} k^2 + \ldots$$

When the band minimum occurs at $k = 0$, the gradient of $E(k)$ is zero at $k = 0$, so, to the lowest order,

$$E(k) = E(0) + \frac{\hbar^2 k^2}{2m^*}, \tag{1.36}$$

where

$$\frac{1}{m^*} \equiv \frac{1}{\hbar^2}\frac{\partial^2 E(k)}{\partial k^2} \tag{1.37}$$

is the *effective mass*. A comparison of eq. (1.36) with eq. (1.6) shows that for electrons near a band minimum, the $E(k)$ relation in a crystal is just like that for free electrons except for a change in curvature of the band. The electron mass is simply replaced by the effective mass. Knowledge of m^* is sometimes the only information from the $E(k)$ characteristic that is required to describe carrier transport.

1.2.1 Model band structure

In real, three-dimensional, semiconductors the Brillouin zone becomes a volume and $E(\mathbf{k})$ generally depends on the direction of $\mathbf{k}$. The simple $E(\mathbf{k})$ model shown in Fig. 1.5a describes both diamond and zincblende crystals such as silicon and GaAs. The conduction band has three minima; one at $\mathbf{k} = 0$ (called the Γ point), another along $\langle 111\rangle$ directions at the boundary of the first Brillouin zone (called L) and a third near the zone boundary along $\langle 100\rangle$ directions (the Δ-line). (The standard notation for labeling lines and points in the Brillouin zone is displayed in Fig. 1.5b.) The model semiconductor has three valence bands – each has a maximum at $\mathbf{k} = 0$. Two of the valence bands are degenerate at $\mathbf{k} = 0$; the third is *split-off* by the spin-orbit interaction.

Table 1.1 lists several band parameters, defined in Fig. 1.5, for a variety of semiconductors. For silicon and germanium, the lowest conduction band minima are along Δ and at L respectively. When the conduction band minimum doesn't occur at the same point as the valence band maximum, the semiconductor is called *indirect*. GaAs is seen to be *direct*, but it also has a minimum at L that is only a few tenths of an electron volt higher. This upper minimum plays an important role for electron transport in GaAs.

When the conduction band minimum or valence band maximum lies at $\mathbf{k} = 0$, $E(\mathbf{k})$ may be approximated as

$$E(\mathbf{k}) = \pm\frac{\hbar^2 k^2}{2m^*}, \tag{1.38}$$

where the positive sign is for the conduction band and the negative sign for the valence band. The energy zero is taken at the band extrema, so $E(\mathbf{k})$ represents the carrier's kinetic energy. Equation (1.38) describes a band whose constant energy surface in $\mathbf{k}$-space is a sphere; the effective mass is isotropic. This simple model is appropriate for the conduction band at Γ and for the split-off valence band, but is often used more widely – whenever rough estimates suffice.

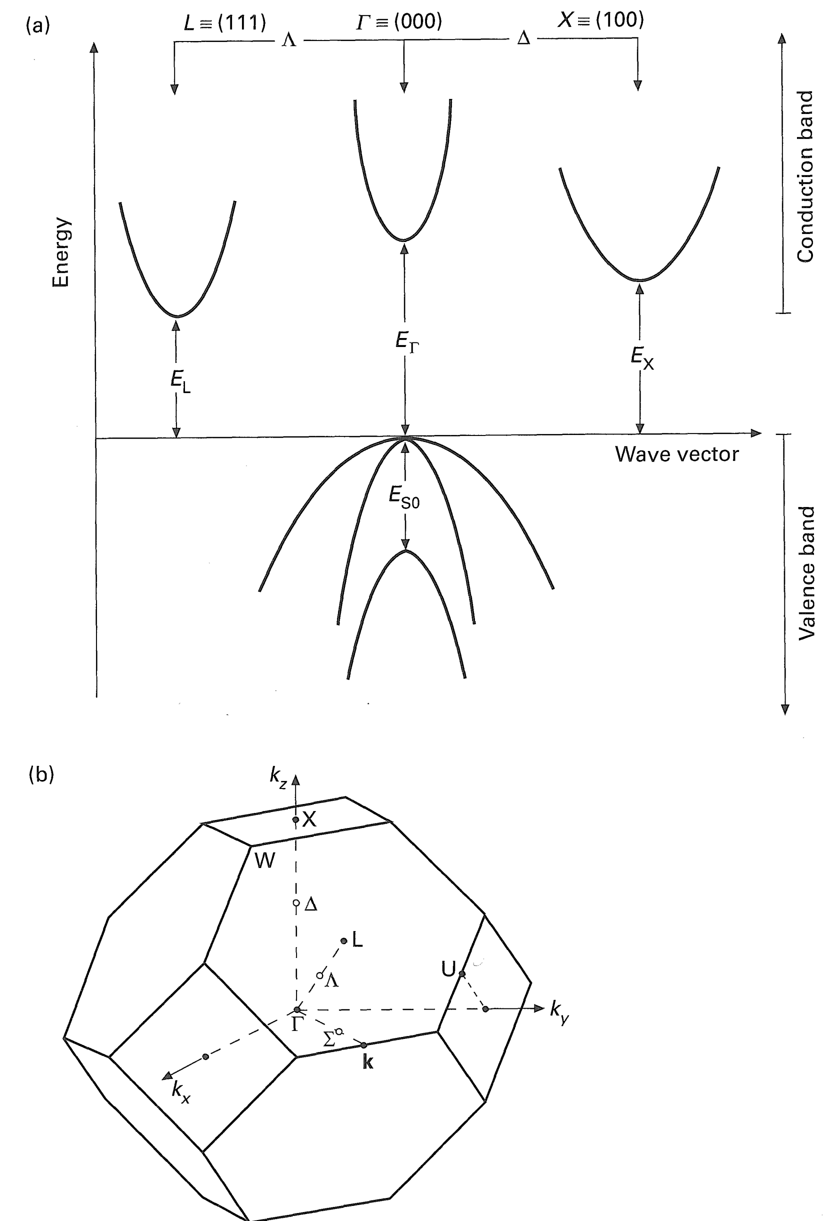


Fig. 1.5 (a) Band structure of the model semiconductor (from Reggiani, L., Chapter 1: General Theory. In *Hot Electron Transport in Semiconductors*. Springer-Verlag, New York, 1985.) (b) Standard notation for labeling high symmetry lines and points in the Brillouin zone for diamond and zincblende crystals. (Reproduced with permission from Springer-Verlag.)

Table 1.1. *Band parameters of common cubic semiconductors (from L. Reggiani, Chapter 1: General Theory. In* Hot Electron Transport in Semiconductors. *Springer-Verlag, New York, 1985. Reproduced with permission from Springer-Verlag.)*

	E_Γ (eV)	E_L (eV)	E_Δ (eV)	E_{so} (eV)	m_ℓ^* (m_0)	m^* (m_0)	m_t^* (m_0)	α (eV^{-1})	$\lvert A\rvert$	$\lvert B\rvert$	$\lvert C\rvert$
C	11.67	12.67	5.45	0.006	1.4	–	0.36	–	3.61	0.18	3.76
Si	4.08	1.87	1.13	0.044	0.98	–	0.19	0.5	4.22	0.78	4.80
Ge	0.89	0.76	0.96	0.29	1.64	–	0.082	0.65	13.35	8.50	13.11
AlP	3.3	3.0	2.1	0.05	–	–	–	–	3.47	0.12	3.98
AlAs	2.95	2.67	2.16	0.28	2.0	–	–	–	4.04	1.56	4.71
AlSb	2.5	2.39	1.6	0.75	1.64	–	0.23	–	4.15	2.02	4.95
GaP	2.7	2.7	2.2	0.08	1.12	–	0.22	–	4.20	1.96	4.65
GaAs	1.42	1.71	1.90	0.34	–	0.067	–	0.64	7.65	4.82	7.71
GaSb	0.67	1.07	1.30	0.77	–	0.045	–	1.36	11.80	8.06	11.71
InP	1.26	2.0	2.3	0.13	–	0.080	–	0.67	6.28	4.16	6.35
InAs	0.35	1.45	2.14	0.38	–	0.023	–	2.73	19.67	16.74	13.96
InSb	0.23	0.98	0.73	0.81	–	0.014	–	5.72	35.08	31.28	22.27
ZnS	3.8	5.3	5.2	0.07	–	0.28	–	0.14	2.54	1.50	2.75
ZnSe	2.9	4.5	4.5	0.43	–	0.14	–	0.26	3.77	2.48	3.87
ZnTe	2.56	3.64	4.26	0.92	–	0.18	–	0.26	3.74	2.14	4.30
CdTe	1.80	3.40	4.32	0.91	–	0.096	–	0.45	5.29	3.78	5.46

For many semiconductors, electrons respond to applied fields with an effective mass that depends on the crystallographic orientation of the field. In common cubic semiconductors, we find that

$$E(\mathbf{k}) = \frac{\hbar^2}{2}\left[\frac{k_\ell^2}{m_\ell^*} + \frac{k_t^2}{m_t^*}\right]. \tag{1.39}$$

Equation (1.39) describes a band with ellipsoidal constant energy surfaces. The effective mass is a tensor – the longitudinal and transverse effective masses, m_ℓ^* and m_t^*, differ. Equation (1.39) describes conduction bands at L and along Δ. Note that there are eight equivalent L points and six equivalent Δ lines in cubic crystals so there are actually many of these valleys in **k**-space.

For high applied fields, carriers may be far above the minimum, and the higher order terms in the Taylor series expansion cannot be ignored. For the conduction band, *nonparabolicity* is often described by a relation of the form

$$E(1 + \alpha E) = \frac{\hbar^2 k^2}{2m^*}, \tag{1.40}$$

where m^* is determined from eq. (1.37) at the minimum. (Equation (1.40) is obtained from approximate solutions to eq. (1.33) derived by $\mathbf{k}\cdot\mathbf{p}$ theory.) For a minimum at $\mathbf{k} = 0$,

$$\alpha_\Gamma = \frac{1}{E_{G\Gamma}}\left(1 - \frac{m_\Gamma^*}{m_0}\right)^2, \tag{1.41}$$

where $E_{G\Gamma}$ is the direct bandgap.

The simple expressions we have presented generally work well for electrons in the conduction band, but the valence bands are much more complex. In $\mathbf{k}\cdot\mathbf{p}$ theory, the shape of $E(\mathbf{k})$ is attributed to interactions between various bands. In wide bandgap semiconductors, the conduction band is well separated from the valence bands, interactions are weak, and the resulting band structure is parabolic. (But at high energies, or in narrow bandgap semiconductors, these interactions become important leading to conduction band nonparabolicity as discussed above.) The light and heavy hole valence bands, however, are degenerate at $\mathbf{k} = 0$, so the interactions are strong and the band shapes become complex. Since the split-off valence band is generally rather well-separated from the other valence bands, we expect its shape to be more nearly parabolic. Because the band structure has such a strong influence on carrier transport, it is important that we develop a descriptive understanding of the valence bands. (For a discussion of how to actually compute these band shapes, consult Datta [1.1] or Singh [1.6].)

The light and heavy hole valence bands in common semiconductors can be described by

$$E(\mathbf{k}) = ak^2[1 \mp g(\theta, \phi)]. \tag{1.42}$$

Such bands have a warped constant energy surface (the $\mp$ refers to the heavy and light hold bands respectively). The angles, θ and ϕ, are the polar and azimuthal angles of **k** with respect to the crystallographic axes. The function $g(\theta, \phi)$ is given by

$$g(\theta, \phi) = \left[b^2 + c^2(\sin^4\theta\cos^2\phi\sin^2\phi + \sin^2\theta\cos^2\theta)\right]^{1/2} \tag{1.43}$$

with

$$a = \frac{\hbar^2|A|}{2m_0}, \quad b = \frac{|B|}{|A|}, \quad c = \frac{|C|}{|A|} \tag{1.44}$$

where A, B, and C are listed in Table 1.1.

Figures 1.6–1.8 show the constant energy surfaces for the heavy, light, and split-off valence bands in Si. As shown in Fig. 1.6, the heavy hole band is warped at low energies, and the shape becomes complicated at higher hole energies. For the light-hole band displayed in Fig. 1.7, the distortion is smaller, and it is even

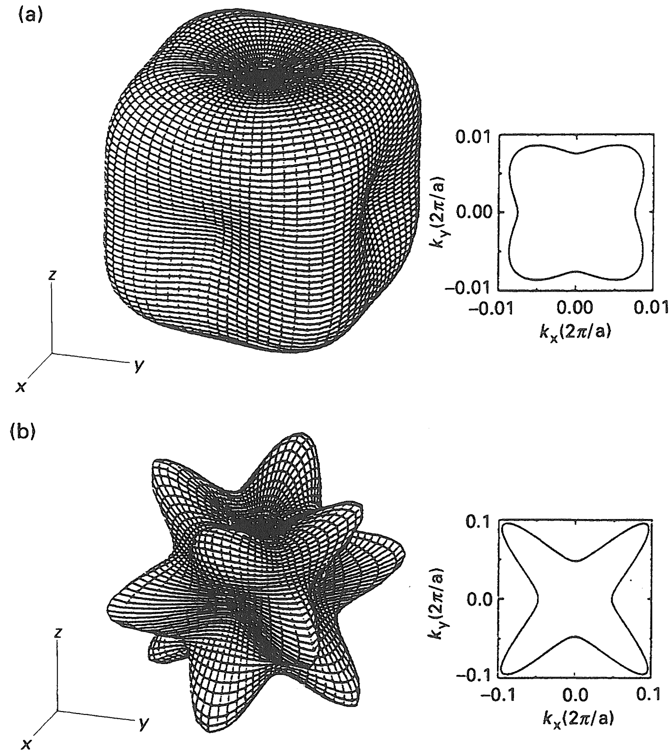


Fig. 1.6 The constant energy surfaces for the heavy hole band in Si. (a) $E = 1\,\text{meV}$, and (b) $E = 40\,\text{meV}$. From Singh [1.6]. (Reproduced with permission of The McGraw-Hill Companies.)

less for the split-off valence band, as shown in Fig. 1.8. For Si, the spin-orbit coupling is small ($\Delta_{SO} = 0.044\,\text{eV}$), so the split-off band can play a role in hole transport, but for most semiconductors, the spin-orbit coupling is much larger, and the split-off band is not typically populated by holes (e.g. in GaAs, $\Delta_{SO} = 0.35\,\text{eV}$). These examples show why hole transport is difficult to treat, even at low fields when the carriers reside near the top of the band. We will generally make use of very simple, spherical and parabolic energy band models, but it is important to recognize that a realistic description of the valence band shape is required for a quantitative treatment of hole transport.

1.2.2 Full band structure

As discussed in the previous section, the conduction band can be approximated as parabolic only near the band minima. For higher energies, we can approximate the conduction band with a nonparabolicity parameter, α, as defined in eq. (1.40). In Section 1.4.2, we will show that a spherical, nonparabolic band provides a reasonable approximation to the density of states in Si up to an energy of

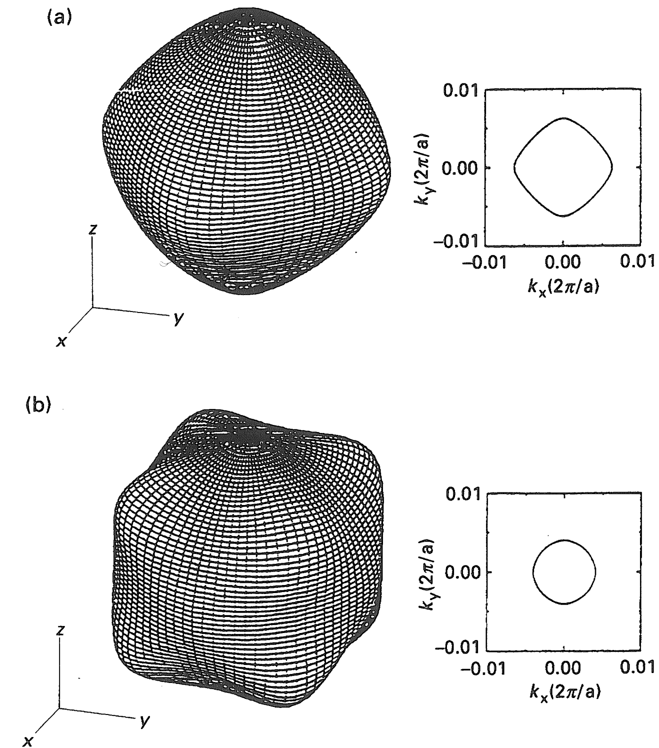


Fig. 1.7 The constant energy surfaces for the light hole band in Si. (a) $E = 1\,\text{meV}$, and (b) $E = 40\,\text{meV}$. From Singh [1.6]. (Reproduced with permission of The McGraw-Hill Companies.)

about ≈ 1–$2\,\text{eV}$. There are important reasons, however, for examining electrons at much higher energies. Impact ionization, for example, involves electrons with a few electron volts of kinetic energy, and in Si MOSFETs an important reliability problem is the injection of electrons from the channel into the gate oxide. The energy barrier at the SiO_2 : Si interface is $\approx 3.1\,\text{eV}$. For such problems, we must abandon simple expressions for $E(\mathbf{k})$ and resort to a numerically-generated table of $E(\mathbf{k})$.

To evaluate $E(\mathbf{k})$ numerically, eq. (1.1) is solved for a bulk semiconductor ($E_{C0}(\mathbf{r}) = 0$) in the absence of scattering ($U_S(\mathbf{r}, t) = 0$). The well-developed art of such calculations is discussed by Singh [1.6]. One popular method is the pseudopotential technique which relies on the fact that the band structure is largely determined by the valence electrons. Accordingly, the effect of the core potential is subtracted out by replacing the actual potential by a *pseudopotential* which reproduces the actual potential between atoms but which is smooth through the ion core. Empirical *form factors* have been derived to fit optical bandgaps at the high symmetry locations. Using this empirical pseudopotential method, the energy band structures of most common semiconductors have been

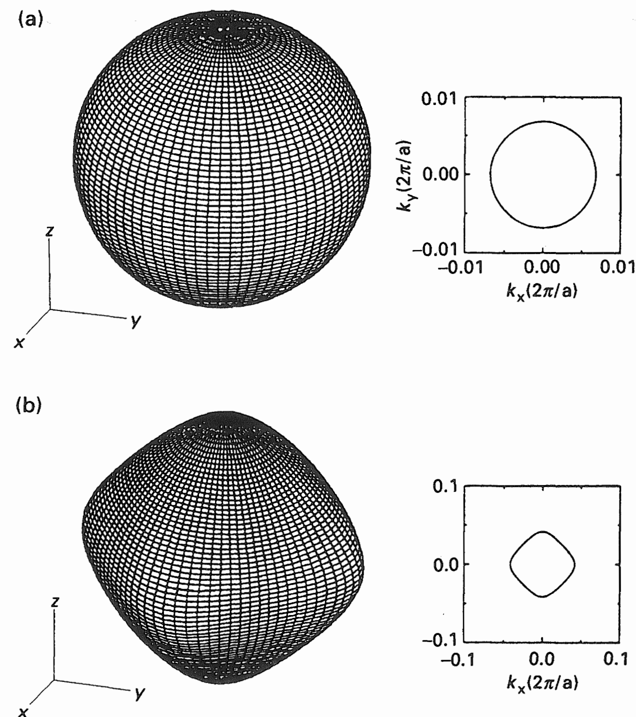


Fig. 1.8 The constant energy surfaces for the split-off hole band in Si. (a) $E = 45\,\text{meV}$, and (b) $E = 84\,\text{meV}$. From Singh [1.6]. (Reproduced with permission of The McGraw-Hill Companies.)

evaluated. The results of such calculations are generally presented as plots of $E(\mathbf{k})$ along the high symmetry lines displayed in Fig. 1.5b. Figures 1.9 and 1.10 show the results for the conduction bands of Si and GaAs [1.7].

Figures 1.9 and 1.10 show the first several conduction bands; the lowest conduction band in each case is indicated by a heavy line. For Si, we see that the lowest conduction band energy is along the x-line (a [100] direction) and occurs at about 85% of the way to the zone boundary. These are the well-known six, equivalent ellipsoidal constant energy surfaces of Si. Note that when electrons gain $\approx 0.13\,\text{eV}$ of kinetic energy, they can cross the zone boundary. More importantly, we see a second conduction band only 0.1 eV above the minimum of the first conduction band. Carriers above $\approx 0.1\,\text{eV}$ in kinetic energy may reside in either of two conduction bands. Note also that the first conduction band minimum at L lies about 1 eV above the lowest first band minimum. Under high electric fields, carriers can gain enough energy to populate these (Ge-like) valleys too.

The ellipsoidal constant energy surfaces at low energy become very complicated at high energies. In Figs. 1.11a–c we examine the constant energy contours in a cross-section of the Brillouin zone. The ellipsoidal constant energy surfaces

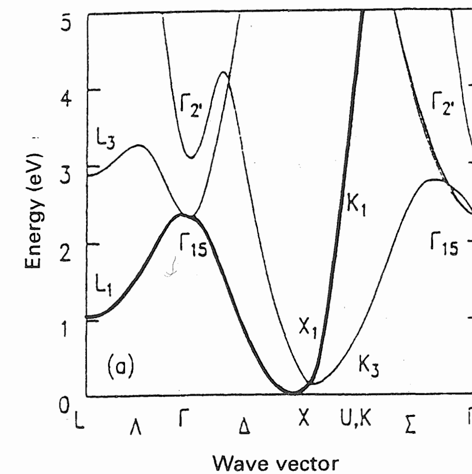


Fig. 1.9 $E(\mathbf{k})$ for the conduction bands of Si. From M. Fischetti [1.7]. (© 1991 IEEE)

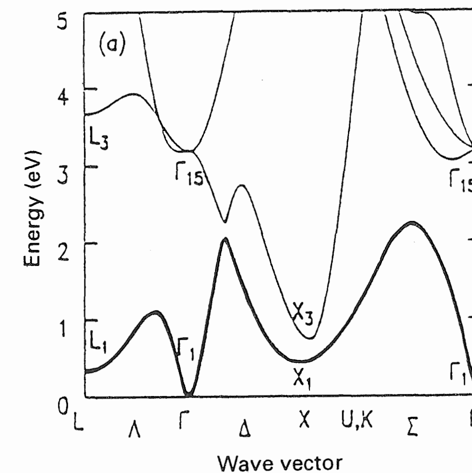


Fig. 1.10 $E(\mathbf{k})$ for the conduction bands of GaAs. From M. Fischetti [1.7]. (© 1991 IEEE)

of the first conduction band are displayed in Fig. 1.11b, and the constant energy surfaces of the second conduction band are displayed in 1.11c. The second conduction band minima lie at the zone boundary, and the constant energy surfaces are more spherical than the first. Under high electric fields, electrons populate the entire Brillouin zone, and Figs. 1.11b and c show that the constant energy surfaces between the minima cannot be described by simple analytical expressions.

The $E(\mathbf{k})$ relations for the conduction bands of GaAs are displayed in Fig. 1.10. In contrast to Si, we see that the second conduction band lies well above the

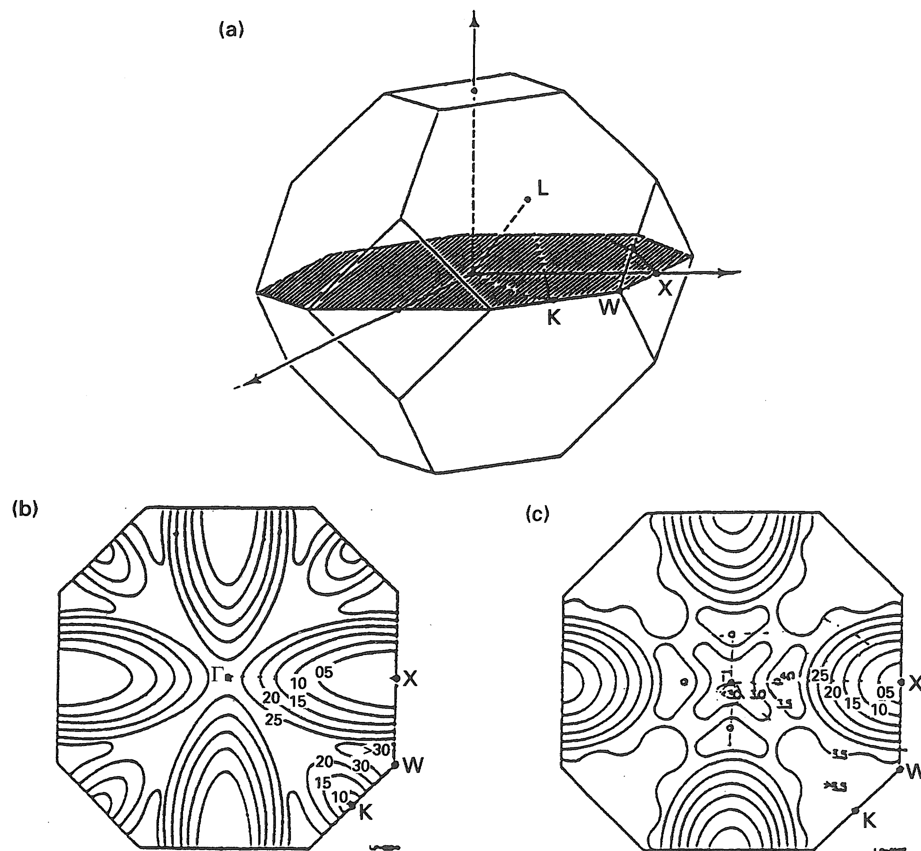


Fig. 1.11 (a) A (100) plane cross-section of the Brillouin zone. From J. Y. Tang and K. Hess, Impact ionization of electrons in silicon (steady state), *J. Appl. Phys.*, **54**(9) 5139–5144, 1983. (b) The (100)-plane contours of constant energy for the first conduction band in Si. From Tang and Hess. (c) The (100)-plane contours of constant energy for the second conduction band in Si. From Tang and Hess. (Reproduced with permission of American Institute of Physics.)

first; we can do a decent job of describing transport in GaAs with just a single conduction band. Note, however, that the first conduction band shows three minima with the lowest at the Γ point. The first conduction band minima at L (the Ge-like valleys) lie only about 0.31 eV above the Γ-valley minima, and the minima along X (the Si-like valleys) are only about 0.5 eV above the Γ-valley minimum. These valleys are easily populated under modest and high electric fields and give GaAs its distinctive transport features.

Because of the increasing importance of high energy carriers in modern devices, the use of full, numerical tables of $E(\mathbf{k})$ is common. When generating such tables, it is important to exploit symmetry to minimize the data to be stored. Consider Fig. 1.12, which shows a location in a cubic coordinate system. For a cube, one symmetry operation is reflection across the x–z plane. From repeated

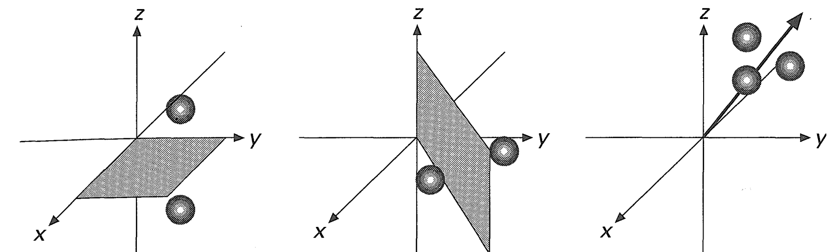


Fig. 1.12 Illustration of the symmetry operations for a cubic lattice. (a) Reflection across a (100) plane, (b) reflection across a (110) plane, and (c) rotation of $3\pi/2$ about a [111] direction.

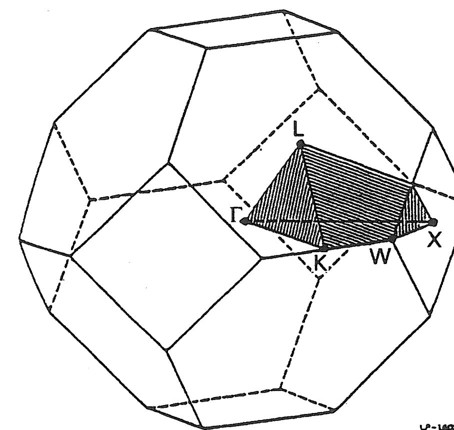


Fig. 1.13 The irreducible wedge of the Brillouin zone. Given $E(\mathbf{k})$ for this 1/48th of the Brillouin zone, $E(\mathbf{k})$ for the entire Brillouin zone is obtained by applying the symmetry operations of the cubic lattice.

reflections across equivalent planes, eight equivalent points can be formed. Another symmetry operation is reflection across a (110) plane, which generates another equivalent point for each of the first eight equivalent points. Finally, one can rotate a cube by $3\pi/2$ about a [111] direction to get three more equivalent points. The result is that for the given point, there are $8 \times 2 \times 3 = 48$ equivalent points. It is sufficient, therefore, to evaluate $E(\mathbf{k})$ in 1/48 of the Brillouin zone and to generate the other points by symmetry operations. The volume commonly used is shown in Fig. 1.13; it is known as the *irreducible wedge* of the Brillouin zone and is defined by

$$\begin{aligned} &0 \le k_z \le k_x \le k_y \le 2\pi/a \\ &k_x + k_y + k_z \le 3\pi/a. \end{aligned} \tag{1.45}$$

When simple, analytical expressions for $E(\mathbf{k})$ (e.g. eq. (1.38)) are used, we shall see that analytical expressions for quantities such as the density of states and

carrier scattering rates can be obtained, but when a full, numerical description of the bandstructure is employed, we must resort to numerical integration.

1.3 Semiconductor heterostructures

One feature of modern semiconductor technology is that the material composition is readily varied as a semiconductor film is grown. This is particularly easy to accomplish in semiconductor alloys, but other combinations of different semiconductors are readily produced. Junctions between two different semiconductors are called *heterojunctions*. More complex, perhaps continuous, compositional variations are referred to as *heterostructures*. In this section, we introduce some general concepts that we will use when we discuss transport in heterostructures.

1.3.1 Band structure of semiconductor alloys

Alloys of two or more semiconductors have many device applications. The alloy $Al_xGa_{1-x}As$, for example, comprises GaAs (a direct gap semiconductor with a bandgap of 1.42 eV) and AlAs (an indirect semiconductor with a bandgap of 2.16 eV). As the AlAs mole faction varies from 0 to 1, the bandgap of the alloy varies from that of GaAs to that of AlAs. The use of such alloys offers an additional degree of freedom to device engineers because both the doping and bandgap can be varied with position. To analyze such devices, the alloy's composition-dependent properties must be known. Adachi [1.9] describes how parameters such as the bandgap, effective masses, and dielectric constant vary with alloy composition in the $Al_xGa_{1-x}As$ system. For other materials, consult Landolt-Börnstein [1.10].

1.3.2 Energy band diagrams for abrupt heterojunctions

To draw energy band diagrams for compositionally nonuniform devices, we need to know more than how the bandgap varies with composition, we must also know how the bands line up at compositional junctions. Figure 1.14 shows the experimentally observed band alignments for $Al_{0.3}Ga_{0.7}As/GaAs$ heterojunction. For $Al_xGa_{1-x}As$ heterojunctions, the offset in conduction bands is found to be about 65% of the difference in band gaps for alloy compositions below about $x = 0.5$ where the band structure is direct.

With modern epitaxial growth techniques, the alloy composition can be varied on an atomic scale to produce structures like that shown in Fig. 1.15a, which consists of a GaAs *quantum well* sandwiched between two $Al_xGa_{1-x}As$ layers.

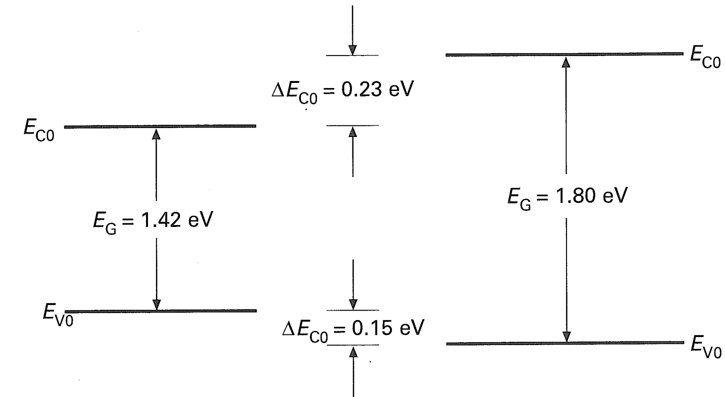


Fig. 1.14 Experimentally observed band alignments for $Al_xGa_{1-x}As$ for $x = 0$ and $x = 0.3$.

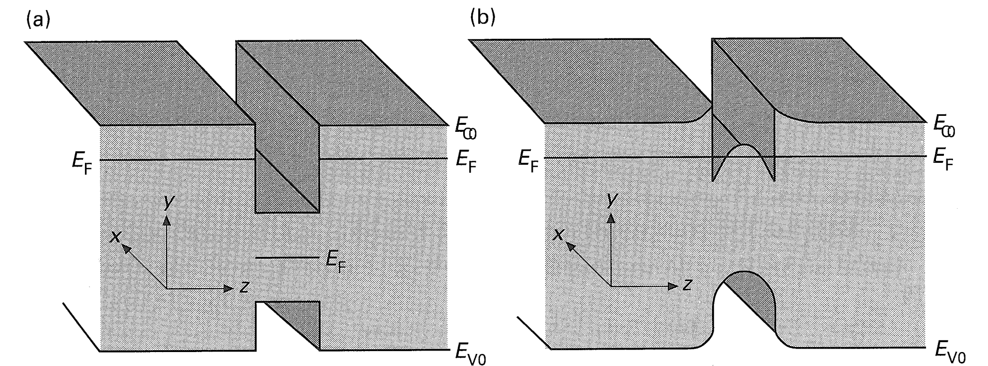


Fig. 1.15 Energy band diagram for an Al/GaAs/GaAs/AlGaAs quantum well structure. (a) For this case, we assume that the electrons are 'frozen' in place in the N-AlGaAs so that they cannot transfer to the *i*-GaAs. (b) A more realistic energy band diagram for the AlGaAs/GaAs/AlGaAs quantum well which displays the effects of mobile charge transfer to the quantum well.

Since the width of the well may be less than 100 Å, carriers within these wells are strongly influenced by quantum effects.

For this example, we assume that the $Al_xGa_{1-x}As$ layers are doped *n*-type and that the GaAs well is undoped. In Fig. 1.15a we have assumed that the electrons are frozen in place so that they cannot move down in energy from the N-AlGaAs to the *i*-GaAs. For this case, the bands are flat, and there is no electric field. A more realistic case is illustrated in Fig. 1.15b where the electrons move from the higher Fermi level to the lower one and establish equilibrium. As a consequence of the charge transfer, the AlGaAs layers are depleted and the GaAs well is accumulated. It is interesting to note that the electrons reside in the undoped quantum well – spatially separated from their parent donors in

the N-AlGaAs. As a result, electrons in the well experience little ionized impurity scattering and have an especially high mobility. The technique is known as *modulation doping* and is the basis for a transistor known as a modulation doped field-effect transistor (MODFET) (or HEMT) in which the electrons in a modulation-doped quantum well comprise the channel of the field-effect transistor. This device, and several other *heterostructure devices* are described in Weisbuch and Vinter [1.11].

1.3.3 Energy band diagrams for continuous compositional variation

For a conventional, homostructure semiconductor device, the conduction and valence band edges move in response to a macroscopic potential set up by space charges. The slope of the conduction or valence band give the electric field. More generally, both $V(x)$ and the composition of the semiconductor vary with position. Since the composition is nonuniform, the crystal periodicity is broken, and one may question the whole concept of energy bands. If the composition varies slowly, however, we may take the band structure at any point to be the band structure of the corresponding bulk semiconductor with the composition at that point (see [1.12–1.14]).

Because the composition is nonuniform, E_{C0} and E_{V0} (and therefore the electron affinity χ and the bandgap E_G) will also be nonuniform. As a result,

$$E_{C0}(z) = E_0 - \chi_S(z) - qV(z) \tag{1.46a}$$

and

$$E_{V0}(z) = E_0 - \chi_S(z) - qV(z) - E_G(z). \tag{1.46b}$$

An energy band for this case might look like Fig. 1.16 which shows a semiconductor with band-bending which is due to both an electric field and to compositional variations. The slope of the conduction band gives the force of an electron, but it is impossible to deduce the electric field from the energy band diagram.

Consider the force acting on an electron in the conduction band,

$$F_e = \frac{-\mathrm{d}E_{C0}}{\mathrm{d}z} = q\frac{\mathrm{d}V(z)}{\mathrm{d}z} + \frac{\mathrm{d}\chi_S}{\mathrm{d}z} \tag{1.47a}$$

and on a hole in the valence band

$$F_h = \frac{+\mathrm{d}E_{V0}}{\mathrm{d}z} = -q\frac{\mathrm{d}V(z)}{\mathrm{d}z} - \frac{\mathrm{d}(\chi_S + E_G)}{\mathrm{d}z}. \tag{1.47b}$$

The force on electrons is not equal in magnitude and opposite in direction to the force on holes, as we would expect for forces due to electric fields. The electric field is only one component of the force on a carrier. Since we are used to

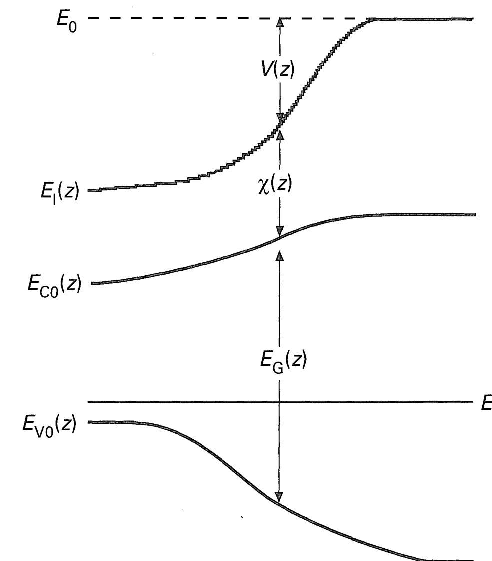


Fig. 1.16 An energy band diagram for a compositionally graded semiconductor.

thinking of electric fields producing forces on carriers, we can define *quasi-electric fields* for electrons by

$$F_e = -q\mathcal{E}(z) - q\mathcal{E}_{QN}(z) \tag{1.48a}$$

and for holes by

$$F_h = +q\mathcal{E}(z) + q\mathcal{E}_{QP}(z). \tag{1.48b}$$

With these definitions we have

$$\mathcal{E}_{QN} = -\frac{1}{q}\frac{\mathrm{d}\chi_S}{\mathrm{d}z} \tag{1.49a}$$

and

$$\mathcal{E}_{QP} = -\frac{1}{q}\frac{\mathrm{d}}{\mathrm{d}z}(\chi_S + E_G). \tag{1.49b}$$

Notice that the quasi-electric field for electrons can differ both in magnitude and direction from the quasi-electric field for holes. These quasi-electric fields give the device designer an additional degree of freedom since they can be controlled by the nonuniform composition. Notice that E_{C0} and E_{V0} are not constrained to be parallel in a heterostructure.

1.4 Counting electron states

Any finite volume of semiconductor will contain a finite number of states derived from the finite number of energy levels in the isolated atoms. To determine the macroscopic properties, the contributions from each occupied state have to be added. Because the number of states is usually very large, it is more convenient to integrate over a range of states in **k**-space or in energy space. To do so, however, we need to know the *density of states* in **k**-space or in energy space.

1.4.1 Density of states in k-space

Although we have drawn $E(k)$ as continuous, in a semiconductor of finite size only a finite number of k's is allowed. Consider a chain of N atoms. Since the precise boundary conditions matter only very near the ends, we impose periodic boundary conditions,

$$\psi(z) = \psi(z + Na), \tag{1.50}$$

for mathematical convenience. From eq. (1.32)

$$\psi(z + Na) = e^{ik(z+Na)}u(z + Na) = e^{ikNa}\psi(z). \tag{1.51}$$

The boundary condition eq. (1.50) then requires

$$kNa = 2\pi\ell \quad \ell = 1, 2, 3, \ldots, N$$

so only discrete values of k are given by

$$k = \frac{2\pi\ell}{Na} \quad \ell = 1, \ldots, N \tag{1.52}$$

are allowed. Since $Na = L$ the sample's length, each state occupies a space $2\pi/L$ in **k**-space. The number of states between k and $k + dk$ on the curves of Fig. 1.4 is $Ldk/2\pi$. In three dimensions the number of states per unit volume of k-space generalizes to $L^3/8\pi^3$. We also need to multiply by two to account for the spin of the two electrons which can occupy a state. We conclude, therefore, that

$$\frac{\text{Number of electron states}}{\text{Volume of } k\text{-space}} = N_k = \frac{\Omega}{4\pi^3} \tag{1.53}$$

where $\Omega = L^3$ is the sample's volume.

We shall frequently have to evaluate sums like

$$\sum_{\mathbf{k}} g(\mathbf{k}),$$

where $g(\mathbf{k})$ is some function of **k** and the sum contains all states in the first Brillouin zone as given in eq. (1.52). It is usually convenient to think of the

$E(\mathbf{k})$ curve as continuous and to integrate rather than sum. To do so, we must properly account for the number of states between **k** and **k** + d**k** as given by eq. (1.53). The result,

$$\boxed{\sum_{\mathbf{k}} g(\mathbf{k}) = N_k \int_{\mathbf{k}} g(\mathbf{k}) d\mathbf{k}}\,, \tag{1.54}$$

is one that we shall often make use of. For device applications, we'll evaluate sums like eq. (1.54) to determine how the carrier density or current density varies with position within the device.

For carriers in a bulk semiconductor, N_k is given by eq. (1.53), and Ω is a purely conceptual box whose dimensions are large compared to an average electron's wavelength but small on the scale of the device. With the use of heterostructures, carriers can be confined in quantum wells, where they are free to move only in two dimensions, or in quantum wires where they can only move in one dimension. Equation (1.53) generalizes to

$$N_k = 2 \times \frac{L^d}{(2\pi)^d}, \tag{1.55}$$

where L is the sample size, d the dimensionality (1, 2, or 3), and the factor of two accounts for spin degeneracy. The integrals are then carried out in one, two, or three dimensions. We shall see many applications of eqs. (1.54) and (1.55).

1.4.2 Density of states in energy space

Equation (1.53) shows that the density of states in **k**-space is constant. We will, however, frequently find it convenient or necessary to deal with the density of states in energy space. Figure 1.17 illustrates the relationship between $N(\mathrm{k})$ and $N(\mathrm{E})$. The states are distributed uniformly in **k**-space, but not in energy space. One way to evaluate $N(E)$ is to construct a histogram. After defining bins of width ΔE, we can scan through all of the allowed **k**-states, evaluate their energy, and increment the count in the appropriate energy bin. Mathematically, the number of states in a range of ΔE about E is

$$N(E)\Delta E = \sum_{\mathbf{k}} \Delta[E - E(\mathbf{k})], \tag{1.56}$$

where the sum is over all states in the Brillouin zone and $\Delta[E - E(\mathbf{k})] = 1$ if $E(\mathbf{k}) - \Delta E/2 < E < E(\mathbf{k}) + \Delta E/2$ and zero otherwise. Letting ΔE approach zero, we obtain

$$g_C(E) = \frac{N(E)}{\Omega} = \frac{1}{\Omega}\sum_{\mathbf{k}'} \delta[E - E(\mathbf{k}')]. \tag{1.57}$$

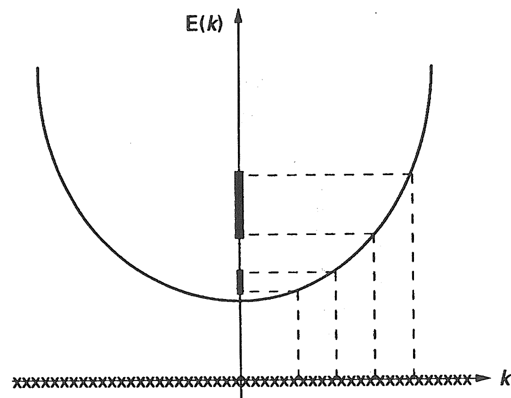


Fig. 1.17 Illustration of the density of states in **k**-space and in energy space. In **k**-space, the density of states is uniform as shown by the x's. A given number of k-states, however, occupies different ranges of energy, as shown by the shaded lines on the energy axis. In this example, $g(E)$ decreases as E increases because we are considering 1D electrons (see eq. (1.63a)).

(In this equation, $\delta(\bullet)$ is actually a Kronecker δ, which becomes a δ-function when the sum is converted to an integral.)

Equation (1.57) can be understood as a count of every state with energy E. We can prove that this is the correct result by evaluating the electron density from

$$n = \frac{1}{\Omega}\sum_{\mathbf{k}} f(\mathbf{k}), \tag{1.58}$$

where the $f(\mathbf{k})$ is the probability that the state at k is occupied. Alternatively, we can evaluate the electron density in energy space from

$$n = \int_{E_{bot}}^{E_{top}} g_C(E) f(E) \mathrm{d}E. \tag{1.59}$$

Using eq. (1.57) for the density of states, we find

$$n = \int_{E_{bot}}^{E_{top}} \frac{1}{\Omega}\sum_{\mathbf{k}'} \delta[E - E(\mathbf{k}')] f(E) \mathrm{d}E. \tag{1.60}$$

By interchanging the order of integration and summation,

$$n = \frac{1}{\Omega}\sum_{\mathbf{k}'} \int_{E_{bot}}^{E_{top}} \delta[E - E(\mathbf{k}')] f(E) \mathrm{d}E, \tag{1.61}$$

we find

$$n = \frac{1}{\Omega}\sum_{\mathbf{k}'} f[E(\mathbf{k}')]. \tag{1.62}$$

Example: Density of states calculation for 3D carriers

A simple example will illustrate the use of eq. (1.57). Recall that the density of states versus energy varies as $E^{1/2}$ for three dimensional electrons. We can obtain this result by evaluating,

$$g_C[E(\mathbf{k})] = \frac{1}{\Omega}\sum_{\mathbf{k}'} \delta[E(\mathbf{k}) - E(\mathbf{k}')].$$

To perform the sum, we convert it to an integral using the prescription given by eq. (1.54) and insert the $E(\mathbf{k})$ relation to find

$$g_C[E(\mathbf{k})] = \frac{1}{4\pi^3}\int_{k'} \delta\left(\frac{\hbar^2 k^2}{2m^*} - \frac{\hbar^2 k'^2}{2m^*}\right) 4\pi k'^2 \mathrm{d}k'.$$

Using $\delta(ax) = \delta(x)/a$, this becomes

$$g_C[E(\mathbf{k})] = \frac{2m^*}{\pi^2\hbar^2}\int_{k'} \delta(k^2 - k'^2) k'^2 \mathrm{d}k'.$$

To integrate a δ-function, we need an expression of the form, $\int \delta(x - x') f(x') \mathrm{d}x' = f(x)$, so letting $x = k^2$, we find

$$g_C(E) = \frac{2m^*}{\pi^2\hbar^2}\int_{k'} \delta(x - x')\frac{k'\mathrm{d}x'}{2} = \frac{m^*}{\pi^2\hbar^2} k = \frac{m^*\sqrt{2m^* E}}{\pi^2\hbar^3},$$

the expected result.

The fact that eq. (1.62) is identical to the correct result, eq. (1.58), verifies that eq. (1.57) is the correct expression for the density of states.

For parabolic energy bands, the density of states goes as $E^{1/2}$, as the example calculation showed. For non-parabolic energy bands, we need to repeat the calculation using eq. (1.40) for $E(k)$. (See homework problem 1.4 for the resulting expression.) As illustrated in Fig. 1.18, nonparabolicity flattens the energy bands, so there are more k-states between E and $E + \mathrm{d}E$ and the density of states increases. In general, it will be necessary to evaluate eq. (1.56) numerically using a table of $E(\mathbf{k})$.

Figure 1.19 compares the density of states in silicon assuming full, numerical energy bands, to that evaluated from the parabolic and nonparabolic expressions. The structure in the full band density of states is a result of the $E(\mathbf{k})$ relation plotted in Fig. 1.9. The sharp drop above ≈ 2 eV results from the fact that the first conduction band extends to only about 2 eV. The parabolic band assumption is seen to apply only to very low energy carriers near the band minima. The non-parabolic energy band assumption provides a rough approximation to almost 2 eV, but for very energetic carriers, the full, numerical density of states must be used.

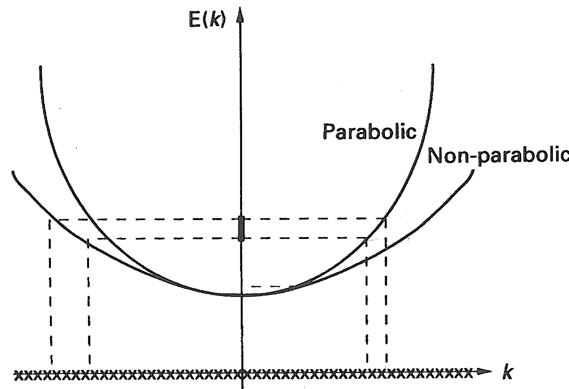


Fig. 1.18 Illustration of how conduction band nonparabolicity flattens the $E(k)$ relation and increases the density of states in energy space. For a given dE, shown in dark on the energy axis, there are more allowed k-states for the nonparabolic energy band.

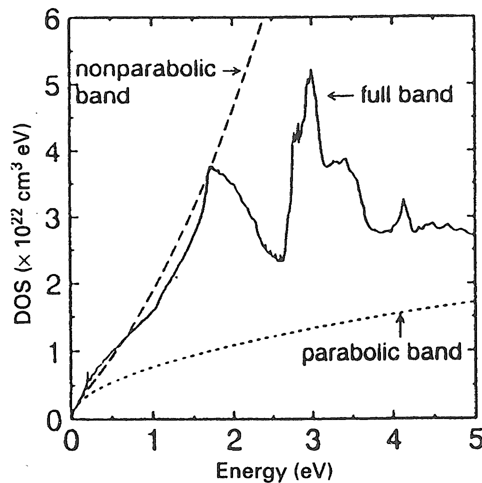


Fig. 1.19 Comparison of the density of states (DOS) for the conduction bands of silicon. The results for parabolic and non-parabolic energy bands are compared to the result using the full numerical description of the energy bands. From Kunikiyo, T. et al., *Journal of Applied Physics,* **75**(1), 297–312, 1994. (Reproduced with permission of American Institute of Physics.)

1.4.3 Density of states for confined carriers

For carriers in a bulk semiconductor, the density-of-states for parabolic energy bands goes as $E^{1/2}$, but for confined carriers, the density of states is altered. Using eqs. (1.54) and (1.55), we find the one-, two-, or three-dimensional density of states as (see homework problem 1.3)

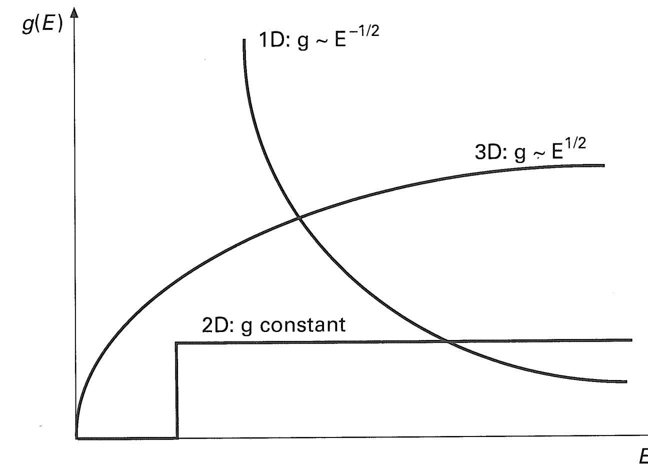


Fig. 1.20 The density of states versus energy for 1, 2, and 3-dimensional carriers with parabolic energy bands. The energy E has been referenced to E_{min}, the conduction band minima for three-dimensional carriers. For two- and one-dimensional carriers, the minimum energy is raised by quantum confinement.

$$g_{1D} = \frac{m^*}{\pi\hbar^2}\frac{2\hbar}{\sqrt{2m^*E}} \tag{1.63a}$$

for one-dimensional carriers, and

$$g_{2D} = \frac{m^*}{\pi\hbar^2} \tag{1.63b}$$

for two-dimensional carriers. These results should be compared to

$$g_{3D} = \frac{m^*}{\pi\hbar^2}\frac{\sqrt{2m^*E}}{\pi\hbar} \tag{1.63c}$$

for three-dimensional carriers. Figure 1.20 sketches the density-of-states versus energy for one-, two-, and three-dimensional carriers. Because carrier confinement is common in modern devices, we shall have to become familiar with evaluating carrier densities, average kinetic energies, scattering rates, etc. in one, two, and three dimensions.

1.5 Electron wave propagation in devices

An electron propagating within a device sees both the crystal potential and those that are applied or built-in to the device. When the applied and built-in fields are absent, only the crystal potential, $U_C(z)$, sketched in Fig. 1.21a is present. The solutions to the wave equation are well known for the crystal potentials of

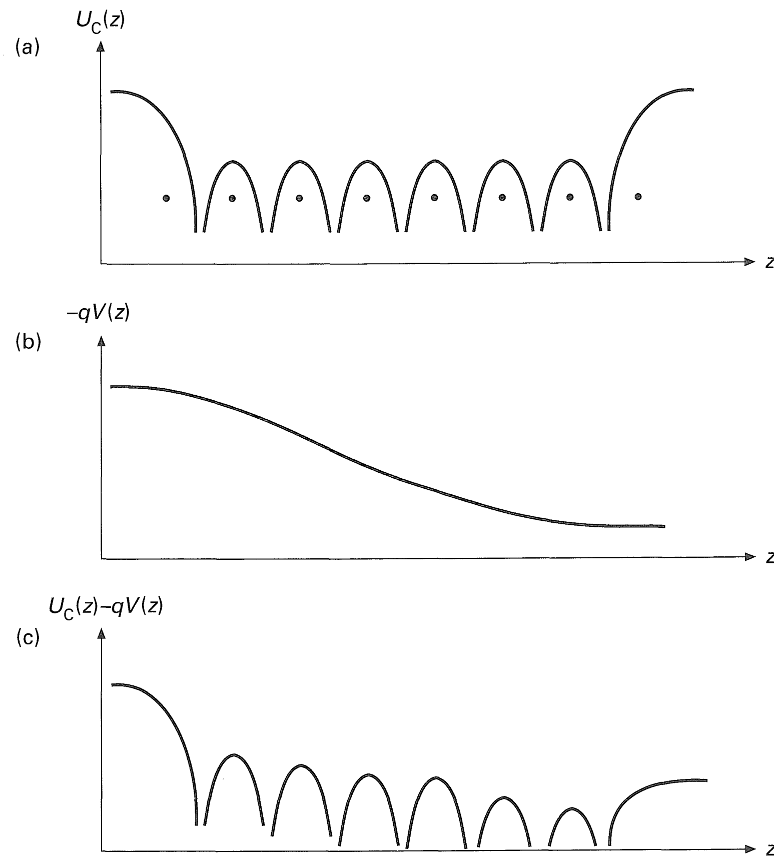


Fig. 1.21 Illustration of the crystal and applied potentials within a semiconductor. (a) The crystal potential versus position, (b) the applied or built-in potential versus position, (c) the total potential versus position.

common semiconductors [1.6]. In devices, however, another potential can be built-in by varying the doping or material composition or imposed by biasing the device. [This is the potential we refer to as $E_{C0}(\mathbf{r})$.] As sketched in Fig. 1.21b, the applied and built-in potentials often vary slowly in comparison to the crystal potential [but with modern epitaxial growth techniques, potentials that vary as rapidly as the crystal potential can also be engineered into the device (recall Fig. 1.15)]. Electrons see both the crystal and applied or built-in potentials as sketched in Fig. 1.21c.

1.5.1 The effective mass equation

When scattering can be neglected, the electron's wave function is found by solving

$$\left[-\frac{\hbar^2}{2m_0}\frac{d^2}{dz^2}+U_C(z)+E_{C0}(z)\right]\Psi(z,t)=i\hbar\frac{\partial\Psi}{\partial t}. \tag{1.64}$$

Since the Bloch function solutions to eq. (1.64) in the absence of $E_{C0}(z)$ are assumed to be known, the question of whether eq. (1.64) can be simplified by using these known solutions arises. The answer is yes, for carriers near the bottom of a simple, spherical, parabolic band, the wave equation can be written as

$$\left[-\frac{\hbar^2}{2m^*}\frac{d^2}{dz^2}+E_{C0}(z)\right]F(z,t)=i\hbar\frac{\partial F}{\partial t} \tag{1.65}$$

where $F(z, t)$ is the *envelope function*, and the actual wave function is

$$\Psi(z,t)\cong F(z,t)u_{k=0}. \tag{1.66}$$

The wave function is the product of a slowly varying envelope function and a rapidly varying Bloch function evaluated at the band minimum. Equation (1.65) is known as the single band *effective mass equation* and represents an enormous simplification of eq. (1.64) because the effects of the complicated crystal potential have been described by a single number, the effective mass. Equation (1.65) applies only when the applied or built-in potential varies slowly on the scale of the crystal potential. This certainly is not the case for the quantum well sketched in Fig. 1.15, but if the well is not too narrow, then an effective mass equation usually provides a good description of the energy levels for electrons within the well. Equation (1.65) also applies only to a parabolic band. The effective mass equation needs to be generalized when the band is non-parabolic or, in the case of valence bands, when several nearby energy bands are coupled. The derivation of the effective mass equation, and its extension to more realistic band structures, are discussed by Datta [1.1].

In devices, the contacts launch electron waves which propagate through the device according to the effective mass equation. A device can be described by specifying its energy band diagram as displayed in Fig. 1.22. The 'contacts' are heavily doped regions where $E_{C0}(z)$ is uniform; they are assumed to be near thermodynamic equilibrium so that each can be described by its own Fermi level. To compute the current through the device, we evaluate the sum

$$J_z=\frac{(-q)}{\Omega}\sum_{\mathbf{k}}\left\{f_L(\mathbf{k})\left(\frac{\hbar k_z}{m^*}\right)T_{LR}(\mathbf{k})-f_R(\mathbf{k})\left(\frac{\hbar k_z}{m^*}\right)T_{RL}(\mathbf{k})\right\}. \tag{1.67}$$

In this equation, $\hbar k_z/m^*$ is the velocity of electrons as they are injected from the contact with wave vector, $\mathbf{k}$, $f_L(\mathbf{k})$ is the Fermi factor for the left contact, which gives the probability that such an electron is injected from the contact, and $T_{LR}(\mathbf{k})$ is the current transmission coefficient for the electron. In Section 1.1, we

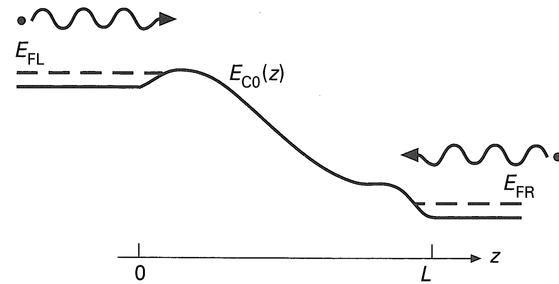


Fig. 1.22 Representation of a device by its energy band profile. Each contact, assumed to be in thermodynamic equilibrium, injects electrons into the device and absorbs electrons incident upon it.

computed $T(\mathbf{k})$ for a simple potential step. For arbitrary potential profiles, the current transmission coefficient is found by numerically solving the effective mass equation. The sum of eq. (1.67) accounts for the current due to all the electrons injected from each of the two contacts, and must, in general, be evaluated numerically.

Device analysis based on solving the effective mass equation is necessary when the potential, $E_{C0}(z)$, varies rapidly so that wave phenomena are important. Vassell et al. [1.8] describes how this technique is applied to devices.

1.5.2 Quantum confinement

Because electrons confined within a small potential well experience a rapidly varying potential, the carriers' wave nature becomes important. For the quantum well illustrated in Fig. 1.23a, electrons are confined in the $\hat{z}$-direction but are free to move in the $\hat{x}-\hat{y}$ plane. For the quantum wire illustrated in Fig. 1.23b, electrons are confined in the $\hat{x}-\hat{y}$ plane but are free to move in the $\hat{z}$-direction. Such structures can be produced with semiconductor heterojunctions, as illustrated in Fig. 1.15 for the quantum well. To describe confined carriers, the three-dimensional effective mass equation,

$$-\frac{\hbar^2}{2m^*}\nabla^2 F(\mathbf{r}) + E_{C0}(\mathbf{r})F(\mathbf{r}) = EF(\mathbf{r}) \tag{1.68}$$

must be solved. Equation (1.68) is readily solved by separating variables. For quantum wells, carriers are free to move in the $\hat{x}-\hat{y}$ plane, so we try solutions of the form

$$F(\mathbf{r}) = \phi(z)\frac{e^{ik_x x}e^{ik_y y}}{\sqrt{L_x L_y}} \tag{1.69}$$

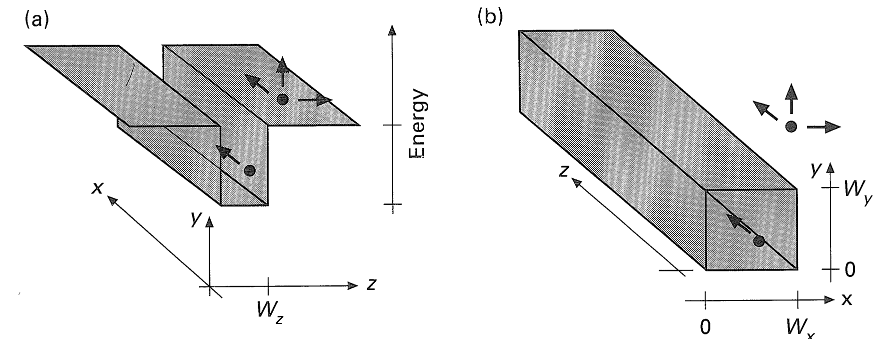


Fig. 1.23 Electrons confined in a quantum well (a) and in a quantum wire (b).

where $A = L_x L_y$ is the cross-sectional area. After substituting eq. (1.69) into eq. (1.68), we find an equation for $\phi(z)$ as

$$\frac{d^2\phi(z)}{dz^2} + k_z^2\phi(z) = 0, \tag{1.70}$$

where

$$k_z^2 = \frac{2m^*}{\hbar^2}[\varepsilon - E_{C0}(z)] \tag{1.71}$$

and

$$\varepsilon = E - \frac{\hbar^2}{2m^*}(k_x^2 + k_y^2). \tag{1.72}$$

Because $\hbar^2(k_x^2 + k_y^2)/2m^*$ is the kinetic energy we associate with motion in the $\hat{x}-\hat{y}$ plane, ε must be the energy associated with confinement in the $\hat{z}$-direction.

Equation (1.70) is identical in form to the simple, one-dimensional wave equation, eq. (1.3). The three-dimensional wave equation consists of a plane wave in the $\hat{x}-\hat{y}$ plane multiplied by a function, $\phi(z)$, which is found by solving an equation that is very similar to the one-dimensional wave equation. If the quantum well is deep, then $\phi(z)$ is given by the infinite well solutions of Section 1.1 as

$$\phi(z) = \sqrt{\frac{2}{W}}\sin k_z z, \tag{1.73}$$

where

$$k_z = n\pi/W \tag{1.74}$$

and

$$E = E_{C0} + \frac{\hbar^2 k_z^2}{2m^*} + \frac{\hbar^2}{2m^*}(k_x^2 + k_y^2)$$

or

$$E = E_{C0} + \varepsilon_n + \frac{\hbar^2}{2m^*}(k_x^2 + k_y^2), \tag{1.75}$$

where ε_n is given by eq. (1.17). Quantum confinement restricts k_z to discrete values, and the energy consists of a component due to confinement in the $\hat{z}$-direction and one due to the free motion in the $\hat{x} - \hat{y}$ plane.

Carriers in quantum wires are treated in a very similar manner. Instead of eq. (1.69), we write the wavefunction as

$$F(\mathbf{r}) = \phi(x, y)\frac{e^{ik_z z}}{\sqrt{L_z}}. \tag{1.76}$$

If the confinement potential is infinite, we find

$$\phi(x, y) = \sqrt{\frac{2}{W_x}}\sin k_x x\sqrt{\frac{2}{W_y}}\sin k_y y \tag{1.77}$$

where

$$k_x = n_x\pi/W_x \tag{1.78a}$$

and

$$k_y = n_y\pi/W_y. \tag{1.78b}$$

For quantum wires, the bottom of subbands are at the energies,

$$\varepsilon_{n_x,n_y} = \frac{\hbar^2\pi^2}{2m^*}\left(\frac{n_x^2}{W_x^2} + \frac{n_y^2}{W_y^2}\right) \tag{1.79}$$

and if the electron is moving in the $\hat{z}$-direction, its total energy is

$$E = E_{C0} + \varepsilon_{n_x,n_y} + \frac{\hbar^2 k_z^2}{2m^*}. \tag{1.80}$$

1.5.2.1 Carrier density relations for confined carriers

For three-dimensional carriers in a bulk semiconductor, there is a simple relation between the equilibrium carrier density and the location of the Fermi level. The corresponding relations for confined carriers are readily derived. Figure 1.24a shows three energy levels, or *subbands*, in a quantum well along with the position of the Fermi level. To compute the density of electrons in the well, we evaluate

$$n = \frac{1}{\Omega}\sum_{k_x,k_y,k_z} f_0[E(k_x, k_y, k_z)] \quad \text{cm}^{-3}, \tag{1.81}$$

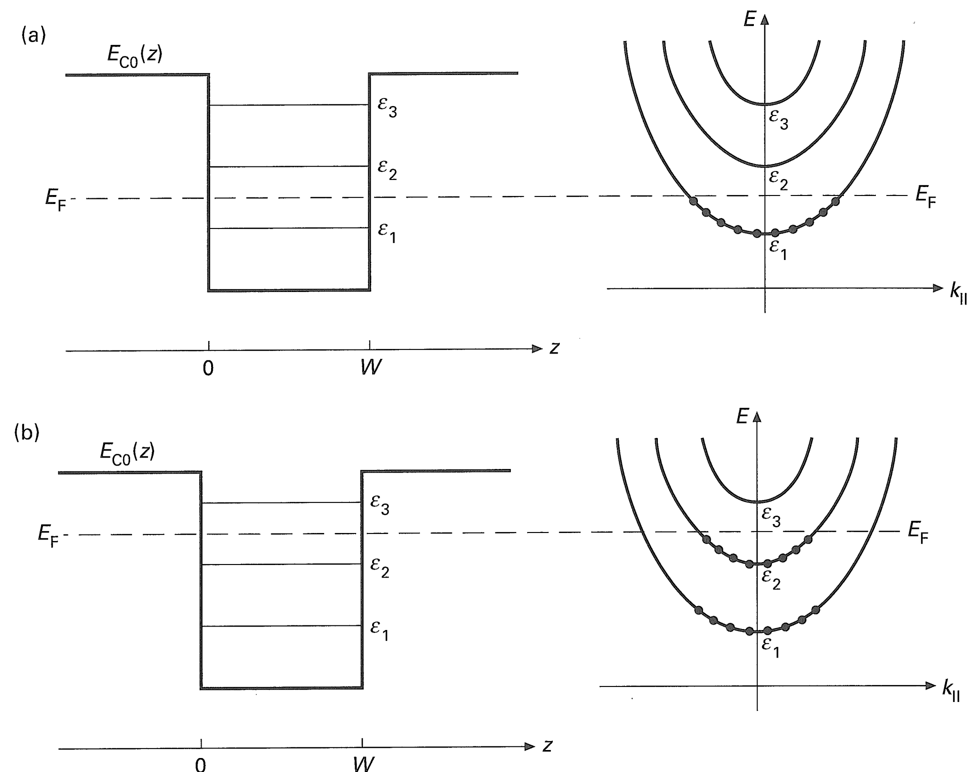


Fig. 1.24 Energy band diagram of a quantum well with the location of three subbands and the Fermi level indicated. (a) The filled circles at the right indicate occupied states and show that only the first subband is occupied. ($T = 0$ K is assumed.) (b) The first two subbands are occupied.

where the sum is over each state in the Brillouin zone, and f_0 gives the probability that the state at energy, E, with crystal momentum, (k_x, k_y, k_z), is occupied.

The sum, eq. (1.81), is easiest to evaluate at $T = 0$ K because then all states below E_F are occupied, and all those above E_F are empty. For the example illustrated in Fig. 1.24a, only states with $k_z = k_{z1}$ are occupied, so eq. (1.81) becomes

$$n = \frac{1}{W}\frac{1}{L_x L_y}\sum_{k_x,k_y} f_0[E(k_x, k_y, k_{z1})]. \tag{1.82}$$

Using eqs. (1.54) and (1.55), we convert the sum over wave vectors in the $\hat{x} - \hat{y}$ plane to an integral as

$$n_S = nW = \frac{1}{2\pi^2}\int_0^\infty f_0(E)2\pi k_\| dk_\| \ \text{cm}^{-2}, \tag{1.83}$$

where n_S is the electron density per unit area, and

$$k_\parallel^2 = k_x^2 + k_y^2. \tag{1.84}$$

Only states below E_F are occupied, so from eq. (1.75) we find that all states above

$$k_\parallel^2 = k_F^2 = \frac{2m^*}{\hbar^2}(E_F - \varepsilon_1) \tag{1.85}$$

are empty. At $T = 0\,\mathrm{K}$, eq. (1.83) becomes

$$n_S = \frac{1}{2\pi^2}\int_0^{k_F} 2\pi k_\parallel \mathrm{d}k_\parallel = \frac{k_F^2}{2\pi}$$

or

$$n_S = \left(\frac{m^*}{\pi\hbar^2}\right)(E_F - \varepsilon_1) = g_{2D}(E_F - \varepsilon_1). \tag{1.86}$$

If the Fermi level lies above two subbands, as it does in Fig. 1.24b, then the sum in eq. (1.81) becomes

$$n_S = \frac{1}{L_x L_y}\sum_{k_x,k_y} f_0[E(k_x,k_y,k_{z1})] + \frac{1}{L_x L_y}\sum_{k_x,k_y} f_0[E(k_x,k_y,k_{z2})]. \tag{1.87}$$

For this case, the carrier density is simply the sum of the contributions from the two subbands,

$$n_S = g_{2D}(E_F - \varepsilon_1) + g_{2D}(E_F - \varepsilon_2). \tag{1.88}$$

The corresponding results for finite temperatures are also readily derived. We begin with eq. (1.81) but use the Fermi function,

$$f_0 = \frac{1}{1 + \mathrm{e}^{(\varepsilon_1 + \hbar^2 k_\parallel^2/2m^* - E_F)/k_B T}}. \tag{1.89}$$

Alternatively, we can work in energy space and evaluate

$$n_S = \int_{\varepsilon_1}^{\infty} \frac{g_{2D}\mathrm{d}E}{1 + \mathrm{e}^{(E - E_F)/k_B T}}. \tag{1.90}$$

In either case, for one occupied subband, we find

$$n_S = g_{2D} k_B T \ln(1 + \mathrm{e}^{(E_F - \varepsilon_1)/k_B T})\ \mathrm{cm}^{-2}, \tag{1.91}$$

which is analogous to

$$n = N_C \mathcal{F}_{1/2}[(E_F - E_C)/k_B T]\ \mathrm{cm}^{-3} \tag{1.92}$$

for three-dimensional electrons. Here, $\mathcal{F}_{1/2}$ is the Fermi–Dirac integral of order 1/2, and

$$N_C = 2\left(\frac{2\pi m^* k_B T}{h^2}\right)^{3/2} \tag{1.93}$$

is the effective density of states. When additional subbands are occupied, eq. (1.91) is easily generalized by adding the contributions from the additional subbands.

Quantum confinement often occurs in modern devices such as heterostructure field effect transistors in which the channel comprises carriers confined in a quantum well [1.10]. Even in the conventional, silicon MOSFET, carriers are confined within a nearly triangular potential well at the oxide–silicon interface. The wave functions of confined electrons are qualitatively different from the plane waves that describe three-dimensional, bulk electrons. There is a close analogy between these confined electrons and electromagnetic waves in a waveguide. The various subbands are analogous to the waveguide modes; occupied subbands correspond to propagating modes, unoccupied subbands to evanescent modes. This analogy can even be exploited to build electron devices analogous to optical or microwave devices.

1.6 Semiclassical electron dynamics

For conventional devices, the applied or built-in potentials vary slowly in comparison to the crystal potential, so that wave phenomena such as reflections and tunneling are absent and electron motion can be described by classical physics. When the potential is nearly constant, the bottom of the band is simply shifted,

$$E(k, z) = E_{C0}(z) + E(k). \tag{1.94}$$

$E_{C0}(z)$ is interpreted as the bottom of the conduction band, and $E(k)$ represents only the kinetic energy. Since $E_{C0} = \text{constant} - qV(z)$, where $V(z)$ is the electrostatic potential, it varies with position only when applied or built-in fields are present.

Equation (1.94) is illustrated in Fig. 1.25. As the electron wave packet (centered at k_0) moves without scattering, its total energy remains constant. From eq. (1.94) we have

$$\frac{\mathrm{d}E(k_0, z)}{\mathrm{d}t} = \frac{\partial E}{\partial k_0}\cdot\frac{\mathrm{d}k_0}{\mathrm{d}t} + \frac{\partial E}{\partial z}\cdot\frac{\mathrm{d}z}{\mathrm{d}t}$$
$$= \upsilon_g \cdot \hbar\frac{\mathrm{d}k_0}{\mathrm{d}t} + \frac{\partial E_{C0}(z)}{\partial z}\cdot \upsilon_g.$$

Because energy must be conserved, $\mathrm{d}E/\mathrm{d}t = 0$ and we conclude that

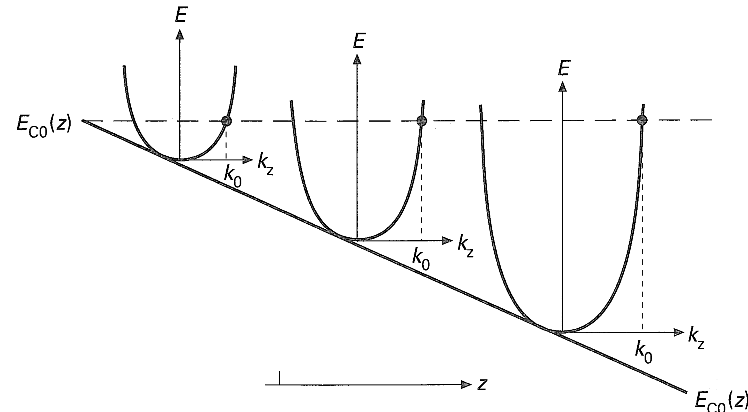


Fig. 1.25 Motion of an electron wave packet centered at $k_z = k_0$ across a region of slowly varying potential. After Datta [1.1].

$$\boxed{\frac{d(\hbar \mathbf{k}_0)}{dt} = -\nabla E_{C0}(z) = \mathbf{F}_e} \tag{1.95a}$$

where $\mathbf{k}_0$ is the wave vector at the center of the wave packet. Because eq. (1.95a) is so similar to the equation of motion for classical particles, with $\hbar\mathbf{k}_0$ playing the role of momentum, $\hbar\mathbf{k}_0$ is termed the *crystal momentum.*

For heterostructures, the effective mass may vary with position, so

$$E_C(z,k) = E_{C0} + \frac{\hbar^2 k^2}{2m^*(z)}.$$

For heterostructures, the equation of motion generalizes to (see homework problem 1.13)

$$\boxed{\frac{d(\hbar \mathbf{k}_0)}{dt} = -\nabla E_{C0}(z) - \nabla\left(\frac{\hbar^2 k^2}{2m^*(z)}\right)} \tag{1.95b}$$

In this case, only the first term represents a real, physical force on carriers.

The analogy of $\hbar\mathbf{k}_0$ to momentum is also apparent from the velocity of a Bloch electron as given by eq. (1.14). For spherical, parabolic bands described by eq. (1.38), we obtain

$$\upsilon_g = \frac{\hbar \mathbf{k}_0}{m^*}, \tag{1.96}$$

which looks like momentum divided by mass. But for nonparabolic bands described by eq. (1.40), the group velocity is

$$\upsilon_g = \frac{\hbar \mathbf{k}_0}{m^*[1 + 2\alpha E(k)]}. \tag{1.97}$$

In the semiclassical view of electron transport, the electron wave packet is treated as a particle; the uncertainty in the momentum is assumed to be small so that the electron's energy is sharply defined, the uncertainty in the electron's position is assumed to be small in comparison to the distance over which applied and built-in potentials vary significantly. The motion of the center of this wave packet is described by eq. (1.95), which looks like the classical relation between force and momentum. The velocity of the electron, eq. (1.14), corresponds to the velocity of a classical particle only for spherical, parabolic bands. This semiclassical treatment of carrier dynamics is the basis for each of the following chapters, but collisions involve rapidly varying potentials and must be treated quantum mechanically.

1.7 Scattering of electrons by the random potential, $U_S(\mathbf{r}, t)$

Bloch waves move through the lattice unimpeded by the crystal potential. Occasionally, however, the electron encounters a perturbation caused when a lattice vibration moves an atom or by impurities or defects which may be present. When an electron encounters such a perturbation it scatters – scattering 'knocks' an electron wave packet centered at k_0 to k_0'. Frequent scattering tends to 'wash out' the interference effects due to the carrier's wave nature. Scattering plays a dominant role in transport, and it is important that we know $S(k_0, k_0')$, the *transition rate* from k_0 to k_0'. We now present a brief derivation of the expression for $S(k_0, k_0')$ in terms of $U_S(z, t)$, the perturbing potential. For a proper derivation, consult a quantum mechanics textbook such as Datta [1.1]. The intent here is to indicate how the result is derived and some of its limitations. We will make extensive use of the result, known as *Fermi's Golden Rule*, to calculate scattering rates for electrons in semiconductors, so the reader should develop a familiarity with its use.

1.7.1 Fermi's Golden Rule

The wave equation, [eq. (1.1)] is written as

$$[H_0 + U_S(z,t)]\Psi(z,t) = i\hbar\frac{\partial\Psi(z,t)}{\partial t}, \tag{1.98}$$

where H_0 is the Hamiltonian operator for the unperturbed problem (the problem without the scattering potential). We assume that the unperturbed problem:

$$H_0\psi_k = E(k)\psi_k \tag{1.99}$$

$$\Psi_k^0(z, t) = \psi_k(z)e^{-iE(k)t/\hbar} \tag{1.100}$$

has been solved. These solutions form a complete, orthonormal set, so we can express the solution to the perturbed problem as linear combinations of them:

$$\Psi(z, t) = \sum_k c_k(t)\Psi_k^0(z, t) = \sum_k c_k(t)\psi_k(z)e^{-iE(k)t/\hbar}. \tag{1.101}$$

Now consider the situation sketched in Fig. 1.26 – an electron wave packet centered at $k = k_0$ enters, interacts with $U_S(z, t)$, and emerges centered at k_0'. At $t = 0$ we have

$$\begin{aligned} c_{k_0}(t=0) &= 1 \\ c_k(t=0) &= 0 \qquad (k \neq k_0). \end{aligned} \tag{1.102}$$

After the scattering event, the probability of finding the electron with wave vector, k_0', is

$$P(k = k_0') = \lim_{t\to\infty} \left|c_{k_0'}(t)\right|^2 \tag{1.103}$$

so the scattering rate from k_0 to k_0' is

$$S(k_0, k_0') = \lim_{t\to\infty} \frac{|c_{k_0'}(t)|^2}{t}. \tag{1.104}$$

(To allow $t \to \infty$ in these expressions, without another collision occurring, collisions must be infrequent.)

To find c_k, we insert eq. (1.101) in eq. (1.98) and obtain

$$U_S(z, t)\sum_k c_k(t)\psi_k e^{-iE(k)t/\hbar} = i\hbar\sum_k \frac{\partial c_k}{\partial t}\psi_k e^{-iE(k)t/\hbar}. \tag{1.105}$$

Next, we multiply both sides by $\psi_{k_0'}^* e^{iE(k_0')t/\hbar}$, integrate over position, and make use of the orthogonality of eigenfunctions, to find

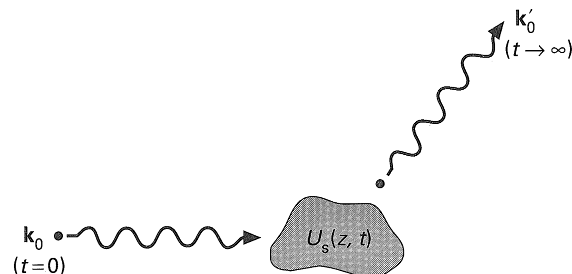


Fig. 1.26 Scattering of a wave packet centered at $k = k_0$ to one centered at $k = k_0'$.

$$i\hbar\frac{\partial c_{k_0'}}{\partial t} = \sum_k H_{k_0'k}c_k(t)e^{i(E(k_0')-E(k))t/\hbar}, \tag{1.106}$$

where

$$\boxed{H_{k_0',k_0}(t) \equiv \int_{-L/2}^{+L/2} \psi_{k_0'}(z)U_S(z, t)\psi_{k_0}(z)dz} \tag{1.107}$$

is the *matrix element* of the scattering potential between states k_0' and k. We have normalized the wavefunctions over a length L, which becomes a volume, Ω, in three dimensions.

Since we assume that the scattering is weak, $c_{k_0} \simeq 1$ for all time, and the other c_k's are always small. With this approximation (the so-called *Born approximation*) the sum in eq. (1.106) can be approximated by one term as

$$i\hbar\frac{\partial c_{k_0'}}{\partial t} = H_{k_0'k_0}(1)e^{i[E(k_0')-E(k_0)]t/\hbar},$$

which can be integrated to find

$$c_{k_0'}(t) = \frac{1}{i\hbar}\int_0^t H_{k_0'k_0}e^{i[E(k_0')-E(k_0)]t/\hbar}dt + c_{k_0'}(0). \tag{1.108}$$

Because the final state k_0' was empty at $t = 0$, $c_{k_0'}(0) = 0$.

Let's specify the time-dependent matrix element as

$$H_{k_0',k_0}(t) = H_{k_0'k_0}^{a,e}e^{\mp i\omega t}. \tag{1.109}$$

(The significance of the a and e superscripts which apply the minus and plus signs respectively will be explained shortly.) With eq. (1.109) for the matrix element, eq. (1.108) can be integrated as

$$c_{k_0'} = \frac{1}{i\hbar}H_{k_0'k_0}^{a,e}\frac{e^{i(E(k_0')-E(k_0)\mp\hbar\omega)t/\hbar} - 1}{i[E(k_0') - E(k_0) \mp \hbar\omega]/\hbar}. \tag{1.110}$$

When we define

$$\Lambda = \left[E(k_0') - E(k_0) \mp \hbar\omega\right]/\hbar, \tag{1.111}$$

then eq. (1.110) can be written as

$$c_{k_0'}(t) = \frac{1}{i\hbar}H_{k_0'k_0}^{a,e}e^{i\Lambda t/2}\frac{\sin(\Lambda t/2)}{\Lambda t/2}t. \tag{1.112}$$

Now, according to eq. (1.104), we find the transition rate as

$$S(k_0, k_0') = \lim_{t\to\infty} \frac{\left|H^{a,e}_{k_0'k_0}\right|^2}{t\hbar^2}\left[\frac{\sin(\Delta t/2)}{\Delta t/2}\right]^2 t^2. \quad (1.113)$$

For large t, the function in brackets is very sharply peaked near the origin and looks like a δ-function. The strength of the δ-function is determined from the area under the curve. Recall that

$$\int_{-\infty}^{\infty} \frac{\sin^2 x}{x^2}\mathrm{d}x = \pi,$$

so $\sin^2 x/x^2$ can be replaced by $\delta(x)$. Using $x = \Delta t/2$, we find the replacement

$$\lim_{t\to\infty} \frac{\sin^2(\Delta t/2)}{(\Delta t/2)^2} = \frac{2\pi}{t}\delta(\Delta) \quad (1.114)$$

which can be inserted in eq. (1.113) to find

$$\boxed{\begin{aligned} S(k_0, k_0') &= \frac{2\pi}{\hbar}\left|H^a_{k_0'k_0}\right|^2 \delta(E(k_0') - E(k_0) - \hbar\omega) \\ &+ \frac{2\pi}{\hbar}\left|H^e_{k_0'k_0}\right|^2 \delta(E(k_0') - E(k_0) + \hbar\omega) \end{aligned}} \quad . \quad (1.115)$$

The δ-function in eq. (1.115) simply expresses conservation of energy and applies when scattering is weak, so that time can approach infinity in eq. (1.113). For frequent scattering, there is an uncertainty in the final energy, given by eq. (1.12), which is known as *collisional broadening*. The first term in eq. (1.115) contributes when $E(k_0') = E(k_0) + \hbar\omega$; an energy of $\hbar\omega$ has been absorbed. The second contributes when $E(k_0') = E(k_0) - \hbar\omega$; an energy of $\hbar\omega$ has been emitted.

Equation (1.115) is the basic result of scattering theory that we will apply to carriers in semiconductors. The result is known as *Fermi's Golden Rule*. To apply the Golden Rule, the scattering potential must be identified so that the matrix element can be evaluated. For electrons in semiconductors, the wave functions for the unperturbed problem are Bloch waves. When the matrix element, eq. (1.107) is evaluated for Bloch waves, one finds [1.5]:

$$H_{k'k} = I(k, k')U_S(k - k') \quad (1.116)$$

where

$$I(k, k') \equiv \int_{\text{cell}} u^*_{k'}(z)u_k(z)\mathrm{d}z \quad (1.117)$$

is called the *overlap integral* (the integral is over a unit cell), and

$$U_S(k - k') = \int_{-L/2}^{+L/2} \frac{\mathrm{e}^{-ik'z}}{\sqrt{L}} U_S(z)\frac{\mathrm{e}^{+ikz}}{\sqrt{L}}\mathrm{d}z. \quad (1.118)$$

For a parabolic band, $I(k, k') \simeq 1$ [1.5] and

$$\boxed{H_{k'k} \simeq \frac{1}{L}\int_{-L/2}^{+L/2} \mathrm{e}^{-ik'z}U_S(z)\mathrm{e}^{ikz}\mathrm{d}z} \quad , \quad (1.119)$$

which is just what we would have obtained from eq. (1.107) using plane waves rather than Bloch waves. When we evaluate scattering rates in Chapter 2, we'll keep the algebra to a minimum by assuming that the energy bands are parabolic and employ eq. (1.119), but for quantitative work, overlap integrals should be considered.

1.7.2 Examples

To illustrate how the Golden Rule is applied to scattering problems, we consider two simple, but illustrative, examples. First, we consider scattering from a δ-function potential, which might approximate a short range scattering potential. Second, we consider a periodic perturbing potential, which might represent, for example, a lattice vibration.

Example: Scattering from a δ-function potential

Consider a perturbing potential of the form

$$U_S(z) = \Delta_0\delta(z). \quad (1.120)$$

From eq. (1.119) we find

$$H_{k'k} = \frac{\Delta_0}{L}, \quad (1.121)$$

and from eq. (1.115), the transition rate becomes

$$S(k, k') = \frac{2\pi}{\hbar}\frac{\Delta_0^2}{L^2}\delta[E(k') - E(k)]. \quad (1.122)$$

This time-independent scattering potential elastically scatters electrons with a transition rate that is proportional to the squared magnitude of the scattering potential. The δ-function potential is an approximate description of ionized impurity scattering when it is strongly screened by free carriers.

Example: Scattering from a periodic potential

As a second example, consider the scattering potential,

$$U_S(z,t) = A_\beta^{a,e} e^{\pm i(\beta z - \omega t)}. \quad (1.123)$$

For plane waves confined to a normalization length, $-L/2 \leq z \leq L/2$,

$$\psi(z) = \frac{1}{\sqrt{L}} e^{ikz}, \quad (1.124)$$

and the matrix element becomes

$$H_{k'k} = \int_{-L/2}^{L/2} \frac{A_\beta^{ae}}{L} e^{i(k-k' \pm \beta)z} dz. \quad (1.125)$$

If the normalization length is long, the oscillating, exponential factor ensures that no net contribution to the integral will result unless $k' = k \pm \beta$, so

$$H_{k'k} = A_\beta^{a,e} \delta_{k',k\pm\beta}, \quad (1.126)$$

where the Kronecker delta, δ_{ij}, is defined to be one if $i = j$ and zero for $i \neq j$. For this scattering potential, the transition rate is

$$S(k,k') = \frac{2\pi}{\hbar} \left| A_\beta^{a,e} \right|^2 \delta[E(k') - E(k) \mp \hbar\omega] \delta_{k',k\pm\beta}, \quad (1.127)$$

which is, again, proportional to the squared magnitude of the perturbing potential. The δ-function in eq. (1.127) states that

$$E(k') = E(k) \pm \hbar\omega, \quad (1.128)$$

which is a statement of conservation of energy. For time-dependent scattering potentials like eq. (1.123), carriers either absorb or emit energy. To satisfy the Kronecker δ,

$$\hbar k' = \hbar k \pm \hbar\beta, \quad (1.129)$$

which we interpret as a statement of conservation of momentum. The scattered momentum has either absorbed or emitted momentum. (The momentum-conserving Kronecker δ was absent in eq. (1.122) because the δ-function scattering potential contained Fourier components with all momenta.) This scattering potential is a good description of the perturbing potential due to lattice vibrations (phonons).

1.8 Lattice vibrations (phonons)

Because much of the scattering in semiconductors is due to lattice vibrations, it is important that we understand their basic properties. If an atom is displaced from its equilibrium position, the bonding forces tend to push it back, so it oscillates about its equilibrium site. Since lattice waves propagate in a periodic medium, they have properties much like those of Bloch waves. Figure 1.27a shows a typical dispersion relation, ω versus β, observed for elastic waves in cubic semiconductors like silicon and gallium arsenide. (We label the wave vector by $\boldsymbol{\beta}$

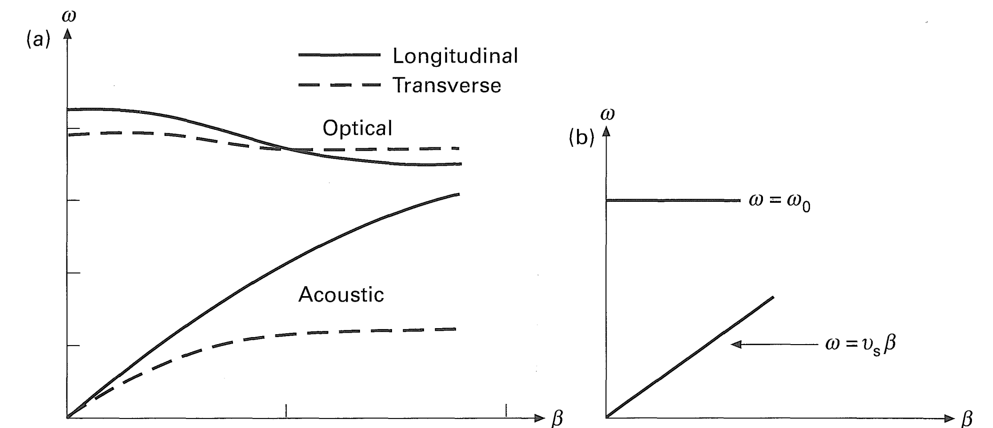


Fig. 1.27 (a) Typical dispersion relation for elastic waves propagating along a high-symmetry direction in cubic semiconductors. (b) Simplified dispersion relation useful when only longitudinal lattice vibrations near the center of the Brillouin zone are considered. After Datta [1.1]. (Reproduced with permission from Addison-Wesley)

rather than **k** to distinguish elastic waves from electron waves.) Six types of elastic wave exist – three *acoustic* modes, and three *optical* modes. Acoustic modes are like sound waves in that adjacent atoms are displaced in the same direction – only the magnitude of the displacement varies from atom to atom. Of the three acoustic modes, one is longitudinal (LA) and two are transverse (TA). For longitudinal waves, atoms are displaced in the direction of propagation; the two transverse modes, in which atoms are displaced in a transverse direction, are degenerate in cubic silicon and GaAs.

In Chapter 2 we shall establish that the scattering of electrons within a valley is due to lattice vibrations with wave vectors very near the origin of the Brillouin zone. For small β, the dispersion relation for acoustic modes can be approximated by

$$\omega(\beta) = \upsilon_s \beta, \quad (1.130)$$

where υ_s is the sound velocity.

Optical modes differ from acoustic modes in that adjacent atoms are displaced out of phase. (The term arises because such vibrations can interact strongly with light.) As shown in Fig. 1.27a, the dispersion relation for optical modes displays relatively little variation with wave vector. When electrons are scattered by optical phonons and remain within the same valley, only small wave vectors are involved and the dispersion relation can be approximated as

$$\omega(\beta) = \omega_0, \quad (1.131)$$

where ω_0 is a constant. Figure 1.27b shows a simplified dispersion relation for acoustic and optical modes that is often used for scattering calculations.

Lattice vibrations are much like the vibrations of a harmonic oscillator, so the energy of each normal mode must be quantized according to

$$E(\beta) = \hbar\omega(\beta)\left(N_\beta + \frac{1}{2}\right). \quad (1.132)$$

The quantum of energy is viewed as a particle called a *phonon*, and the number of phonons is given by the Bose–Einstein factor as

$$N_\beta = \frac{1}{e^{\hbar\omega(\beta)/k_B T_L} - 1}. \quad (1.133)$$

For

$$\hbar\omega(\beta) \ll k_B T_L$$

eq. (1.133) reduces to

$$N_\beta \simeq \frac{k_B T_L}{\hbar\omega(\beta)}, \quad (1.134)$$

which is known as *equipartition* and is usually valid for acoustic phonons – except at very low temperatures. Equation (1.134) is easy to understand; $k_B T_L$ is the thermal energy and $\hbar\omega_\beta$ is the energy of the phonon at β, so eq. (1.134) just tells us how many phonons are needed to account for the thermal energy. In Chapter 2, we shall describe how phonons, both acoustic and optical, scatter carriers.

1.9 Summary

A simple approach for treating carrier motion within conventional devices has been outlined. This semiclassical approach treats carriers as particles whose dynamics, between collisions, are governed by eq. (1.14) and eq. (1.95), which are analogous to Newton's Laws. Carrier scattering, however, is treated by quantum mechanics using Fermi's Golden Rule. The semiclassical approach is applicable when the applied and built-in potentials vary slowly on the scale of an electron's wavelength. Room-temperature, thermal average electrons in silicon have a wavelength of about 120 Å and about 240 Å in GaAs, so the semiclassical approach may be questioned in ultra-small devices. Many devices contain quantum wells, and the carriers within such wells clearly display their wave nature. Quantum confinement alters the wavefunctions of electrons confined in potential wells, but transport within the confined region can often be described semiclassically. Our focus in this text is on the semiclassical transport of three dimensional carriers, but we shall also from time to time consider the transport of carriers confined in quantum wells and wires. An introduction to quantum transport, in which the electron's wave nature is essential, is contained in Chapter 9.

References and further reading

The quantum mechanical foundations which underlie device analysis are thoroughly discussed by Datta in a volume of the Modular Series on Solid State Devices.

1.1 Datta, S. *Quantum Phenomena*, Vol. VIII of the Modular Series on Solid State Devices. Addison-Wesley, Reading, Mass., 1989.

An introductory quantum mechanics text discusses the principles of wave mechanics; my own favorite is

1.2 Bohm, D. *Quantum Theory*. Prentice-Hall, Englewood Cliffs, NJ, 1951.

The basics of energy band theory and phonons are treated in introductory solid-state physics texts such as

1.3 Ashcroft, N. W. and Mermin, N. D. *Solid-State Physics*. Saunders College, Philadelphia, PA, 1976.

1.4 Harrison, W. *Solid State Theory*. Dover, New York, 1980.

Nag derives eq. (1.116), the matrix element for Bloch waves, and discusses the evaluation of overlap integrals.

1.5 Nag, B. *Electron Transport in Compound Semiconductors*. Springer-Verlag, New York, 1980.

For descriptions of the energy band structure of important semiconductors, and of the $\mathbf{k} \cdot \mathbf{p}$ method for approximating $E(k)$ near a band minima, consult

1.6 Singh, J. *Physics of Semiconductors and their Heterostructures*. McGraw-Hill, New York, 1993.

The complete, numerically evaluated bandstructures of several common semiconductors are presented in

1.7 Fischetti, M. V. Monte Carlo simulation of transport in technologically significant semiconductors of the diamond and zinc-blende structures – Part I: homogeneous transport. *IEEE Transactions in Electron Devices*, **38**, 634–49, 1991.

Applications of the effective mass equation to 'quantum devices' are described in

1.8 Vassell, M., Lee, J. and Lockwood, H. Multibarrier tunneling in $Ga_{1-x}Al_x$/GaAs heterostructures, *Journal of Applied Physics*, **54**, 5206–13, 1983.

Adachi describes the properties of the alloy, $Al_xGa_{1-x}As$

1.9 Adachi, S. GaAs, AlAs, and $Al_xGa_{1-x}As$: material parameters for use in research and device applications. *Journal of Applied Physics*, **58**, R1–R29, 1985.

For information on materials parameters for other semiconductors, consult

1.10 Landolt-Börnstein, *Numerical Data and Functional Relationships in Science and Technology*. New Series, Vol 17a, Physics of Group IV Elements and III-V Compounds, ed. by O. Madelung. Springer, Berlin, 1982.

Many modern devices make use of heterostructures. For an introduction to these structures and devices, good starting points are

1.11 Weisbuch, C. and Vinter, B. *Quantum Semiconductor Structures.* Academic Press, Inc., Boston, 1991.

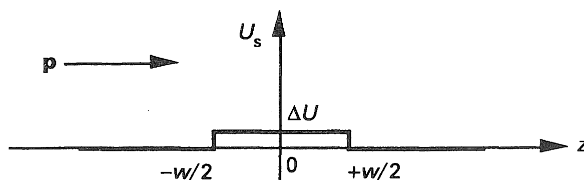
1.12 Marshak, A. H. and van Vliet, K. M. Electrical currents in solids with position-dependent band structure. *Solid-State Electronics*, **21**, 417–28, 1978.

1.13 Marshak, A. H. and van Vliet, K. M. Carrier densities and emitter efficiency in degenerate materials with position-dependent band structure. *Solid State Electronics*, **21**, 429–34, 1978.

1.14 Van Vliet, C. M., Nainaparampil, J. J. and Marshak, A. H. Ehrenfest derivation of the mean forces acting in materials with non-uniform band structure: a canonical approach. *Solid-State Electronics*, **38**, 217–23, 1995.

Problems

1.1 Assume the scattering potential shown below and assume that electrons are free to move in the z direction only.



(a) Work out an expression for the transition rate, $S(\mathbf{p}, \mathbf{p}')$ for one-dimensional electrons. Be sure to normalize the wavefunction over a length L.
(b) An incident electron with crystal momentum $\mathbf{p}$ can only make a transition to one different state, $\mathbf{p}'$. What is that state?
(c) Explain what would happen if the sign of ΔU were to change.

1.2 Consider the effect of a perturbing potential that is constant in both space and time,

$$U_S(z, t) = U_0$$

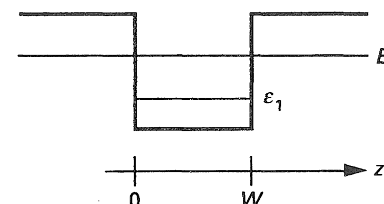
and answer the following questions.
(a) Obtain an expression for the transition rate, $S(k, k')$.
(b) Interpret your answer to part (a). What does your result imply about the motion of electrons through regions of uniform potential?

1.3 The densities of states in one, two, and three dimensions can each be expressed as the sum in **k**-space as given by eq. (1.57).
(a) Evaluate the two-dimensional density of states, and show that the result is eq. (1.63b).
(b) Evaluate the density of states for one-dimensional electrons and show that the result is eq. (1.63a).

1.4 When evaluating the density of states in energy space, we have assumed parabolic energy bands, but energy bands are typically nonparabolic.

(a) For three-dimensional carriers with parabolic energy bands, the density of states goes as $E^{1/2}$, as given by eq. (1.63c). Work out the corresponding result for 3D carriers with nonparabolic energy bands as given by eq. (1.40). Show that the density of states for nonparabolic energy bands is the parabolic band result multiplied by $\sqrt{1+\alpha E}(1+2\alpha E)$.
(b) For two-dimensional carriers with parabolic energy bands, the density of states is constant, as given by eq. (1.63b). Work out the corresponding result for 2D carriers with nonparabolic energy bands as given by eq. (1.40). Show that the density of states for nonparabolic energy bands is the parabolic band result multiplied by $(1+2\alpha E)$.

1.5 The following problem concerns electrons in a quantum well of width, W, with one subband at $E=\varepsilon_1$. Assume equilibrium conditions and that the Fermi level is located above ε_1. Answer the following questions assuming parabolic energy bands.



(a) Write an expression, involving sums over momentum space, which gives the average kinetic energy (due to its motion in the x–y plane) per electron.
(b) Convert the sum to an integral over momentum or **k**-space.
(c) Write an expression, involving an integral over energy space, which gives the average kinetic energy per electron.
(d) Assume $T = 0$K and evaluate the average kinetic energy per electron. You may work in either energy or momentum space.

1.6 Obtain an expression for the concentration per unit area of electrons in a quantum well as a function of the Fermi level position, $n_S(E_F)$. You should assume that the temperature is finite, and do not assume that the semiconductor is nondegenerate.
(a) Find $n_S(E_F)$ when one subband is occupied. For this part, you should work in energy-space using eq. (1.90).
(b) Repeat part (a) but do the work in **k**-space using eq. (1.83). For both parts (a) and (b), show that the answer is eq. (1.91).
(c) What is the result when two subbands are occupied?
(d) Explain how the results depend on the shape of the quantum well (i.e., does eq. (1.91) hold for parabolic or triangular quantum wells? What changes, if anything?).

1.7 For an infinite depth GaAs quantum well of width $W = 200$ Å at $T = 0$ K,
(a) How many subbands are occupied if $n_S = 5\times10^{11}$ cm^{-2}?
(b) How many if $n_S = 5\times10^{12}$ cm^{-2}?

1.8 For the quantum well problems, we have been asking about the total number of electrons per unit area within the well – not how the electrons are distributed within the well. The carrier density as a function of position within the well is found from

$$n(z) = \sum_{k_x,k_y,k_z} f(E)\phi^*(z)\phi(z)$$

where

$$\phi(z) = \sqrt{\frac{2}{W}} \sin\frac{n\pi z}{W}.$$

(a) Show that the carrier concentration within the well is

$$n(z) = \frac{2m^* k_B T}{\pi\hbar^2 W} \sum_{n=1}^{\infty} \sin^2\frac{n\pi z}{W} \ln(1 + e^{-\eta_n})$$

where

$$\eta_n = \frac{\left(\frac{\hbar^2 n^2 \pi^2}{2m^* W^2} - E_F\right)}{k_B T}.$$

Hint: It's easiest to perform the integral over k_x and k_y in polar coordinates.

(b) Compute and plot $n(z)$ versus z for quantum wells of width $W = 50$, 100, and 500 Å. Assume that $m^* = 0.067m_0$, $E_{C0} - E_F = 0.3$ eV, $T = 300$ K.

(c) Compare the results of part (b) with the classical result.

1.9 Consider a 100 Å wide GaAs quantum well, and assume that $\varepsilon_1 = 56$ meV and $\varepsilon_2 = 225$ meV above the bottom of the quantum well. If the Fermi level is 100 meV above the bottom of the well, then

(a) What is n_S at $T = 300$ K?

(b) If the electrons were considered to be three-dimensional, what n_S would be computed?

1.10 The quantum well that confines carriers at the AlGaAs/GaAs interface in a MODFET is approximately triangular. The first two energy levels are given by

$$E_1 = \gamma_1 n_S^{2/3}$$

and

$$E_2 = \gamma_2 n_S^{2/3}$$

where

$$\gamma_1 = 2.5 \times 10^{-12}\ \text{eV} - m^{4/3}$$

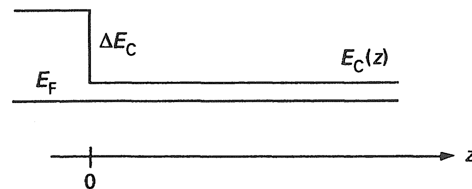
and

$$\gamma_2 = 3.2 \times 10^{-12}\ \text{eV} - m^{4/3}.$$

(See M. Shur, *GaAs Devices and Circuits*, p. 519, Plenum Press, New York, 1987, for a derivation of this result.) If we require that only one subband be occupied, what is the maximum number of electrons per square centimeter that can be accommodated in the well at $T = 0$ K?

1.11 Verify the results, eqs. (1.77)–(1.80) for electrons in a quantum wire.

1.12 Assume $E_C(z)$ is as follows:



(a) Compute $\phi^*_{P_z}(z)\phi_{P_z}(z)$ assuming $\Delta E_C = \infty$.

(b) Compute $n(z)$ for $x > 0$, approximate $f_R(k)$ using Boltzmann statistics.

(c) Sketch $n(z)$ and compare it with the classical value. Show that differences occur when z is with Λ of $z = 0$.

$$\left(\Lambda = \sqrt{\frac{2\hbar^2}{m^* k_B T}}\right)$$

1.13 Prove that the equation of motion for a semiconductor heterostructure in which m^* varies slowly with position is given by eq. (1.95b).

1.14 In the so-called tight binding method for computing bandstructures, the $E(k)$ relation for a one-dimensional lattice is given by

$$E(k) = A - B\cos(ka),$$

where A and B are constants, and a is the lattice spacing. Using this band structure for one-dimensional electrons, answer the following questions.

(a) Plot the $E(k)$ relation for $-\pi/a \le k \le \pi/a$.

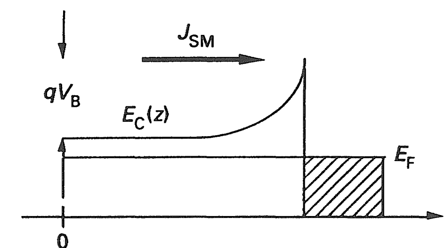
(b) Plot the velocity, $\upsilon(k)$, for $-\pi/a \le k \le \pi/a$.

(c) Assuming that the electric field is $-\mathcal{E}_0$, how long would it take an electron to travel from $k = 0$ to $k = \pi/2a$?

(d) Given the answer, T_0, from part (c), how far would the electron go in this time?

(e) Compute the density of states, $g(E)$, for this one-dimensional semiconductor.

1.15 Consider a metal-semiconductor barrier as shown below:



(a) Write an expression, involving a sum, for the electron current injected from the semiconductor to the metal, J_{SM}.

(b) Convert the sum to an integral. Be sure to show the limits of integration.

(c) Sketch the transmission coefficient versus k_z expected from (1) quantum mechanical and (2) classical considerations.

(d) Set up the problem for computing the classical (thermionic emission) current J_{SM}. Show the formula that has to be integrated, but do not integrate it.

(e) Evaluate the integral and show that the result is the expected thermionic emission relation.