Chapter 1
Principles of Semiconductor Physics

The understanding of electrochemical processes at semiconductor electrodes naturally depends on the knowledge of semiconductor physics. This chapter presents a brief introduction to this field; only those subjects relevant to semiconductor electrochemistry are included here. For detailed information, the reader is referred to the standard textbooks on semiconductor physics by Kittel [1], Smith [2], Moss [3], and Pankove [4].

1.1
Crystal Structure

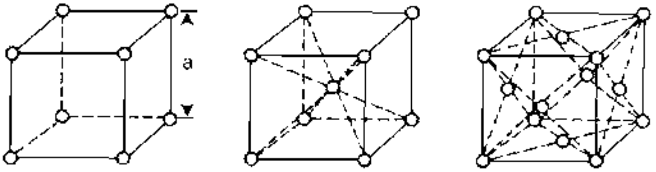
A crystalline solid can be described by three vectors $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$, so that the crystal structure remains invariant under translation through any vector that is the sum of integral multiples of these vectors. Accordingly, the direct lattice sites can be defined by the set

$$\boldsymbol{R} = m\boldsymbol{a} + n\boldsymbol{b} + p\boldsymbol{c} \tag{1.1}$$

where m, n, and p are integers [1].

Various unit cells of crystal structures are shown in Figure 1.1. Most of the important semiconductors have diamond or zincblende lattice structures which belong to the tetrahedral phases, that is, each atom is surrounded by four equidistant nearest neighbors. The diamond and zincblende lattices can be considered as two interpenetrating face-centered cubic (f.c.c.) lattices. In the case of a diamond lattice structure, such as silicon, all the atoms are silicon. In a zincblende lattice structure, such as gallium arsenide (the so-called III–V compound), one sublattice is gallium and the other is arsenic. Most other III–V compounds also crystallize in the zincblende structure [5]. Various II–VI compounds, such as CdS, crystallize in the wurtzite structure, and others in the rock salt structure (not shown). The wurtzite lattice can be considered as two interpenetrating hexagonal close-packed lattices. In the case of CdS, for example, the sublattices are composed of cadmium and sulfur. The wurtzite structure has a tetrahedral arrangement of four equidistant nearest neighbors, similar to a zincblende structure. The lattice constants and structures of the most important semiconductors are given in Appendix A.3.

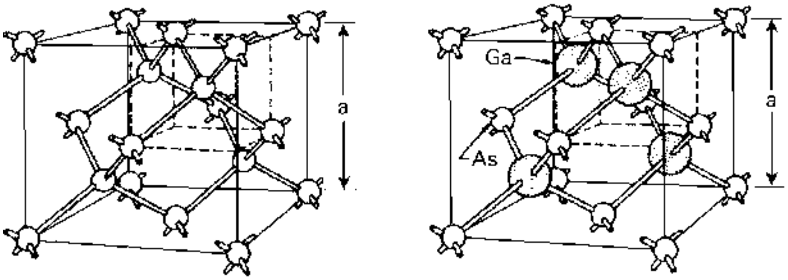
Semiconductor Electrochemistry, Zweite Auflage. Rüdiger Memming.



Simple cubic
(P, etc)

Body-centrered cubic
(Na, W, etc)

Face-centered cubic
(Al, Au, etc)



Diamond
(C, Ge, Si, etc)

Zincblende
(GaAs, GaP, etc)

Figure 1.1 Important unit cells (taken from [7]).

It is also common to define a set of reciprocal lattice vectors $\boldsymbol{a}^*$, $\boldsymbol{b}^*$, $\boldsymbol{c}^*$, such as

$$\boldsymbol{a}^* = 2\pi\frac{\boldsymbol{b}\cdot\boldsymbol{c}}{\boldsymbol{a}\cdot\boldsymbol{b}\cdot\boldsymbol{c}};\quad \boldsymbol{b}^* = 2\pi\frac{\boldsymbol{c}\cdot\boldsymbol{a}}{\boldsymbol{a}\cdot\boldsymbol{b}\cdot\boldsymbol{c}};\quad \boldsymbol{c}^* = 2\pi\frac{\boldsymbol{a}\cdot\boldsymbol{b}}{\boldsymbol{a}\cdot\boldsymbol{b}\cdot\boldsymbol{c}} \tag{1.2}$$

so that $\boldsymbol{a}\cdot\boldsymbol{a}^* = 2\pi$; $\boldsymbol{a}\cdot\boldsymbol{b}^* = 0$ and so on. The general reciprocal lattice vector is given by

$$\boldsymbol{G} = h\boldsymbol{a}^* + k\boldsymbol{b}^* + l\boldsymbol{c}^* \tag{1.3}$$

where h, k, l are integers.

According to the definitions given by Eqs. (1.1)–(1.3), the product $\boldsymbol{G}\cdot\boldsymbol{R} = 2\pi\times$ integer. Therefore, each vector of the reciprocal lattice is normal to a set of planes in the direct lattice, and the volume V_c^* of a unit cell of the reciprocal lattice is related to the volume of the direct lattice V_c by

$$V_c = \frac{(2\pi)^3}{V_c} \tag{1.4}$$

where $V_c = \boldsymbol{a}\cdot\boldsymbol{b}\cdot\boldsymbol{c}$.

It is convenient to characterize the various planes in a crystal by using the Miller indices h, k, l. They are determined by first finding the intercepts of the plane with the basis axis in terms of the lattice constants, and then taking the reciprocals of

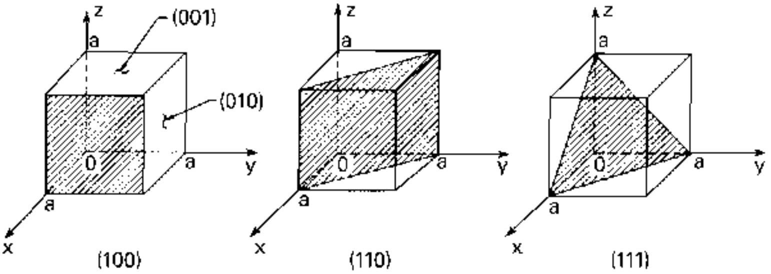


Figure 1.2 Miller indices of some important planes in a cubic crystal.

these numbers and reducing them to the smallest three integers having the same ratio. The three integers are written in parentheses (hkl) as Miller indices for a single plane or a set of parallel planes. One example is given in Figure 1.2 where the Miller indices of some planes in a cubic crystal are shown. Planes that intercepted, for example, the x-axis on the negative side would be characterized by ($\bar{h}kl$). For directions perpendicular to the corresponding planes, one uses the Miller indices in brackets, that is, [hkl].

Some physical properties of semiconductor electrodes depend on the orientation of the crystal, and surface properties vary from one crystal plane to the other. It is, therefore, very important in studies of surface and interface effects that the proper surface is selected. A semiconductor crystal can be cut by sawing or by cleavage.

Cleavage, in particular, is a common technique for preparing clean surfaces in an ultrahigh vacuum. Unfortunately, however, only a few surface planes can be exposed by cleavage. The easiest planes in silicon and germanium are (111) and their equivalents. In contrast, gallium arsenide cleaves on (110) planes. Accordingly, the most interesting planes, which consist of a Ga surface (111) or an As surface ($\bar{1}\bar{1}\bar{1}$), cannot be produced by cleavage.

1.2
Energy Levels in Solids

Before the energy bands of semiconductors can be described, the following basic quantities must be introduced.

A free electron in space can be described by classical relations as well as by quantum mechanical methods. Combining both methods, the wavelength λ of the electron wave is related to the momentum p by

$$\lambda = \frac{h}{p} = \frac{h}{mv} \tag{1.5}$$

in which h is the Planck constant, m is the electron mass, and v is the electron velocity. The electron wave can also be described by the wave vector defined by the relation

$$k = \frac{2\pi}{\lambda} \tag{1.6}$$

Combining Eqs. (1.5) and (1.6), one obtains

$$k = \frac{2\pi}{h} p \tag{1.7}$$

The kinetic energy of a free electron is then given by

$$E = \frac{1}{2} m v^2 = \frac{h^2}{8\pi^2 m} k^2 \tag{1.8}$$

The parabolic relation between the energy and the wave vector k is illustrated in Figure 1.3.

In a metal, the electrons are not completely free. A quantum mechanical treatment of the problem leads to the consequence that not all energy values are allowed. The corresponding wave vectors are now given by

$$k = \frac{\pi n}{L} \tag{1.9}$$

in which L is the length of a metal cube and n is any nonzero integer. Inserting Eq. (1.9) into Eq. (1.8), one obtains

$$E = \frac{h^2}{8mL^2} n^2 \tag{1.10}$$

The relation between the energy and the wave vector is still parabolic, but the energy of the electron can only attain certain values. Since, however, the range of the allowed k values is proportional to the reciprocal value of L, the range of the energy values is very small for a reasonable size of metal, so that the E–k dependence is still a quasi-continuum.

The band structure of crystalline solids is usually obtained by solving the Schrödinger equation of an approximate one-electron problem. In the case of

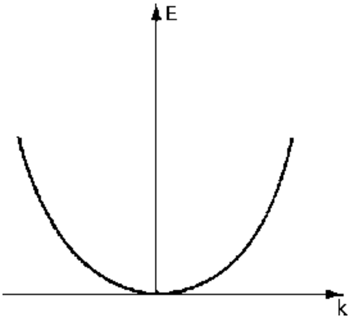


Figure 1.3 Parabolic dependence of the energy of a free electron vs wave vector.

nonmetallic materials, such as semiconductors and insulators, there are essentially no free electrons. This problem is taken care of by the Bloch theorem. This important theorem assumes a potential energy profile $V(\boldsymbol{r})$ being periodic with the periodicity of the lattice. In this case the Schrödinger equation is given by

$$\left[-\frac{h^2}{2\pi^2}\nabla^2 + V(r)\right] \Psi_k k(r) = E_k \Psi_k(r) \tag{1.11}$$

The solution to this equation is of the form

$$\psi_k(r) = \mathrm{e}^{\mathrm{j}kr} U_\mathrm{n}(\boldsymbol{kr}) \tag{1.12}$$

where $U_\mathrm{n}(\boldsymbol{k}, \boldsymbol{r})$ is periodic in $\boldsymbol{r}$ with the periodicity of the direct lattice, and n is the band index. Restricting the problem to the one-dimensional case, the lattice constant is a, b, or c (see Eq. (1.1)). If N is an integral number of unit lattice cells, then $k = \pi/a$ is the maximum value of k for $n = N$. This maximum occurs at the edge of the so-called Brillouin zone. A Brillouin zone is the volume of k space containing all the values of k up to π/a. Larger values of k lead only to a repetition of the first Brillouin zone.

Accordingly, it is only useful to determine the band structure within the first Brillouin zone. The solution of the Schrödinger equation (see Eqs. (1.11) and (1.12)) leads to two energy bands separated by an energy gap, as shown in Figure 1.4. The energy profile of the conduction band (upper curve) still appears parabolic (at least near the minimum), but it may deviate considerably from a parabolic E–k relation. In order to continue to use the relation derived for free electrons (Eq. (1.8)), the electron mass is adjusted to provide a good fit. We then have, instead of Eq. (1.8),

$$E = \frac{h^2}{8\pi^2 m^*} k^2 \tag{1.13}$$

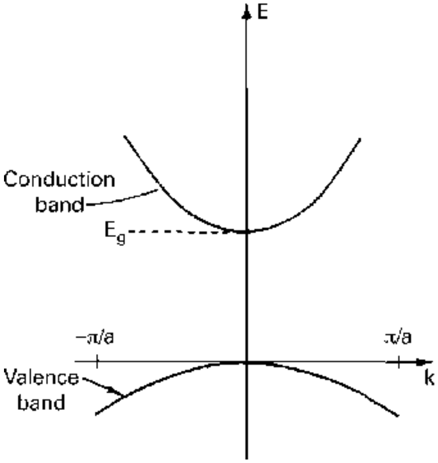


Figure 1.4 Electron energy vs wave vector in a semiconductor.

in which m^* is the effective mass. Differentiating this equation, the effective mass is given by

$$m^* = \frac{h^2}{4\pi^2} \frac{1}{\mathrm{d}^2E/\mathrm{d}k^2} \tag{1.14}$$

This means that the effective mass is determined by the second derivative of the E–k curve, that is, by its curvature. From this, it follows that the width of an energy band is larger for a small value of m^* and smaller for a large m^* value. The width can be determined by optical investigation and the effective mass by cyclotron resonance measurements.

According to Eq. (1.14), the negative curvature of the valence band would mean a negative electron mass, which is physically not acceptable. It has therefore been concluded that occupied orbitals in the valence band correspond to holes. A hole acts in an applied electric or magnetic field as though it were a particle with a positive charge. This concept has been experimentally proved by Hall measurements (see Section 1.6). However, it only makes sense if nearly all energy states are filled by electrons. It should be further mentioned that the effective mass of holes may be different from that of electrons. A selection of values is listed in Appendix A.4.

The band structure of solids has been studied theoretically by various research groups. In most cases, it is rather complex as shown for Si and GaAs in Figure 1.5. The band structure, $E(k)$, is a function of the three-dimensional wave vector within the Brillouin zone. The latter depends on the crystal structure and corresponds to the unit cell of the reciprocal lattice. One example is the Brillouin zone of a diamond type of crystal structure (C, Si, Ge), as shown in Figure 1.6. The diamond lattice can also be considered as two penetrating f.c.c. lattices. In the case of silicon, all cell atoms are Si. The main crystal directions, $\Gamma \rightarrow L([111])$, $\Gamma \rightarrow X([100])$ and $\Gamma \rightarrow K([110])$, where Γ is the center, are indicated in the Brillouin zone by the dashed lines in Figure 1.6. Crystals of the zincblende structure, such as GaAs, can be described in the same way. Here one sublattice consists of Ga atoms and the other of As atoms. The band structure, $E(k)$, is usually plotted along particular directions within the Brillouin zone, for instance from the center Γ along the [111] and [100] directions as shown in Figure 1.5.

In all semiconductors, there is a forbidden energy region or gap in which energy states cannot exist. Energy bands are only permitted above and below this energy gap. The upper bands are called the conduction bands, while the lower ones are called the valence bands. The bandgaps of a variety of semiconductors are listed in Appendix A.4.

According to Figure 1.5, the conduction as well as the valence band consists of several bands. Some valence bands are degenerated around $k = 0$ (the Γ point). Since the curvature differs from one band to another, each band is associated with a different effective mass (see also Appendix A.4). Rather flat energy profiles correspond to heavy holes (high effective mass), and steep profiles to light holes (small effective mass). In the case of GaAs, the maxima of all valence bands and the minimum of the lowest conduction band occur at $k = 0$, that is, in the center of the Brillouin zone (Γ point) (Figure 1.5). The corresponding bandgap

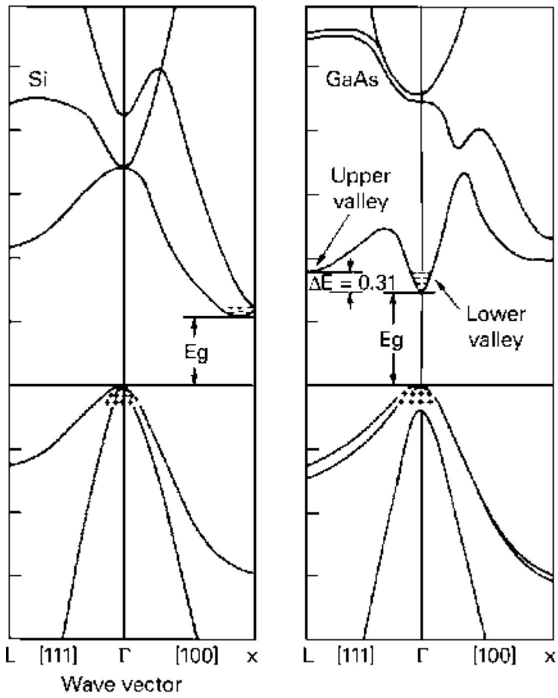


Figure 1.5 Energy band structure of Si and GaAs. Compare with Figure 1.4 (after [11]).

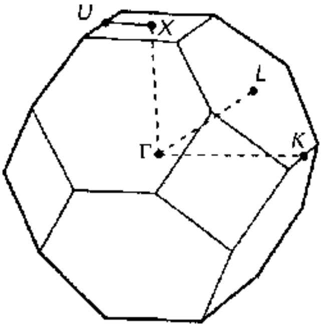


Figure 1.6 Brillouin zone for face-centered cubic lattices with high symmetry points labeled (after [6]).

(1.4 eV for GaAs) is indicated. In the band structure of many semiconductors, however, the lowest minimum of the conduction band occurs at a different wave vector ($k \neq 0$) from the maximum of the valence band ($k = 0$). For instance, in the case of silicon, the lowest minimum of the conduction band occurs at the edge of the Brillouin zone (X point) (Figure 1.5). If both the conduction band minimum and the valence band maximum occur at $k = 0$, the energy difference, E_g, is a so-called direct bandgap. If the lowest conduction band minimum is found at $k \neq 0$,

E_g is termed an indirect bandgap. The consequences of these differences in band structure will be discussed in Section 1.3. During the course of this book, only the lowest edge of the conduction band (E_c) and the upper edge of the valence band (E_v) are considered (as illustrated in Figure 1.9 in Section 1.3).

1.3
Optical Properties

The simplest method for probing the band structure of semiconductors is to measure the absorption spectrum. The absorption coefficient, α, is defined as

$$\alpha = \frac{1}{d} \ln \frac{I_0}{I} \tag{1.15}$$

in which d is the thickness of the sample, and I and I_0 are the transmitted and the incident light intensities, respectively. Since the refractive index of semiconductors is frequently quite high, accurate measurements require the determination of the transmission coefficient, T, as well as the reflection coefficient, R. For normal incidence, they are given by

$$T = \frac{(1 - R^2)\exp(-4\pi d/\lambda)}{1 - R^2\exp(-8\pi d/\lambda)} \tag{1.16}$$

$$R = \frac{(1 - n)^2 + k^2}{(1 + n)^2 + k^2} \tag{1.17}$$

where λ is the wavelength, n is the refractive index, and k is the absorption constant. The latter is related to the absorption coefficient α by

$$\alpha = \frac{4\pi k}{\lambda} \tag{1.18}$$

By analyzing the T and λ or the R and λ data at normal incidence, or by measuring R and T at different angles of incidence, both n and α can be obtained.

The fundamental absorption refers to a band-to-band excitation which can be recognized by a steep rise in absorption when the photon energy of the incident light goes through this range. Since, however, optical transitions must follow certain selection rules, the determination of the energy gap from absorption measurements is not a straightforward procedure.

Since the momentum of photons, h/λ, is small compared with the crystal momentum, h/a (a is the lattice constant), the momentum of electrons should be conserved during the absorption of photons. The absorption coefficient $\alpha(h\nu)$ for a given photon energy is proportional to the probability, P, for transition from the initial to the final state and to the density of electrons in the initial state as well as to the density of empty final states. On this basis, a relation between the absorption coefficient α and the photon energy E_{ph} can be derived [2, 4]. For a direct band–

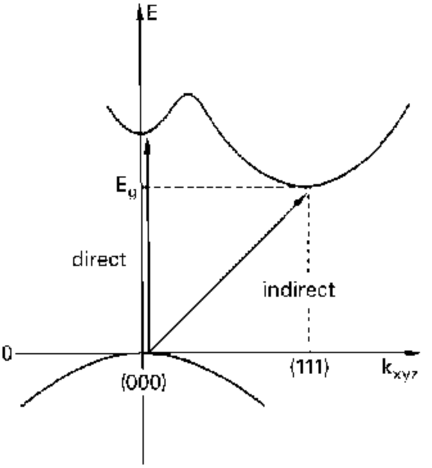


Figure 1.7 Optical transitions in semiconductors with an indirect bandgap.

band transition, for which the momentum remains constant (see Figure 1.7), it has been obtained for a parabolic energy structure (near the absorption edge):

$$\alpha \sim (E_{\mathrm{ph}} - E_{\mathrm{g}})^{1/2} \tag{1.19}$$

in which E_{g} is the bandgap. Accordingly, a plot of $(\alpha\ E_{\mathrm{ph}})^2$ vs E_{ph} should yield a straight line and E_{g} can be determined from the intercept. However, this procedure does not always yield a straight line. Therefore, some scientists define E_{g} at that photon energy where $\alpha = 10^4\ \mathrm{cm}^{-1}$. High α values of up to $10^6\ \mathrm{cm}^{-1}$ have been found for direct transitions. Electrons excited into higher energy levels of the conduction band (transition 1a in Figure 1.9) are thermalized to the lower edge of the conduction band within about 10^{-12} to 10^{-13} s.

As already mentioned in the previous section, the lowest minimum in the conduction band energy frequently occurs not at $k = 0$, but at other wave numbers as shown for silicon in Figure 1.5. The law of conservation of momentum excludes here the possibility of the absorption of a photon of energy close to the bandgap. Photon absorption becomes possible, however, if a phonon supplies the missing momentum to the electron as illustrated in Figure 1.7. Since such an indirect transition requires a “three-body” collision (photon, electron, phonon) which occurs less frequently than a “two-body” collision, the absorption coefficient will be considerably smaller for semiconductors with an indirect gap. This becomes obvious when the absorption spectra of semiconductors are measured; a selection is given in Figure 1.8. For instance, GaAs and $\mathrm{CuInSe_2}$ provide examples of a direct bandgap, that is, the absorption coefficient rises steeply near the bandgap and reaches very high values. Si and GaP provide typical examples of an indirect transition. In the case of Si, one can recognize in Figure 1.8 that α remains at a very low level for a large range of photon energies (GaP not shown). For indirect

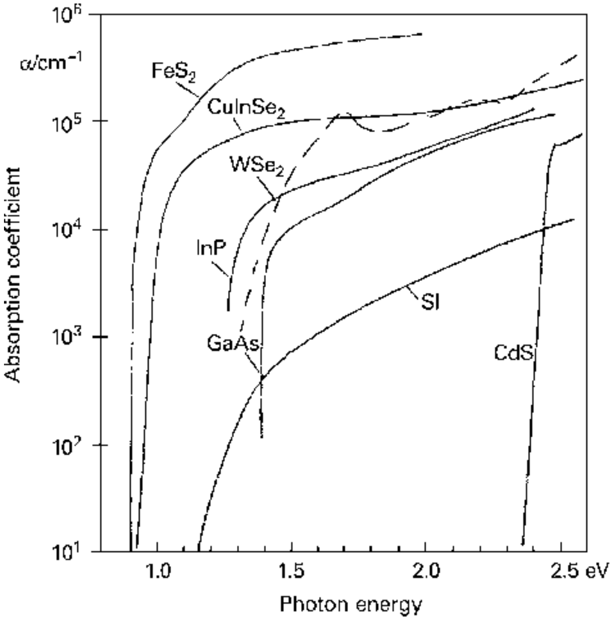


Figure 1.8 Absorption spectra of various semiconductors.

transitions, the relation between α and E_{ph} is given in [2, 7] as

$$\alpha \sim (E_{\mathrm{ph}} - E_{\mathrm{g}})^2 \tag{1.20}$$

The interpretation of the interband transition is based on a single-particle model, although in the final state two particles, an electron and a hole, exist. In some semiconductors, however, a quasi one-particle state, an exciton, is formed upon excitation [4, 8]. Such an exciton represents a bound state, formed by an electron and a hole, as a result of their Coulomb attraction, that is, it is a neutral quasi-particle, which can move through the crystal. Its energy state is close to the conduction band (transition 3 in Figure 1.9), and it can be split into an independent electron and a hole by thermal excitation. Therefore, a sharp absorption peak just below the bandgap energy can usually only be observed at low temperatures, whereas at room temperature only the typical band–band transition is visible in the absorption spectrum. The situation is different in organic crystals [9] and also for small semiconductor particles (see Chapter 9).

Various other electronic transitions are possible upon light excitation. Besides the band–band transitions, an excitation of an electron from a donor state or an impurity level into the conduction band is feasible (transition 2 in Figure 1.9). However, since the impurity concentration is very small, the absorption cross-section and therefore the corresponding absorption coefficient will be smaller by many orders of magnitude than that for a band–band transition. At lower photon energies, that is, at $E_{\mathrm{ph}} \ll E_{\mathrm{g}}$, an absorption increase with decreasing E_{ph} has frequently been observed for heavily doped semiconductors. This absorption has

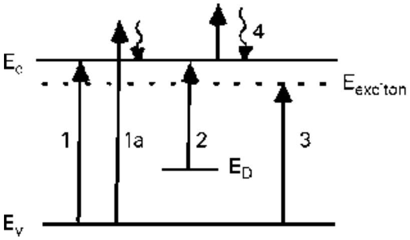
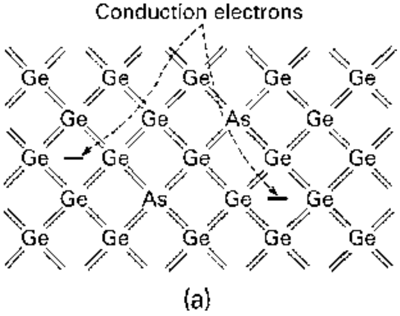


Figure 1.9 Optical transitions in a semiconductor.

been related to an intraband transition (transition 4 in Figure 1.9), and is approximately described by the Drude theory [4]. This free carrier absorption increases with the carrier density. It is negligible for carrier densities below about 10^{18} cm^{-3}.

1.4
Density of States and Carrier Concentrations

Semiconductor single crystals grown from extremely pure material exhibit a low conductivity because of low carrier density. The latter can be increased by orders of magnitude by doping the material. The principal effect of doping is illustrated



Holes

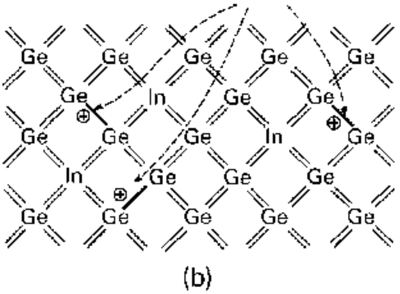


Figure 1.10 Doping of a semiconductor crystal (Ge): (a) n-type doping; (b) p-type doping.

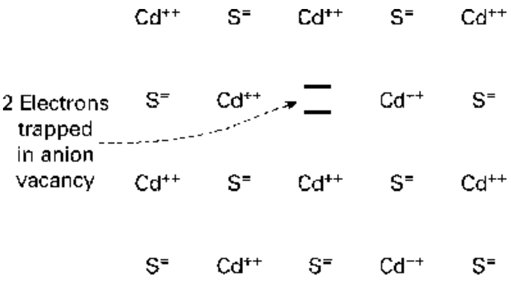


Figure 1.11 Imperfections in a compound semiconductor (CdS).

in Figure 1.10, taking germanium as an example. Figure 1.10a shows intrinsic Ge which contains a negligibly small amount of impurities. Each Ge atom shares its four valence electrons with the four neighboring atoms forming covalent bonds. By doping the material with arsenic, n-type germanium is formed (Figure 1.10a). The arsenic atom with five valence electrons has replaced a Ge atom and an electron is donated to the lattice. This additional electron occupies one level in the conduction band. Similarly, p-type germanium is made by doping with a trivalent atom such as indium. This atom with three valence electrons substitutes for a Ge atom, an additional electron is transferred to indium leaving a positive hole in the Ge lattice (Figure 1.10b). In principle, compound semiconductors are doped in the same way. In this case, however, doping can also occur by unstoichiometry, as illustrated for n-type CdS in Figure 1.11. The bonding is partly ionic and additional free electrons occur if a sulfur atom is missing; if a sulfur vacancy, V_s, is formed the material becomes n-type.

The additional electrons and holes occupy energy states in the conduction and valence bands, respectively. Before discussing the rules of occupation of energy levels, the energy distribution of the available energy states must first be derived, as follows.

In momentum space, the density of allowed points is uniform. Assuming that the surfaces of constant energy are spherical, then the volume of k space between spheres of energy E and $E + \Delta E$ is $4\pi k^2\,\mathrm{d}k$ (see [4]). Since a single level occupies a volume of $8\pi^3/V$ (V = crystal volume) in momentum space and there are two states per level, the density of states is given by

$$N(E)\,\mathrm{d}E = \frac{8\pi k^2}{8\pi^3}\,\mathrm{d}k = \frac{k^2}{\pi^2}\,\mathrm{d}k \tag{1.21}$$

It has been assumed here that the volume is unity (e.g., 1 cm^3). Inserting Eq. (1.13), one obtains

$$N(E)\,\mathrm{d}E = \frac{1}{2\pi^2 \hbar^3}(2m^*)^{3/2}E^{1/2}\,\mathrm{d}E \tag{1.22}$$

where E is measured with respect to the band edge. This equation is valid for the conduction and valence bands. The energy states can be occupied by electrons in the conduction band and by holes in the valence band. According to Eq. (1.22),

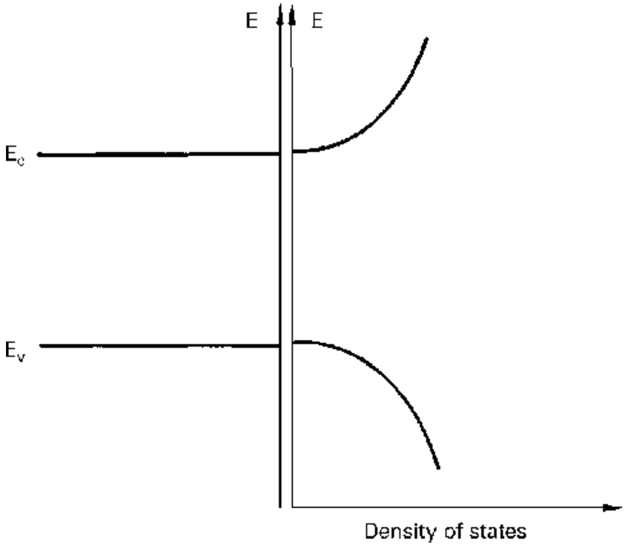


Figure 1.12 Density of energy states near the band edges of a semiconductor vs energy.

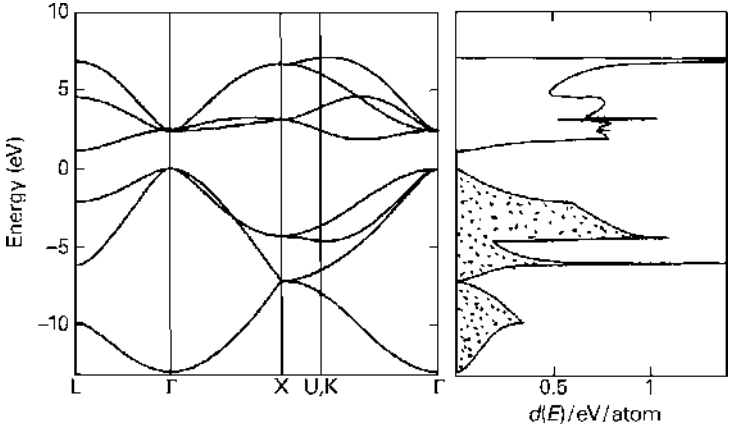


Figure 1.13 Band structure and energy distribution of the density of states of silicon, as calculated for a large energy range. Dotted distribution is occupied by electrons (from [6]).

the density of states per energy interval increases with the square root of the energy from the bottom of the corresponding band edge as illustrated in Figure 1.12 (electron and hole energies have opposite signs). Since the reduced masses may be different for electrons and holes, the slopes of the curves are also different. These curves are based on the parabolic shape of the E–k relation as assumed near the minimum. The density of states looks very different when it is measured over a much larger energy range as shown in Figure 1.13.

The total density of energy states up to a certain energy level is obtained by integration of Eq. (1.22). The result is

$$N(E) = \frac{1}{3\pi^2 h^3}(2m^*)^{3/2}E^{3/2} \tag{1.23}$$

1.4.1
Intrinsic Semiconductors

The number of electrons occupying levels in the conduction band is given by

$$n = \int_{E_c}^{\infty} N(E) f(E)\,\mathrm{d}E \tag{1.24}$$

in which $f(E)$ is the Fermi–Dirac distribution as given by

$$f(E) = \frac{1}{1 + \exp((E - E_F)/(kT))} \tag{1.25}$$

where E_F is the Fermi level. The integral in Eq. (1.24) cannot be solved analytically. Nevertheless, the integral must have a limited value because the density of states increases with increasing energy whereas $f(E)$ decreases. Equation (1.24) can only be solved by assuming that $(E-E_F)/(kT) \gg 1$. In this case, the following has been obtained [2]:

$$n = N_c \exp\left(-\frac{E_c - E_F}{kT}\right) \tag{1.26}$$

in which N_c is the density of energy states within few kT above the conduction band edge and is given by

$$N_c = \frac{2(2\pi m_e^* kT)^{3/2}}{h^3} \tag{1.27}$$

According to Eq. (1.27), one obtains $N_c \approx 5 \times 10^{19}\,\mathrm{cm}^{-3}$ for the density of states within $1\,kT$ above the lower edge of the conduction band, assuming an effective mass of $m^* = 1 \times m_0$ (m_0 = electron mass in free space). Since semiconductors with doping of less than $1 \times 10^{19}\,\mathrm{cm}^{-3}$ are used in most investigations and applications, the majority of the energy levels remain empty.

Similarly, we can obtain the hole density near the top of the valence band. We have, then

$$p = \int_{-\infty}^{E_v} N(E)(1 - f(E))\,\mathrm{d}E \tag{1.28}$$

Using the same approximations as above, we obtain

$$p = N_v \exp\left(\frac{E_v - E_F}{kT}\right) \tag{1.29}$$

where the density of states, N_v, around the top of the valence band is given by

$$N_\mathrm{v} = \frac{2(2\pi m_\mathrm{h}^* kT)^{3/2}}{h^3} \tag{1.30}$$

in which m_h^* is the effective hole mass.

In order to preserve charge neutrality in an intrinsic semiconductor, the electron and hole densities must be equal. The position of the Fermi level can then be calculated from Eqs. (1.26) and (1.29). We then have

$$\begin{aligned} E_\mathrm{F} &= \frac{E_\mathrm{c} + E_\mathrm{v}}{2} + \frac{kT}{2} \ln\left(\frac{N_\mathrm{v}}{N_\mathrm{c}}\right) \\ &= \frac{E_\mathrm{c} + E_\mathrm{v}}{2} + \frac{kT}{2} \ln\left(\frac{m_\mathrm{h}^*}{m_\mathrm{e}^*}\right)^{3/2} \end{aligned} \tag{1.31}$$

Accordingly, the Fermi level E_F is close to the middle of the energy gap, or for $m_\mathrm{e}^* = m_\mathrm{h}^*$ it is exactly at the middle of the gap. The intrinsic carrier density can be obtained by multiplying Eqs. (1.26) and (1.29), that is,

$$np = N_\mathrm{c} N_\mathrm{v} \exp\left(-\frac{E_\mathrm{g}}{kT}\right) = n_\mathrm{i}^2 \tag{1.32}$$

The product of n and p is constant and the corresponding concentration is $n = p = n_\mathrm{i}$, that is, is the intrinsic electron density. Equation (1.32) is called the "mass law" of electrons and holes, in comparison with chemical equilibria in solutions. The intrinsic concentration can be calculated from Eq. (1.32) if the densities of states are known. Assuming that $m_\mathrm{e}^*/m_0 = 1$, then $n_\mathrm{i} \approx 10^{11}\ \mathrm{cm}^{-3}$ for a bandgap of $E_\mathrm{g} = 1$ eV, that is, n_i is a very small quantity. In the case of intrinsic material, the electron hole pairs are created entirely by thermal excitation. Since this excitation becomes very small for large bandgaps, n_i decreases with increasing bandgaps as proved by Eq. (1.32). Equation (1.32) which is also valid for doped semiconductors, is of great importance because when one carrier density (e.g., n) is known then the other (here p) can be calculated. Examples are given in Appendix A.5.

1.4.2
Doped Semiconductors

When a semiconductor is doped with donor or acceptor atoms (see Figure 1.10), then corresponding energy levels are introduced within the forbidden zone, as shown on the left side of Figure 1.14. The donor level is usually close to the conduction band and the acceptor level close to the valence band. A donor level is defined as being neutral if filled by an electron, and positive if empty. An acceptor level is neutral if empty, and negative if filled by an electron. Depending on the distance of the donor and acceptor levels with respect to the corresponding bands, electrons are thermally excited into the conduction band and holes into the valence band.

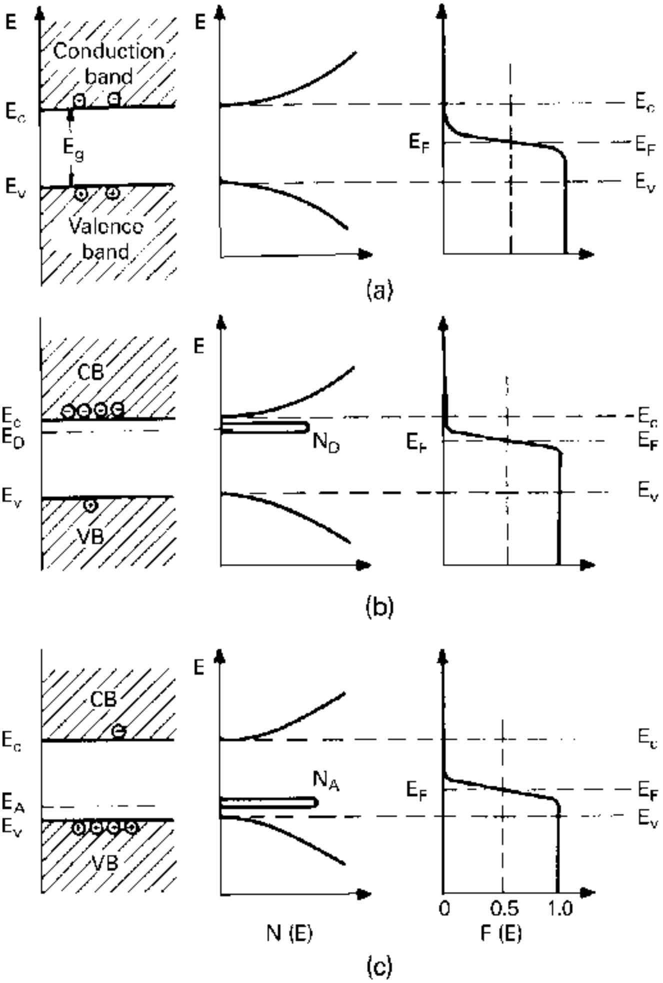


Figure 1.14 Band diagram, density of states, and Fermi distribution.

In the presence of impurities, the Fermi level must adjust itself to preserve charge neutrality. The latter is given for an n-type semiconductor by

$$n = N_D^+ + p \tag{1.33}$$

in which N_D^+ is the density of ionized donors. The latter is related to the occupied donor density N_D by the Fermi function, that is,

$$N_D^+ = (1 - f)N_D = N_D \left[1 - \frac{1}{1 + \exp((E_D - E_F)/(kT))}\right] \tag{1.34}$$

Introducing Eqs. (1.28), (1.30), and (1.34) into Eq. (1.33), the Fermi level, E_F, can be calculated. According to Eq. (1.34), it is clear that all donors are completely ionized if the Fermi level occurs below the donor level, as shown on the right side of Figure 1.14. On the other hand, if the donor concentration is increased then the electron density also rises. In this case, E_F may be located between E_c and E_D, but then not all of the more highly concentrated donors are ionized. Similar relations can be derived for acceptor states in a p-type semiconductor.

At extremely high impurity concentrations, the Fermi level may pass the band edge. In this case, the semiconductor becomes degenerated, and most of the relations derived above are no longer applicable. The semiconductor then shows a metal-like behavior.

1.5
Carrier Transport Phenomena

When an electric field of strength $\mathcal{E}$ is applied across a crystal, electrons and holes are forced to move in the material. The corresponding current density is given by

$$j = \sigma\mathcal{E} \tag{1.35}$$

in which σ is the conductivity, the reciprocal value of the resistivity ρ. For semiconductors with both electrons and holes as carriers, the conductivity is determined by

$$\sigma = e(\mu_\mathrm{n} n + \mu_\mathrm{p} p) \tag{1.36}$$

in which e is the elementary charge, and μ_n and μ_p are the mobilities of electrons and holes, respectively. For doped semiconductors, the first or second term within the brackets dominates. According to Eq. (1.36), the conductivity can be varied by many orders of magnitude by increasing the doping.

The mobility is a material constant. Values for some typical semiconductors are given in Appendix A.5. Electron and hole mobilities are typically in the range between 1 and 1000 $\mathrm{cm^2\,V^{-1}s^{-1}}$ These values are many orders of magnitude higher than the mobility of molecules and ions in solution ($\sim 10^{-4}$–$10^{-3}\,\mathrm{cm^2\,V^{-1}\,s^{-1}}$). The presence of acoustic phonons and ionized impurities leads to carrier scattering which can significantly affect the mobility. The mobility μ_i (μ_n or μ_p), determined by interaction with acoustic phonons, as shown in [2, 7], is given by

$$\mu_\mathrm{i} \sim (m^*)^{-5/2}T^{-3/2} \tag{1.37}$$

Accordingly, the mobility decreases with temperature. The mobility influenced by scattering of electrons (or holes) at ionized impurities can be described [2, 7] by

$$\mu_\mathrm{i} \sim (m^*)^{-1/2}N_\mathrm{i}^{-1}T^{3/2} \tag{1.38}$$

in which N_i is the density of impurity centers (N_D or N_A). In contrast with phonon scattering, mobility increases with temperature for impurity scattering. This makes it possible to distinguish experimentally between these two scattering

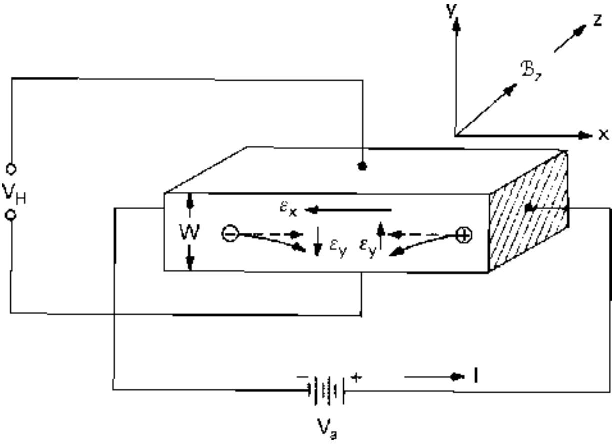


Figure 1.15 Arrangement for measuring carrier concentrations by the Hall effect (from [7]).

processes. For the complete equations and their derivation, the reader is referred to [2, 7].

The carrier diffusion coefficient, D_n for electrons and D_p for holes, is another important parameter associated with mobility. It is given by

$$D_n = \frac{kT}{e}\mu_n\,; \quad D_p = \frac{kT}{e}\mu_p \tag{1.39}$$

It should be emphasized that a carrier transport can only be described by Ohm's law (Eq. (1.35)) if sufficient empty energy levels exist in the corresponding energy band and a minimum carrier density is present in the material. On the other hand, in the case of an intrinsic high bandgap semiconductor, the carrier density may be negligible so that only those carriers carry the current which are injected into the crystal via one contact. In this case, we have a space charge limited current which is proportional to $\mathcal{E}^2$ (Child's law).

The most common method for measuring the conductivity is the four-point probe technique [10]. Here a small current I is passed through the outer two probes and the voltage V is measured between the inner two probes (s is the distance between two probes). When such a measurement is performed with a semiconductor disk of diameter $2r$ and a thickness w, the resistivity is given by

$$\rho = \frac{\pi}{\ln 2}\frac{V}{I}w \tag{1.40}$$

provided that $2r \gg s$. The advantage of this method is that the conductivity can be measured without there being ohmic contacts between the semiconductor and the outer probes.

In order to measure the carrier concentration directly, a method is applied which uses the Hall effect. The simplest setup is shown in Figure 1.15. Here a voltage is applied to a semiconducting sample in the x-direction and a magnetic field is applied along the z-direction. The resulting Lorentz effect causes a force

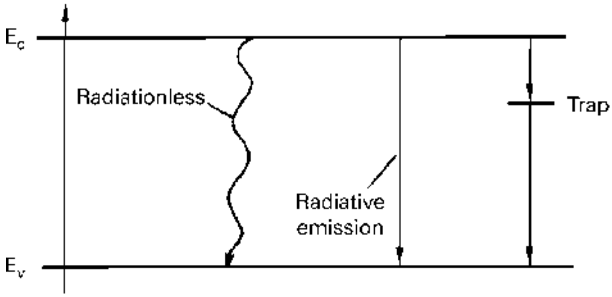


Figure 1.16 Excitation and recombination of electrons.

on the charge carriers, in the y-axis; this leads to an accumulation of electrons at the top side of an n-type sample and of holes at the bottom of a p-type sample. This effect causes a voltage V_H in the y-direction which is given by

$$V_\mathrm{H} = R_\mathrm{H} I_x \mathcal{B}_z w \tag{1.41}$$

in which $\mathcal{B}_z$ is the magnetic field, w the thickness of the sample, and R_H the Hall coefficient.

The latter is defined by

$$R_\mathrm{H} = -r\frac{1}{en}(\text{n-type})\,; \quad R_\mathrm{H} = r\frac{1}{ep}(\text{p-type}) \tag{1.42}$$

where r is a constant depending on the scattering mechanism [2]. The corresponding mobility values can be obtained by using Eq. (1.36).

1.6
Excitation and Recombination of Charge Carriers

If the equilibrium of a semiconductor is disturbed by excitation of an electron from the valence to the conduction band, the system tends to return to its equilibrium state. Various recombination processes are illustrated in Figure 1.16. For example, the electron may directly recombine with a hole. The excess energy may be transmitted by emission of a photon (radiative process) or the recombination may occur in a radiationless fashion. The energy may also be transferred to another free electron or hole (Auger process). Radiative processes associated with direct electron–hole recombination occur mainly in semiconductors with a direct bandgap, because the momentum is conserved (see also Section 1.2). In this case, the corresponding emission occurs at a high quantum yield. The recombination rate is given by

$$R_{np} = C_0 np \tag{1.43}$$

in which C_0 is a constant. During excitation the carrier density is increased by Δn and Δp, where $\Delta n = \Delta p$. Taking an n-type material as an example ($n_0 \gg p_0$), and using light intensities which are such that $\Delta n \ll n_0$ and $\Delta p \gg p_0$, then we have

$$R_{np} = C_0 n_0 \Delta p \tag{1.44}$$

The lifetime is defined as $\tau = \Delta n / R_{np}$, so that

$$\tau = \frac{1}{C_0 n_0} \tag{1.45}$$

Accordingly, the recombination rate as well as the lifetime of a band–band recombination process depends strongly on the carrier density.

In the case of semiconductors with an indirect gap, recombination occurs primarily via deep traps (Figure 1.16). Here, an electron is first captured by the trap; in a second step the trapped electron recombines with the hole. It has been found that the recombination probability is much higher for a two-step process than for a single recombination process. This two-step process can be analyzed as follows.

The trapping rate for electrons from the conduction band into traps is proportional to the electron density in the conduction band and to the number of empty traps. We have then

$$R_\mathrm{C} = C_\mathrm{n}(1 - f_\mathrm{t}) N_\mathrm{t} n \tag{1.46}$$

in which N_t is the trap density, f_t denotes the fraction of traps occupied by electrons and C_n is given by

$$C_\mathrm{n} = \gamma_\mathrm{n} v_\mathrm{th} \tag{1.47}$$

where γ_n is the electron capture cross-section and v_th the carrier thermal velocity equal to $(3kT/m^*)^{1/2}$. The rate of excitation of electrons from the trap into the conduction band is given by

$$R_\mathrm{e} = C'_\mathrm{n} f_\mathrm{t} N_\mathrm{t} \tag{1.48}$$

C'_n can be related to C_n by analyzing the equilibrium state which is determined by $R_\mathrm{c} = R_\mathrm{e}$. Applying this condition, we have

$$C'_\mathrm{n} = C_\mathrm{n} \frac{n_0 \left(1 - f_\mathrm{t}^0\right)}{f_\mathrm{t}^0} = C_\mathrm{n} n_1 \tag{1.49}$$

in which n_0 and f_t^0 are the electron density and the fraction of occupied traps at equilibrium, respectively, with f_t^0 being given by

$$f_\mathrm{t}^0 = \frac{1}{1 + \exp((E_\mathrm{t} - E_\mathrm{F})/(kT))} \tag{1.50}$$

The carrier density for a Fermi level located just at the trap level ($E_\mathrm{F} = E_t$) is given by

$$n_1 = N_\mathrm{c} \exp\left(\frac{E_\mathrm{c} - E_\mathrm{t}}{kT}\right) \tag{1.51}$$

Combining Eqs. (1.28) and (1.51), one can show that n_1 is equal to the second term in Eq. (1.49), that is,

$$n_1 = n_0 \left(1 - f_{\mathrm{t}}^0\right) \left(f_{\mathrm{t}}^0\right)^{-1} \tag{1.52}$$

Using Eqs. (1.46)–(1.49), we can derive the overall flow of electrons into the traps as given by

$$R_{\mathrm{n}} = \gamma_{\mathrm{n}} v_{\mathrm{th}} N_{\mathrm{t}} [(1 - f_{\mathrm{t}}) n - f_{\mathrm{t}} n_1] \tag{1.53}$$

By analogy, a similar expression can be derived for the net capture rate of holes, R_{p}. We then have

$$R_{\mathrm{p}} = \gamma_{\mathrm{p}} v_{\mathrm{th}} N_{\mathrm{t}} [f_{\mathrm{t}} p - (1 - f_{\mathrm{t}}) p_1] \tag{1.54}$$

In the case of stationary illumination, the electron and hole flow must be equal ($R = R_{\mathrm{n}} = R_{\mathrm{p}}$). Applying this condition to Eqs. (1.53) and (1.54), f_{t} can be determined. Inserting the resulting equation into Eq. (1.50), one obtains

$$R = \frac{\gamma_{\mathrm{n}} \gamma_{\mathrm{p}} v_{\mathrm{th}} (np - n_{\mathrm{i}}^2) N_{\mathrm{t}}}{\gamma_{\mathrm{n}} (n + n_1) + \gamma_{\mathrm{p}} (p + p_1)} \tag{1.55}$$

This is the so-called Shockley–Read equation describing recombination via traps. It also plays an important role in the description of recombination processes via surface states, as discussed in Chapter 2. In the above equation one may also replace n_1 and p_1 by the relations

$$n_1 = n_{\mathrm{i}} \exp\left(\frac{E_{\mathrm{t}} - E_{\mathrm{i}}}{kT}\right) \tag{1.56a}$$

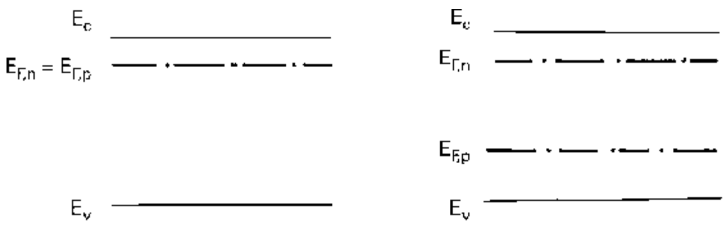
$$p_1 = n_{\mathrm{i}} \exp\left(-\frac{E_{\mathrm{t}} - E_{\mathrm{i}}}{kT}\right) \tag{1.56b}$$

which can be derived using Eqs. (1.28) and (1.30).

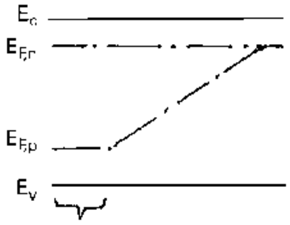
There are various techniques for measuring the lifetime of excited carriers, which cannot be described here. Details are given by Sze [7].

1.7
Fermi Levels under Nonequilibrium Conditions

At equilibrium, the Fermi level, that is, the electrochemical potential is constant throughout the semiconductor sample (Figure 1.17a). In addition, the density of electrons and holes can be calculated simultaneously from Eqs. (1.28) and (1.30) if the position of the Fermi level within the bandgap is known. If the thermal equilibrium is disturbed, for instance by light excitation, then the electron and hole densities are increased to above their equilibrium value and we have $np > n_{\mathrm{i}}^2$.



(a) At equilibrium (b) Under illumination



(c) Local excitation

Figure 1.17 Quasi–Fermi levels of electrons and holes (a), at equilibrium (b), and (c) under illumination.

Accordingly, the electron and hole density are not determined by the same Fermi level. It is useful to define quasi-Fermi levels, $E_{\mathrm{F,n}}$ and $E_{\mathrm{F,p}}$, one for electrons and another for holes, as given by

$$E_{\mathrm{F,n}} = E_{\mathrm{c}} - \ln\left(\frac{N_{\mathrm{c}}}{n}\right) \tag{1.57a}$$

$$E_{\mathrm{F,p}} = E_{\mathrm{v}} + \ln\left(\frac{N_{\mathrm{v}}}{p}\right) \tag{1.57b}$$

so that formally the original relations between carrier densities and Fermi level remain the same.

Let us consider a light excitation of electrons and holes ($\Delta n = \Delta p$) within a doped n-type semiconductor so that $\Delta n \ll n_0$ and $\Delta p \gg p_0$. Then the Fermi level of electrons, $E_{\mathrm{F,n}}$, remains unchanged with respect to the equilibrium case, whereas that of holes, $E_{\mathrm{F,p}}$, is shifted considerably downwards, as illustrated in Figure 1.17 b. In many cases, however, the excitation of electron–hole pairs occurs locally near the sample surface because the penetration of light is small. Then the splitting of the quasi-Fermi levels is large near the surface. Since the carriers diffuse out of the excitation range and recombine, the quasi-Fermi level of holes varies with distance from the excitation area (Fig. 1.17 c).

The quasi-Fermi levels play an important role in processes at the semiconductor–liquid interface, because the relative position of the quasi-Fermi level with respect to that in solution yields the thermodynamic force which drives an electrochemical reaction (see Section 7.3.5(b)).