CHAPTER 5

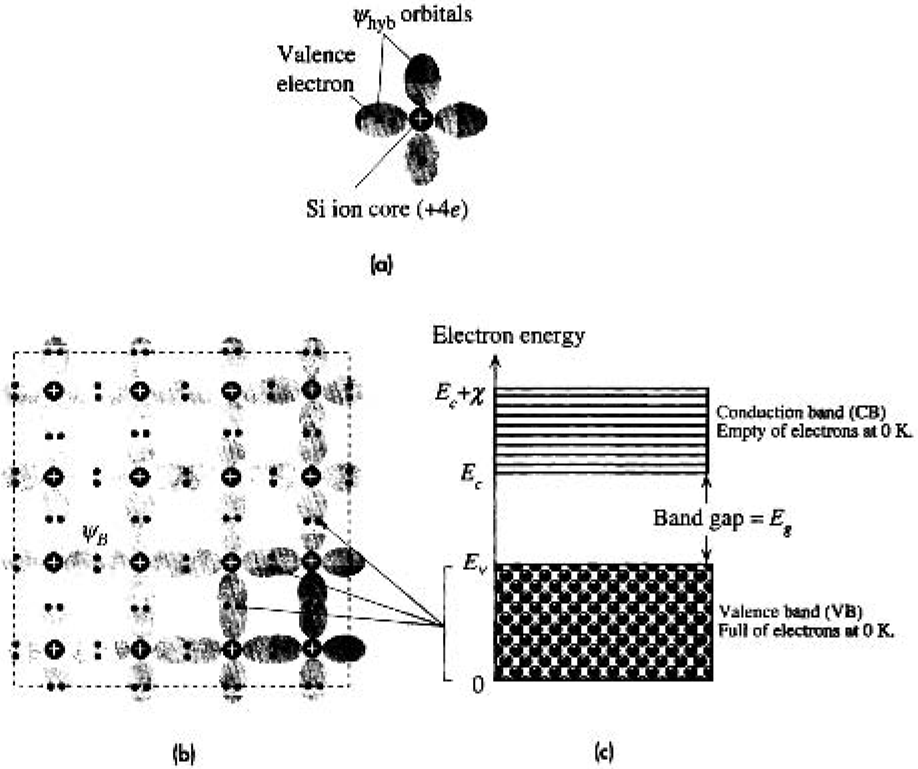
Semiconductors

In this chapter we develop a basic understanding of the properties of intrinsic and extrinsic semiconductors. Although most of our discussions and examples will be based on Si, the ideas are applicable to Ge and to the compound semiconductors such as GaAs, InP, and others. By intrinsic Si we mean an ideal perfect crystal of Si that has no impurities or crystal defects such as dislocations and grain boundaries. The crystal thus consists of Si atoms perfectly bonded to each other in the diamond structure. At temperatures above absolute zero, we know that the Si atoms in the crystal lattice will be vibrating with a distribution of energies. Even though the average energy of the vibrations is at most $3kT$ and incapable of breaking the Si–Si bond, a few of the lattice vibrations in certain crystal regions may nonetheless be sufficiently energetic to "rupture" a Si–Si bond. When a Si–Si bond is broken, a "free" electron is created that can wander around the crystal and also contribute to electrical conduction in the presence of an applied field. The broken bond has a missing electron that causes this region to be positively charged. The vacancy left behind by the missing electron in the bonding orbital is called a **hole.** An electron in a neighboring bond can readily tunnel into this broken bond and fill it, thereby effectively causing the hole to be displaced to the original position of the tunneling electron. By electron tunneling from a neighboring bond, holes are therefore also free to wander around the crystal and also contribute to electrical conduction in the presence of an applied field. In an intrinsic semiconductor, the number of thermally generated electrons is equal to the number of holes (broken bonds). In an extrinsic semiconductor, impurities are added to the semiconductor that can contribute either excess electrons or excess holes. For example, when an impurity such as arsenic is added to Si, each As atom acts as a donor and contributes a free electron to the crystal. Since these electrons do not come from broken bonds, the numbers of electrons and holes are not equal in an extrinsic semiconductor, and the As-doped Si in this example will have excess electrons. It will be an n-type Si since electrical conduction will be mainly due to the motion of electrons. It is also possible to obtain a p-type Si crystal in which hole concentration is in excess of the electron concentration due to, for example, boron doping.

5.1 INTRINSIC SEMICONDUCTORS

5.1.1 SILICON CRYSTAL AND ENERGY BAND DIAGRAM

The electronic configuration of an isolated Si atom is [Ne]$3s^2p^2$. However, in the vicinity of other atoms, the $3s$ and $3p$ energy levels are so close that the interactions result in the *four* orbitals $\psi(3s)$, $\psi(3p_x)$, $\psi(3p_y)$, and $\psi(3p_z)$ mixing together to form *four* new hybrid orbitals (called ψ_{hyb}) that are symmetrically directed as far away from each other as possible (toward the corners of a tetrahedron). In two dimensions, we can simply view the orbitals pictorially as in Figure 5.1a. The four hybrid orbitals, ψ_{hyb}, each have one electron so that they are half-occupied. Therefore, a ψ_{hyb} orbital of one Si atom can overlap a ψ_{hyb} orbital of a neighboring Si atom to form a covalent bond with two spin-paired electrons. In this manner one Si atom bonds with four other Si atoms by overlapping the half-occupied ψ_{hyb} orbitals, as illustrated in Figure 5.1b.



(a)

(b)

(c)

Figure 5.1

(a) A simplified two-dimensional illustration of a Si atom with four hybrid orbitals ψ_{hyb}. Each orbital has one electron.

(b) A simplified two-dimensional view of a region of the Si crystal showing covalent bonds.

(c) The energy band diagram at absolute zero of temperature.

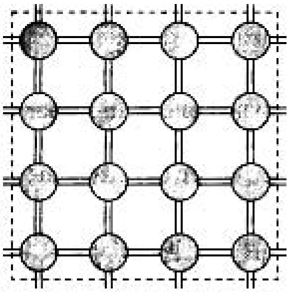


Figure 5.2 A two-dimensional pictorial view of the Si crystal showing covalent bonds as two lines where each line is a valence electron.

Each Si–Si bond corresponds to a bonding orbital, ψ_B, obtained by overlapping two neighboring ψ_{hyb} orbitals. Each bonding orbital (ψ_B) has two spin-paired electrons and is therefore *full*. Neighboring Si atoms can also form covalent bonds with other Si atoms, thus forming a three-dimensional network of Si atoms. The resulting structure is the Si crystal in which each Si atom bonds with four Si atoms in a tetrahedral arrangement. The crystal structure is that of a *diamond*, which was described in Chapter 1. We can imagine the Si crystal in two dimensions as depicted in Figure 5.1b. The electrons in the covalent bonds are the valence electrons.

The energy band diagram of the silicon crystal is shown in Figure 5.1c.[1] The vertical axis is the electron energy in the crystal. The valence band (VB) contains those electronic states that correspond to the overlap of bonding orbitals (ψ_B). Since all the bonding orbitals (ψ_B) are full with valence electrons in the crystal, the VB is also full with these valence electrons at a temperature of absolute zero. The conduction band (CB) contains electronic states that are at higher energies, those corresponding to the overlap of antibonding orbitals. The CB is separated from the VB by an energy gap E_g, called the **bandgap.** The energy level E_v marks the top of the VB and E_c marks the bottom of the CB. The energy distance from E_c to the vacuum level, the width of the CB, is called the **electron affinity** χ. The general energy band diagram in Figure 5.1c applies to all crystalline semiconductors with appropriate changes in the energies.

The electrons shown in the VB in Figure 5.1c are those in the covalent bonds between the Si atoms in Figure 5.1b. An electron in the VB, however, is not localized to an atomic site but extends throughout the whole solid. Although the electrons appear localized in Figure 5.1b, at the bonding orbitals between the Si atoms this is not, in fact, true. In the crystal, the electrons can tunnel from one bond to another and exchange places. If we were to work out the wavefunction of a valence electron in the Si crystal, we would find that it extends throughout the whole solid. This means that the electrons in the covalent bonds are indistinguishable. We cannot label an electron from the start and say that the electron is in the covalent bond between these two atoms.

We can crudely represent the silicon crystal in two dimensions as shown in Figure 5.2. Each covalent bond between Si atoms is represented by two lines corresponding to two spin-paired electrons. Each line represents a valence electron.

---

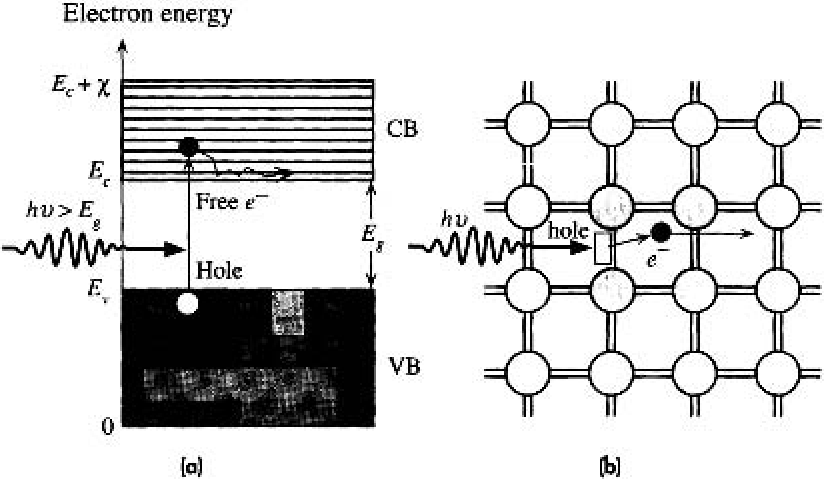
[1] The formation of energy bands in the silicon crystal was described in detail in Chapter 4.

5.1.2 ELECTRONS AND HOLES

The only empty electronic states in the silicon crystal are in the CB (Figure 5.1c). An electron placed in the CB is free to move around the crystal and also respond to an applied electric field because there are plenty of neighboring empty energy levels. An electron in the CB can easily gain energy from the field and move to higher energy levels because these states are empty. Generally we can treat an electron in the CB as if it were free within the crystal with certain modifications to its mass, as explained later in Section 5.1.3.

Since the only empty states are in the CB, the excitation of an electron from the VB requires a minimum energy of E_g. Figure 5.3a shows what happens when a photon of energy $h\nu > E_g$ is incident on an electron in the VB. This electron absorbs the incident photon and gains sufficient energy to surmount the energy gap E_g and reach the CB. Consequently, a free electron and a "hole," corresponding to a missing electron in the VB, are created. In some semiconductors such as Si and Ge, the photon absorption process also involves lattice vibrations (vibrations of the Si atoms), which we have not shown in Figure 5.3b.

Although in this specific example a photon of energy $h\nu > E_g$ creates an electron–hole pair, this is not necessary. In fact, in the absence of radiation, there is an electron–hole generation process going on in the sample as a result of **thermal generation.** Due to thermal energy, the atoms in the crystal are constantly vibrating, which corresponds to the bonds between the Si atoms being periodically deformed. In a certain region, the atoms, at some instant, may be moving in such a way that a bond becomes overstretched, as pictorially depicted in Figure 5.4. This will result in the overstretched bond rupturing and hence releasing an electron into the CB (the electron effectively



(a)

(b)

Figure 5.3

(a) A photon with an energy greater than E_g can excite an electron from the VB to the CB.

(b) When a photon breaks a Si–Si bond, a free electron and a hole in the Si–Si bond are created.

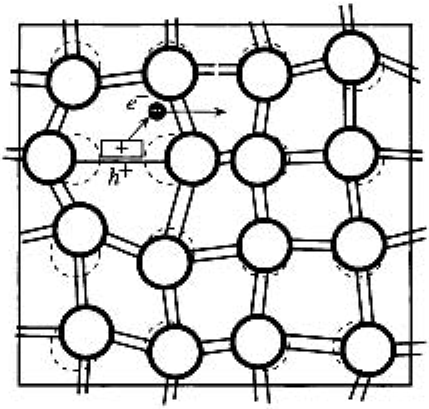


Figure 5.4 Thermal vibrations of atoms can break bonds and thereby create electron–hole pairs.

becomes "free"). The empty electronic state of the missing electron in the bond is what we call a **hole** in the valence band. The free electron, which is in the CB, can wander around the crystal and contribute to the electrical conduction when an electric field is applied. The region remaining around the hole in the VB is positively charged because a charge of $-e$ has been removed from an otherwise neutral region of the crystal. This hole, denoted as h^+, can also wander around the crystal as if it were free. This is because an electron in a neighboring bond can "jump," that is, tunnel, into the hole to fill the vacant electronic state at this site and thereby create a hole at its original position. This is effectively equivalent to the hole being displaced in the opposite direction, as illustrated in Figure 5.5a. This single step can reoccur, causing the hole to be further displaced. As a result, the hole moves around the crystal as if it were a free positively charged entity, as pictured in Figure 5.5a to d. Its motion is quite independent from that of the original electron. When an electric field is applied, the hole will drift in the direction of the field and hence contribute to electrical conduction. It is now apparent that there are essentially two types of charge carriers in semiconductors: *electrons* and *holes.* A hole is effectively an empty electronic state in the VB that behaves as if it were a positively charged "particle" free to respond to an applied electric field.

When a wandering electron in the CB meets a hole in the VB, the electron has found an empty state of lower energy and therefore occupies the hole. The electron falls from the CB to the VB to fill the hole, as depicted in Figure 5.5e and f. This is called **recombination** and results in the annihilation of an electron in the CB and a hole in the VB. The excess energy of the electron falling from CB to VB in certain semiconductors such as GaAs and InP is emitted as a photon. In Si and Ge the excess energy is lost as lattice vibrations (heat).

It must be emphasized that the illustrations in Figure 5.5 are pedagogical pictorial visualizations of hole motion based on classical notions and cannot be taken too seriously, as discussed in more advanced texts (see also Section 5.11). We should remember that the electron has a wavefunction in the crystal that is extended and not localized, as the pictures in Figure 5.5 imply. Further, the hole is a concept that corresponds to an empty valence band wavefunction that normally has an electron. Again, we cannot localize the hole to a particular site, as the pictures in Figure 5.5 imply.

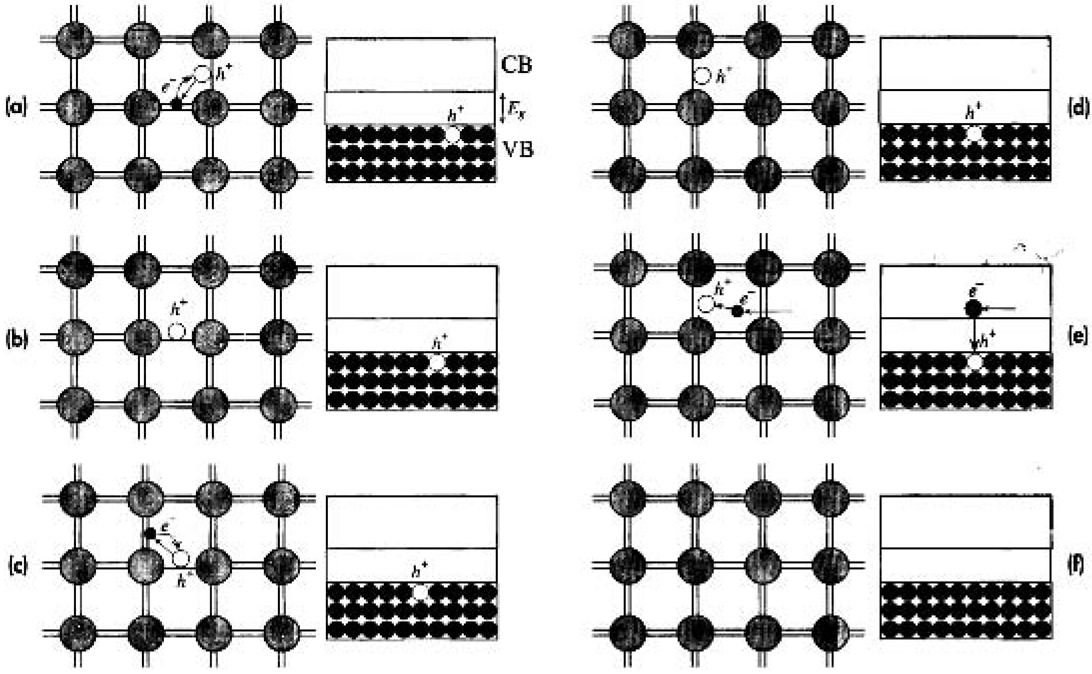


Figure 5.5 A pictorial illustration of a hole in the valence band wandering around the crystal due to the tunneling of electrons from neighboring bonds.

5.1.3 CONDUCTION IN SEMICONDUCTORS

When an electric field is applied across a semiconductor as shown in Figure 5.6, the energy bands bend. The total electron energy E is $KE + PE$, but now there is an additional electrostatic PE contribution that is not constant in an applied electric field. A uniform electric field $\mathcal{E}_x$ implies a linearly decreasing potential $V(x)$, by virtue of $(dV/dx) = -\mathcal{E}_x$, that is, $V = -Ax + B$. This means that the PE, $-eV(x)$, of the electron is now $eAx - eB$, which increases linearly across the sample. All the energy levels and hence the energy bands must therefore tilt up in the x direction, as shown in Figure 5.6, in the presence of an applied field.

Under the action of $\mathcal{E}_x$, the electron in the CB moves to the left and immediately starts gaining energy from the field. When the electron collides with a thermal vibration of a Si atom, it loses some of this energy and thus "falls" down in energy in the CB. After the collision, the electron starts to accelerate again, until the next collision, and so on. We recognize this process as the drift of the electron in an applied field, as illustrated in Figure 5.6. The drift velocity v_{de} of the electron is $\mu_e \mathcal{E}_x$ where μ_e is the drift mobility of the electron. In a similar fashion, the holes in the VB also drift in an applied field, but here the drift is along the field. Notice that when a hole gains energy, it moves "down" in the VB because the potential energy of the hole is of opposite sign to that of the electron.

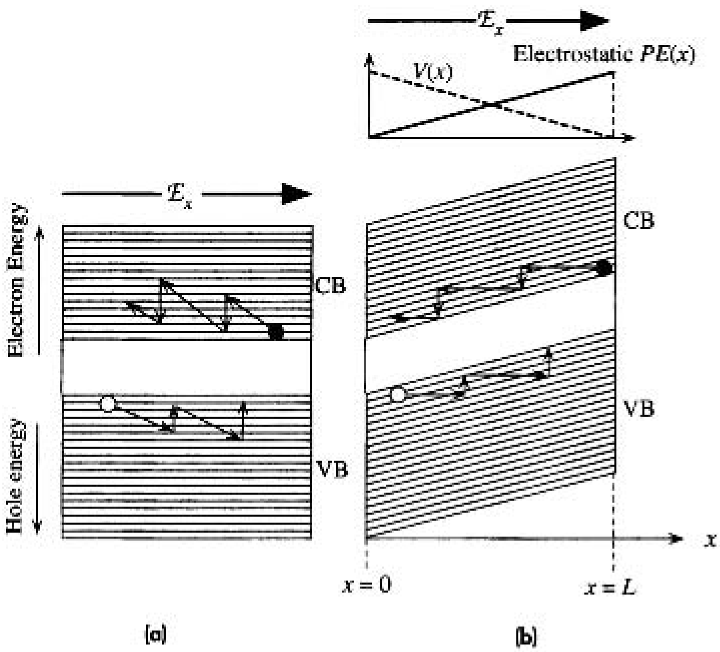


Figure 5.6 When an electric field is applied, electrons in the CB and holes in the VB can drift and contribute to the conductivity.

(a) A simplified illustration of drift in $\mathcal{E}_x$.

(b) Applied field bends the energy bands since the electrostatic PE of the electron is $-eV(x)$ and $V(x)$ decreases in the direction of $\mathcal{E}_x$, whereas PE increases.

Since both electrons and holes contribute to electrical conduction, we may write the current density J, from its definition, as

$$J = env_{de} + epv_{dh} \qquad [5.1]$$

where n is the electron concentration in the CB, p is the hole concentration in the VB, and v_{de} and v_{dh} are the drift velocities of electrons and holes in response to an applied electric field $\mathcal{E}_x$. Thus,

Electron and hole drift velocities

$$v_{de} = \mu_e \mathcal{E}_x \qquad \text{and} \qquad v_{dh} = \mu_h \mathcal{E}_x \qquad [5.2]$$

where μ_e and μ_h are the electron and hole drift mobilities. In Chapter 2 we derived the drift mobility μ_e of the electrons in a conductor as

$$\mu_e = \frac{e\tau_e}{m_e} \qquad [5.3]$$

where τ_e is the mean free time between scattering events and m_e is the electronic mass. The ideas on electron motion in metals can also be applied to the electron motion in the CB of a semiconductor to rederive Equation 5.3. We must, however, use an effective mass m_e^* for the electron in the crystal rather than the mass m_e in free space. A "free" electron in a crystal is not entirely free because as it moves it interacts with the potential energy (PE) of the ions in the solid and therefore experiences various internal forces. The effective mass m_e^* accounts for these internal forces in such a way that we can relate the acceleration a of the electron in the CB to an external force F_{ext} (e.g., $-e\mathcal{E}_x$) by $F_{\text{ext}} = m_e^* a$ just as we do for the electron in vacuum by $F_{\text{ext}} = m_e a$. In applying the

$F_{\text{ext}} = m_e^* a$ type of description to the motion of the electron, we are assuming, of course, that the effective mass of the electron can be calculated or measured experimentally. It is important to remark that the true behavior is governed by the solution of the Schrödinger equation in a periodic lattice (crystal) from which it can be shown that we can indeed describe the inertial resistance of the electron to acceleration in terms of an effective mass m_e^*. The effective mass depends on the interaction of the electron with its environment within the crystal.

We can now speculate on whether the hole can also have a mass. As long as we view mass as resistance to acceleration, that is, inertia, there is no reason why the hole should not have a mass. Accelerating the hole means accelerating electrons tunneling from bond to bond in the opposite direction. Therefore it is apparent that the hole will have a nonzero finite inertial mass because otherwise the smallest external force will impart an infinite acceleration to it. If we represent the effective mass of the hole in the VB by m_h^*, then the hole drift mobility will be

$$\mu_h = \frac{e\tau_h}{m_h^*} \tag{5.4}$$

where τ_h is the mean free time between scattering events for holes.

Taking Equation 5.1 for the current density further, we can write the **conductivity of a semiconductor** as

Conductivity of a semiconductor

$$\sigma = en\mu_e + ep\mu_h \tag{5.5}$$

where n and p are the electron and hole concentrations in the CB and VB, respectively. This is a general equation valid for all semiconductors.

5.1.4 ELECTRON AND HOLE CONCENTRATIONS

The general equation for the conductivity of a semiconductor, Equation 5.5, depends on n, the electron concentration, and p, the hole concentration. How do we determine these quantities? We follow the procedure schematically shown in Figure 5.7a to d in which the density of states is multiplied by the probability of a state being occupied and integrated over the entire CB for n and over the entire VB for p.

We define $g_{\text{cb}}(E)$ as the **density of states** in the CB, that is, the number of states per unit energy per unit volume. The probability of finding an electron in a state with energy E is given by the Fermi–Dirac function $f(E)$, which is discussed in Chapter 4. Then $g_{\text{cb}}(E)f(E)$ is the actual number of electrons per unit energy per unit volume $n_E(E)$ in the CB. Thus,

$$n_E\, dE = g_{\text{cb}}(E)f(E)\, dE$$

is the number of electrons in the energy range E to $E + dE$. Integrating this from the bottom (E_c) to the top ($E_c + \chi$) of the CB gives the electron concentration n, number of electrons per unit volume, in the CB. In other words,

$$n = \int_{E_c}^{E_c+\chi} n_E(E)\, dE = \int_{E_c}^{E_c+\chi} g_{\text{cb}}(E)f(E)\, dE$$

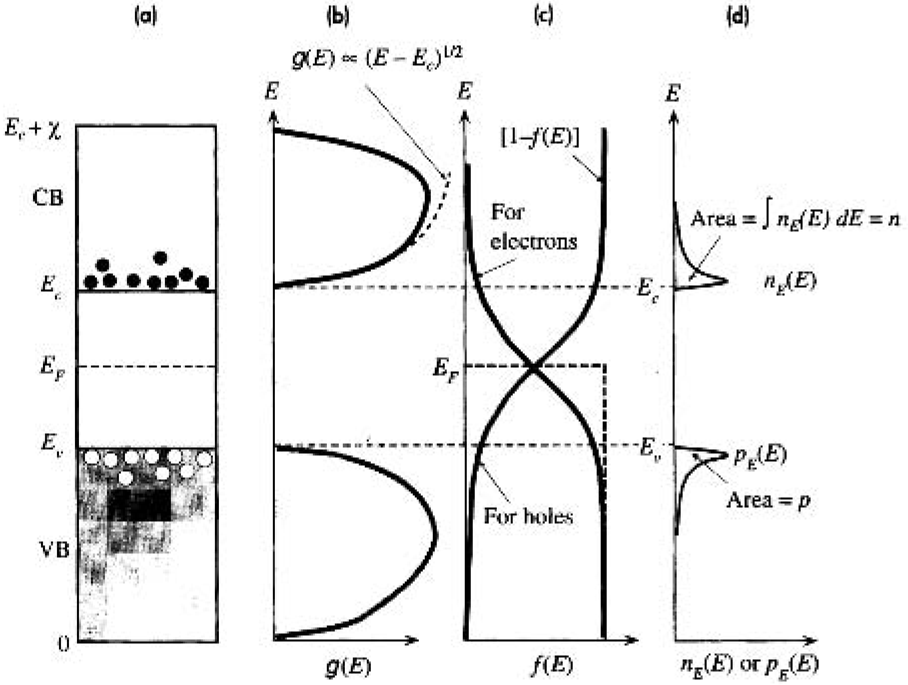


Figure 5.7

(a) Energy band diagram.

(b) Density of states (number of states per unit energy per unit volume).

(c) Fermi–Dirac probability function (probability of occupancy of a state).

(d) The product of $g(E)$ and $f(E)$ is the energy density of electrons in the CB (number of electrons per unit energy per unit volume). The area under $n_E(E)$ versus E is the electron concentration.

We will assume that $(E_c - E_F) \gg kT$ (*i.e.*, E_F is at least a few kT below E_c) so that

$$f(E) \approx \exp[-(E - E_F)/kT]$$

We are thus replacing Fermi–Dirac statistics by Boltzmann statistics and thereby inherently assuming that the number of electrons in the CB is far less than the number of states in this band.

Further, we will take the upper limit to be $E = \infty$ rather than $E_c + \chi$ since $f(E)$ decays rapidly with energy so that $g_{cb}(E)f(E) \to 0$ near the top of the band. Furthermore, since $g_{cb}(E)f(E)$ is significant only close to E_c, we can use

Density of states in conduction band

$$g_{cb}(E) = \frac{(\pi 8\sqrt{2})m_e^{*3/2}}{h^3}(E - E_c)^{1/2}$$

for an electron in a three-dimensional PE well without having to consider the exact form of $g_{cb}(E)$ across the whole band. Thus

$$n \approx \frac{(\pi 8\sqrt{2})m_e^{*3/2}}{h^3}\int_{E_c}^{\infty}(E - E_c)^{1/2}\exp\left[-\frac{(E - E_F)}{kT}\right]dE$$

which leads to

Electron concentration in CB

$$n = N_c \exp\left[-\frac{(E_c - E_F)}{kT}\right] \tag{5.6}$$

where

Effective density of states at CB edge

$$N_c = 2\left(\frac{2\pi m_e^* kT}{h^2}\right)^{3/2} \tag{5.7}$$

The result of the integration in Equation 5.6 seems to be simple, but it is an approximation as it assumes that $(E_c - E_F) \gg kT$. N_c is a temperature-dependent constant, called the **effective density of states at the CB edge.** Equation 5.6 can be interpreted as follows. If we take all the states in the conduction band and replace them with an effective concentration N_c (number of states per unit volume) at E_c and then multiply this simply by the Boltzmann probability function, $f(E_c) = \exp[-(E_c - E_F)/kT]$, we obtain the concentration of electrons at E_c, that is, in the conduction band. N_c is thus an effective density of states at the CB band edge.

We can carry out a similar analysis for the concentration of holes in the VB. Multiplying the density of states $g_{vb}(E)$ in the VB with the probability of occupancy by a hole $[1 - f(E)]$, that is, the probability that an electron is absent, gives p_E, the hole concentration per unit energy. Integrating this over the VB gives the hole concentration

$$p = \int_0^{E_v} p_E \, dE = \int_0^{E_v} g_{vb}(E)[(1 - f(E)]\, dE$$

With the assumption that E_F is a few kT above E_v, the integration simplifies to

Hole concentration in VB

$$p = N_v \exp\left[-\frac{(E_F - E_v)}{kT}\right] \tag{5.8}$$

where N_v is the effective density of states at the VB edge and is given by

Effective density of states at VB edge

$$N_v = 2\left(\frac{2\pi m_h^* kT}{h^2}\right)^{3/2} \tag{5.9}$$

We can now see the virtues of studying the density of states $g(E)$ as a function of energy E and the Fermi–Dirac function $f(E)$. Both were central factors in deriving the expressions for n and p. There are no specific assumptions in our derivations, except for E_F being a few kT away from the band edges, which means that Equations 5.6 and 5.8 are generally valid.

The general equations that determine the free electron and hole concentrations are thus given by Equations 5.6 and 5.8. It is interesting to consider the product np,

$$np = N_c \exp\left[-\frac{(E_c - E_F)}{kT}\right] N_v \exp\left[-\frac{(E_F - E_v)}{kT}\right] = N_c N_v \exp\left[-\frac{(E_c - E_v)}{kT}\right]$$

or

$$np = N_c N_v \exp\left(-\frac{E_g}{kT}\right) \tag{5.10}$$

where $E_g = E_c - E_v$ is the bandgap energy. First, we note that this is a general expression in which the right-hand side, $N_cN_v\exp(-E_g/kT)$, is a constant that depends on the temperature and the material properties, for example, E_g, and not on the position of the Fermi level. In the special case of an intrinsic semiconductor, $n = p$, which we can denote as n_i, the **intrinsic concentration,** so that $N_cN_v\exp(-E_g/kT)$ must be n_i^2. From Equation 5.10 we therefore have

$$np = n_i^2 = N_cN_v\exp\left(-\frac{E_g}{kT}\right) \qquad [5.11]$$

Mass action law

This is a general equation that is valid as long as we have thermal equilibrium. External excitation, such as photogeneration, is excluded. It states that the product np is a temperature-dependent constant. If we somehow increase the electron concentration, then we inevitably reduce the hole concentration. The constant n_i has a special significance because it represents the free electron and hole concentrations in the intrinsic material.

An **intrinsic semiconductor** is a pure semiconductor crystal in which the electron and hole concentrations are equal. By pure we mean virtually no impurities in the crystal. We should also exclude crystal defects that may capture carriers of one sign and thus result in unequal electron and hole concentrations. Clearly in a pure semiconductor, electrons and holes are generated in pairs by thermal excitation across the bandgap. It must be emphasized that Equation 5.11 is generally valid and therefore applies to both intrinsic and nonintrinsic ($n \neq p$) semiconductors.

When an electron and hole meet in the crystal, they "recombine." The electron falls in energy and occupies the empty electronic state that the hole represents. Consequently, the broken bond is "repaired," but we lose two free charge carriers. **Recombination** of an electron and hole results in their annihilation. In a semiconductor we therefore have thermal generation of electron–hole pairs by thermal excitation from the VB to the CB, and we also have recombination of electron–hole pairs that removes them from their conduction and valence bands, respectively. The rate of recombination R will be proportional to the number of electrons and also to the number of holes. Thus

$$R \propto np$$

The rate of generation G will depend on how many electrons are available for excitation at E_v, that is, N_v; how many empty states are available at E_c, that is, N_c; and the probability that the electron will make the transition, that is, $\exp(-E_g/kT)$, so that

$$G \propto N_cN_v\exp\left(-\frac{E_g}{kT}\right)$$

Since in thermal equilibrium we have no continuous increase in n or p, we must have the rate of generation equal to the rate of recombination, that is, $G = R$. This is equivalent to Equation 5.11.

In sketching the diagrams in Figure 5.7a to d to illustrate the derivation of the expressions for n and p (in Equations 5.6 and 5.8), we assumed that the Fermi level E_F is somewhere around the middle of the energy bandgap. This was not an assumption in the mathematical derivations but only in the sketches. From Equations 5.6 and 5.8 we also note that the position of Fermi level is important in determining the electron and hole concentrations. It serves as a "mathematical crank" to determine n and p.

We first consider an intrinsic semiconductor, $n = p = n_i$. Setting $p = n_i$ in Equation 5.8, we can solve for the Fermi energy in the intrinsic semiconductor, E_{Fi}, that is,

$$N_v \exp\left[-\frac{(E_{Fi} - E_v)}{kT}\right] = (N_c N_v)^{1/2} \exp\left(-\frac{E_g}{2kT}\right)$$

which leads to

Fermi energy in intrinsic semiconductor

$$E_{Fi} = E_v + \frac{1}{2}E_g - \frac{1}{2}kT \ln\left(\frac{N_c}{N_v}\right) \qquad [5.12]$$

Furthermore, substituting the proper expressions for N_c and N_v we get

Fermi energy in intrinsic semiconductor

$$E_{Fi} = E_v + \frac{1}{2}E_g - \frac{3}{4}kT \ln\left(\frac{m_e^*}{m_h^*}\right) \qquad [5.13]$$

It is apparent from these equations that if $N_c = N_v$ or $m_e^* = m_h^*$, then

$$E_{Fi} = E_v + \frac{1}{2}E_g$$

that is, E_{Fi} is right in the middle of the energy gap. Normally, however, the effective masses will not be equal and the Fermi level will be slightly shifted down from midgap by an amount $\frac{3}{4}kT \ln(m_e^*/m_h^*)$, which is quite small compared with $\frac{1}{2}E_g$. For Si and Ge, the hole effective mass (for density of states) is slightly smaller than the electron effective mass, so E_{Fi} is slightly below the midgap.

The condition $np = n_i^2$ means that if we can somehow increase the electron concentration in the CB over the intrinsic value—for example, by adding impurities into the Si crystal that donate additional electrons to the CB—we will then have $n > p$. The semiconductor is then called **n-type.** The Fermi level must be closer to E_c than E_v, so that

$$E_c - E_F < E_F - E_v$$

and Equations 5.6 and 5.8 yield $n > p$. The np product always yields n_i^2 in thermal equilibrium in the absence of external excitation, for example, illumination.

It is also possible to have an excess of holes in the VB over electrons in the CB, for example, by adding impurities that remove electrons from the VB and thereby generate holes. In that case E_F is closer to E_v than to E_c. A semiconductor in which $p > n$ is called a **p-type semiconductor.** The general band diagrams with the appropriate Fermi levels for intrinsic, n-type, and p-type semiconductors (*e.g.*, i-Si, n-Si, and p-Si, respectively) are illustrated in Figure 5.8a to c.

It is apparent that if we know where E_F is, then we have effectively determined n and p by virtue of Equations 5.6 and 5.8. We can view E_F as a *material property* that is related to the concentration of charge carriers that contribute to electrical conduction. Its significance, however, goes beyond n and p. It also determines the energy needed to remove an electron from the semiconductor. The energy difference between the vacuum level (where the electron is free) and E_F is the **work function** Φ of the semiconductor, the energy required to remove an electron even though there are no electrons at E_F in a semiconductor.

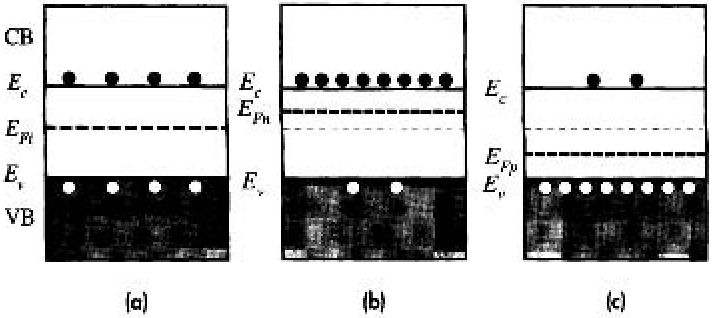


Figure 5.8 Energy band diagrams for (a) intrinsic, (b) n-type, and (c) p-type semiconductors.
In all cases, $np = n_i^2$.

The Fermi level can also be interpreted in terms of the potential energy per electron for electrical work similar to the interpretation of electrostatic PE. Just as $e\,\Delta V$ is the electrical work involved in taking a charge e across a potential difference ΔV, any difference in E_F in going from one end of a material (or system) to another is available to do an amount ΔE_F of external work. A corollary to this is that if electrical work is done on the material, for example, by passing a current through it, then the Fermi level is not uniform in the material. ΔE_F then represents the work done per electron. For a material in thermal equilibrium and not subject to any external excitation such as illumination or connections to a voltage supply, the Fermi level in the material must therefore be uniform, $\Delta E_F = 0$.

What is the average energy of an electron in the conduction band of a semiconductor? Also, what is the mean speed of an electron in the conduction band? We note that the concentration of electrons with energies E to $E + dE$ is $n_E(E)\,dE$ or $g_{cb}(E)f(E)\,dE$. Thus the average energy of electrons in the CB, by definition of the mean, is

$$\overline{E}_{\text{CB}} = \frac{1}{n}\int_{\text{CB}} E g_{cb}(E) f(E)\,dE$$

where the integration must be over the CB. Substituting the proper expressions for $g_{cb}(E)$ and $f(E)$ in the integrand and carrying out the integration from E_c to the top of the band, we find the very simple result that

$$\overline{E}_{\text{CB}} = E_c + \frac{3}{2}kT \qquad [5.14]$$

Average electron energy in CB

Thus, an electron in the conduction band has an average energy of $\frac{3}{2}kT$ *above* E_c. Since we know that an electron at E_c is "free" in the crystal, $\frac{3}{2}kT$ must be its average kinetic energy.

This is just like the average kinetic energy of gas atoms (such as He atoms) in a tank assuming that the atoms (or the "particles") do not interact, that is, they are independent. We know from the kinetic theory that the statistics of a collection of independent gas atoms obeys the classical Maxwell–Boltzmann description with an average energy given by $\frac{3}{2}kT$. We should also recall that the description of electron statistics in a metal involves the Fermi–Dirac function, which is based on the Pauli exclusion principle. In a metal the average energy of the conduction electron is $\frac{3}{5}E_F$ and, for all practical purposes, temperature independent. We see that the collective electron behavior is completely different in the two solids. We can explain the difference by noting that the conduction band in a

Table 5.1 Selected typical properties of Ge, Si, and GaAs at 300 K

	E_g (eV)	χ (eV)	N_c (cm^{-3})	N_v (cm^{-3})	n_i (cm^{-3})	μ_e (cm^2 V^{-1} s^{-1})	μ_h (cm^2 V^{-1} s^{-1})	m_e^*/m_e	m_h^*/m_e	ε_r
Ge	0.66	4.13	1.04×10^{19}	6.0×10^{18}	2.3×10^{13}	3900	1900	0.12*a* 0.56*b*	0.23*a* 0.40*b*	16
Si	1.10	4.01	2.8×10^{19}	1.2×10^{19}	1.0×10^{10}	1350	450	0.26*a* 1.08*b*	0.38*a* 0.60*b*	11.9
GaAs	1.42	4.07	4.7×10^{17}	7×10^{18}	2.1×10^{6}	8500	400	0.067*a*,*b*	0.40*a* 0.50*b*	13.1

NOTE: Effective mass related to conductivity (labeled *a*) is different than that for density of states (labeled *b*). In numerous textbooks, n_i is taken as 1.45×10^{10} cm^{-3} and is therefore the most widely used value of n_i for Si, though the correct value is actually 1.0×10^{10} cm^{-3}. (M. A. Green, *J. Appl. Phys.*, **67**, 2944, 1990.)

semiconductor is only scarcely populated by electrons, which means that there are many more electronic states than electrons and thus the likelihood of two electrons trying to occupy the same electronic state is practically nil. We can then neglect the Pauli exclusion principle and use the Boltzmann statistics. This is not the case for metals where the number of conduction electrons and the number of states are comparable in magnitude.

Table 5.1 is a comparative table of some of the properties of the important semiconductors, Ge, Si, and GaAs.

EXAMPLE 5.1 **INTRINSIC CONCENTRATION AND CONDUCTIVITY OF Si** Given that the density of states related effective masses of electrons and holes in Si are approximately $1.08m_e$ and $0.60m_e$, respectively, and the electron and hole drift mobilities at room temperature are 1350 and 450 cm^2 V^{-1} s^{-1}, respectively, calculate the intrinsic concentration and intrinsic resistivity of Si.

SOLUTION

We simply calculate the effective density of states N_c and N_v by

$$N_c = 2\left(\frac{2\pi m_e^* kT}{h^2}\right)^{3/2} \qquad \text{and} \qquad N_v = 2\left(\frac{2\pi m_h^* kT}{h^2}\right)^{3/2}$$

Thus

$$N_c = 2\left[\frac{2\pi(1.08 \times 9.1 \times 10^{-31}\ \text{kg})(1.38 \times 10^{-23}\ \text{J K}^{-1})\ (300\ \text{K})}{(6.63 \times 10^{-34}\ \text{J s})^2}\right]^{3/2}$$

$$= 2.81 \times 10^{25}\ \text{m}^{-3} \qquad \text{or} \qquad 2.81 \times 10^{19}\ \text{cm}^{-3}$$

and

$$N_v = 2\left[\frac{2\pi(0.60 \times 9.1 \times 10^{-31}\ \text{kg})(1.38 \times 10^{-23}\ \text{J K}^{-1})(300\ \text{K})}{(6.63 \times 10^{-34}\ \text{J s})^2}\right]^{3/2}$$

$$= 1.16 \times 10^{25}\ \text{m}^{-3} \qquad \text{or} \qquad 1.16 \times 10^{19}\ \text{cm}^{-3}$$

The intrinsic concentration is

$$n_i = (N_c N_v)^{1/2} \exp\left(-\frac{E_g}{2kT}\right)$$

so that

$$n_i = [(2.81 \times 10^{19}\ \text{cm}^{-3})(1.16 \times 10^{19}\ \text{cm}^{-3})]^{1/2}\exp\left[-\frac{(1.10\ \text{eV})}{2(300\ \text{K})(8.62 \times 10^{-5}\ \text{eV K}^{-1})}\right]$$
$$= 1.0 \times 10^{10}\text{cm}^{-3}$$

The conductivity is

$$\sigma = en\mu_e + ep\mu_h = en_i(\mu_e + \mu_h)$$

that is,

$$\sigma = (1.6 \times 10^{-19}\ \text{C})(1.0 \times 10^{10}\ \text{cm}^{-3})(1350 + 450\ \text{cm}^2\ \text{V}^{-1}\ \text{s}^{-1})$$
$$= 2.9 \times 10^{-6}\ \Omega^{-1}\ \text{cm}^{-1}$$

The resistivity is

$$\rho = \frac{1}{\sigma} = 3.5 \times 10^5\ \Omega\ \text{cm}$$

Although we calculated $n_i = 1.0 \times 10^{10}$ cm^{-3}, the most widely used n_i value in the literature has been 1.45×10^{10} cm^{-3}. The difference arises from a number of factors but, most importantly, from what exact value of the effective hole mass should be used in calculating N_v. Henceforth we will simply use[2] $n_i = 1.0 \times 10^{10}$ cm^{-3}, which seems to be the "true" value.

EXAMPLE 5.2

MEAN SPEED OF ELECTRONS IN THE CB Estimate the mean speed of electrons in the conduction band of Si at 300 K. If a is the magnitude of lattice vibrations, then the kinetic theory predicts $a^2 \propto T$; or stated differently, the mean energy associated with lattice vibrations (proportional to a^2) increases with kT. Given the temperature dependence of the mean speed of electrons in the CB, what should be the temperature dependence of the drift mobility? The effective mass of an electron in the conduction band is $0.26m_e$.

SOLUTION

The fact that the average KE, $\frac{1}{2}m_e^* v_e^2$, of an electron in the CB of a semiconductor is $\frac{3}{2}kT$ means that the effective mean speed v_e must be

$$v_e = \left(\frac{3kT}{m_e^*}\right)^{1/2} = \left[\frac{(3 \times 1.38 \times 10^{-23} \times 300)}{(0.26 \times 9.1 \times 10^{-31})}\right]^{1/2} = 2.3 \times 10^5\ \text{m s}^{-1}$$

The effective mean speed v_e is called the **thermal velocity** v_{th} of the electron.

The mean free time τ of the electron between scattering events due to thermal vibrations of the atoms is inversely proportional to both the mean speed v_e of the electron and the scattering cross section of the thermal vibrations, that is,

$$\tau \propto \frac{1}{v_e(\pi a^2)}$$

where a is the amplitude of the atomic thermal vibrations. But, $v_e \propto T^{1/2}$ and $(\pi a^2) \propto kT$, so that $\tau \propto T^{-3/2}$ and consequently $\mu_e \propto T^{-3/2}$.

Experimentally μ_e is not exactly proportional to $T^{-3/2}$ but to $T^{-2.4}$, a higher power index. The effective mass used in the density of states calculations is actually different than that used in transport calculations such as the mean speed, drift mobility, and so on.

---

[2] The correct value appears to be 1.0×10^{10} cm^{-3} as discussed by M. A. Green (*J. Appl. Phys.*, **67**, 2944, 1990) and A. B. Sproul and M. A. Green (*J. Appl. Phys.*, **70**, 846, 1991).

5.2 EXTRINSIC SEMICONDUCTORS

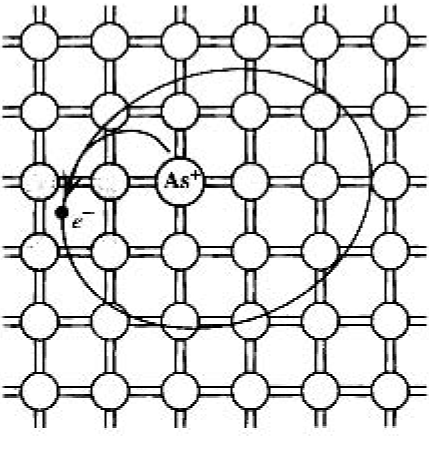
By introducing small amounts of impurities into an otherwise pure Si crystal, it is possible to obtain a semiconductor in which the concentration of carriers of one polarity is much in excess of the other type. Such semiconductors are referred to as **extrinsic semiconductors** vis-à-vis the intrinsic case of a pure and perfect crystal. For example, by adding pentavalent impurities, such as arsenic, which have a valency of more than four, we can obtain a semiconductor in which the electron concentration is much larger than the hole concentration. In this case we will have an *n*-type semiconductor. If we add trivalent impurities, such as boron, which have a valency of less than four, then we find that we have an excess of holes over electrons. We now have a *p*-type semiconductor. How do impurities change the concentrations of holes and electrons in a semiconductor?

5.2.1 *n*-TYPE DOPING

Consider what happens when small amounts of a pentavalent (valency of 5) element from Group V in the Periodic Table, such as As, P, Sb, are introduced into a pure Si crystal. We only add small amounts (*e.g.*, one impurity atom for every million host atoms) because we wish to surround each impurity atom by millions of Si atoms, thereby forcing the impurity atoms to bond with Si atoms in the same diamond crystal structure. Arsenic has five valence electrons, whereas Si has four. Thus when an As atom bonds with four Si atoms, it has one electron left unbonded. It cannot find a bond to go into, so it is left orbiting around the As atom, as illustrated in Figure 5.9. The As^+ ionic center with an electron e^- orbiting it is just like a hydrogen atom in a silicon environment. We can easily calculate how much energy is required to free this electron away from the As site, thereby ionizing the As impurity. Had this been a hydrogen atom in free space, the energy required to remove the electron from its ground state (at $n = 1$) to far away from the positive center would have been given by $-E_n$ with $n = 1$. The binding energy of the electron in the H atom is thus

$$E_b = -E_1 = \frac{m_e e^4}{8\varepsilon_o^2 h^2} = 13.6\ \text{eV}$$

Figure 5.9 Arsenic-doped Si crystal.
The four valence electrons of As allow it to bond just like Si, but the fifth electron is left orbiting the As site. The energy required to release the free fifth electron into the CB is very small.



If we wish to apply this to the electron around an As^+ core in the Si crystal environment, we must use $\varepsilon_r\varepsilon_o$ instead of ε_o, where ε_r is the relative permittivity of silicon, and also the effective mass of the electron m_e^* in the silicon crystal. Thus, the binding energy of the electron to the As^+ site in the Si crystal is

Electron binding energy at a donor

$$E_b^{\mathrm{Si}} = \frac{m_e^* e^4}{8\varepsilon_r^2\varepsilon_o^2 h^2} = (13.6\ \mathrm{eV})\left(\frac{m_e^*}{m_e}\right)\left(\frac{1}{\varepsilon_r^2}\right) \qquad [5.15]$$

With $\varepsilon_r = 11.9$ and $m_e^* \approx \frac{1}{3}m_e$ for silicon, we find $E_b^{\mathrm{Si}} = 0.032$ eV, which is comparable with the average thermal energy of atomic vibrations at room temperature, $\sim 3kT$ (~ 0.07 eV). Thus, the fifth valence electron can be readily freed by thermal vibrations of the Si lattice. The electron will then be "free" in the semiconductor, or, in other words, it will be in the CB. The energy required to excite the electron to the CB is therefore 0.032 eV. The addition of As atoms introduces localized electronic states at the As sites because the fifth electron has a localized wavefunction, of the hydrogenic type, around As^+. The energy E_d of these states is 0.032 eV below E_c because this is how much energy is required to take the electron away into the CB. Thermal excitation by the lattice vibrations at room temperature is sufficient to ionize the As atom, that is, excite the electron from E_d into the CB. This process creates free electrons but immobile As^+ ions, as shown in the energy band diagram of an n-type semiconductor in Figure 5.10. Because the As atom donates an electron into the CB, it is called a **donor atom.** E_d is the electron energy around the donor atom. E_d is close to E_c, so the spare fifth electron from the dopant can be readily donated to the CB. If N_d is the donor atom concentration in the crystal, then provided that $N_d \gg n_i$, at room temperature the electron concentration in the CB will be nearly equal to N_d, that is $n \approx N_d$. The hole concentration will be $p = n_i^2/N_d$, which is less than the intrinsic concentration because a few of the large number of electrons in the CB recombine with holes in the VB so as to maintain $np = n_i^2$. The conductivity will then be

n-type conductivity

$$\sigma = eN_d\mu_e + e\left(\frac{n_i^2}{N_d}\right)\mu_h \approx eN_d\mu_e \qquad [5.16]$$

At low temperatures, however, not all the donors will be ionized and we need to know the probability, denoted as $f_d(E_d)$, of finding an electron in a state with energy

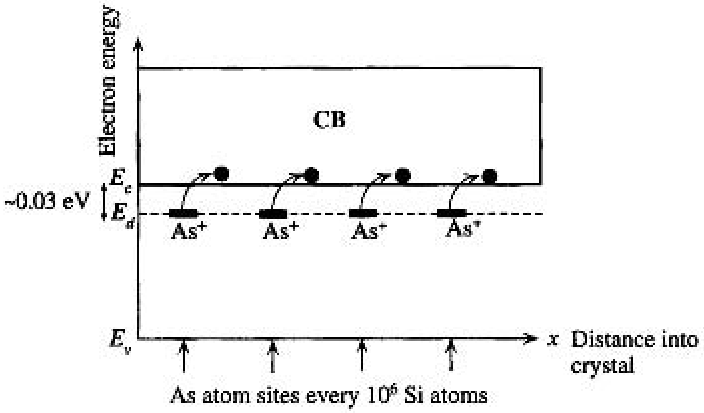


Figure 5.10 Energy band diagram for an n-type Si doped with 1 ppm As.

There are donor energy levels just below E_c around As^+ sites.

E_d at a donor. This probability function is similar to the Fermi–Dirac function $f(E_d)$ except that it has a factor of $\frac{1}{2}$ multiplying the exponential term,

Occupation probability at a donor

$$f_d(E_d) = \frac{1}{1 + \frac{1}{2}\exp\left[\frac{(E_d - E_F)}{kT}\right]} \qquad [5.17]$$

The factor $\frac{1}{2}$ is due to the fact that the electron state at the donor can take an electron with spin either up or down but not both[3] (once the donor has been occupied, a second electron cannot enter this site). Thus, the number of ionized donors at a temperature T is given by

$$\begin{aligned} N_d^+ &= N_d \times (\text{probability of not finding an electron at } E_d) \\ &= N_d[1 - f_d(E_d)] \\ &= \frac{N_d}{1 + 2\exp\left[\frac{(E_F - E_d)}{kT}\right]} \end{aligned} \qquad [5.18]$$

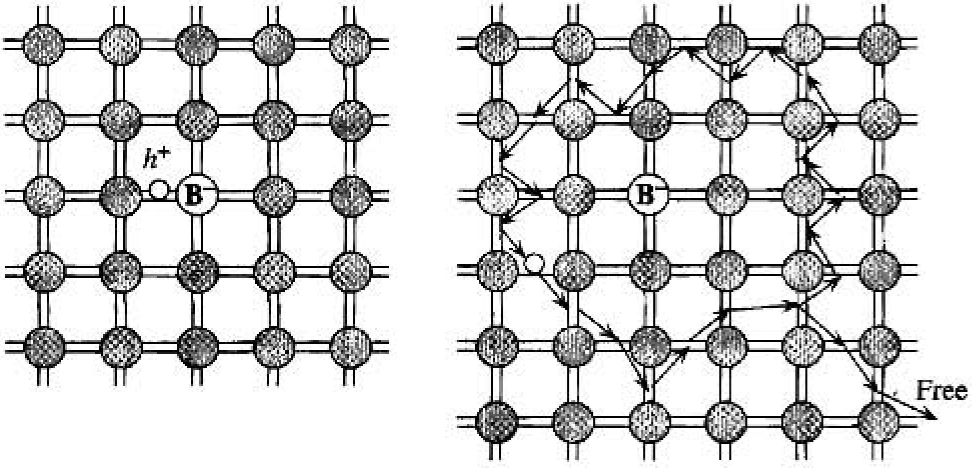
5.2.2 *p*-TYPE DOPING

We saw that introducing a pentavalent atom into a Si crystal results in *n*-type doping because the fifth electron cannot go into a bond and escapes from the donor into the CB by thermal excitation. By similar arguments, we should anticipate that doping a Si crystal with a trivalent atom (valency of 3) such as B, Al, Ga, or In will result in a *p*-type Si crystal. We consider doping Si with small amounts of B as shown in Figure 5.11a. Because B has only three valence electrons, when it shares them with four neighboring Si atoms, one of the bonds has a missing electron, which of course is a hole. A nearby electron can tunnel into this hole and displace the hole further away from the boron atom. As the hole moves away, it gets attracted by the negative charge left behind on the boron atom and therefore takes an orbit around the B^- ion, as shown in Figure 5.11b. The binding energy of this hole to the B^- ion can be calculated using the hydrogenic atom analogy as in the *n*-type Si case. This binding energy turns out to be very small, ~0.05 eV, so at room temperature the thermal vibrations of the lattice can free the hole away from the B^- site. A free hole, we recall, exists in the VB. The escape of the hole from the B^- site involves the B atom *accepting* an electron from a neighboring Si–Si bond (from the VB), which effectively results in the hole being displaced away and its eventual escape to freedom in the VB. The B atom introduced into the Si crystal therefore acts as an electron acceptor and, because of this, it is called an **acceptor impurity.** The electron accepted by the B atom comes from a nearby bond. On the energy band diagram, an electron leaves the VB and gets accepted by a B atom, which becomes negatively charged. This process leaves a hole in the VB that is free to wander away, as illustrated in Figure 5.12.

It is apparent that doping a silicon crystal with a trivalent impurity results in a *p*-type material. We have many more holes than electrons for electrical conduction

---

[3] The proof can be found in advanced solid-state physics texts.



(a) (b)

Figure 5.11 Boron-doped Si crystal.

B has only three valence electrons. When it substitutes for a Si atom, one of its bonds has an electron missing and therefore a hole, as shown in (a). The hole orbits around the B^- site by the tunneling of electrons from neighboring bonds, as shown in (b). Eventually, thermally vibrating Si atoms provide enough energy to free the hole from the B^- site into the VB, as shown.

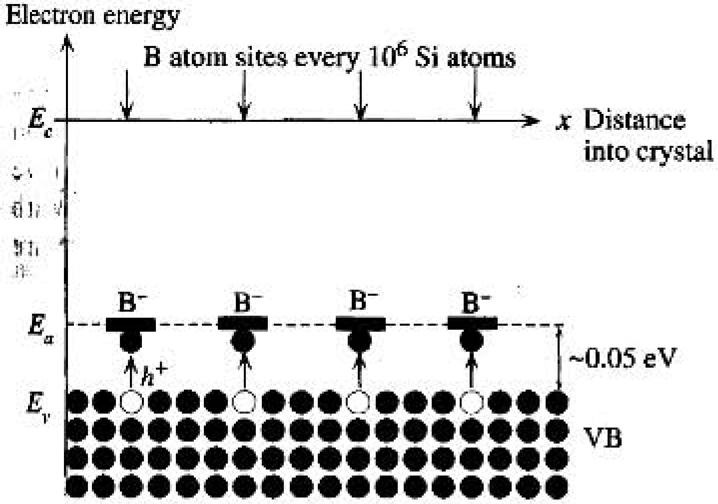


Figure 5.12 Energy band diagram for a *p*-type Si doped with 1 ppm B.

There are acceptor energy levels E_a just above E_v around B^- sites. These acceptor levels accept electrons from the VB and therefore create holes in the VB.

since the negatively charged B atoms are immobile and hence cannot contribute to the conductivity. If the concentration of acceptor impurities N_a in the crystal is much greater than the intrinsic concentration n_i, then at room temperature all the acceptors would have been ionized and thus $p \approx N_a$. The electron concentration is then determined by the mass action law, $n = n_i^2/N_a$, which is much smaller than p, and consequently the conductivity is simply given by $\sigma = eN_a\mu_h$.

Typical ionization energies for donor and acceptor atoms in the silicon crystal are summarized in Table 5.2.

Table 5.2 Examples of donor and acceptor ionization energies (eV) in Si

Donors			Acceptors		
P	As	Sb	B	Al	Ga
0.045	0.054	0.039	0.045	0.057	0.072

5.2.3 COMPENSATION DOPING

What happens when a semiconductor contains both donors and acceptors? **Compensation doping** is a term used to describe the doping of a semiconductor with both donors and acceptors to control the properties. For example, a p-type semiconductor doped with N_a acceptors can be converted to an n-type semiconductor by simply adding donors until the concentration N_d exceeds N_a. The effect of donors compensates for the effect of acceptors and vice versa. The electron concentration is then given by $N_d - N_a$ provided the latter is larger than n_i. When both acceptors and donors are present, what essentially happens is that electrons from donors recombine with the holes from the acceptors so that the mass action law $np = n_i^2$ is obeyed. Remember that we cannot simultaneously increase the electron and hole concentrations because that leads to an increase in the recombination rate that returns the electron and hole concentrations to satisfy $np = n_i^2$. When an acceptor atom accepts a valence band electron, a hole is created in the VB. This hole then recombines with an electron from the CB. Suppose that we have more donors than acceptors. If we take the initial electron concentration as $n = N_d$, then the recombination between the electrons from the donors and N_a holes generated by N_a acceptors results in the electron concentration reduced by N_a to $n = N_d - N_a$. By a similar argument, if we have more acceptors than donors, the hole concentration becomes $p = N_a - N_d$, with electrons from N_d donors recombining with holes from N_a acceptors. Thus there are two compensation effects:

Compensation doping

1. More donors: $N_d - N_a \gg n_i$ $\quad n = (N_d - N_a)$ and $\quad p = \dfrac{n_i^2}{(N_d - N_a)}$
2. More acceptors: $N_a - N_d \gg n_i$ $\quad p = (N_a - N_d)$ and $\quad n = \dfrac{n_i^2}{(N_a - N_d)}$

These arguments assume that the temperature is sufficiently high for donors and acceptors to have been ionized. This will be the case at room temperature. At low temperatures, we have to consider donor and acceptor statistics and the charge neutrality of the whole crystal, as in Example 5.8.

EXAMPLE 5.3 **RESISTIVITY OF INTRINSIC AND DOPED Si** Find the resistance of a 1 cm^3 pure silicon crystal. What is the resistance when the crystal is doped with arsenic if the doping is 1 in 10^9, that is, 1 part per billion (ppb) (note that this doping corresponds to one foreigner living in China)? Given data: Atomic concentration in silicon is 5×10^{22} cm^{-3}, $n_i = 1.0 \times 10^{10}$ cm^{-3}, $\mu_e = 1350$ cm^2 V^{-1} s^{-1}, and $\mu_h = 450$ cm^2 V^{-1} s^{-1}.

SOLUTION

For the intrinsic case, we apply

$$\sigma = en\mu_e + ep\mu_h = en(\mu_e + \mu_h)$$

so

$$\sigma = (1.6 \times 10^{-19}\ \text{C})(1.0 \times 10^{10}\ \text{cm}^{-3})(1350 + 450\ \text{cm}^2\ \text{V}^{-1}\ \text{s}^{-1})$$
$$= 2.88 \times 10^{-6}\ \Omega^{-1}\ \text{cm}^{-1}$$

Since $L = 1$ cm and $A = 1$ cm^2, the resistance is

$$R = \frac{L}{\sigma A} = \frac{1}{\sigma} = 3.47 \times 10^5\ \Omega \qquad \text{or} \qquad 347\ \text{k}\Omega$$

When the crystal is doped with 1 in 10^9, then

$$N_d = \frac{N_{\text{Si}}}{10^9} = \frac{5 \times 10^{22}}{10^9} = 5 \times 10^{13}\ \text{cm}^{-3}$$

At room temperature all the donors are ionized, so

$$n = N_d = 5 \times 10^{13}\ \text{cm}^{-3}$$

The hole concentration is

$$p = \frac{n_i^2}{N_d} = \frac{(1.0 \times 10^{10})^2}{(5 \times 10^{13})} = 2.0 \times 10^6\ \text{cm}^{-3} \ll n_i$$

Therefore,

$$\sigma = en\mu_e = (1.6 \times 10^{-19}\ \text{C})(5 \times 10^{13}\ \text{cm}^{-3})(1350\ \text{cm}^2\ \text{V}^{-1}\ \text{s}^{-1})$$
$$= 1.08 \times 10^{-2}\ \Omega^{-1}\ \text{cm}^{-1}$$

Further,

$$R = \frac{L}{\sigma A} = \frac{1}{\sigma} = 92.6\ \Omega$$

Notice the drastic fall in the resistance when the crystal is doped with only 1 in 10^9 atoms.

Doping the silicon crystal with boron instead of arsenic, but still in amounts of 1 in 10^9, means that $N_a = 5 \times 10^{13}$ cm^{-3}, which results in a conductivity of

$$\sigma = ep\mu_h = (1.6 \times 10^{-19}\ \text{C})(5 \times 10^{13}\ \text{cm}^{-3})(450\ \text{cm}^2\ \text{V}^{-1}\ \text{s}^{-1})$$
$$= 3.6 \times 10^{-3}\ \Omega^{-1}\ \text{cm}^{-1}$$

Therefore,

$$R = \frac{L}{\sigma A} = \frac{1}{\sigma} = 278\ \Omega$$

The reason for a higher resistance with p-type doping compared with the same amount of n-type doping is that $\mu_h < \mu_e$.

EXAMPLE 5.4

COMPENSATION DOPING An n-type Si semiconductor containing 10^{16} phosphorus (donor) atoms cm^{-3} has been doped with 10^{17} boron (acceptor) atoms cm^{-3}. Calculate the electron and hole concentrations in this semiconductor.

SOLUTION

This semiconductor has been compensation doped with excess acceptors over donors, so

$$N_a - N_d = 10^{17} - 10^{16} = 9 \times 10^{16}\ \text{cm}^{-3}$$

This is much larger than the intrinsic concentration $n_i = 1.0 \times 10^{10}\ \text{cm}^{-3}$ at room temperature, so

$$p = N_a - N_d = 9 \times 10^{16}\ \text{cm}^{-3}$$

The electron concentration

$$n = \frac{n_i^2}{p} = \frac{(1.0 \times 10^{10}\ \text{cm}^{-3})^2}{(9 \times 10^{16}\ \text{cm}^{-3})} = 1.1 \times 10^3\ \text{cm}^{-3}$$

Clearly, the electron concentration and hence its contribution to electrical conduction is completely negligible compared with the hole concentration. Thus, by excessive boron doping, the n-type semiconductor has been converted to a p-type semiconductor.

EXAMPLE 5.5

THE FERMI LEVEL IN n- AND p-TYPE Si An n-type Si wafer has been doped uniformly with 10^{16} antimony (Sb) atoms cm^{-3}. Calculate the position of the Fermi energy with respect to the Fermi energy E_{Fi} in intrinsic Si. The above n-type Si sample is further doped with 2×10^{17} boron atoms cm^{-3}. Calculate the position of the Fermi energy with respect to the Fermi energy E_{Fi} in intrinsic Si. (Assume that $T = 300$ K, and $kT = 0.0259$ eV.)

SOLUTION

Sb gives n-type doping with $N_d = 10^{16}\ \text{cm}^{-3}$, and since $N_d \gg n_i\ (= 1.0 \times 10^{10}\ \text{cm}^{-3})$, we have

$$n = N_d = 10^{16}\ \text{cm}^{-3}$$

For intrinsic Si,

$$n_i = N_c \exp\left[-\frac{(E_c - E_{Fi})}{kT}\right]$$

whereas for doped Si,

$$n = N_c \exp\left[-\frac{(E_c - E_{Fn})}{kT}\right] = N_d$$

where E_{Fi} and E_{Fn} are the Fermi energies in the intrinsic and n-type Si. Dividing the two expressions,

$$\frac{N_d}{n_i} = \exp\left[\frac{(E_{Fn} - E_{Fi})}{kT}\right]$$

so that

$$E_{Fn} - E_{Fi} = kT \ln\left(\frac{N_d}{n_i}\right) = (0.0259\ \text{eV}) \ln\left(\frac{10^{16}}{1.0 \times 10^{10}}\right) = 0.36\ \text{eV}$$

When the wafer is further doped with boron, the acceptor concentration is

$$N_a = 2 \times 10^{17}\ \text{cm}^{-3} > N_d = 10^{16}\ \text{cm}^{-3}$$

The semiconductor is compensation doped and compensation converts the semiconductor to p-type Si. Thus

$$p = N_a - N_d = (2 \times 10^{17} - 10^{16}) = 1.9 \times 10^{17}\ \text{cm}^{-3}$$

For intrinsic Si,

$$n_i = N_v \exp\left[-\frac{(E_{Fi} - E_v)}{kT}\right]$$

whereas for doped Si,

$$p = N_v \exp\left[-\frac{(E_{Fp} - E_v)}{kT}\right] = N_a - N_d$$

where E_{Fi} and E_{Fp} are the Fermi energies in the intrinsic and p-type Si, respectively. Dividing the two expressions,

$$\frac{p}{n_i} = \exp\left[-\frac{(E_{Fp} - E_{Fi})}{kT}\right]$$

so that

$$E_{Fp} - E_{Fi} = -kT \ln\left(\frac{p}{n_i}\right) = -(0.0259 \text{ eV}) \ln\left(\frac{1.9 \times 10^{17}}{1.0 \times 10^{10}}\right)$$
$$= -0.43 \text{ eV}$$

ENERGY BAND DIAGRAM OF AN *n*-TYPE SEMICONDUCTOR CONNECTED TO A VOLTAGE SUPPLY **EXAMPLE 5.6** Consider the energy band diagram for an n-type semiconductor that is connected to a voltage supply of V and is carrying a current. The applied voltage drops uniformly along the semiconductor, so the electrons in the semiconductor now also have an imposed electrostatic potential energy that decreases toward the positive terminal, as depicted in Figure 5.13. The whole band structure, the CB and the VB, therefore tilts. When an electron drifts from A toward

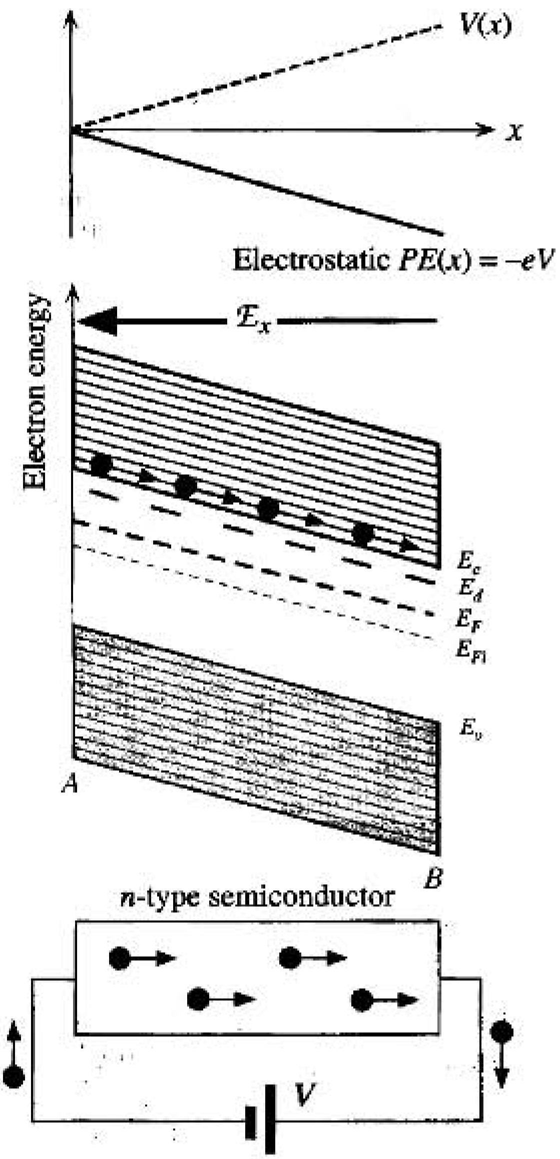


Figure 5.13 Energy band diagram of an n-type semiconductor connected to a voltage supply of V volts.

The whole energy diagram tilts because the electron now also has an electrostatic potential energy.

B, its PE decreases because it is approaching the positive terminal. The Fermi level E_F is above that for the intrinsic case, E_{Fi}.

We should remember that an important property of the Fermi level is that a change in E_F within a system is available externally to do electrical work. As a corollary we note that when electrical work is done on the system, for example, when a battery is connected to a semiconductor, then E_F is not uniform throughout the whole system. A change in E_F within a system ΔE_F is equivalent to electrical work per electron or eV. E_F therefore follows the electrostatic PE behavior, and the change in E_F from one end to the other, $E_F(A) - E_F(B)$, is just eV, the energy expended in taking an electron through the semiconductor, as shown in Figure 5.13. Electron concentration in the semiconductor is uniform, so $E_c - E_F$ must be constant from one end to the other. Thus the CB, VB, and E_F all bend by the same amount.

5.3 TEMPERATURE DEPENDENCE OF CONDUCTIVITY

So far we have been calculating conductivities and resistivities of doped semiconductors at room temperature by simply assuming that $n \approx N_d$ for n-type and $p \approx N_a$ for p-type doping, with the proviso that the concentration of dopants is much greater than the intrinsic concentration n_i. To obtain the conductivity at other temperatures we have to consider two factors: the temperature dependence of the carrier concentration and the drift mobility.

5.3.1 CARRIER CONCENTRATION TEMPERATURE DEPENDENCE

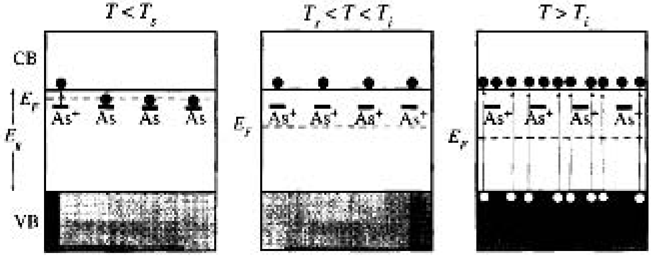
Consider an n-type semiconductor doped with N_d donors per unit volume where $N_d \gg n_i$. We take the semiconductor down to very low temperatures until its conductivity is practically nil. At this temperature, the donors will *not* be ionized because the thermal vibrational energy is insufficiently small. As the temperature is increased, some of the donors become ionized and donate their electrons to the CB, as shown in Figure 5.14a. The Si–Si bond breaking, that is, thermal excitation from E_v to E_c, is unlikely because it takes too much energy. Since the donor ionization energy $\Delta E = E_c - E_d$ is very small ($\ll E_g$), thermal generation involves exciting electrons from E_d to E_c. The electron concentration at low temperatures is given by the expression

$$n = \left(\frac{1}{2}N_c N_d\right)^{1/2} \exp\left(-\frac{\Delta E}{2kT}\right) \qquad [5.19]$$

similar to the intrinsic case, that is,

$$n = (N_c N_v)^{1/2} \exp\left(-\frac{E_g}{2kT}\right) \qquad [5.20]$$

Equation 5.20 is valid when thermal generation occurs across the bandgap E_g from E_v to E_c. Equation 5.19 is the counterpart of Equation 5.20 taking into account that at low temperatures the excitation is from E_d to E_c (across ΔE) and that instead of N_v, we have N_d as the number of available electrons. The numerical factor $\frac{1}{2}$ in



(a) $T = T_1$ (b) $T = T_2$ (c) $T = T_3$

Figure 5.14

(a) Below T_s, the electron concentration is controlled by the ionization of the donors.

(b) Between T_s and T_i, the electron concentration is equal to the concentration of donors since they would all have ionized.

(c) At high temperatures, thermally generated electrons from the VB exceed the number of electrons from ionized donors and the semiconductor behaves as if intrinsic.

Equation 5.19 arises because donor occupation statistics is different by this factor from the usual Fermi–Dirac function, as mentioned earlier.

As the temperature is increased further, eventually all the donors become ionized and the electron concentration is equal to the donor concentration, that is, $n = N_d$, as depicted in Figure 5.14b. This state of affairs remains unchanged until very high temperatures are reached, when thermal generation across the bandgap begins to dominate. At very high temperatures, thermal vibrations of the atoms will be so strong that many Si–Si bonds will be broken and thermal generation across E_g will dominate. The electron concentration in the CB will then be mainly due to thermal excitation from the VB to the CB, as illustrated in Figure 5.14c. But this process also generates an equal concentration of holes in the VB. Accordingly, the semiconductor behaves as if it were intrinsic. The electron concentration at these temperatures will therefore be equal to the intrinsic concentration n_i, which is given by Equation 5.20.

The dependence of the electron concentration on temperature thus has three regions:

1. **Low-temperature range ($T < T_s$).** The increase in temperature at these low temperatures ionizes more and more donors. The donor ionization continues until we reach a temperature T_s, called the **saturation temperature,** when all donors have been ionized and we have saturation in the concentration of ionized donors. The electron concentration is given by Equation 5.19. This temperature range is often referred to as the **ionization range.**

2. **Medium-temperature range ($T_s < T < T_i$).** Since nearly all the donors have been ionized in this range, $n = N_d$. This condition remains unchanged until $T = T_i$, when n_i, which is temperature dependent, becomes equal to N_d. It is this

temperature range $T_s < T < T_i$ that utilizes the n-type doping properties of the semiconductor in pn junction device applications. This temperature range is often referred to as the **extrinsic range.**

3. High-temperature range ($T > T_i$). The concentration of electrons generated by thermal excitation across the bandgap n_i is now much larger than N_d, so the electron concentration $n = n_i(T)$. Furthermore, as excitation occurs from the VB to the CB, the hole concentration $p = n$. This temperature range is referred to as the **intrinsic range.**

Figure 5.15 shows the behavior of the electron concentration with temperature in an n-type semiconductor. By convention we plot $\ln(n)$ versus the reciprocal temperature T^{-1}. At low temperatures, $\ln(n)$ versus T^{-1} is almost a straight line with a slope $-(\Delta E/2k)$, since the temperature dependence of $N_c^{1/2}(\propto T^{3/4})$ is negligible compared with the $\exp(-\Delta E/2kT)$ part in Equation 5.19. In the high-temperature range, however, the slope is quite steep and almost $-E_g/2k$ since Equation 5.20 implies that

$$n \propto T^{3/2}\exp\left(-\frac{E_g}{2kT}\right)$$

and the exponential part again dominates over the $T^{3/2}$ part. In the intermediate range, n is equal to N_d and practically independent of the temperature.

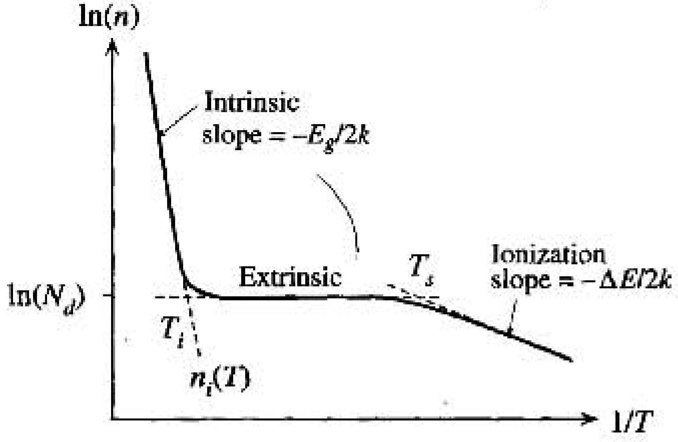


Figure 5.15 The temperature dependence of the electron concentration in an n-type semiconductor.

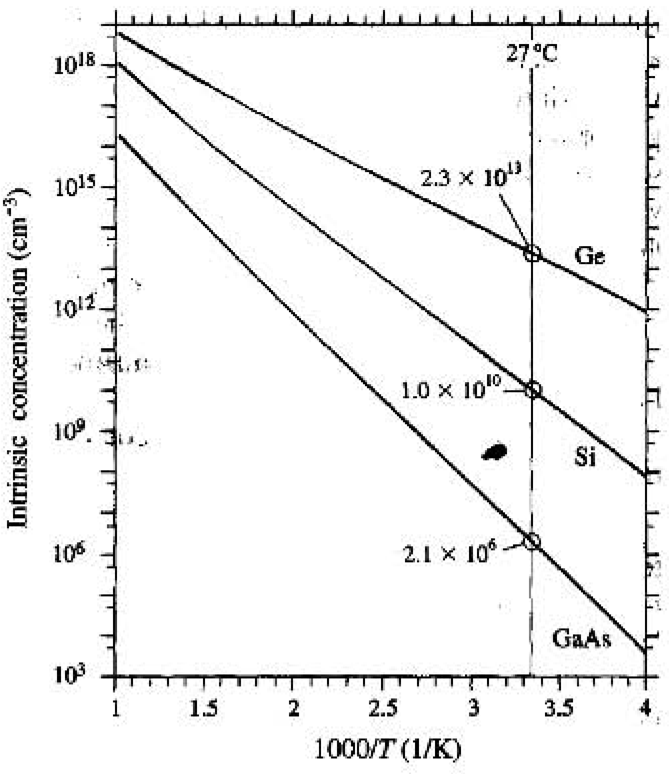


Figure 5.16 The temperature dependence of the intrinsic concentration.

Figure 5.16 displays the temperature dependence of the intrinsic concentration in Ge, Si, and GaAs as $\log(n_i)$ versus $1/T$ where the slope of the lines is, of course, a measure of the bandgap energy E_g. The $\log(n_i)$ versus $1/T$ graphs can be used to find, for example, whether the dopant concentration at a given temperature is more than the intrinsic concentration. As we will find out in Chapter 6, the reverse saturation current in a *pn* junction diode depends on n_i^2, so Figure 5.16 also indicates how this saturation current varies with temperature.

EXAMPLE 5.7

SATURATION AND INTRINSIC TEMPERATURES An *n*-type Si sample has been doped with 10^{15} phosphorus atoms cm^{-3}. The donor energy level for P in Si is 0.045 eV below the conduction band edge energy.

a. Estimate the temperature above which the sample behaves as if intrinsic.

b. Estimate the lowest temperature above which most of the donors are ionized.

SOLUTION

Remember that $n_i(T)$ is highly temperature dependent, as shown in Figure 5.16 so that as T increases, eventually at $T \approx T_i$, n_i becomes comparable to N_d. Beyond T_i, $n_i(T > T_i) \gg N_d$. Thus we need to solve

$$n_i(T_i) = N_d = 10^{15}\ \text{cm}^{-3}$$

From the $\log(n_i)$ versus $10^3/T$ graph for Si in Figure 5.16, when $n_i = 10^{15}\ \text{cm}^{-3}$, $(10^3/T_i) \approx 1.85$, giving $T_i \approx 541$ K or 268 °C.

We will assume that most of the donors are ionized, say at $T \approx T_s$, where the extrinsic and the extrapolated ionization lines intersect in Figure 5.15:

$$n = \left(\frac{1}{2}N_c N_d\right)^{1/2} \exp\left(-\frac{\Delta E}{2kT_s}\right) \approx N_d$$

This is the temperature at which the ionization behavior intersects the extrinsic region. In the above equation, $N_d = 10^{15}\ \text{cm}^{-3}$, $\Delta E = 0.045$ eV, and $N_c \propto T^{3/2}$, that is,

$$N_c(T_s) = N_c(300\ \text{K})\left(\frac{T_s}{300}\right)^{3/2}$$

Clearly, then, the equation can only be solved numerically. Similar equations occur in a wide range of physical problems where one term has the strongest temperature dependence. Here, $\exp(-\Delta E/kT_s)$ has the strongest temperature dependence. First assume N_c is that at 300 K, $N_c = 2.8 \times 10^{19}\ \text{cm}^{-3}$, and evaluate T_s,

$$T_s = \frac{\Delta E}{k \ln\left(\dfrac{N_c}{2N_d}\right)} = \frac{0.045\ \text{eV}}{(8.62 \times 10^{-5}\ \text{eV K}^{-1}) \ln\left[\dfrac{2.8 \times 10^{19}\ \text{cm}^{-3}}{2(1.0 \times 10^{15}\ \text{cm}^{-3})}\right]} = 54.7\ \text{K}$$

At $T = 54.7$ K,

$$N_c(54.7\ \text{K}) = N_c(300\ \text{K})\left(\frac{54.7}{300}\right)^{3/2} = 2.18 \times 10^{18}\ \text{cm}^{-3}$$

With this new N_c at a lower temperature, the improved T_s is 74.6 K. Since we only need an estimate of T_s, the extrinsic range of this semiconductor is therefore from about 75 to 541 K or −198 to about 268 °C.

EXAMPLE 5.8

TEMPERATURE DEPENDENCE OF THE ELECTRON CONCENTRATION By considering the mass action law, charge neutrality within the crystal, and occupation statistics of electronic states, we can show that at the lowest temperatures the electron concentration in an n-type semiconductor is given by

Electron concentration in the ionization region

$$n = \left(\frac{1}{2}N_c N_d\right)^{1/2} \exp\left(-\frac{\Delta E}{2kT}\right)$$

where $\Delta E = E_c - E_d$. Furthermore, at the lowest temperatures, the Fermi energy is midway between E_d and E_c.

There are only a few physical principles that must be considered to arrive at the effect of doping on the electron and hole concentrations. For an n-type semiconductor, these are

1. **Charge carrier statistics.**

$$n = N_c \exp\left[-\frac{(E_c - E_F)}{kT}\right] \tag{1}$$

2. **Mass action law.**

$$np = n_i^2 \tag{2}$$

3. **Electrical neutrality of the crystal.** We must have the same number of positive and negative charges:

$$p + N_d^+ = n \tag{3}$$

where N_d^+ is the concentration of *ionized* donors.

4. **Statistics of ionization of the dopants.**

$$N_d^+ = N_d \times \text{(probability of not finding an electron at } E_d) = N_d[1 - f_d(E_d)]$$

$$= \frac{N_d}{1 + 2\exp\left[\frac{(E_F - E_d)}{kT}\right]} \tag{4}$$

Solving Equations 1 to 4 for n will give the dependence of n on T and N_d. For example, from the mass action law, Equation 2, and the charge neutrality condition, Equation 3, we get

$$\frac{n_i^2}{n} + N_d^+ = n$$

This is a quadratic equation in n. Solving this equation gives

$$n = \frac{1}{2}(N_d^+) + \left[\frac{1}{4}(N_d^+)^2 + n_i^2\right]^{1/2}$$

Clearly, this equation should give the behavior of n as a function of T and N_d when we also consider the statistics in Equation 4. In the low-temperature region ($T < T_s$), n_i^2 is negligible in the expression for n and we have

$$n = N_d^+ = \frac{N_d}{1 + 2\exp\left[\frac{(E_F - E_d)}{kT}\right]} \approx \frac{1}{2}N_d \exp\left[-\frac{(E_F - E_d)}{kT}\right]$$

But the statistical description in Equation 1 is generally valid, so multiplying the low-temperature region equation by Equation 1 and taking the square root eliminates E_F from the expression, giving

$$n = \left(\frac{1}{2}N_cN_d\right)^{1/2} \exp\left[-\frac{(E_c - E_d)}{2kT}\right]$$

To find the location of the Fermi energy, consider the general expression

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

which must now correspond to n at low temperatures. Equating the two and rearranging to obtain E_F we find

$$E_F = \frac{E_c + E_d}{2} + \frac{1}{2}kT \ln\left(\frac{N_d}{2N_c}\right)$$

which puts the Fermi energy near the middle of $\Delta E = E_c - E_d$ at low temperatures.

5.3.2 DRIFT MOBILITY: TEMPERATURE AND IMPURITY DEPENDENCE

The temperature dependence of the drift mobility follows two distinctly different temperature variations. In the high-temperature region, it is observed that the drift mobility is limited by scattering from lattice vibrations. As the magnitude of atomic vibrations increases with temperature, the drift mobility decreases in the fashion $\mu \propto T^{-3/2}$. However, at low temperatures the lattice vibrations are not sufficiently strong to be the major limitation to the mobility of the electrons. It is observed that at low temperatures the scattering of electrons by ionized impurities is the major mobility limiting mechanism and $\mu \propto T^{3/2}$, as we will show below.

We recall from Chapter 2 that the electron drift mobility μ depends on the mean free time τ between scattering events via

$$\mu = \frac{e\tau}{m_e^*} \qquad [5.21]$$

in which

$$\tau = \frac{1}{Sv_{\text{th}}N_s} \qquad [5.22]$$

where S is the cross-sectional area of the scatterer; v_{th} is the mean speed of the electrons, called the **thermal velocity;** and N_s is the number of scatterers per unit volume. If a is the amplitude of the atomic vibrations about the equilibrium, then $S = \pi a^2$. As the temperature increases, so does the amplitude a of the lattice vibrations following $a^2 \propto T$ behavior, as shown in Chapter 2. An electron in the CB is free to wander around and therefore has only *KE*. We also know that the mean kinetic energy per electron in the CB is $\frac{3}{2}kT$, just as if the kinetic molecular theory could be applied to all those electrons in the CB. Therefore,

$$\frac{1}{2}m_e^*v_{\text{th}}^2 = \frac{3}{2}kT$$

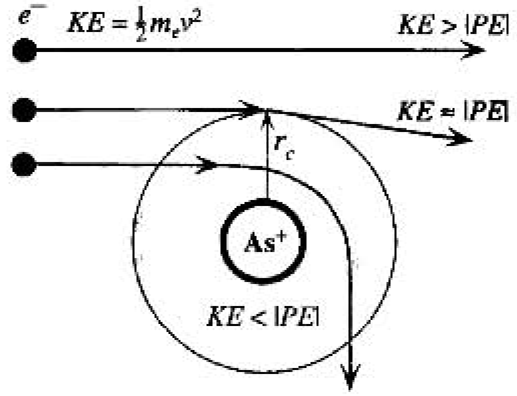


Figure 5.17 Scattering of electrons by an ionized impurity.

so that $v_{\mathrm{th}} \propto T^{1/2}$. Thus the mean time τ_L between scattering events from lattice vibrations is

$$\tau_L = \frac{1}{(\pi a^2)v_{\mathrm{th}}N_s} \propto \frac{1}{(T)(T^{1/2})} \propto T^{-3/2}$$

Lattice-scattering-limited mobility

which leads to a **lattice vibration scattering limited mobility,** denoted as μ_L, of the form

$$\mu_L \propto T^{-3/2} \qquad [5.23]$$

At low temperatures, scattering of electrons by thermal vibrations of the lattice will not be as strong as the electron scattering brought about by ionized donor impurities. As an electron passes by an ionized donor As$^+$, it is attracted and thus deflected from its straight path, as schematically shown in Figure 5.17. This type of scattering of an electron is what limits the drift mobility at low temperatures.

The *PE* of an electron at a distance r from an As$^+$ ion is due to the Coulombic attraction, and its magnitude is given by

$$|PE| = \frac{e^2}{4\pi\varepsilon_o\varepsilon_r r}$$

If the *KE* of the electron approaching an As$^+$ ion is larger than its *PE* at distance r from As$^+$, then the electron will essentially continue without feeling the *PE* and therefore without being deflected, and we can say that it has not been scattered. Effectively, due to its high *KE*, the electron does not feel the Coulombic pull of the donor. On the other hand, if the *KE* of the electron is less than its *PE* at r from As$^+$, then the *PE* of the Coulombic interaction will be so strong that the electron will be strongly deflected. This is illustrated in Figure 5.17. The critical radius r_c corresponds to the case when the electron is just scattered, which is when $KE \approx |PE(r_c)|$. But average $KE = \frac{3}{2}kT$, so at $r = r_c$

$$\frac{3}{2}kT = |PE(r_c)| = \frac{e^2}{4\pi\varepsilon_o\varepsilon_r r_c}$$

from which $r_c = e^2/(6\pi\varepsilon_o\varepsilon_r kT)$. As the temperature increases, the scattering radius decreases. The scattering cross section $S = \pi r_c^2$ is thus given by

$$S = \frac{\pi e^4}{(6\pi\varepsilon_o\varepsilon_r kT)^2} \propto T^{-2}$$

Incorporating $v_{\text{th}} \propto T^{1/2}$ as well, the temperature dependence of the mean scattering time τ_I between impurities, from Equation 5.22, must be

$$\tau_I = \frac{1}{Sv_{\text{th}}N_I} \propto \frac{1}{(T^{-2})(T^{1/2})N_I} \propto \frac{T^{3/2}}{N_I}$$

where N_I is the concentration of ionized impurities (all ionized impurities including donors and acceptors). Consequently, the **ionized impurity scattering limited mobility** from Equation 5.21 is

Ionized impurity scattering limited mobility

$$\mu_I \propto \frac{T^{3/2}}{N_I} \qquad [5.24]$$

Note also that μ_I decreases with increasing ionized dopant concentration N_I, which itself may be temperature dependent. Indeed, at the lowest temperatures, below the saturation temperature T_s, N_I will be strongly temperature dependent because not all the donors would have been fully ionized.

The overall temperature dependence of the drift mobility is then, simply, the reciprocal additions of the μ_I and μ_L by virtue of Matthiessen's rule, that is,

Effective mobility

$$\frac{1}{\mu_e} = \frac{1}{\mu_I} + \frac{1}{\mu_L} \qquad [5.25]$$

so the scattering process having the lowest mobility determines the overall (effective) drift mobility.

The experimental temperature dependence of the electron drift mobility in both Ge and Si is shown in Figure 5.18 as a log–log plot for various donor concentrations. The slope on this plot corresponds to the index n in $\mu_e \propto T^n$. The simple theoretical sketches in the insets show how μ_L and μ_I from Equations 5.23 and 5.24 depend on the temperature. For Ge, at low doping concentrations (*e.g.*, $N_d = 10^{13}\ \text{cm}^{-3}$), the experiments indicate a $\mu_e \propto T^{-1.5}$ type of behavior, which is in agreement with μ_e determined by μ_L in Equation 5.23. Curves for Si at low-level doping (μ_I negligible)

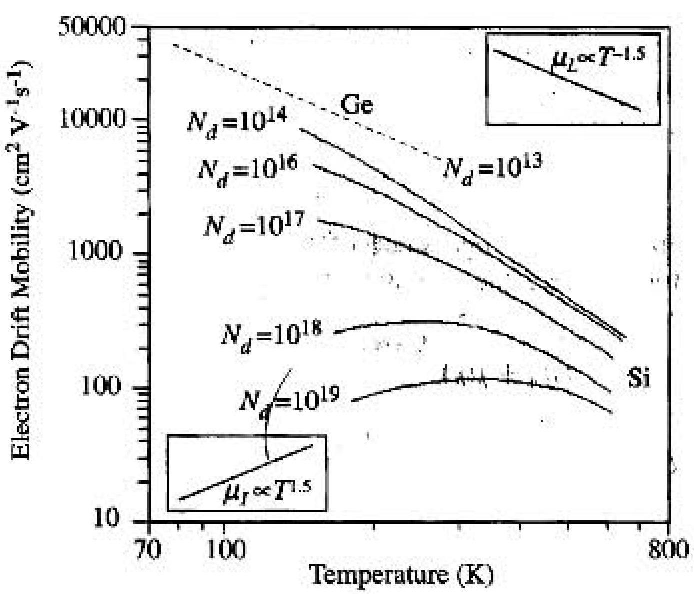


Figure 5.18 Log–log plot of drift mobility versus temperature for *n*-type Ge and *n*-type Si samples.
Various donor concentrations for Si are shown. N_d are in cm^{-3}. The upper right inset is the simple theory for lattice limited mobility, whereas the lower left inset is the simple theory for impurity scattering limited mobility.

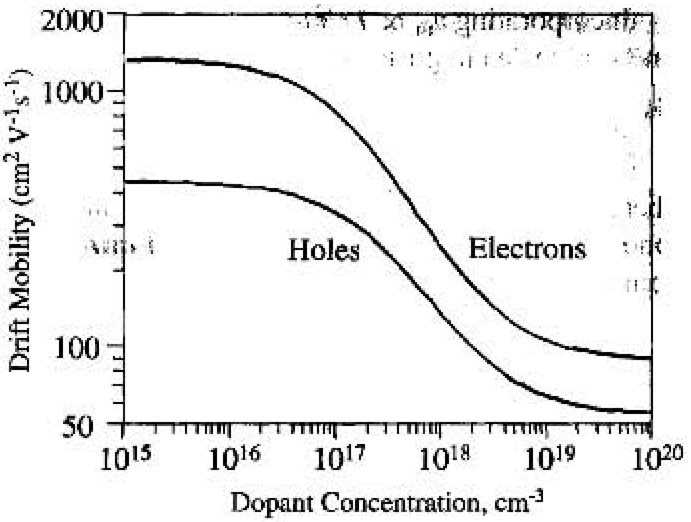


Figure 5.19 The variation of the drift mobility with dopant concentration in Si for electrons and holes at 300 K.

at high temperatures, however, exhibit a $\mu_e \propto T^{-2.5}$ type of behavior rather than $T^{-1.5}$, which can be accounted for in a more rigorous theory. As the donor concentration increases, the drift mobility decreases by virtue of μ_I getting smaller. At the highest doping concentrations and at low temperatures, the electron drift mobility in Si exhibits almost a $\mu_e \propto T^{3/2}$ type of behavior. Similar arguments can be extended to the temperature dependence of the hole drift mobility.

The dependences of the room temperature electron and hole drift mobilities on the dopant concentration for Si are shown in Figure 5.19 where, as expected, past a certain amount of impurity addition, the drift mobility is overwhelmingly controlled by μ_I in Equation 5.25.

5.3.3 CONDUCTIVITY TEMPERATURE DEPENDENCE

The conductivity of an extrinsic semiconductor doped with donors depends on the electron concentration and the drift mobility, both of which have been determined above. At the lowest temperatures in the ionization range, the electron concentration depends exponentially on the temperature by virtue of

Electron concentration in ionization region

$$n = \left(\frac{1}{2}N_c N_d\right)^{1/2} \exp\left[-\frac{(E_c - E_d)}{2kT}\right]$$

which then also dominates the temperature dependence of the conductivity. In the intrinsic range at the highest temperatures, the conductivity is dominated by the temperature dependence of n_i since

$$\sigma = en_i(\mu_e + \mu_h)$$

and n_i is an exponential function of temperature in contrast to $\mu \propto T^{-3/2}$. In the extrinsic temperature range, $n = N_d$ and is constant, so the conductivity follows the temperature dependence of the drift mobility. Figure 5.20 shows schematically the semilogarithmic plot of the conductivity against the reciprocal temperature where through the extrinsic range σ exhibits a broad "S" due to the temperature dependence of the drift mobility.

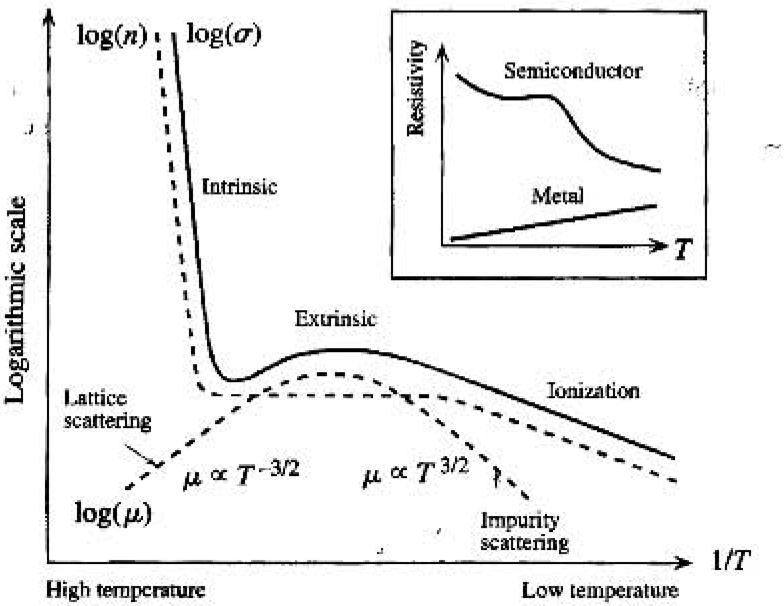


Figure 5.20 Schematic illustration of the temperature dependence of electrical conductivity for a doped (*n*-type) semiconductor.

COMPENSATION-DOPED Si

EXAMPLE 5.9

a. A Si sample has been doped with 10^{17} arsenic atoms cm^{-3}. Calculate the conductivity of the sample at 27 °C (300 K) and at 127 °C (400 K).

b. The above *n*-type Si sample is further doped with 9×10^{16} boron atoms cm^{-3}. Calculate the conductivity of the sample at 27 °C and 127 °C.

SOLUTION

a. The arsenic dopant concentration, $N_d = 10^{17}\ \text{cm}^{-3}$, is much larger than the intrinsic concentration n_i, which means that $n = N_d$ and $p = (n_i^2/N_d) \ll n$ and can be neglected. Thus $n = 10^{17}\ \text{cm}^{-3}$ and the electron drift mobility at $N_d = 10^{17}\ \text{cm}^{-3}$ is 800 $\text{cm}^2\ \text{V}^{-1}\ \text{s}^{-1}$ from the drift mobility versus dopant concentration graph in Figure 5.19, so

$$\sigma = en\mu_e + ep\mu_h = eN_d\mu_e$$

$$= (1.6 \times 10^{-19}\ \text{C})(10^{17}\ \text{cm}^{-3})(800\ \text{cm}^2\ \text{V}^{-1}\,\text{s}^{-1}) = 12.8\ \Omega^{-1}\ \text{cm}^{-1}$$

At $T = 127\ ^\circ\text{C} = 400\ \text{K}$,

$$\mu_e \approx 420\ \text{cm}^2\ \text{V}^{-1}\ \text{s}^{-1}$$

(from the μ_e versus T graph in Figure 5.18). Thus

$$\sigma = eN_d\mu_e = 6.72\ \Omega^{-1}\ \text{cm}^{-1}$$

b. With further doping we have $N_a = 9 \times 10^{16}\ \text{cm}^{-3}$, so from the compensation effect

$$N_d - N_a = 1 \times 10^{17} - 9 \times 10^{16} = 10^{16}\ \text{cm}^{-3}$$

Since $N_d - N_a \gg n_i$, we have an *n*-type material with $n = N_d - N_a = 10^{16}\ \text{cm}^{-3}$. But the drift mobility now is about $\sim 600\ \text{cm}^2\ \text{V}^{-1}\ \text{s}^{-1}$ because, even though $N_d - N_a$ is now $10^{16}\ \text{cm}^{-3}$ and not $10^{17}\ \text{cm}^{-3}$, all the donors and acceptors are still ionized and hence still scatter the charge carriers. The recombination of electrons from the donors and holes from the acceptors does not alter the fact that at room temperature all the dopants will be ionized.

Effectively, the compensation effect is as if all electrons from the donors were being accepted by the acceptors. Although with compensation doping the net electron concentration is $n = N_d - N_a$, the drift mobility scattering is determined by $(N_d + N_a)$, which in this case is $10^{17} + 9 \times 10^{16}\ \text{cm}^{-3} = 1.9 \times 10^{17}\ \text{cm}^{-3}$, which gives an electron drift mobility of ~600 cm^2 V^{-1} s^{-1} at 300 K and ~400 cm^2 V^{-1} s^{-1} at 400 K. Then, neglecting the hole concentration $p = n_i^2/(N_d - N_a)$, we have

At 300 K, $\sigma = e(N_d - N_a)\mu_e \approx (1.6 \times 10^{-19}\ \text{C})(10^{16}\ \text{cm}^{-3})(600\ \text{cm}^2\ \text{V}^{-1}\ \text{s}^{-1})$

$= 0.96\ \Omega^{-1}\ \text{cm}^{-1}$

At 400 K, $\sigma = e(N_d - N_a)\mu_e \approx (1.6 \times 10^{-19}\ \text{C})(10^{16}\ \text{cm}^{-3})(400\ \text{cm}^2\ \text{V}^{-1}\ \text{s}^{-1})$

$= 0.64\ \Omega^{-1}\ \text{cm}^{-1}$

5.3.4 DEGENERATE AND NONDEGENERATE SEMICONDUCTORS

The general exponential expression for the concentration of electron in the CB,

$$n \approx N_c \exp\left[-\frac{(E_c - E_F)}{kT}\right] \qquad [5.26]$$

is based on replacing Fermi–Dirac statistics with Boltzmann statistics, which is only valid when E_c is several kT above E_F. In other words, we assumed that the number of states in the CB far exceeds the number of electrons there, so the likelihood of two electrons trying to occupy the same state is almost nil. This means that the Pauli exclusion principle can be neglected and the electron statistics can be described by the Boltzmann statistics. N_c is a measure of the density of states in the CB. The Boltzmann expression for n is valid only when $n \ll N_c$. Those semiconductors for which $n \ll N_c$ and $p \ll N_v$ are termed **nondegenerate semiconductors.** They essentially follow all the discussions above and exhibit all the normal semiconductor properties outlined above.

When the semiconductor has been excessively doped with donors, then n may be so large, typically 10^{19}–10^{20} cm^{-3}, that it may be comparable to or greater than N_c. In that case the Pauli exclusion principle becomes important in the electron statistics and we have to use the Fermi–Dirac statistics. Equation 5.26 for n is then no longer valid. Such a semiconductor exhibits properties that are more metal-like than semiconductor-like; for example, the resistivity follows $\rho \propto T$. Semiconductors that have $n > N_c$ or $p > N_v$ are called **degenerate semiconductors.**

The large carrier concentration in a degenerate semiconductor is due to its heavy doping. For example, as the donor concentration in an n-type semiconductor is increased, at sufficiently high doping levels, the donor atoms become so close to each other that their orbitals overlap to form a narrow energy band that overlaps and becomes part of the conduction band. E_c is therefore slightly shifted down and E_g becomes slightly narrower. The valence electrons from the donors fill the band from E_c. This situation is reminiscent of the valence electrons filling overlapping energy bands in a metal. In a degenerate n-type semiconductor, the Fermi level is therefore within the CB, or above E_c just like E_F is within the band in a metal. The

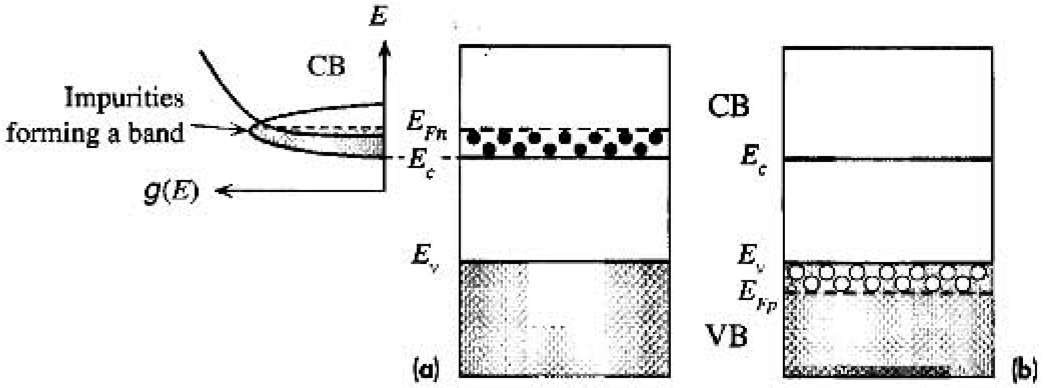


Figure 5.21

(a) Degenerate *n*-type semiconductor. Large number of donors form a band that overlaps the CB.

(b) Degenerate *p*-type semiconductor.

majority of the states between E_c and E_F are full of electrons as indicated in Figure 5.21. In the case of a *p*-type degenerate semiconductor, the Fermi level lies in the VB below E_v. It should be emphasized that one cannot simply assume that $n = N_d$ or $p = N_a$ in a degenerate semiconductor because the dopant concentration is so large that they interact with each other. Not all dopants are able to become ionized, and the carrier concentration eventually reaches a saturation typically around $\sim 10^{20}$ cm^{-3}. Furthermore, the mass action law $np = n_i^2$ is not valid for degenerate semiconductors.

Degenerate semiconductors have many important uses. For example, they are used in laser diodes, zener diodes, and ohmic contacts in ICs, and as metal gates in many microelectronic MOS devices.

5.4 RECOMBINATION AND MINORITY CARRIER INJECTION

5.4.1 DIRECT AND INDIRECT RECOMBINATION

Above absolute zero of temperature, the thermal excitation of electrons from the VB to the CB continuously generates free electron–hole pairs. It should be apparent that in equilibrium there should be some annihilation mechanism that returns the electron from the CB down to an empty state (a hole) in the VB. When a free electron, wandering around in the CB of a crystal, "meets" a hole, it falls into this low-energy empty electronic state and fills it. This process is called **recombination.** Intuitively, recombination corresponds to the free electron finding an incomplete bond with a missing electron. The electron then enters and completes this bond. The free electron in the CB and the free hole in the VB are consequently annihilated. On the energy band diagram, the recombination process is represented by returning the electron from the CB (where it is free) into a hole in the VB (where it is in a bond). Figure 5.22

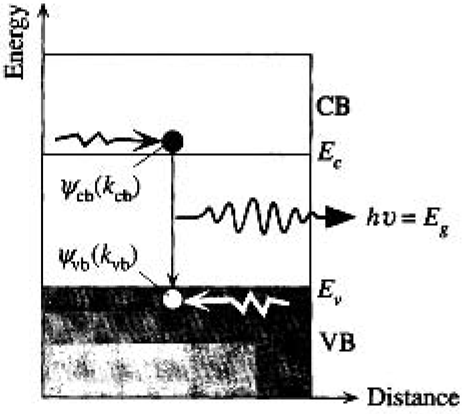
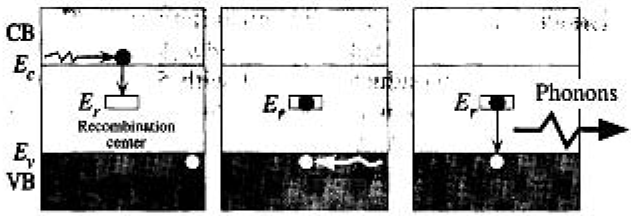


Figure 5.22 Direct recombination in GaAs. $k_{cb} = k_{vb}$ so that momentum conservation is satisfied.

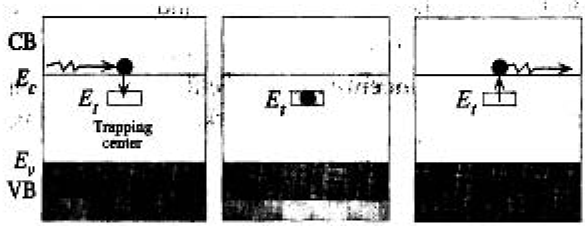
shows a direct recombination mechanism, for example, as it occurs in GaAs, in which a free electron recombines with a free hole when they meet at one location in the crystal. The excess energy of the electron is lost as a photon of energy $h\nu = E_g$. In fact, it is this type of recombination that results in the emitted light from light emitting diodes (LEDs).

The recombination process between an electron and a hole, like every other process in nature, must obey the momentum conservation law. The wavefunction of an electron in the CB, $\psi_{cb}(k_{cb})$, will have a certain momentum $\hbar k_{cb}$ associated with the wavevector k_{cb} and, similarly, the electron wavefunction $\psi_{vb}(k_{vb})$ in the VB will have a momentum $\hbar k_{vb}$ associated with the wavevector k_{vb}. Conservation of linear momentum during recombination requires that when the electron drops from the CB to the VB, its wavevector should remain the same, $k_{vb} = k_{cb}$. For the elemental semiconductors, Si and Ge, the electronic states $\psi_{vb}(k_{vb})$ with $k_{vb} = k_{cb}$ are right in the middle of the VB and are therefore fully occupied. Consequently, there are no empty states in the VB that can satisfy $k_{vb} = k_{cb}$, and so direct recombination in Si and Ge is next to impossible. For some compound semiconductors, such as GaAs and InSb, for example, the states with $k_{vb} = k_{cb}$ are right at the top of the valence band, so they are essentially empty (contain holes). Consequently, an electron in the CB of GaAs can drop down to an empty electronic state at the top of the VB and maintain $k_{vb} = k_{cb}$. Thus **direct recombination** is highly probable in GaAs, and it is this very reason that makes GaAs an LED material.

In elemental semiconductor crystals, for example, in Si and Ge, electrons and holes usually recombine through recombination centers. A recombination center increases the probability of recombination because it can "take up" any momentum difference between a hole and electron. The process essentially involves a third body, which may be an impurity atom or a crystal defect. The electron is captured by the recombination center and thus becomes localized at this site. It is "held" at the center until some hole arrives and recombines with it. In the energy band diagram picture shown in Figure 5.23a, the recombination center provides a localized electronic state below E_c in the bandgap, which is at a certain location in the crystal. When an electron approaches the center, it is captured. The electron is then localized and bound to this



(a) Recombination



(b) Trapping

Figure 5.23 Recombination and trapping.

(a) Recombination in Si via a recombination center that has a localized energy level at E_r in the bandgap, usually near the middle.

(b) Trapping and detrapping of electrons by trapping centers. A trapping center has a localized energy level in the bandgap.

center and "waits" there for a hole with which it can recombine. In this recombination process, the energy of the electron is usually lost to lattice vibrations (as "sound") via the "recoiling" of the third body. Emitted lattice vibrations are called phonons. A **phonon** is a quantum of energy associated with atomic vibrations in the crystal analogous to the photon.

Typical recombination centers, besides the donor and acceptor impurities, might be metallic impurities and crystal defects such as dislocations, vacancies, or interstitials. Each has its own peculiar behavior in aiding recombination, which will not be described here.

It is instructive to mention briefly the phenomenon of charge carrier **trapping** since in many devices this can be the main limiting factor on the performance. An electron in the conduction band can be captured by a localized state, just like a recombination center, located in the bandgap, as shown in Figure 5.23b. The electron falls into the trapping center at E_t and becomes temporarily removed from the CB. At a later time, due to an incident energetic lattice vibration, it becomes excited back into the CB and is available for conduction again. Thus trapping involves the temporary removal of the electron from the CB, whereas in the case of recombination, the electron is permanently removed from the CB since the capture is followed by recombination with a hole. We can view a trap as essentially being a flaw in the crystal that results in the creation of a localized electronic state, around the flaw site, with an energy in the bandgap. A charge carrier passing by the flaw can be captured and lose its freedom. The flaw can be an impurity or a crystal imperfection in the same way as a recombination center. The only difference is that when a charge carrier is captured at a recombination site, it has no possibility of escaping again because the center aids recombination. Although Figure 5.23b illustrates an electron trap, similar arguments also apply to hole traps, which are normally closer to E_v. In general, flaws and defects that give localized states near the middle of the bandgap tend to act as recombination centers.

5.4.2 MINORITY CARRIER LIFETIME

Consider what happens when an n-type semiconductor, doped with 5×10^{16} cm^{-3} donors, is uniformly illuminated with appropriate wavelength light to photogenerate electron–hole pairs (EHPs), as shown in Figure 5.24. We will now define thermal equilibrium majority and minority carrier concentrations in an extrinsic semiconductor. In general, the subscript n or p is used to denote the type of semiconductor, and o to refer to thermal equilibrium in the dark.

In an n-type semiconductor, electrons are the majority carriers and holes are the minority carriers

> n_{no} is defined as the **majority carrier concentration** (electron concentration in an n-type semiconductor) in thermal equilibrium in the dark. These electrons, constituting the majority carriers, are thermally ionized from the donors.
>
> p_{no} is termed the **minority carrier concentration** (hole concentration in an n-type semiconductor) in thermal equilibrium in the dark. These holes that constitute the minority carriers are thermally generated across the bandgap.

In both cases the subscript no refers to an n-type semiconductor and thermal equilibrium conditions, respectively. Thermal equilibrium means that the mass action law is obeyed and $n_{no}p_{no} = n_i^2$.

When we illuminate the semiconductor, we create *excess* EHPs by photogeneration. Suppose that the electron and hole concentrations at any instant are denoted by n_n and p_n, which are defined as the *instantaneous* majority (electron) and minority (hole) concentrations, respectively. At any instant and at any location in the semiconductor, we define the departure from the equilibrium by **excess concentrations** as follows:

> Δn_n is the *excess* electron (majority carrier) concentration: $\Delta n_n = n_n - n_{no}$
>
> Δp_n is the *excess* hole (minority carrier) concentration: $\Delta p_n = p_n - p_{no}$

Under illumination, at any instant, therefore

$$n_n = n_{no} + \Delta n_n \qquad \text{and} \qquad p_n = p_{no} + \Delta p_n$$

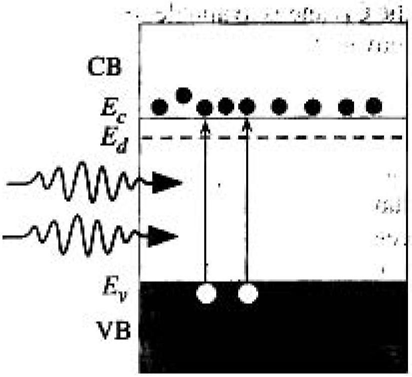


Figure 5.24 Low-level photoinjection into an n-type semiconductor in which $\Delta n_n < n_{no}$.

Photoexcitation creates EHPs or an equal number of electrons and holes, as shown in Figure 5.24, which means that

$$\Delta p_n = \Delta n_n$$

and obviously the mass action law is not obeyed: $n_n p_n \neq n_i^2$. It is worth remembering that

$$\frac{dn_n}{dt} = \frac{d\Delta n_n}{dt} \quad \text{and} \quad \frac{dp_n}{dt} = \frac{d\Delta p_n}{dt}$$

since n_{no} and p_{no} depend only on temperature.

Let us assume that we have "weak" illumination, which causes, say, only a 10 percent change in n_{no}, that is,

$$\Delta n_n = 0.1 n_{no} = 0.5 \times 10^{16}\,\text{cm}^{-3}$$

Then

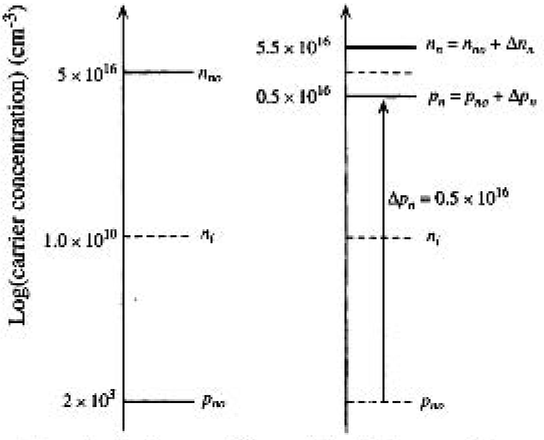
$$\Delta p_n = \Delta n_n = 0.5 \times 10^{16}\,\text{cm}^{-3}$$

Figure 5.25 shows a single-axis plot of the majority (n_n) and minority (p_n) concentrations in the dark and in light. The scale is logarithmic to allow large orders of magnitude changes to be recorded. Under illumination, the minority carrier concentration is

$$p_n = p_{no} + \Delta p_n = 2.0 \times 10^3 + 0.5 \times 10^{16} \approx 0.5 \times 10^{16} = \Delta p_n$$

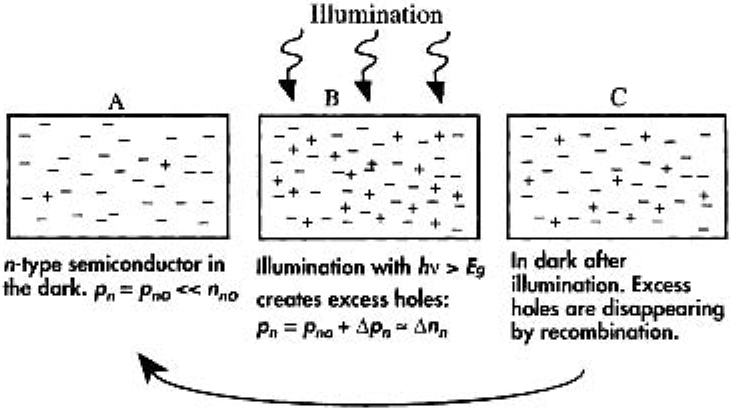
That is, $p_n \approx \Delta p_n$, which shows that although n_n changes by only 10 percent, p_n changes *drastically*, that is, by a factor of ~10^{12}.

Figure 5.26 shows a pictorial view of what is happening inside an n-type semiconductor when light is switched on at a certain time and then later switched off again. Obviously when the light is switched off, the condition $p_n = \Delta p_n$ (state B in Figure 5.26) must eventually revert back to the dark case (state A) where $p_n = p_{no}$. In other words, the excess minority carriers Δp_n and excess majority carriers Δn_n must



(a) In the dark: $np = n_i^2$ **(b) In light: $np \neq n_i^2$**

Figure 5.25 Low-level injection in an n-type semiconductor does not significantly affect n_n but drastically affects the minority carrier concentration p_n.



n-type semiconductor in the dark. $p_n = p_{no} \ll n_{no}$

Illumination with $h\nu > E_g$ creates excess holes: $p_n = p_{no} + \Delta p_n \approx \Delta n_n$

In dark after illumination. Excess holes are disappearing by recombination.

Figure 5.26 Illumination of an *n*-type semiconductor results in excess electron and hole concentrations.

After the illumination, the recombination process restores equilibrium; the excess electrons and holes simply recombine.

be removed. This removal occurs by recombination. Excess holes recombine with the electrons available and disappear. This, however, takes time because the electrons and holes have to find each other. In order to describe the rate of recombination, we introduce a temporal quantity, denoted by τ_h and called the **minority carrier lifetime (mean recombination time),** which is defined as follows: τ_h is the average time a hole exists in the VB from its generation to its recombination, that is, the mean time the hole is free before recombining with an electron. An alternative and equivalent definition is that $1/\tau_h$ is the average probability per unit time that a hole will recombine with an electron. We must remember that the recombination process occurs through recombination centers, so the recombination time τ_h will depend on the concentration of these centers and their effectiveness in capturing the minority carriers. Once a minority carrier has been captured by a recombination center, there are many majority carriers available to recombine with it, so τ_h in an indirect process is independent of the majority carrier concentration. This is the reason for defining the recombination time as a minority carrier lifetime.

If the minority carrier recombination time is, say, 10 s, and if there are some 1000 excess holes, then it is clear that these excess holes will be disappearing at a rate of $1000/10 \text{ s} = 100$ per second. The rate of recombination of excess minority carriers is simply $\Delta p_n/\tau_h$. At any instant, therefore,

$$\begin{array}{c} \text{Rate of increase in excess} \\ \text{hole concentration} \end{array} = \begin{array}{c} \text{Rate of} \\ \text{photogeneration} \end{array} - \begin{array}{c} \text{Rate of recombination} \\ \text{of excess holes} \end{array}$$

If G_{ph} is the rate of photogeneration, then clearly the net rate of change of Δp_n is

Excess minority carrier concentration

$$\frac{d\Delta p_n}{dt} = G_{\text{ph}} - \frac{\Delta p_n}{\tau_h} \qquad [5.27]$$

This is a general expression that describes the time evolution of the excess minority carrier concentration given the photogeneration rate G_{ph}, the minority carrier lifetime τ_h, and the initial condition at $t = 0$. The only assumption is weak injection ($\Delta p_n < n_{no}$).

We should note that the recombination time τ_h depends on the semiconductor material, impurities, crystal defects, temperature, and so forth, and there is no typical value to quote. It can be anywhere from nanoseconds to seconds. Later it will be shown that certain applications require a short τ_h, as in fast switching of *pn* junctions, whereas others require a long τ_h, for example, persistent luminescence.

EXAMPLE 5.10

PHOTORESPONSE TIME Sketch the hole concentration when a step illumination is applied to an *n*-type semiconductor at time $t = 0$ and switched off at time $t = t_{off} (\gg \tau_h)$.

SOLUTION

We use Equation 5.27 with G_{ph} = constant in $0 \le t \le t_{off}$. Since Equation 5.27 is a first-order differential equation, integrating it we simply find

$$\ln\left[G_{ph} - \left(\frac{\Delta p_n}{\tau_h}\right)\right] = -\frac{t}{\tau_h} + C_1$$

where C_1 is the integration constant. At $t = 0$, $\Delta p_n = 0$, so $C_1 = \ln G_{ph}$. Therefore the solution is

$$\Delta p_n(t) = \tau_h G_{ph}\left[1 - \exp\left(-\frac{t}{\tau_h}\right)\right] \qquad 0 \le t < t_{off} \qquad [5.28]$$

We see that as soon as the illumination is turned on, the minority carrier concentration rises exponentially toward its steady-state value $\Delta p_n(\infty) = \tau_h G_{ph}$. This is reached after a time $t > \tau_h$.

At the instant the illumination is switched off, we assume that $t_{off} \gg \tau_h$ so that from Equation 5.28,

$$\Delta p_n(t_{off}) = \tau_h G_{ph}$$

We can define t' to be the time measured from $t = t_{off}$, that is, $t' = t - t_{off}$. Then

$$\Delta p_n(t' = 0) = \tau_h G_{ph}$$

Solving Equation 5.27 with $G_{ph} = 0$ in $t > t_{off}$ or $t' > 0$, we get

$$\Delta p_n(t') = \Delta p_n(0)\exp\left(-\frac{t'}{\tau_h}\right)$$

where $\Delta p_n(0)$ is actually an integration constant that is equivalent to the boundary condition on Δp_n at $t' = 0$. Putting $t' = 0$ and $\Delta p_n = \tau_h G_{ph}$ gives

$$\Delta p_n(t') = \tau_h G_{ph}\exp\left(-\frac{t'}{\tau_h}\right) \qquad [5.29]$$

We see that the excess minority carrier concentration decays exponentially from the instant the light is switched off with a time constant equal to the minority carrier recombination time. The time evolution of the minority carrier concentration is sketched in Figure 5.27.

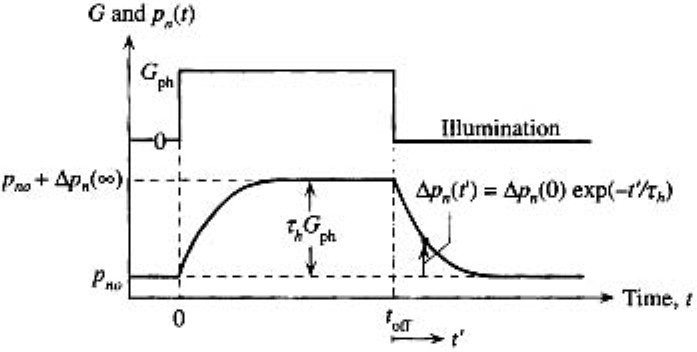


Figure 5.27 Illumination is switched on at time $t = 0$ and then off at $t = t_{off}$.
The excess minority carrier concentration $\Delta p_n(t)$ rises exponentially to its steady-state value with a time constant τ_h. From t_{off}, the excess minority carrier concentration decays exponentially to its equilibrium value.

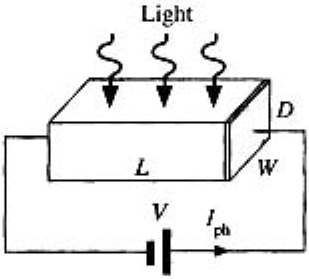


Figure 5.28 A semiconductor slab of length L, width W, and depth D is illuminated with light of wavelength λ. I_{ph} is the steady-state photocurrent.

EXAMPLE 5.11

PHOTOCONDUCTIVITY Suppose that a direct bandgap semiconductor with no traps is illuminated with light of intensity $I(\lambda)$ and wavelength λ that will cause photogeneration as shown in Figure 5.28. The area of illumination is $A = (L \times W)$, and the thickness (depth) of the semiconductor is D. If η is the **quantum efficiency** (number of free EHPs generated per absorbed photon) and τ is the recombination lifetime of the photogenerated carriers, show that the **steady-state photoconductivity,** defined as

$$\Delta\sigma = \sigma(\text{in light}) - \sigma(\text{in dark})$$

is given by

Steady-state photo-conductivity

$$\Delta\sigma = \frac{e\eta I\lambda\tau(\mu_e + \mu_h)}{hcD} \qquad [5.30]$$

A photoconductive cell has a CdS crystal 1 mm long, 1 mm wide, and 0.1 mm thick with electrical contacts at the end, so the receiving area of radiation is 1 mm^2, whereas the area of each contact is 0.1 mm^2. The cell is illuminated with a blue radiation of wavelength 450 nm and intensity 1 mW/cm^2. For unity quantum efficiency and an electron recombination time of 1 ms, calculate

a. The number of EHPs generated per second

b. The photoconductivity of the sample

c. The photocurrent produced if 50 V is applied to the sample

Note that a CdS photoconductor is a direct bandgap semiconductor with an energy gap $E_g = 2.6$ eV, electron mobility $\mu_e = 0.034$ m^2 V^{-1} s^{-1}, and hole mobility $\mu_h = 0.0018$ m^2 V^{-1} s^{-1}.

SOLUTION

If Γ_{ph} is the number of photons arriving per unit area per unit second (the photon flux), then $\Gamma_{ph} = I/h\nu$ where I is the light intensity (energy flowing per unit area per second) and $h\nu$ is the energy per photon. The quantum efficiency η is defined as the number of free EHPs generated per absorbed photon. Thus, the number of EHPs generated *per unit volume per second*, the photogeneration rate per unit volume G_{ph} is given by

$$G_{ph} = \frac{\eta A \Gamma_{ph}}{AD} = \frac{\eta\left(\dfrac{I}{h\nu}\right)}{D} = \frac{\eta I \lambda}{hcD}$$

In the steady state,

$$\frac{d\Delta n}{dt} = G_{ph} - \frac{\Delta n}{\tau} = 0$$

so

$$\Delta n = \tau G_{ph} = \frac{\tau \eta I \lambda}{hcD}$$

But, by definition,

$$\Delta\sigma = e\mu_e\,\Delta n + e\mu_h\,\Delta p = e\,\Delta n(\mu_e + \mu_h)$$

since electrons and holes are generated in pairs, $\Delta n = \Delta p$. Thus, substituting for Δn in the $\Delta\sigma$ expression, we get Equation 5.30:

$$\Delta\sigma = \frac{e\eta I \lambda \tau(\mu_e + \mu_h)}{hcD}$$

a. The photogeneration rate per unit time is not G_{ph}, which is per unit time per unit volume. We define EHP_{ph} as the total number of EHPs photogenerated per unit time in the whole volume (AD). Thus

$$\begin{aligned}
\mathrm{EHP}_{ph} &= \text{Total photogeneration rate} \\
&= (AD)G_{ph} = (AD)\frac{\eta I \lambda}{hcD} = \frac{A\eta I\lambda}{hc} \\
&= [(10^{-3} \times 10^{-3}\ \mathrm{m}^2)(1)(10^{-3} \times 10^4\ \mathrm{J\,s^{-1}\,m^{-2}})(450 \times 10^{-9}\ \mathrm{m})] \\
&\quad \div [(6.63 \times 10^{-34}\ \mathrm{J\,s})(3 \times 10^8\ \mathrm{m\,s^{-1}})] \\
&= 2.26 \times 10^{13}\ \mathrm{EHP\,s^{-1}}
\end{aligned}$$

b. From Equation 5.30,

$$\Delta\sigma = \frac{e\eta I \lambda \tau(\mu_e + \mu_h)}{hcD}$$

That is

$$\Delta\sigma = \frac{(1.6 \times 10^{-19}\ \mathrm{C})(1)(10^{-3} \times 10^4\ \mathrm{J\,s^{-1}\,m^{-2}})(450 \times 10^{-9}\ \mathrm{m})(1 \times 10^{-3}\ \mathrm{s})(0.0358\ \mathrm{m^2\,V^{-1}\,s^{-1}})}{(6.63 \times 10^{-34}\ \mathrm{J\,s})(3 \times 10^8\ \mathrm{ms^{-1}})(0.1 \times 10^{-3}\ \mathrm{m})}$$

$$= 1.30\ \Omega^{-1}\mathrm{m}^{-1}$$

c. Photocurrent density will be

$$\Delta J = \mathcal{E}\,\Delta\sigma = (1.30\ \Omega^{-1}\,\mathrm{m}^{-1})(50\ \mathrm{V}/10^{-3}\ \mathrm{m}) = 6.50 \times 10^4\ \mathrm{A\,m^{-2}}$$

Thus the photocurrent

$$\Delta I = A\,\Delta J = (10^{-3} \times 0.1 \times 10^{-3}\ \text{m}^2)(6.50 \times 10^4\ \text{A m}^{-2})$$
$$= 6.5 \times 10^{-3}\ \text{A} \quad \text{or} \quad 6.5\ \text{mA}$$

We assumed that all the incident radiation is absorbed.

5.5 DIFFUSION AND CONDUCTION EQUATIONS, AND RANDOM MOTION

It is well known that, by virtue of their random motion, gas particles diffuse from high-concentration regions to low-concentration regions. When a perfume bottle is opened at one end of a room, the molecules diffuse out from the bottle and, after a while, can be smelled at the other end of the room. Whenever there is a concentration gradient of particles, there is a net diffusional motion of particles in the direction of decreasing concentration. The origin of diffusion lies in the random motion of particles. To quantify particle flow, we define the **particle flux** Γ just like current, as the number of particles (not charges) crossing unit area per unit time. Thus if ΔN particles cross an area A in time Δt, then, by definition, the particle flux is

Definition of particle flux

$$\Gamma = \frac{\Delta N}{A\,\Delta t} \tag{5.31}$$

Clearly if the particles are charged with a charge Q ($-e$ for electrons and $+e$ for holes), then the electric current density J, which is basically a charge flux, is related to the particle flux Γ by

Definition of current density

$$J = Q\Gamma \tag{5.32}$$

Suppose that the electron concentration at some time t in a semiconductor decreases in the x direction and has the profile $n(x, t)$ shown in Figure 5.29a. This may have been achieved, for example, by photogeneration at one end of a semiconductor. We will assume that the electron concentration changes only in the x direction so that the diffusion of electrons can be simplified to a one-dimensional problem as depicted in Figure 5.29a. We know that in the absence of an electric field, the electron motion is random and involves scattering from lattice vibrations and impurities. Suppose that ℓ is the mean free path in the x direction and τ is the mean free time between the scattering events. The electron moves a mean distance ℓ in the $+x$ or $-x$ direction and then it is scattered and changes direction. Its mean speed along x is $v_x = \ell/\tau$. Let us evaluate the flow of electrons in the $+x$ and $-x$ directions through the plane at x_o and hence find the net flow in the $+x$ direction.

We can divide the x axis into hypothetical segments of length ℓ so that each segment corresponds to a mean free path. Going across a segment, the electron experiences one scattering process. Consider what happens during one mean free time, the time it takes for the electrons to move across a segment toward the left or right. Half of the electrons in $(x_o - \ell)$ would be moving toward x_o and the other half away from x_o, and in time τ half of them will reach x_o and cross as shown in Figure 5.29b. If n_1 is the concentration of electrons at $x_o - \frac{1}{2}\ell$, then the number of electrons moving toward the right to

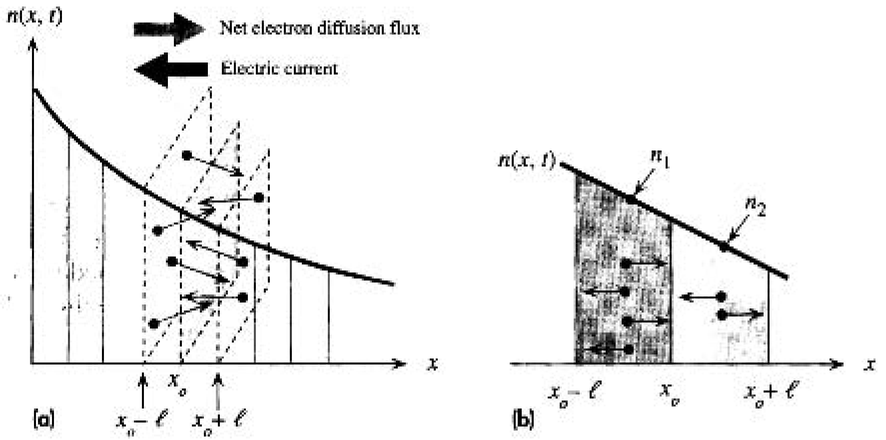


Figure 5.29

(a) Arbitrary electron concentration $n(x, t)$ profile in a semiconductor. There is a net diffusion (flux) of electrons from higher to lower concentrations.

(b) Expanded view of two adjacent sections at x_o. There are more electrons crossing x_o coming from the left $(x_o - \ell)$ than coming from the right $(x_o + \ell)$.

cross x_o is $\frac{1}{2}n_1 A\ell$ where A is the cross-sectional area and hence $A\ell$ is the volume of the segment. Similarly half of the electrons in $(x_o + \ell)$ would be moving toward the left and in time τ would reach x_o. Their number is $\frac{1}{2}n_2 A\ell$ where n_2 is the concentration at $x_o + \frac{1}{2}\ell$. The net number of electrons crossing x_o per unit time per unit area in the $+x$ direction is the electron flux Γ_e,

$$\Gamma_e = \frac{\frac{1}{2}n_1 A\ell - \frac{1}{2}n_2 A\ell}{A\tau}$$

that is,

$$\Gamma_e = -\frac{\ell}{2\tau}(n_2 - n_1) \qquad [5.33]$$

As far as calculus of variations is concerned, the mean free path ℓ is small, so we can calculate $n_2 - n_1$ from the concentration gradient using

$$n_2 - n_1 \approx \left(\frac{dn}{dx}\right)\Delta x = \left(\frac{dn}{dx}\right)\ell$$

We can now write the flux in Equation 5.33 in terms of the concentration gradient as

$$\Gamma_e = -\frac{\ell^2}{2\tau}\left(\frac{dn}{dx}\right)$$

or

$$\Gamma_e = -D_e\frac{dn}{dx} \qquad [5.34]$$

Fick's first law

where the quantity $(\ell^2/2\tau)$ has been defined as the diffusion coefficient of electrons and denoted by D_e. Thus, the net electron flux Γ_e at a position x is proportional to the concentration gradient and the diffusion coefficient. The steeper this gradient, the larger the flux Γ_e. In fact, we can view the concentration gradient dn/dx as the driving force for the diffusion flux, just like the electric field $-(dV/dx)$ is the driving force for the electric current: $J = \sigma\mathcal{E} = -\sigma(dV/dx)$.

Equation 5.34 is called **Fick's first law** and represents the relationship between the net particle flux and the driving force, which is the concentration gradient. It is the counterpart of Ohm's law for diffusion. D_e has the dimensions of $\mathrm{m^2\,s^{-1}}$ and is a measure of how readily the particles (in this case, electrons) diffuse in the medium. Note that Equation 5.34 gives the electron flux Γ_e at a position x where the electron concentration gradient is dn/dx. Since from Figure 5.29, the slope dn/dx is a negative number, Γ_e in Equation 5.34 comes out positive, which indicates that the flux is in the positive x direction. The electric current (conventional current) due to the diffusion of electrons to the right will be in the negative direction by virtue of Equation 5.32. Representing this electric current density due to diffusion as $J_{D,e}$ we can write

Electron diffusion current density

$$J_{D,e} = -e\Gamma_e = eD_e\frac{dn}{dx} \qquad [5.35]$$

In the case of a hole concentration gradient, as shown in Figure 5.30, the hole flux $\Gamma_h(x)$ is given by

$$\Gamma_h = -D_h\frac{dp}{dx}$$

where D_h is the hole diffusion coefficient. Putting in a negative number for the slope dp/dx, as shown in Figure 5.30, results in a positive hole flux (in the positive x direction), which in turn implies a diffusion current density toward the right. The current density due to hole diffusion is given by

Hole diffusion current density

$$J_{D,h} = e\Gamma_h = -eD_h\frac{dp}{dx} \qquad [5.36]$$

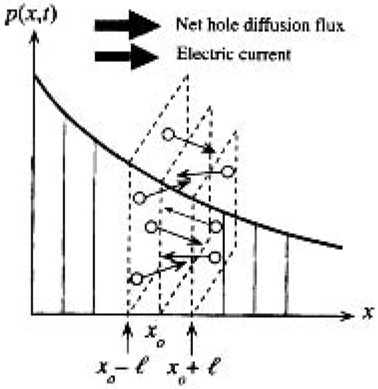


Figure 5.30 Arbitrary hole concentration $p(x, t)$ profile in a semiconductor.

There is a net diffusion (flux) of holes from higher to lower concentrations. There are more holes crossing x_o coming from the left $(x_o - \ell)$ than coming from the right $(x_o + \ell)$.

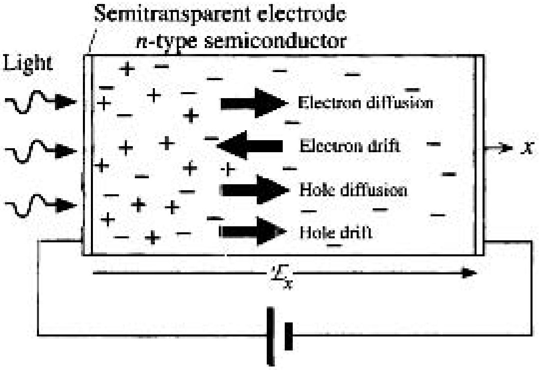


Figure 5.31 When there is an electric field and also a concentration gradient, charge carriers move both by diffusion and drift.

Suppose that there is also a positive electric field $\mathcal{E}_x$ acting along $+x$ in Figures 5.29 and 5.30. A practical example is shown in Figure 5.31 in which a semiconductor is sandwiched between two electrodes, the left one semitransparent. By connecting a battery to the electrodes, an applied field of $\mathcal{E}_x$ is set up in the semiconductor along $+x$. The left electrode is continuously illuminated, so excess EHPs are generated at this surface that give rise to concentration gradients in n and p. The applied field imposes an electrical force on the charges, which then try to drift. Holes drift toward the right and electrons toward the left. Charge motion then involves both drift and diffusion. The total current density due to the electrons drifting, driven by $\mathcal{E}_x$, and also diffusing, driven by dn/dx, is then given by adding Equation 5.35 to the usual electron drift current density,

Total electron current due to drift and diffusion

$$J_e = en\mu_e\mathcal{E}_x + eD_e\frac{dn}{dx} \qquad [5.37]$$

We note that as $\mathcal{E}_x$ is along x, so is the drift current (first term), but the diffusion current (second term) is actually in the opposite direction by virtue of a negative dn/dx.

Similarly, the hole current due to holes drifting and diffusing, Equation 5.36, is given by

Total hole current due to drift and diffusion

$$J_h = ep\mu_h\mathcal{E}_x - eD_h\frac{dp}{dx} \qquad [5.38]$$

In this case the drift and diffusion currents are in the same direction.

We mentioned that the diffusion coefficient is a measure of the ease with which the diffusing charge carriers move in the medium. But drift mobility is also a measure of the ease with which the charge carriers move in the medium. The two quantities are related through the **Einstein relation,**

Einstein relation

$$\frac{D_e}{\mu_e} = \frac{kT}{e} \quad \text{and} \quad \frac{D_h}{\mu_h} = \frac{kT}{e} \qquad [5.39]$$

In other words, the diffusion coefficient is proportional to the temperature and mobility. This is a reasonable expectation since increasing the temperature will increase the mean speed and thus accelerate diffusion. The randomizing effect against diffusion in one particular direction is introduced by the scattering of the carriers from lattice vibrations, impurities, and so forth, so that the longer the mean free path between scattering events, the larger the diffusion coefficient. This is examined in Example 5.12.

We equated the diffusion coefficient D to $\ell^2/2\tau$ in Equation 5.34. Our analysis, as represented in Figure 5.29, is oversimplified because we simply assumed that all electrons move a distance ℓ before scattering and all are free for a time τ. We essentially assumed that all those at a distance ℓ from x_o and moving toward x_o cross the plane exactly in time τ. This assumption is not entirely true because scattering is a stochastic process and consequently not all electrons moving toward x_o will cross it even in the segment of thickness ℓ. A rigorous statistical analysis shows that the diffusion coefficient is given by

Diffusion coefficient

$$D = \frac{\ell^2}{\tau} \tag{5.40}$$

EXAMPLE 5.12

THE EINSTEIN RELATION Using the relation between the drift mobility and the mean free time τ between scattering events and the expression for the diffusion coefficient $D = \ell^2/\tau$, derive the Einstein relation for electrons.

SOLUTION

In one dimension, for example, along x, the diffusion coefficient for electrons is given by $D_e = \ell^2/\tau$ where ℓ is the mean free path along x and τ is the mean free time between scattering events for electrons. The mean free path $\ell = v_x\tau$, where v_x is the mean (or effective) speed of the electrons along x. Thus,

$$D_e = v_x^2\tau$$

In the conduction band and in one dimension, the mean KE of electrons is $\frac{1}{2}kT$, so $\frac{1}{2}kT = \frac{1}{2}m_e^*v_x^2$ where m_e^* is the effective mass of the electron in the CB. This gives

$$v_x^2 = \frac{kT}{m_e^*}$$

Substituting for v_x in the D_e equation, we get,

$$D_e = \frac{kT\tau}{m_e^*} = \frac{kT}{e}\left(\frac{e\tau}{m_e^*}\right)$$

Further, we know from Chapter 2 that the electron drift mobility μ_e is related to the mean free time τ via $\mu_e = e\tau/m_e^*$, so we can substitute for τ to obtain

$$D_e = \frac{kT}{e}\mu_e$$

which is the Einstein relation. We assumed that Boltzmann statistics, that is, $v_x^2 = kT/m_e^*$ is applicable, which, of course, is true for the conduction band electrons in a semiconductor but not for the conduction electrons in a metal. Thus, the Einstein relation is only valid for electrons and holes in a nondegenerate semiconductor and certainly not valid for electrons in a metal.

EXAMPLE 5.13

DIFFUSION COEFFICIENT OF ELECTRONS IN Si Calculate the diffusion coefficient of electrons at 27 °C in n-type Si doped with 10^{15} As atoms cm^{-3}.

SOLUTION

From the μ_e versus dopant concentration graph, the electron drift mobility μ_e with 10^{15} cm^{-3} of dopants is about 1300 cm^2 V^{-1} s^{-1}, so

$$D_e = \frac{\mu_e kT}{e} = (1300 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1})(0.0259 \text{ V}) = 33.7 \text{ cm}^2 \text{ s}^{-1}$$

EXAMPLE 5.14

BUILT-IN POTENTIAL DUE TO DOPING VARIATION Suppose that due to a variation in the amount of donor doping in a semiconductor, the electron concentration is nonuniform across the semiconductor, that is, $n = n(x)$. What will be the potential difference between two points in the semiconductors where the electron concentrations are n_1 and n_2? If the donor profile in an n-type semiconductor is $N(x) = N_o \exp(-x/b)$, where b is a characteristic of the exponential doping profile, evaluate the built-in field $\mathcal{E}_x$. What is your conclusion?

SOLUTION

Consider a nonuniformly doped n-type semiconductor in which immediately after doping the donor concentration, and hence the electron concentration, decreases toward the right. Initially, the sample is neutral everywhere. The electrons will immediately diffuse from higher- to lower-concentration regions. But this diffusion accumulates *excess* electrons in the right region and exposes the positively charged donors in the left region, as depicted in Figure 5.32. The electric field between the accumulated negative charges and the exposed donors prevents further accumulation. Equilibrium is reached when the diffusion toward the right is just balanced by the drift of electrons toward the left. The total current in the sample must be zero (it is an open circuit),

$$J_e = en\mu_e\mathcal{E}_x + eD_e\frac{dn}{dx} = 0$$

But the field is related to the potential difference by $\mathcal{E}_x = -(dV/dx)$, so

$$-en\mu_e\frac{dV}{dx} + eD_e\frac{dn}{dx} = 0$$

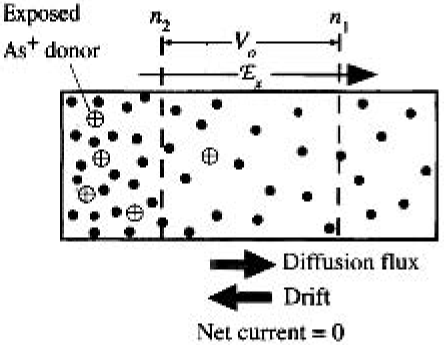


Figure 5.32 Nonuniform doping profile results in electron diffusion toward the less concentrated regions.

This exposes positively charged donors and sets up a built-in field $\mathcal{E}_x$. In the steady state, the diffusion of electrons toward the right is balanced by their drift toward the left.

We can now use the Einstein relation $D_e/\mu_e = kT/e$ to eliminate D_e and μ_e and then cancel dx and integrate the equation,

$$\int_{V_1}^{V_2} dV = \frac{kT}{e}\int_{n_1}^{n_2}\frac{dn}{n}$$

Integrating, we obtain the potential difference between points 1 and 2,

Built-in potential and concentration

$$V_2 - V_1 = \frac{kT}{e}\ln\left(\frac{n_2}{n_1}\right) \qquad [5.41]$$

To find the built-in field, we will assume that (and this is a reasonable assumption) the diffusion of electrons toward the right has not drastically upset the original $n(x) = N_d(x)$ variation because the field builds up quickly to establish equilibrium. Thus

$$n(x) \approx N_d(x) = N_o\exp\left(-\frac{x}{b}\right)$$

Substituting into the equation for $J_e = 0$, and again using the Einstein relation, we obtain $\mathcal{E}_x$ as

Built-in field

$$\mathcal{E}_x = \frac{kT}{be} \qquad [5.42]$$

Note: As a result of the fabrication process, the base region of a bipolar transistor has nonuniform doping, which can be approximated by an exponential $N_d(x)$. The resulting electric field $\mathcal{E}_x$ in Equation 5.42 acts to drift minority carriers faster and therefore speeds up the transistor operation as discussed in Chapter 6.

5.6 CONTINUITY EQUATION[4]

5.6.1 TIME-DEPENDENT CONTINUITY EQUATION

Many semiconductor devices operate on the principle that excess charge carriers are injected into a semiconductor by external means such as illumination or an applied voltage. The injection of carriers upsets the equilibrium concentration. To determine the carrier concentration at any point at any instant we need to solve **the continuity equation,** which is based on accounting for the total charge at that location in the semiconductor. Consider an n-type semiconductor slab as shown in Figure 5.33 in which the hole concentration has been upset along the x axis from its equilibrium value p_{no} by some external means.

Consider an infinitesimally thin elemental volume $A\,\delta x$ as in Figure 5.33 in which the hole concentration is $p_n(x, t)$. The current density at x due to holes flowing into the volume is J_h and that due to holes flowing out at $x + \delta x$ is $J_h + \delta J_h$. There is a change in the hole current density J_h; that is, $J_h(x, t)$ is not uniform along x. (Recall that the total current will also have a component due to electrons.) We assume that $J_h(x, t)$ and $p_n(x, t)$ do not change across the cross section along the y or z directions. If δJ_h is

---

[4] This section may be skipped without loss of continuity. (No pun intended.)

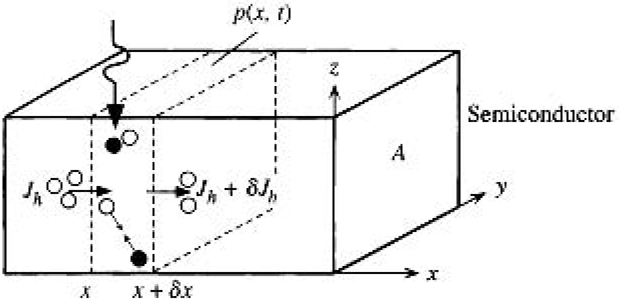


Figure 5.33 Consider an elemental volume $A\,\delta x$ in which the hole concentration is $p(x, t)$.

negative, then the current leaving the volume is less than that entering the volume, which leads to an increase in the hole concentration in $A\,\delta x$. Thus,

$$\frac{1}{A\,\delta x}\left(\frac{-A\,\delta J_h}{e}\right) = \text{Rate of increase in hole concentration due to the change in } J_h \tag{5.43}$$

The negative sign ensures that negative δJ_h leads to an increase in p_n. Recombination taking place in $A\,\delta x$ removes holes from this volume. In addition, there may also be photogeneration at x at time t. Thus,

The *net* rate of increase in the hole concentration p_n in $A\,\delta x$
= Rate of increase due to decrease in J_h − Rate of recombination + Rate of photogeneration

Continuity equation for holes

$$\frac{\partial p_n}{\partial t} = -\frac{1}{e}\left(\frac{\partial J_h}{\partial x}\right) - \frac{p_n - p_{no}}{\tau_h} + G_{\text{ph}} \tag{5.44}$$

where τ_h is the hole recombination time (lifetime), G_{ph} is the photogeneration rate at x at time t, and we used $\partial J_h/\partial x$ for $\delta J_h/\delta x$ since J_h depends on x and t.

Equation 5.44 is called the **continuity equation** for holes. The current density J_h is given by diffusion and drift components in Equations 5.37 and 5.38. There is a similar expression for electrons as well, but the negative sign multiplying $\partial J_e/\partial x$ is changed to positive (the charge e is negative for electrons).

The solutions of the continuity equation depend on the initial and boundary conditions. Many device scientists and engineers have solved Equation 5.44 for various semiconductor problems to characterize the behavior of devices. In most cases numerical solutions are necessary as analytical solutions are not mathematically tractable. As a simple example, consider uniform illumination of the surface of a semiconductor with suitable electrodes at its end as in Figure 5.28. Photogeneration and current density do not vary with distance along the sample length, so $\partial J_h/\partial x = 0$. If Δp_n is the excess concentration, $\Delta p_n = p_n - p_{no}$, then the time derivative of p_n in Equation 5.44 is the same as Δp_n. Thus, the continuity equation becomes

Continuity equation with uniform photogeneration

$$\frac{\partial \Delta p_n}{\partial t} = -\frac{\Delta p_n}{\tau_h} + G_{\text{ph}} \tag{5.45}$$

which is identical to the semiquantitatively derived Equation 5.27 from which photoconductivity was calculated in Example 5.11.

5.6.2 STEADY-STATE CONTINUITY EQUATION

For certain problems, the continuity equation can be further simplified. Consider, for example, the continuous illumination of one end of an n-type semiconductor slab by light that is absorbed in a very small thickness x_o at the surface as depicted in Figure 5.34a. There is no bulk photogeneration, so $G_{\text{ph}} = 0$. Suppose we are interested in the **steady-state** behavior; then the time derivative would be zero in Equation 5.44 to give,

Steady-state continuity equation for holes

$$\frac{1}{e}\left(\frac{\partial J_h}{\partial x}\right) = -\frac{p_n - p_{no}}{\tau_h} \qquad [5.46]$$

The hole current density J_h would have diffusion and drift components. If we assume that the electric field is very small, we can use Equation 5.38 with $\mathcal{E} \approx 0$ in Equation 5.46. Further, since the excess concentration $\Delta p_n(x) = p_n(x) - p_{no}$, we obtain,

Steady-state continuity equation with $\mathcal{E} = 0$

$$\frac{d^2 \Delta p_n}{dx^2} = \frac{\Delta p_n}{L_h^2} \qquad [5.47]$$

where, by definition, $L_h = \sqrt{D_h \tau_h}$ and is called the **diffusion length of holes.** Equation 5.47 describes the **steady-state** behavior of minority carrier concentration in a semiconductor under time-invariant excitation. When the appropriate boundary conditions are also included, its solution gives the *spatial* dependence of the excess minority carrier concentration $\Delta p_n(x)$.

In Figure 5.34a, both excess electrons and holes are photogenerated at the surface, but the percentage increase in the concentration of holes is much more dramatic since

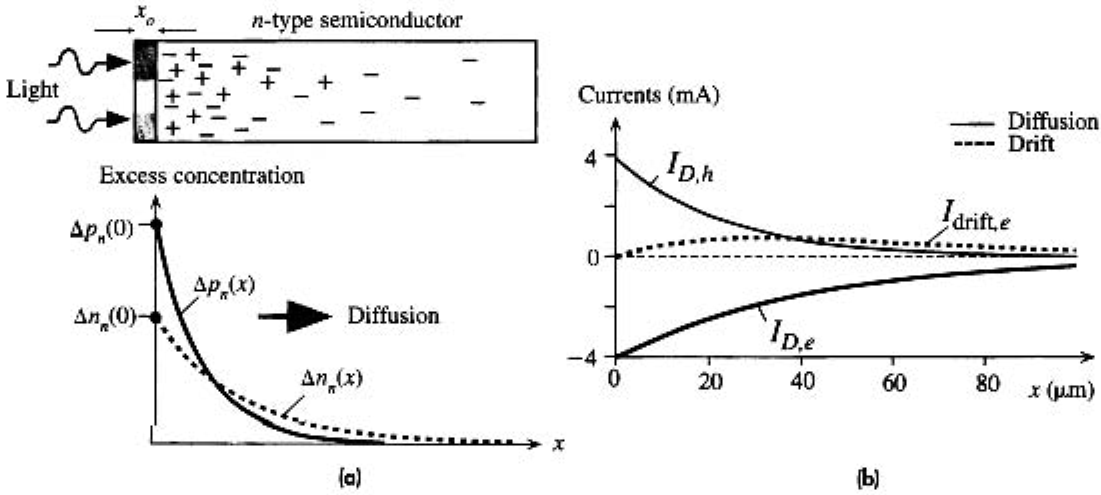


Figure 5.34

(a) Steady-state excess carrier concentration profiles in an n-type semiconductor that is continuously illuminated at one end.

(b) Majority and minority carrier current components in open circuit. Total current is zero.

$p_{no} \ll n_{no}$. We will assume **weak injection,** that is, $\Delta p_n \ll n_{no}$. Suppose that illumination is such that it causes the excess hole concentration at $x = 0$ to be $\Delta p_n(0)$. As holes diffuse toward the right, they meet electrons and recombine as a result of which the hole concentration $p_n(x)$ decays with distance into the semiconductor. If the bar is very long, then far away from the injection end we would expect p_n to be equal to the thermal equilibrium concentration p_{no}. The solution of Equation 5.47 with these boundary conditions shows that $\Delta p_n(x)$ decays *exponentially* as

Minority carrier concentration, long bar

$$\Delta p_n(x) = \Delta p_n(0)\exp\left(-\frac{x}{L_h}\right) \qquad [5.48]$$

This decay in the hole concentration results in a hole diffusion current $I_{D,h}(x)$ that has the same spatial dependence. Thus, if A is the cross-sectional area, the hole current is

Hole diffusion current

$$I_h \approx I_{D,h} = -AeD_h\frac{dp_n(x)}{dx} = \frac{AeD_h}{L_h}\Delta p_n(0)\exp\left(-\frac{x}{L_h}\right) \qquad [5.49]$$

We find $\Delta p_n(0)$ as follows. Under steady state, the holes generated per unit time in x_o must be removed by the hole current (at $x = 0$) at the *same* rate. Thus,

$$Ax_oG_{\text{ph}} = \frac{1}{e}I_{D,h}(0) = \frac{AD_h}{L_h}\Delta p_n(0)$$

or

$$\Delta p_n(0) = x_oG_{\text{ph}}\left(\frac{\tau_h}{D_h}\right)^{1/2} \qquad [5.50]$$

Similarly, electrons photogenerated in x_o diffuse toward the bulk, but their diffusion coefficient D_e and length L_e are larger than those for holes. The excess electron concentration Δn_n decays as

Majority carrier concentration, long bar

$$\Delta n_n(x) = \Delta n_n(0)\exp\left(-\frac{x}{L_e}\right) \qquad [5.51]$$

where $L_e = \sqrt{D_e\tau_h}$ and $\Delta n_n(x)$ decays more slowly than $\Delta p_n(x)$ as $L_e > L_h$. (Note that $\tau_e = \tau_h$.) The electron diffusion current $I_{D,e}$ is

Electron diffusion current

$$I_{D,e} = AeD_e\frac{dn_n(x)}{dx} = -\frac{AeD_e}{L_e}\Delta n_n(0)\exp\left(-\frac{x}{L_e}\right) \qquad [5.52]$$

The field at the surface is zero. Under steady state, the electrons generated per unit time in x_o must be removed by the electron current at the *same* rate. Thus, similarly to Equation 5.50,

$$\Delta n_n(0) = x_oG_{\text{ph}}\left(\frac{\tau_h}{D_e}\right)^{1/2} \qquad [5.53]$$

so that

$$\frac{\Delta p_n(0)}{\Delta n_n(0)} = \left(\frac{D_e}{D_h}\right)^{1/2} \qquad [5.54]$$

which is greater than unity for Si.

Table 5.3 Currents in an infinite slab illuminated at one end for weak injection near the surface

Currents at	Minority Diffusion $I_{D,h}$ (mA)	Minority Drift $I_{\text{drift},h}$ (mA)	Majority Diffusion $I_{D,e}$ (mA)	Majority Drift $I_{\text{drift},e}$ (mA)	Field $\mathcal{E}$ (V cm^{-1})
$x = 0$	3.94	0	−3.94	0	0
$x = L_e$	0.70	0.0022	−1.45	0.75	0.035

It is apparent that the hole and electron diffusion currents are in *opposite* directions. At the surface, the electron and hole diffusion currents are equal and opposite, so the total current is zero. As apparent from Equations 5.49 and 5.52, the hole diffusion current decays more rapidly than the electron diffusion current, so there must be some electron drift to keep the total current zero. The electrons are majority carriers which means that even a small field can cause a marked majority carrier drift current. If $I_{\text{drift},e}$ is the electron drift current, then in an open circuit the total current $I = I_{D,h} + I_{D,e} + I_{\text{drift},e} = 0$, so

Electron drift current

$$I_{\text{drift},e} = -I_{D,h} - I_{D,e} \tag{5.55}$$

The electron drift current increases with distance, so the total current I at every location is zero. It must be emphasized that there must be some field $\mathcal{E}$ in the sample, however small, to provide the necessary drift to balance the currents to zero. The field can be found from $I_{\text{drift},e} \approx Aen_{no}\mu_e\mathcal{E}$, inasmuch as n_{no} does not change significantly (weak injection),

Electric field

$$\mathcal{E} = \frac{I_{\text{drift},e}}{Aen_{no}\mu_e} \tag{5.56}$$

The hole drift current due to this field is

Hole drift current

$$I_{\text{drift},h} = Ae\mu_h p_n(x)\mathcal{E} \tag{5.57}$$

and it will be negligibly small as $p_n \ll n_{no}$.

We can use actual values to gauge magnitudes. Suppose that $A = 1\ \text{mm}^2$ and $N_d = 10^{16}\ \text{cm}^{-3}$ so that $n_{no} = N_d = 10^{16}\ \text{cm}^{-3}$ and $p_{no} = n_i^2/N_d = 1 \times 10^4\ \text{cm}^{-3}$. The light intensity is adjusted to yield $\Delta p_n(0) = 0.05n_{no} = 5 \times 10^{14}\ \text{cm}^{-3}$: *weak injection*. Typical values at 300 K for the material properties in this N_d-doped n-type Si would be $\tau_h = 480\ \text{ns}$, $\mu_e = 1350\ \text{cm}^2\ \text{V}^{-1}\ \text{s}^{-1}$, $D_e = 34.9\ \text{cm}^2\ \text{s}^{-1}$, $L_e = 0.0041\ \text{cm} = 41\ \mu\text{m}$, $\mu_h = 450\ \text{cm}^2\ \text{V}^{-1}\ \text{s}^{-1}$, $D_h = 11.6\ \text{cm}^2\ \text{s}^{-1}$, $L_h = 0.0024\ \text{cm} = 24\ \mu\text{m}$. We can now calculate each current term using the Equations 5.49, 5.52, 5.55 and 5.57 above as shown in Figure 5.34b. The actual values at two locations, $x = 0$ and $x = L_e = 41\ \mu\text{m}$, are shown in Table 5.3.[5]

---

[5] The reader may have observed that the currents in Table 5.3 do not add exactly to zero. The analysis here is only approximate and, further, it was based on neglecting the hole drift current and taking the field as nearly zero to use Equation 5.47 in deriving the carrier concentration profiles. Note that hole drift current is much smaller than the other current components.

EXAMPLE 5.15

INFINITELY LONG SEMICONDUCTOR ILLUMINATED AT ONE END Find the minority carrier concentration profile $p_n(x)$ in an infinite n-type semiconductor that is illuminated continuously at one end as in Figure 5.34. Assume that photogeneration occurs near the surface. Show that the mean distance diffused by the minority carriers before recombination is L_h.

SOLUTION

Continuous illumination means that we have steady-state conditions and thus Equation 5.47 can be used. The general solution of this second-order differential equation is

$$\Delta p_n(x) = A\exp\left(-\frac{x}{L_h}\right) + B\exp\left(\frac{x}{L_h}\right) \qquad [5.58]$$

where A and B are constants that have to be found from the boundary conditions. For an infinite bar, at $x = \infty$, $\Delta p_n(\infty) = 0$ gives $B = 0$. At $x = 0$, $\Delta p_n = \Delta p_n(0)$ so $A = \Delta p_n(0)$. Thus, the excess (photoinjected) hole concentration at position x is

$$\Delta p_n(x) = \Delta p_n(0)\exp\left(-\frac{x}{L_h}\right) \qquad [5.59]$$

which is shown in Figure 5.34a. To find the mean position of the photoinjected holes, we use the definition of the "mean," that is,

$$\bar{x} = \frac{\int_0^\infty x\,\Delta p_n(x)\,dx}{\int_0^\infty \Delta p_n(x)\,dx}$$

Substituting for $\Delta p_n(x)$ from Equation 5.59 and carrying out the integration gives $\bar{x} = L_h$. We conclude that the **diffusion length** L_h is the average distance diffused by the minority carriers before recombination. As a corollary, we should infer that $1/L_h$ is the mean probability per unit distance that the hole recombines with an electron.

5.7 OPTICAL ABSORPTION

We have already seen that a photon of energy $h\nu$ greater than E_g can be absorbed in a semiconductor, resulting in the excitation of an electron from the valence band to the conduction band, as illustrated in Figure 5.35. The average energy of electrons in the conduction band is $\frac{3}{2}kT$ above E_c (average kinetic energy is $\frac{3}{2}kT$), which means that the electrons are very close to E_c. If the photon energy is much larger than the bandgap energy E_g, then the excited electron is not near E_c and has to lose the extra energy $h\nu - E_g$ to reach thermal equilibrium. The excess energy $h\nu - E_g$ is lost to lattice vibrations as heat as the electron is scattered from one atomic vibration to another. This process is called **thermalization.** If, on the other hand, the photon energy $h\nu$ is less than the bandgap energy, the photon will not be absorbed and we can say that the semiconductor is transparent to wavelengths longer than hc/E_g provided that there are no energy states in the bandgap. There, of course, will be reflections occurring at the air/semiconductor surface due to the change in the refractive index.

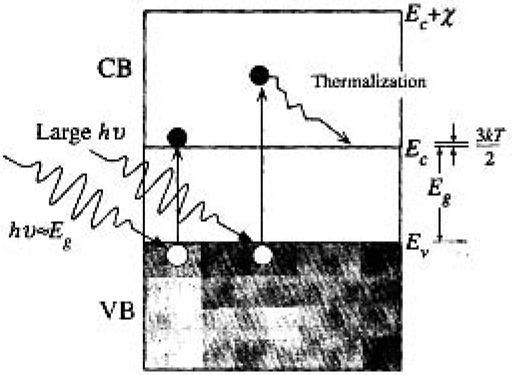


Figure 5.35 Optical absorption generates electron–hole pairs.

Energetic electrons must lose their excess energy to lattice vibrations until their average energy is $\frac{3}{2}kT$ in the CB.

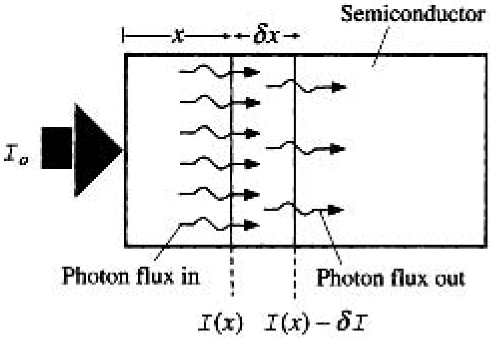


Figure 5.36 Absorption of photons within a small elemental volume of width δx.

Suppose that I_o is the intensity of a beam of **photons** incident on a semiconductor material. Thus, I_o is the energy incident per unit area per unit time. If Γ_{ph} is the photon flux, then

$$I_o = h\nu\Gamma_{\text{ph}}$$

When the photon energy is greater than E_g, photons from the incident radiation will be absorbed by the semiconductor. The absorption of photons requires the excitation of valence band electrons, and there are only so many of them with the right energy *per unit volume*. Consequently, absorption depends on the thickness of the semiconductor. Suppose that $I(x)$ is the light intensity at x and δI is the change in the light intensity in the small elemental volume of thickness δx at x due to photon absorption, as illustrated in Figure 5.36. Then δI will depend on the number of photons arriving at this volume $I(x)$ and the thickness δx. Thus

$$\delta I = -\alpha I\,\delta x$$

where α is a proportionality constant that depends on the photon energy and hence wavelength, that is, $\alpha = \alpha(\lambda)$. The negative sign ensures that δI is a reduction. The constant α as defined by this equation is called the **absorption coefficient** of the semiconductor. It is therefore defined by

Definition of absorption coefficient

$$\alpha = -\frac{\delta I}{I\,\delta x} \tag{5.60}$$

which has the dimensions of length^{-1} (m^{-1}).

When we integrate Equation 5.60 for illumination with constant wavelength light, we get the **Beer–Lambert law,** the transmitted intensity decreases exponentially with the thickness,

Beer–Lambert law

$$I(x) = I_o\exp(-\alpha x) \tag{5.61}$$

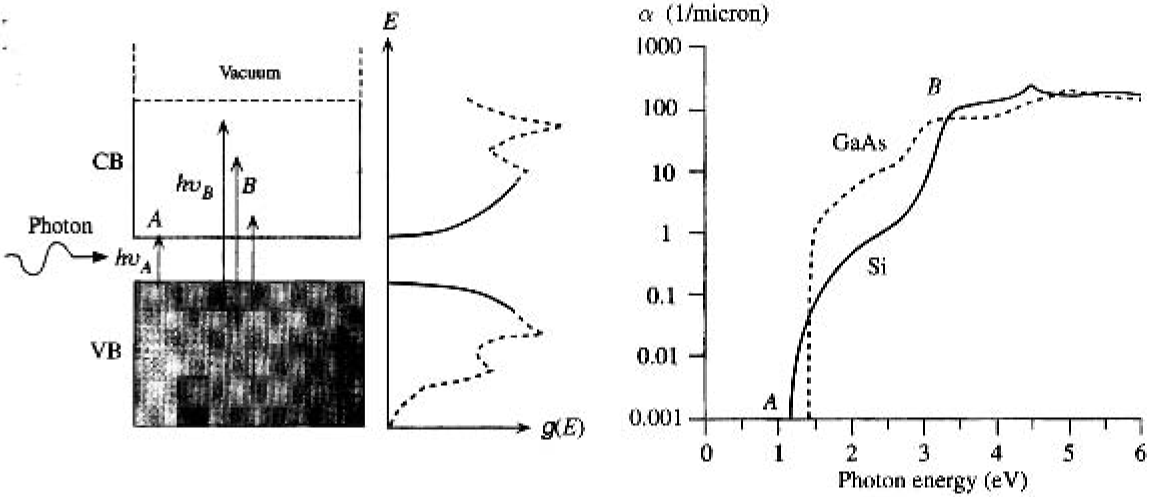


Figure 5.37 The absorption coefficient α depends on the photon energy $h\nu$ and hence on the wavelength.
Density of states increases from band edges and usually exhibits peaks and troughs. Generally α increases with the photon energy greater than E_g because more energetic photons can excite electrons from populated regions of the VB to numerous available states deep in the CB.

As apparent from Equation 5.61, over a distance $x = 1/\alpha$, the light intensity falls to a value $0.37I_o$; that is, it decreases by 63 percent. This distance over which 67 percent of the photons are absorbed is called the **penetration depth,** denoted by $\delta = 1/\alpha$.

The absorption coefficient depends on the photon absorption processes occurring in the semiconductor. In the case of **band-to-band (interband) absorption,** α increases rapidly with the photon energy $h\nu$ above E_g as shown for Si ($E_g = 1.1$ eV) and GaAs ($E_g = 1.42$ eV) in Figure 5.37. Notice that α is plotted on a logarithmic scale. The general trend of the α versus $h\nu$ behavior can be intuitively understood from the density of states diagram also shown in the same figure.

Density of states $g(E)$ represents the number of states per unit energy per unit volume. We assume that the VB states are filled and the CB states are empty since the number of electrons in the CB is much smaller than the number of states in this band ($n \ll N_c$). The photon absorption process increases when there are more VB states available as more electrons can be excited. We also need available CB states into which the electrons can be excited, otherwise the electrons cannot find empty states to fill. The probability of photon absorption depends on both the density of VB states *and* the density of CB states. For photons of energy $h\nu_A = E_g$, the absorption can only occur from E_v to E_c where the VB and CB densities of states are low and thus the absorption coefficient is small, which is illustrated as A in Figure 5.37. For photon energies $h\nu_B$, which can take electrons from very roughly the middle region of the VB to the middle of the CB, the densities of states are large and α is also large as indicated by B in Figure 5.37. Furthermore, there are more choices of excitation for the $h\nu_B$ photon as illustrated by the three arrows in the figure. At even higher photon energies, photon absorption can of course excite electrons from the VB into vacuum. In reality, the density of states $g(E)$ of a real crystalline semiconductor is much more complicated with various sharp peaks and troughs on the density of states function, shown as dashed curves in $g(E)$ in Figure 5.37, particularly away from the band edges. In addition, the absorption process has to satisfy the conservation of momentum and quantum mechanical transition rules which means that certain transitions from the CB to the VB will be more favorable than others. For example, GaAs is a **direct bandgap** semiconductor, so photon absorption can lead directly to the excitation of an electron from the CB to the VB for photon energies just above E_g just as direct recombination of an electron and hole results in photon emission. Si is an **indirect bandgap** semiconductor. Just as direct electron and hole recombination is not possible in silicon, the electron excitation from states near E_v to states near E_c must be accompanied by the emission or absorption of lattice vibrations, and hence the absorption is less efficient; α versus $h\nu$ for GaAs rises more sharply than that for Si above E_g as apparent in Figure 5.37. At sufficiently high photon energies, it is possible to excite electrons directly from the VB to the CB in Si and this gives the sharp rise in α versus $h\nu$ before B in Figure 5.37. (Band-to-band absorption is further discussed in Chapter 9.)

EXAMPLE 5.16 **PHOTOCONDUCTIVITY OF A THIN SLAB** Modify the photoconductivity expression

$$\Delta\sigma = \frac{e\eta I_o \lambda \tau(\mu_e + \mu_h)}{hcD}$$

derived for a direct bandgap semiconductor in Figure 5.28 to take into account that some of the light intensity is transmitted through the material.

SOLUTION

If we assume that all the photons are absorbed (there is no transmitted light intensity), then the photoconductivity expression is

$$\Delta\sigma = \frac{e\eta I_o \lambda \tau(\mu_e + \mu_h)}{hcD}$$

But, in reality, $I_o \exp(-\alpha D)$ is the transmitted intensity through the specimen with thickness D, so absorption is determined by the intensity lost in the material $I_o[1 - \exp(-\alpha D)]$, which means that $\Delta\sigma$ must be accordingly scaled down to

$$\Delta\sigma = \frac{e\eta I_o[1 - \exp(-\alpha D)]\lambda \tau(\mu_e + \mu_h)}{hcD}$$

EXAMPLE 5.17 **PHOTOGENERATION IN GaAs AND THERMALIZATION** Suppose that a GaAs sample is illuminated with a 50 mW HeNe laser beam (wavelength 632.8 nm) on its surface. Calculate how much power is dissipated as heat in the sample during thermalization. Give your answer as mW. The energy bandgap E_g of GaAs is 1.42 eV.

SOLUTION

Suppose P_L is the power in the laser beam; then $P_L = IA$, where I is the intensity of the beam and A is the area of incidence. The photon flux, photons arriving per unit area per unit time, is

$$\Gamma_{\text{ph}} = \frac{I}{h\nu} = \frac{P_L}{Ah\nu}$$

so the number of EHPs generated per unit time is

$$\frac{dN}{dt} = \Gamma_{\text{ph}} A = \frac{P_L}{h\nu}$$

These carriers *thermalize*—lose their excess energy as lattice vibrations (heat) via collisions with the lattice—so eventually their average kinetic energy becomes $\frac{3}{2}kT$ above E_g as depicted in Figure 5.35. Remember that we assume that electrons in the CB are nearly free, so they must obey the kinetic theory and hence have an average kinetic energy of $\frac{3}{2}kT$. The average energy of the electron is then $E_g + \frac{3}{2}kT \approx 1.46$ eV. The excess energy

$$\Delta E = h\nu - \left(E_g + \frac{3}{2}kT\right)$$

is lost to the lattice as heat, that is, lattice vibrations. Since each electron loses an amount of energy ΔE as heat, the heat power generated is

$$P_H = \left(\frac{dN}{dt}\right)\Delta E = \left(\frac{P_L}{h\nu}\right)(\Delta E)$$

The incoming photon has an energy $h\nu = hc/\lambda = 1.96$ eV, so

$$P_H = \frac{(50\,\text{mW})(1.96\,\text{eV} - 1.46\,\text{eV})}{1.96\,\text{eV}} = 12.76\,\text{mW}$$

Notice that in this example, and also in Figure 5.35, we have assigned the excess energy $\Delta E = h\nu - E_g - \frac{3}{2}kT$ to the electron rather than share it between the electron and the hole that is photogenerated. This assumption depends on the ratio of the electron and hole effective masses, and hence depends on the semiconductor material. It is approximately true in GaAs because the electron is much lighter than the hole, almost 10 times, and consequently the absorbed photon is able to "impart" a much higher kinetic energy to the electron than to the hole; $h\nu - E_g$ is used in the photogeneration, and the remainder goes to impart kinetic energy to the photogenerated electron hole pair.

5.8 PIEZORESISTIVITY

When a mechanical stress is applied to a semiconductor sample, as shown in Figure 5.38a, it is found that the resistivity of the semiconductor changes by an amount that depends on the stress.[6] **Piezoresistivity** is the change in the resistivity of a semiconductor (indeed, any material), due to an applied stress. **Elastoresistivity** refers to the change in the resistivity due to an induced strain in the substance. Since the application of stress invariably leads to strain, piezoresistivity and elastoresistivity refer to

---

[6] *Mechanical* stress is defined as the applied force per unit area, $\sigma_m = F/A$, and the resulting strain ε_m is the fractional change in the length of a sample caused by σ_m; $\varepsilon_m = \delta L/L$, where L is the sample length. The two are related through the elastic modulus Y; $\sigma_m = Y\varepsilon_m$. Subscript m is used to distinguish the stress σ_m and strain ε_m from the conductivity σ and permittivity ε.

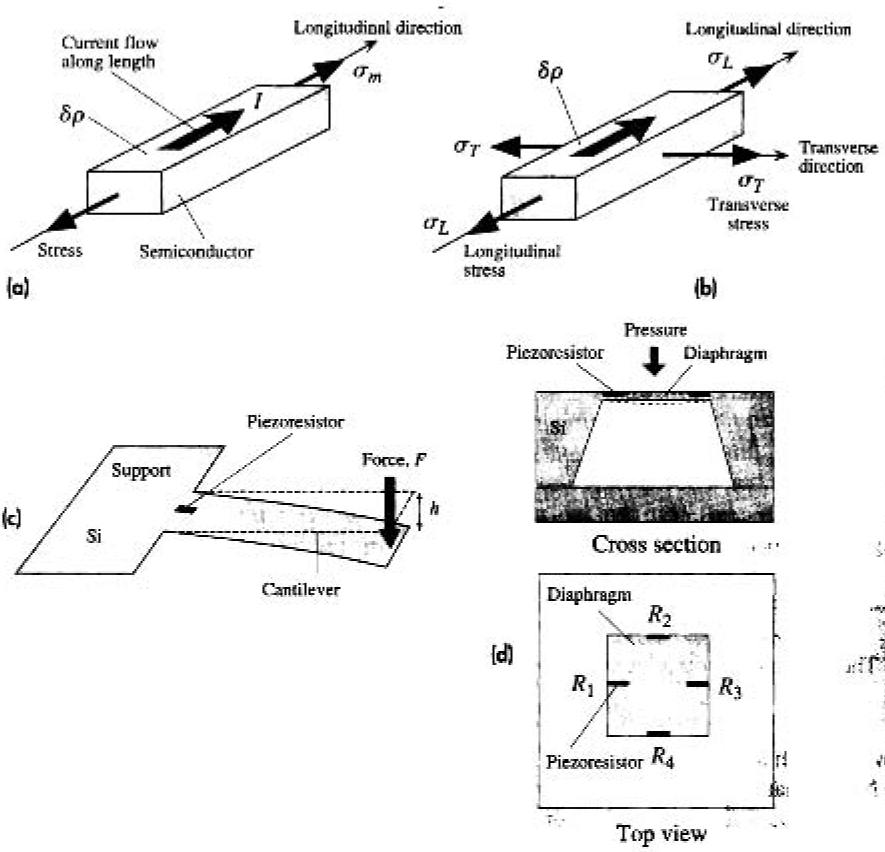


Figure 5.38 Piezoresistivity and its applications.

(a) Stress σ_m along the current (longitudinal) direction changes the resistivity by $\delta\rho$.

(b) Stresses σ_L and σ_T cause a resistivity change.

(c) A force applied to a cantilever bends it. A piezoresistor at the support end (where the stress is large) measures the stress, which is proportional to the force.

(d) A pressure sensor has four piezoresistors R_1, R_2, R_3, R_4 embedded in a diaphragm. The pressure bends the diaphragm, which generates stresses that are sensed by the four piezoresistors.

the same phenomenon. Piezoresistivity is fruitfully utilized in a variety of useful sensor applications such as force, pressure and strain gauges, accelerometers, and microphones.

The change in the resistivity may be due to a change in the concentration of carriers or due to a change in the drift mobility of the carriers, both of which can be modified by a strain in the crystal. Typically, in an extrinsic or doped semiconductor, the concentration of carriers does not change as significantly as the drift mobility; the piezoresistivity is then associated with the change in the mobility. For example, in an n-type Si, the change in the electron mobility μ_e with mechanical strain ε_m, $d\mu_e/d\varepsilon_m$, is of the order of 10^5 cm^2 V^{-1} s^{-1}, so that a strain of 0.015 percent will result in a change in the mobility that is about 1 percent, and a similar change in the resistivity, which is readily measurable. In this case, the change in the mobility μ_e is due to the induced strain changing the effective mass m_e^* which then modifies μ_e. (Recall that $\mu_e = e\tau/m_e^*$, where τ is the mean scattering time.)

The change in the resistivity $\delta\rho$ has been shown to be proportional to the induced strain in the crystal and hence proportional to the applied stress σ_m. The fractional change $\delta\rho/\rho$ can be written as

$$\frac{\delta\rho}{\rho} = \pi\sigma_m \qquad [5.62]$$

Piezoresistivity

where π is a constant called the **piezoresistive coefficient;** π has the units of 1/stress, *e.g.*, m^2/N or 1/Pa. The piezoresistive coefficient π depends on the type of doping, p-or n-type; the dopant concentration; the temperature; and the crystallographic direction. A stress along a certain direction in a crystal, for example, along the length of a semiconductor crystal, will change the resistivity not only in the same direction but also in transverse directions. We know from elementary mechanics that a strain in one direction is accompanied by a transverse strain, as implied by the Poisson ratio, so it is not unexpected that a stress in one direction will also modify the resistivity in a transverse direction. Thus, the change in the resistivity of a semiconductor in a "longitudinal" direction, taken as the direction of current flow, is due to stresses in the longitudinal and transverse directions. If σ_L is the stress along a longitudinal direction, the direction of current flow, and σ_T is the stress along a transverse direction, as in Figure 5.38b, then, generally, the fractional change in the resistivity along the current flow direction (longitudinal direction) is given by

$$\frac{\delta\rho}{\rho} = \pi_L\sigma_L + \pi_T\sigma_T \qquad [5.63]$$

Piezoresistivity

where π_L is the piezoresistive coefficient along a longitudinal direction (different for p- and n-type Si), and π_T is the piezoresistive coefficient in the transverse direction.

The piezoresistive effect is actually more complicated than what we have implied. In reality, we have to consider six types of stresses, three uniaxial stresses along the x, y, and z directions (*e.g.*, trying to pull the crystal along in three independent directions) and three shear stresses (*e.g.*, trying to shear the crystal in three independent ways). In very simple terms, a change in the resistivity $(\delta\rho/\rho)_i$ along a particular direction i (an arbitrary direction) can be induced by a stress σ_j along another direction j (which may or may not be identical to i). The two, $(\delta\rho/\rho)_i$ and σ_j, are then related through a piezoresistivity coefficient denoted by π_{ij}. Consequently, the full description of piezoresistivity involves tensors, and the piezoresistivity coefficients π_{ij} form the elements of this tensor; a treatment beyond the scope of this book. Nonetheless, it is useful to be able to calculate π_L and π_T from various tabulated piezoresistivity coefficients π_{ij}, without having to learn tensors. It turns out that it is sufficient to identify three *principal piezoresistive coefficients* to describe the piezoresistive effect in cubic crystals, which are denoted as π_{11}, π_{12}, and π_{44}. From the latter set we can easily calculate π_L and π_T for a crystallographic direction of interest; the relevant equations can be found in advanced textbooks.

Advances in silicon fabrication technologies and micromachining (ability to fabricate micromechanical structures) have now enabled various piezoresistive silicon microsensors to be developed that have a wide range of useful applications. Figure 5.38c shows a very simple Si microcantilever in which an applied force F to the free end bends the cantilever; the tip of the cantilever is deflected by a distance h. According to elementary mechanics, this deflection induces a maximum stress σ_m that is at the surface, at the support end, of the cantilever. A properly placed piezoresistor at this end can be used to measure this stress σ_m, and hence the deflection or the force. The piezoresistor is implanted by selectively diffusing dopants into the Si cantilever at the support end. Obviously, we need to relate the deflection h of the cantilever tip to the stress σ_m, which is well described in mechanics. In addition, h is proportional to the applied force F through a factor that depends on the elastic modulus and the geometry of the cantilever. Thus, we can measure both the displacement (h) and force (F).

Another useful application is in pressure sensors, which are commercially available. Again, the structure is fabricated from Si. A very thin elastic membrane, called a *diaphragm*, has four piezoresistors embedded, by appropriate dopant diffusion, on its surface as shown in Figure 5.38d. Under pressure, the Si diaphragm deforms elastically, and the stresses that are generated by this deformation cause the resistance of the piezoresistors to change. There are four piezoresistors because the four are connected in a Wheatstone bridge arrangement for better signal detection. The diaphragm area is typically 1 mm × 1 mm, and the thickness is 20 μm. There is no doubt that recent advances in micromachining have made piezoresistivity an important topic for a variety of sensor applications.

EXAMPLE 5.18

PIEZORESISTIVE STRAIN GAUGE Suppose that we apply a stress σ_L along the length, taken along the [110] direction, of a p-type silicon crystal sample. We will measure the resistivity along this direction by passing a current along the length and measuring the voltage drop between two fixed points as in Figure 5.38a. The stress σ_L along the length will result in a strain ε_L along the same length given by $\varepsilon_L = \sigma_L/Y$, where Y is the elastic modulus. From Equation 5.63 the change in the resistivity is

$$\frac{\Delta\rho}{\rho} = \pi_L\sigma_L + \pi_T\sigma_T = \pi_L Y\varepsilon_L$$

where we have ignored the presence of any transverse stresses; $\sigma_T \approx 0$. These transverse stresses depend on how the piezoresistor is used, that is, whether it is allowed to contract laterally. If the resistor cannot contract, it must be experiencing a transverse stress. In any event, for the particular direction of interest, [110], the Poisson ratio is very small (less than 0.1), and we can simply neglect any σ_T. Clearly, we can find the strain ε_L from the measurement of $\Delta\rho/\rho$, which is the principle of the strain gauge. The **gauge factor** G of a strain gauge measures the sensitivity of the gauge in terms of the fractional change in the resistance per unit strain,

Semiconductor strain gauge

$$G = \frac{\left(\dfrac{\Delta R}{R}\right)}{\left(\dfrac{\Delta L}{L}\right)} \approx \frac{\left(\dfrac{\Delta\rho}{\rho}\right)}{\varepsilon_L} \approx Y\pi_L$$

where we have assumed that ΔR is dominated by $\Delta\rho$, since the effects from geometric changes in the sample shape can be ignored compared with the piezoresistive effect in semiconductors.

Using typical values for a p-type Si piezoresistor which has a length along [110], $Y \approx 170$ GPa, $\pi_L \approx 72 \times 10^{-11}$ Pa^{-1}, we find $G \approx 122$. This is much greater than $G \approx 1.7$ for metal resistor–based strain gauges. In most metals, the fractional change in the resistance $\Delta R/R$ is due to the geometric effect, the sample becoming elongated and narrower, whereas in semiconductors it is due to the piezoresistive effect.

5.9 SCHOTTKY JUNCTION

5.9.1 SCHOTTKY DIODE

We consider what happens when a metal and an n-type semiconductor are brought into contact. In practice, this process is frequently carried out by the evaporation of a metal onto the surface of a semiconductor crystal in vacuum.

The energy band diagrams for the metal and the semiconductor are shown in Figure 5.39. The work function, denoted as Φ, is the energy difference between the vacuum level and the Fermi level. The vacuum level defines the energy where the electron is free from that particular solid and where the electron has zero KE.

For the metal, the work function Φ_m is the minimum energy required to remove an electron from the solid. In the metal there are electrons at the Fermi level E_{Fm}, but in the

John Bardeen, Walter Schottky, and Walter Brattain. Walter H. Schottky (1886–1976) obtained his PhD from the University of Berlin in 1912. He made many distinct contributions to physical electronics. He invented the screen grid vacuum tube in 1915, and the tetrode vacuum tube in 1919 while at Siemens. The Schottky junction theory was formulated in 1938. He also made distinct contributions to thermal and shot noise in devices. His book *Thermodynamik* was published in 1929 and included an explanation of the Schottky defect (Chapter 1).

SOURCE: AIP Emilio Segre Visual Archives, Brattain Collection.

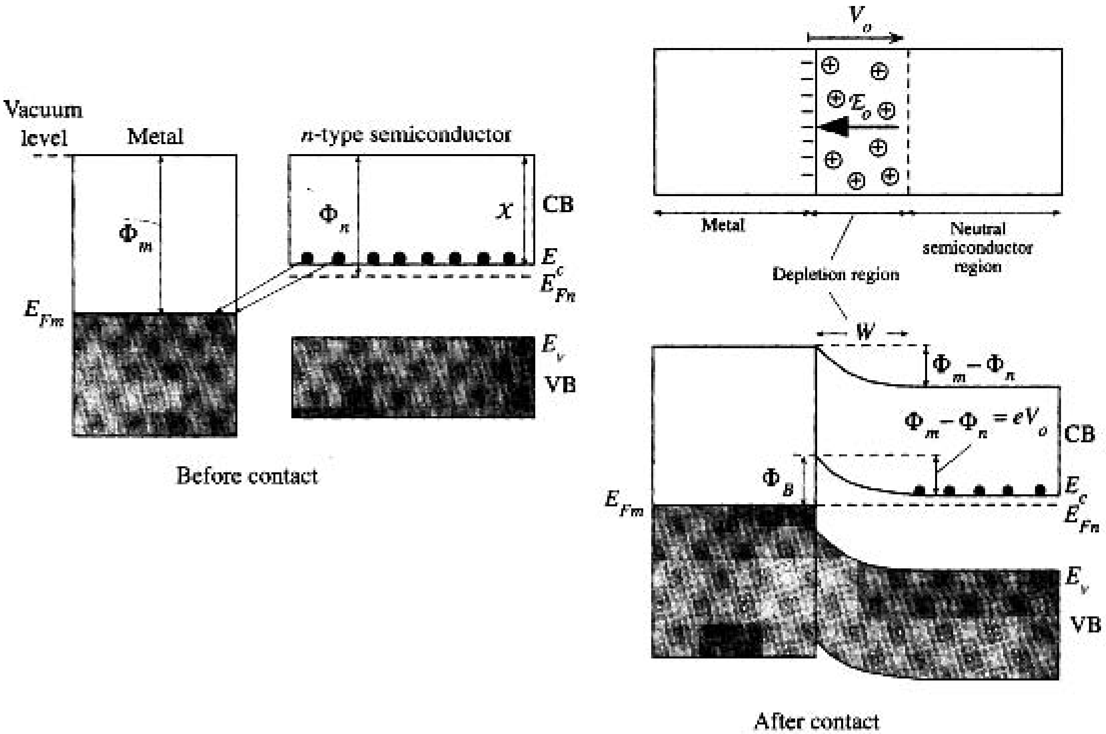


Figure 5.39 Formation of a Schottky junction between a metal and an n-type semiconductor when $\Phi_m > \Phi_n$.

semiconductor there are none at E_{Fn}. Nonetheless, the semiconductor work function Φ_n still represents the energy required to remove an electron from the semiconductor. It may be thought that the minimum energy required to remove an electron from the semiconductor is simply the electron affinity χ, but this is not so. Thermal equilibrium requires that only a certain fraction of all the electrons in the semiconductor should be in the CB at a given temperature. When an electron is removed from the conduction band, then thermal equilibrium can be maintained only if an electron is excited from the VB to CB, which involves absorbing heat (energy) from the environment; thus it takes more energy than simply χ. We will not derive the effective thermal energy required to remove an electron but state that, as for a metal, this is equal to Φ_n, even though there are no electrons at E_{Fn}. In fact, the thermionic emission of electrons from a heated semiconductor is also described by the Richardson–Dushman expression in Equation 4.37 but with Φ representing the work function of the semiconductor, Φ_n in the present n-type case. (In contrast, the minimum *photon energy* required to remove an electron from a semiconductor above absolute zero would be the electron affinity.)

We assume that $\Phi_m > \Phi_n$, the work function of the metal is greater than that of the semiconductor. When the two solids come into contact, the more energetic electrons in the CB of the semiconductor can readily tunnel into the metal in search of lower empty energy levels (just above E_{Fm}) and accumulate near the surface of the metal, as illustrated in Figure 5.39. Electrons tunneling from the semiconductor leave behind an electron-depleted region of width W in which there are exposed positively charged donors, in other words, net positive space charge. The contact potential, called the **built-in potential** V_o, therefore develops between the metal and the semiconductor. There is obviously also a **built-in electric field** $\mathcal{E}_o$ from the positive charges to the negative charges on the metal surface. Eventually this built-in potential reaches a value that prevents further accumulation of electrons at the metal surface and an equilibrium is reached. The value of the built-in voltage V_o is the same as that in the metal–metal junction case in Chapter 4, namely, $(\Phi_m - \Phi_n)/e$. The **depletion region** has been depleted of free carriers (electrons) and hence contains the exposed positive donors. This region thus constitutes a **space charge layer** (SCL) in which there is a nonuniform internal field directed from the semiconductor to the metal surface. The maximum value of this built-in field is denoted as $\mathcal{E}_o$ and occurs right at the metal–semiconductor junction (this is where there are a maximum number of field lines from positive to negative charges).

The Fermi level throughout the whole solid, the metal and semiconductor in contact, must be uniform in equilibrium. Otherwise, a change in the Fermi level ΔE_F going from one end to the other end will be available to do external (electrical) work. Thus, E_{Fm} and E_{Fn} line up. The W region, however, has been depleted of electrons, so in this region $E_c - E_{Fn}$ must increase so that n decreases. The bands must bend to increase $E_c - E_{Fn}$ toward the junction, as depicted in Figure 5.39. Far away from the junction, we, of course, still have an n-type semiconductor. The bending is just enough for the vacuum level to be continuous and changing by $\Phi_m - \Phi_n$ from the semiconductor to the metal, as this much energy is needed to take an electron across from the semiconductor to the metal. The *PE* barrier for electrons moving from the metal to the semiconductor is called the **Schottky barrier height** Φ_B, which is given by

Schottky barrier

$$\Phi_B = \Phi_m - \chi = eV_o + (E_c - E_{Fn}) \quad [5.64]$$

which is greater than eV_o.

Under open circuit conditions, there is no net current flowing through the metal–semiconductor junction. The number of electrons thermally emitted over the *PE* barrier Φ_B from the metal to the semiconductor is equal to the number of electrons thermally emitted over eV_o from the semiconductor to the metal. Emission probability depends on the *PE* barrier for emission through the Boltzmann factor. There are two current components due to electrons flowing through the junction. The current due to electrons being thermally emitted from the metal to the CB of the semiconductor is

$$J_1 = C_1 \exp\left(-\frac{\Phi_B}{kT}\right) \quad [5.65]$$

where C_1 is some constant, whereas the current due to electrons being thermally emitted from the CB of the semiconductor to the metal is

$$J_2 = C_2 \exp\left(-\frac{eV_o}{kT}\right) \quad [5.66]$$

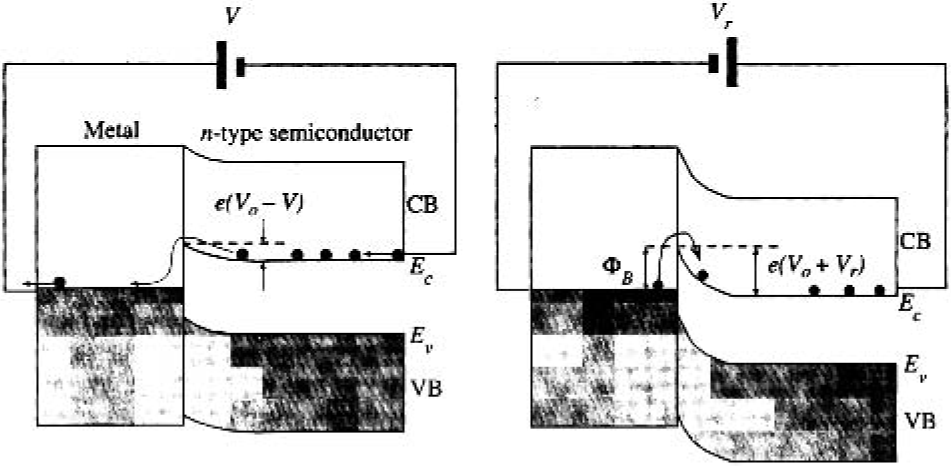
where C_2 is some constant different than C_1.

In equilibrium, that is, open circuit conditions in the dark, the currents are equal but in the reverse directions:

$$J_{\text{open circuit}} = J_2 - J_1 = 0$$

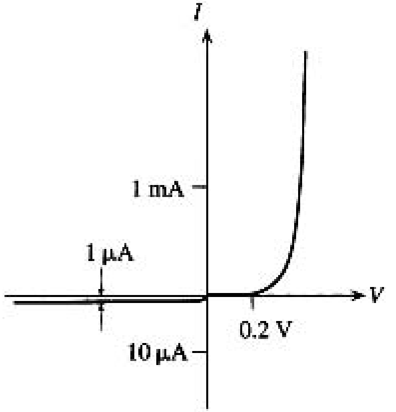
Under forward bias conditions, the semiconductor side is connected to the negative terminal, as depicted schematically in Figure 5.40a. Since the depletion region W has a much larger resistance than the neutral n-region (outside W) and the metal side, nearly all the voltage drop is across the depletion region. The applied bias is in the opposite direction to the built-in voltage V_o. Thus V_o is reduced to $V_o - V$. Φ_B remains unchanged. The semiconductor band diagram outside the depletion region has been effectively shifted up with respect to the metal side by an amount eV because

$$PE = \text{Charge} \times \text{Voltage}$$



(a) Forward-biased Schottky junction. Electrons in the CB of the semiconductor can easily overcome the small *PE* barrier to enter the metal.

(b) Reverse-biased Schottky junction. Electrons in the metal cannot easily overcome the *PE* barrier Φ_B to enter the semiconductor.



(c) $I-V$ characteristics of a Schottky junction exhibits rectifying properties (negative current axis is in microamps).

Figure 5.40 The Schottky junction.

The charge is negative but so is the voltage connected to the semiconductor, as shown in Figure 5.40a.

The *PE* barrier for thermal emission of electrons from the semiconductor to the metal is now $e(V_o - V)$. The electrons in the CB can now readily overcome the *PE* barrier to the metal.

The current J_2^{for}, due to the electron emission from the semiconductor to the metal, is now

$$J_2^{\text{for}} = C_2 \exp\left[-\frac{e(V_o - V)}{kT}\right] \qquad [5.67]$$

Since Φ_B is the same, J_1 remains unchanged. The net current is then

$$J = J_2^{\text{for}} - J_1 = C_2 \exp\left[-\frac{e(V_o - V)}{kT}\right] - C_2 \exp\left(-\frac{eV_o}{kT}\right)$$

or

$$J = C_2 \exp\left(-\frac{eV_o}{kT}\right)\left[\exp\left(\frac{eV}{kT}\right) - 1\right]$$

giving

$$J = J_o\left[\exp\left(\frac{eV}{kT}\right) - 1\right] \qquad [5.68]$$

Schottky junction

where J_o is a constant that depends on the material and surface properties of the two solids. In fact, examination of the above steps shows that J_o is also J_1 in Equation 5.65.

When the Schottky junction is reverse biased, then the positive terminal is connected to the semiconductor, as illustrated in Figure 5.40b. The applied voltage V_r drops across the depletion region since this region has very few carriers and is highly resistive. The built-in voltage V_o thus increases to $V_o + V_r$. Effectively, the semiconductor band diagram is shifted down with respect to the metal side because the charge is negative but the voltage is positive and *PE* = Charge × Voltage. The *PE* barrier for thermal emission of electrons from the CB to the metal becomes $e(V_o + V_r)$, which means that the corresponding current component becomes

$$J_2^{\text{rev}} = C_2 \exp\left[-\frac{e(V_o + V_r)}{kT}\right] \ll J_1 \qquad [5.69]$$

Since generally V_o is typically a fraction of a volt and the reverse bias is more than a few volts, $J_2^{\text{rev}} \ll J_1$ and the reverse bias current is essentially limited by J_1 only and is very small. Thus, under reverse bias conditions, the current is primarily due to the thermal emission of electrons over the barrier Φ_B from the metal to the CB of the semiconductor as determined by Equation 5.65. Figure 5.40c illustrates the *I–V* characteristics of a typical Schottky junction. The *I–V* characteristics exhibit rectifying properties, and the device is called a **Schottky diode.**

Equation 5.68, which is derived for forward bias conditions, is also valid under reverse bias by making *V* negative, that is, $V = -V_r$. Furthermore, it turns out to be applicable not only to Schottky-type metal–semiconductor junctions but also to junctions between a p-type and an n-type semiconductor, pn junctions, as we will show in Chapter 6. Under a forward bias V_f, which is greater than 25 mV at room temperature, the forward current is simply

Schottky junction forward bias

$$J_f = J_o \exp\left(\frac{eV_f}{kT}\right) \qquad V_f > \frac{kT}{e} \tag{5.70}$$

It should be mentioned that it is also possible to obtain a Schottky junction between a metal and a p-type semiconductor. This arises when $\Phi_m < \Phi_p$, where Φ_p is the work function for the p-type semiconductor.

5.9.2 SCHOTTKY JUNCTION SOLAR CELL

The built-in field in the depletion region of the Schottky junction allows this type of device to function as a photovoltaic device and also as a photodetector. We consider a Schottky device that has a thin metal film (usually Au) deposited onto an n-type semiconductor. The energy band diagram is shown in Figure 5.41. The metal is sufficiently thin (~10 nm) to allow light to reach the semiconductor.

For photon energies greater than E_g, EHPs are generated in the depletion region in the semiconductor, as indicated in Figure 5.41. The field in this region separates the EHPs and drifts the electrons toward the semiconductor and holes toward the metal. When an electron reaches the neutral n-region, there is now one extra electron there and therefore an additional negative charge. This end therefore becomes more negative with respect to the situation in the dark or the equilibrium situation. When a hole reaches the metal, it recombines with an electron and reduces the effective charge there by one electron, thus making it more positive relative to its dark state. Under open circuit conditions, therefore, a voltage develops across the Schottky junction device with the metal end positive and semiconductor end negative.

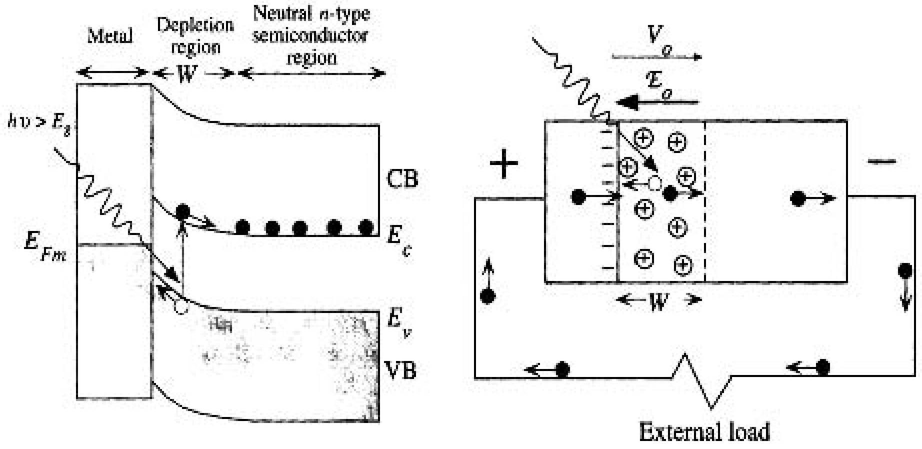


Figure 5.41 The principle of the Schottky junction solar cell.

The photovoltaic explanation in terms of the energy band diagram is simple. At the point of photogeneration, the electron finds itself at a *PE* slope as E_c is decreasing toward the semiconductor, as shown in Figure 5.41. It has no option but to roll down the slope just as a ball that is let go on a slope would roll down the slope to decrease its gravitational *PE*. Recall that there are many more empty states in the CB than electrons, so there is nothing to prevent the electron from rolling down the CB in search of lower energy. When the electron reaches the neutral region (flat E_c region), it upsets the equilibrium there. There is now an additional electron in the CB and this side acquires a negative charge. If we remember that hole energy increases downward on the energy band diagram, then similar arguments also apply to the photogenerated hole in the VB, which rolls down its own *PE* slope to reach the surface of the metal and recombine with an electron there.

If the device is connected to an external load, then the extra electron in the neutral *n*-region is conducted through the external leads, through the load, toward the metal side, where it replenishes the lost electron in the metal. As long as photons are generating EHPs, the flow of electrons around the external circuit will continue and there will be photon energy to electrical energy conversion. Sometimes it is useful to think of the neutral *n*-type semiconductor region as a "conductor," an extension of the external wire (except that the *n*-type semiconductor has a higher resistivity). As soon as the photogenerated electron crosses the depletion region, it reaches a conductor and is conducted around the external circuit to the metal side to replenish the lost electron there.

For photon energies less than E_g, the device can still respond, providing that the $h\nu$ can excite an electron from E_{Fm} in the metal over the *PE* barrier Φ_B into the CB, from where the electron will roll down toward the neutral *n*-region. In this case, $h\nu$ must only be greater than Φ_B.

If the Schottky junction diode is reverse-biased, as shown in Figure 5.42, then the reverse bias V_r increases the built-in potential V_o to $V_o + V_r$ $(V_r \gg V_o)$. The internal field increases to substantially high values. This has the advantage of increasing the drift velocity of the EHPs $(v_d = \mu_d\mathcal{E})$ in the depletion region and therefore

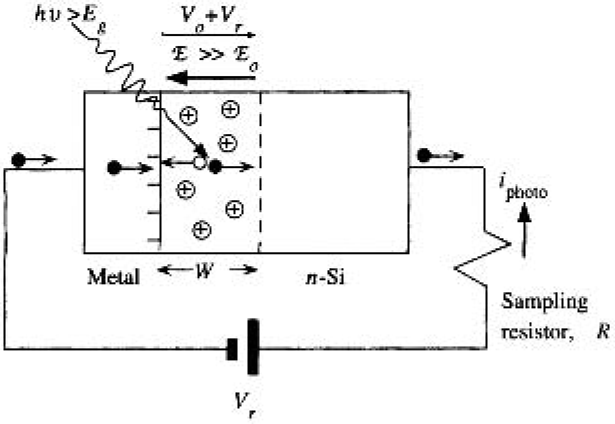


Figure 5.42 Reverse-biased Schottky photodiodes are frequently used as fast photodetectors.

shortening the transit time required to cross the depletion width. The device responds faster and is useful as a fast photodetector. The photocurrent i_{photo} in the external circuit is due to the drift of photogenerated carriers in the depletion region and can be readily measured.

EXAMPLE 5.19

Reverse saturation current in Schottky junction

THE SCHOTTKY DIODE The reverse saturation current J_o in the Schottky junction, as expressed in Equation 5.68, is the same current that is given by the Richardson–Dushman equation for thermionic emission over a potential barrier $\Phi(=\Phi_B)$ derived in Chapter 4. J_o is given by

$$J_o = B_e T^2 \exp\left(-\frac{\Phi_B}{kT}\right)$$

where B_e is the effective Richardson constant that depends on the characteristics of the metal–semiconductor junction. B_e for metal–semiconductor junctions, among other factors, depends on the density of states related effective mass of the thermally emitted carriers in the semiconductor. For example, for a metal to n-Si junction, B_e is about 110 A cm^{-2} K^{-2}, and for a metal to p-Si junction, which involves holes, B_e is about 30 A cm^{-2} K^{-2}.

a. Consider a Schottky junction diode between W (tungsten) and n-Si, doped with 10^{16} donors cm^{-3}. The cross-sectional area is 1 mm^2. Given that the electron affinity χ of Si is 4.01 eV and the work function of W is 4.55 eV, what is the theoretical barrier height Φ_B from the metal to the semiconductor?

b. What is the built-in voltage V_o with no applied bias?

c. Given that the experimental barrier height Φ_B is about 0.66 eV, what is the reverse saturation current and the current when there is a forward bias of 0.2 V across the diode?

SOLUTION

a. From Figure 5.39, it is clear that the barrier height Φ_B is

$$\Phi_B = \Phi_m - \chi = 4.55\text{ eV} - 4.01\text{ eV} = 0.54\text{ eV}$$

The experimental value is around 0.66 eV, which is greater than the theoretical value due to various effects at the metal–semiconductor interface arising from dangling bonds, defects, and so forth. For example, dangling bonds give rise to what are called *surface states* within the bandgap of the semiconductor that can capture electrons and modify the Schottky energy band diagram. (The energy band diagram in Figure 5.39 represents an ideal junction with no surface states.) Further, in some cases, such as Pt on n-Si, the experimental value can be lower than the theoretical value.

b. We can find $E_c - E_{Fn}$ in Figure 5.39 from

$$n = N_d = N_c \exp\left(-\frac{E_c - E_{Fn}}{kT}\right)$$

$$10^{16}\text{ cm}^{-3} = (2.8\times 10^{19}\text{ cm}^{-3})\exp\left(-\frac{E_c - E_{Fn}}{0.026\text{ eV}}\right)$$

which gives $\Delta E = E_c - E_{Fn} = 0.206$ eV. Thus, the built-in potential V_o can be found from

$$V_o = \frac{\Phi_B}{e} - \frac{E_c - E_{Fn}}{e} = 0.54\text{ V} - 0.206\text{ V} = 0.33\text{ V}$$

c. If A is the cross-sectional area, 0.01 cm^2, taking B_e to be 110 A K^{-2} cm^{-2}, and using the experimental value for the barrier height Φ_B, the saturation current is

$$I_o = AB_eT^2\exp\left(-\frac{\Phi_B}{kT}\right) = (0.01)(110)(300^2)\exp\left(-\frac{0.66\text{ eV}}{0.026\text{ eV}}\right)$$

$$= 9.36\times10^{-7}\text{ A} \qquad \text{or} \qquad 0.94\,\mu\text{A}$$

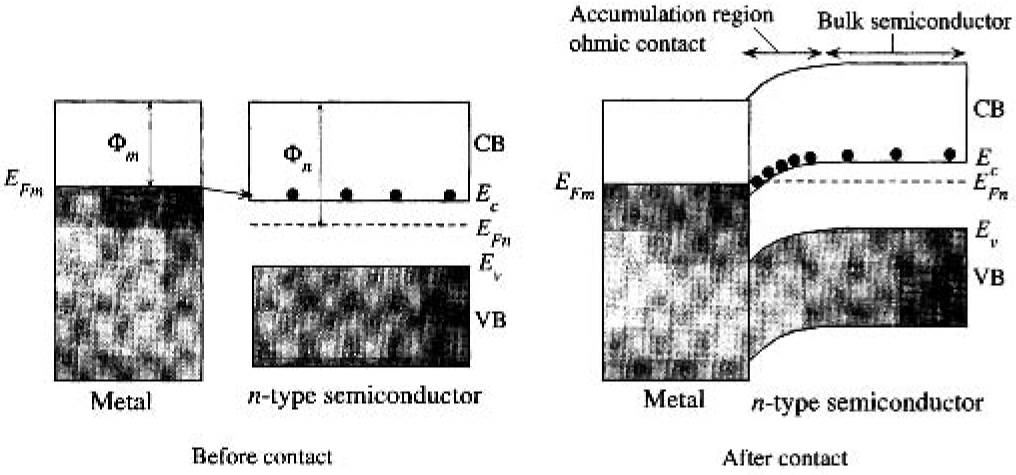
When the applied voltage is V_f, the forward current I_f is

$$I_f = I_o\left[\exp\left(\frac{V_f}{kT}\right)-1\right] = (0.94\,\mu\text{A})\left[\exp\left(\frac{0.2}{0.026}\right)-1\right] = 2.0\text{ mA}$$

5.10 OHMIC CONTACTS AND THERMOELECTRIC COOLERS

An **ohmic contact** is a junction between a metal and a semiconductor that does not limit the current flow. The current is essentially limited by the resistance of the semiconductor outside the contact region rather than the thermal emission rate of carriers across a potential barrier at the contact. In the Schottky diode, the I–V characteristics were determined by the thermal emission rate of carriers across the contact. It should be mentioned that, contrary to intuition, when we talk about an ohmic contact, we do not generally infer a linear I–V characteristic for the ohmic contact itself. We only imply that the contact does not limit the current flow.

Figure 5.43 shows the formation of an ohmic contact between a metal and an n-type semiconductor. The work function of the metal Φ_m is smaller than the work function Φ_n of the semiconductor. There are more energetic electrons in the metal than



Before contact

After contact

Figure 5.43 When a metal with a smaller work function than an n-type semiconductor is put into contact with the n-type semiconductor, the resulting junction is an ohmic contact in the sense that it does not limit the current flow.

in the CB, which means that the electrons (around E_{Fm}) tunnel into the semiconductor in search of lower energy levels, which they find around E_c, as indicated in Figure 5.43. Consequently, many electrons pile in the CB of the semiconductor near the junction. Equilibrium is reached when the accumulated electrons in the CB of the semiconductor prevent further electrons tunneling from the metal. Put more rigorously, equilibrium is reached when the Fermi level is uniform across the whole system from one end to the other.

The semiconductor region near the junction in which there are excess electrons is called the **accumulation region.** To show the increase in n, we draw the semiconductor energy bands bending downward to decrease $E_c - E_{Fn}$, which increases n. Going from the far end of the metal to the far end of the semiconductor, there are always conduction electrons. In sharp contrast, the depletion region of the Schottky junction separates the conduction electrons in the metal from those in the semiconductor. It can be seen from the contact in Figure 5.43 that the conduction electrons immediately on either side of the junction (at E_{Fm} and E_c) have about the same energy and therefore there is no barrier involved when they cross the junction in either direction under the influence of an applied field.

It is clear that the excess electrons in the accumulation region increase the conductivity of the semiconductor in this region. When a voltage is applied to the structure, the voltage drops across the higher resistance region, which is the bulk semiconductor region. Both the metal and the accumulation region have comparatively high concentrations of electrons compared with the bulk of the semiconductor. The current is therefore determined by the resistance of the bulk region. The current density is then simply $J = \sigma\mathcal{E}$ where σ is the conductivity of the semiconductor in the bulk and $\mathcal{E}$ is the applied field in this region.

One of the interesting and important applications of semiconductors is in **thermoelectric,** or **Peltier,** devices, which enable small volumes to be cooled by direct currents. Whenever a dc current flows through a contact between two dissimilar materials, heat is either released or absorbed in the contact region, depending on the direction of the current. Suppose that there is a dc current flowing from an n-type semiconductor to a metal through an ohmic contact, as depicted in Figure 5.44a. Then electrons are flowing from the metal to the CB of the semiconductor. We only consider the contact region where the Peltier effect occurs. Current is carried by electrons near the Fermi level E_{Fm} in the metal. These electrons then cross over into the CB of the semiconductor and when they reach the end of the contact region, their energy is E_c plus average KE (which is $\frac{3}{2}kT$). There is therefore an increase in the average energy $(PE + KE)$ per electron in the contact region. The electron must therefore absorb heat from the environment (lattice vibrations) to gain this energy as it drifts through the junction. Thus, the passage of an electron from the metal to the CB of an n-type semiconductor involves the absorption of heat at the junction.

When the current direction is from the metal to the n-type semiconductor, the electrons flow from the CB of the semiconductor to the Fermi level of the metal as they pass through the contact. Since E_{Fm} is lower than E_c, the passing electron has to lose energy, which it does to lattice vibrations as heat. Thus, the passage of a CB electron from the n-type semiconductor to the metal involves the release of heat at the junction, as indicated in Figure 5.44b.

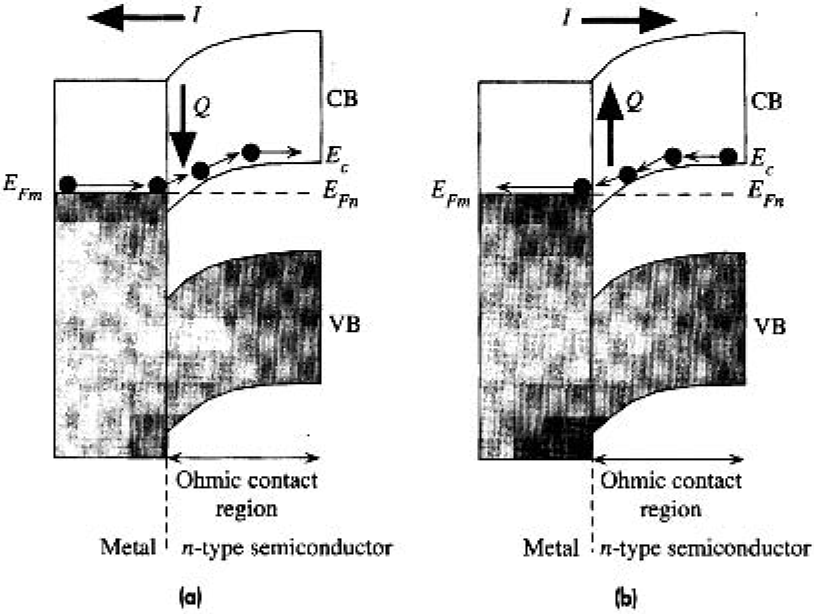


Figure 5.44

(a) Current from an *n*-type semiconductor to the metal results in heat absorption at the junction.

(b) Current from the metal to an *n*-type semiconductor results in heat release at the junction.

It is apparent that depending on the direction of the current flow through a junction between a metal and an *n*-type semiconductor, heat is either absorbed or released at the junction. Although we considered current flow between a metal and an *n*-type semiconductor through an ohmic contact, this thermoelectric effect is a general phenomenon that occurs at a junction between any two dissimilar materials. It is called the **Peltier effect** after its discoverer. In the case of metal–*p*-type semiconductor junctions, heat is absorbed for current flowing from the metal to the *p*-type semiconductor and heat is released in the other direction. Thermoelectric effects occurring at metal–semiconductor junctions are summarized in Figure 5.45. It is important not to confuse the Peltier effect with the Joule heating of the semiconductor and the metal. Joule heating, which we simply call I^2R (or $J^2\rho$) heating, arises from the finite resistivity of the material. It is due to the conduction electrons losing their energy gained from the field to lattice vibrations when they become scattered by such vibrations, as discussed in Chapter 2.

It is self-evident that when a current flows through a semiconductor sample with metal contacts at its ends, as depicted in Figure 5.45, one of the contacts will always absorb heat and the other will always release heat. The contact where heat is absorbed will be cooled and is called the cold junction, whereas the other contact, where heat is released, will warm up and is called the hot junction. One can use the cold junction to

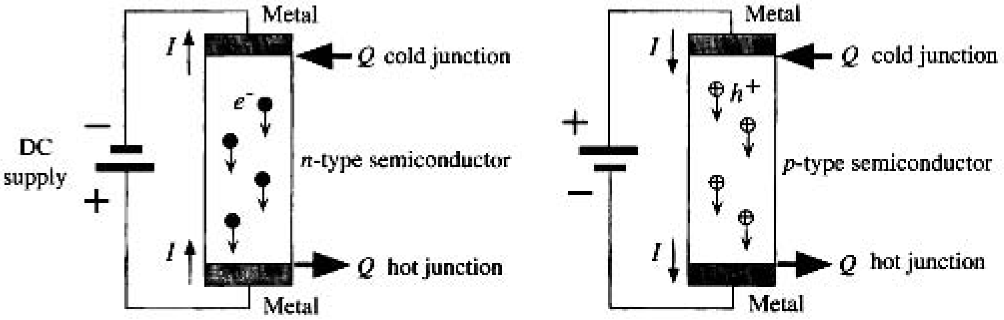


Figure 5.45 When a dc current is passed through a semiconductor to which metal contacts have been made, one junction absorbs heat and cools (the cold junction) and the other releases heat and warms (the hot junction).

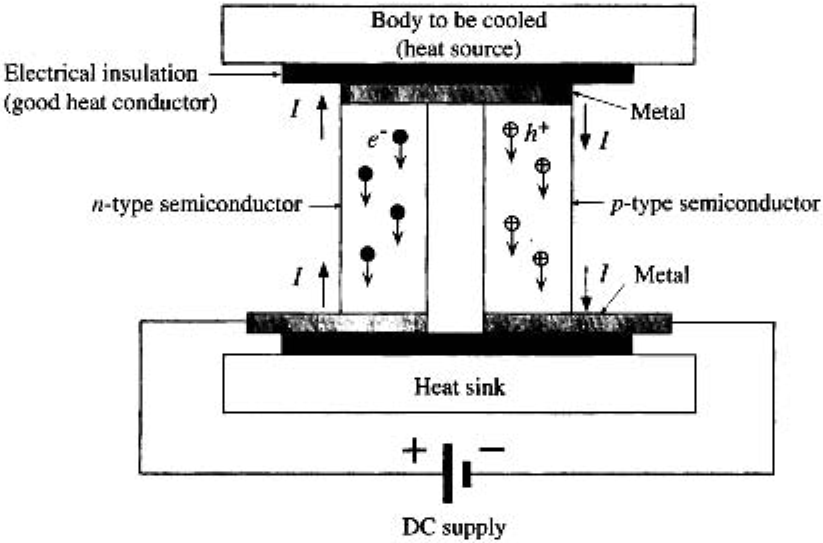


Figure 5.46 Cross section of a typical thermoelectric cooler.

cool another body, providing that the heat generated at the hot junction can be removed from the semiconductor sufficiently quickly to reduce its conduction through the semiconductor to the cold junction. Furthermore, there will always be the Joule heating (I^2R) of the whole semiconductor sample since the bulk will always have a finite resistance.

A simplified schematic diagram of a practical single-element thermoelectric cooling device is shown in Figure 5.46. It uses two semiconductors, one *n*-type and the other *p*-type, each with ohmic contacts. The current direction therefore has opposite thermoelectric effects. On one side, the semiconductors share the same metal

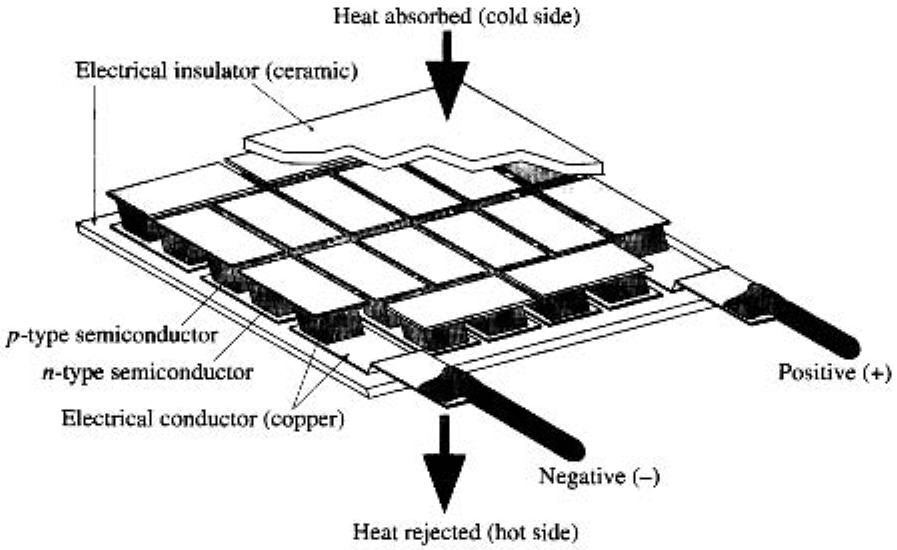


Figure 5.47 Typical structure of a commercial thermoelectric cooler.

electrode. Effectively, the structure is an *n*-type and a *p*-type semiconductor connected in series through a common metal electrode. Typically, either Bi_2Te_3, Bi_2Se_3, or Sb_2Te_3 is used as the semiconductor material with copper usually as the metal electrode.

The current flowing through the *n*-type semiconductor to the common metal electrode causes heat absorption, which cools this junction and hence the metal. The same current then enters the *p*-type semiconductor and causes heat absorption at this junction, which cools the same metal electrode. Thus the common metal electrode is cooled at both ends. The other ends of the semiconductors are hot junctions. They are connected to a large heat sink to remove the heat and thus prevent heat conduction through the semiconductors toward the cold junctions. The other face of the common metal electrode is in contact, through a thin ceramic plate (electrical insulator but thermal conductor), with the body to be cooled. In commercial Peltier devices, many of these elements are connected in series, as illustrated in Figure 5.47, to increase the cooling efficiency.

EXAMPLE 5.20

THE PELTIER COEFFICIENT Consider the motion of electrons across an ohmic contact between a metal and an *n*-type semiconductor and hence show that the rate of heat generation Q' at the contact is approximately

$$Q' = \pm \Pi I$$

where Π, called the **Peltier coefficient** between the two materials, is given by

$$\Pi = \frac{1}{e}\left[(E_c - E_{Fn}) + \frac{3}{2}kT\right]$$

where $E_c - E_{Fn}$ is the energy separation of E_c from the Fermi level in the n-type semiconductor. The sign depends on the convention used for heat liberation or absorption.

SOLUTION

We consider Figure 5.44a, which shows only the ohmic contact region between a metal and an n-type semiconductor when a current is passing through it. The majority of the applied voltage drops across the bulk of the semiconductor because the contact region, or the accumulation region, has an accumulation of electrons in the CB. The current is limited by the bulk resistance of the semiconductor. Thus, in the contact region we can take the Fermi level to be almost undisturbed and hence uniform, $E_{Fm} \approx E_{Fn}$. In the bulk of the metal, a conduction electron is at around E_{Fm} (same as E_{Fn}), whereas just at the end of the contact region in the semiconductor it is at E_c plus an average KE of $\frac{3}{2}kT$. The energy difference is the heat absorbed per electron going through the contact region. Since I/e is the rate at which electrons are flowing through the contact,

$$\text{Rate of energy absorption} = \left[\left(E_c + \frac{3}{2}kT\right) - E_{Fm}\right]\left(\frac{I}{e}\right)$$

or

$$Q' = \left[\frac{(E_c - E_{Fn}) + \frac{3}{2}kT}{e}\right] I = \Pi I$$

so the Peltier coefficient is approximately given by the term in the square brackets. A more rigorous analysis gives Π as

$$\Pi = \frac{1}{e}[(E_c - E_{Fn}) + 2kT]$$

ADDITIONAL TOPICS

5.11 DIRECT AND INDIRECT BANDGAP SEMICONDUCTORS

E–k Diagrams We know from quantum mechanics that when the electron is within a potential well of size L, its energy is quantized and given by

$$E_n = \frac{(\hbar k_n)^2}{2m_e}$$

where the wavevector k_n is essentially a quantum number determined by

$$k_n = \frac{n\pi}{L}$$

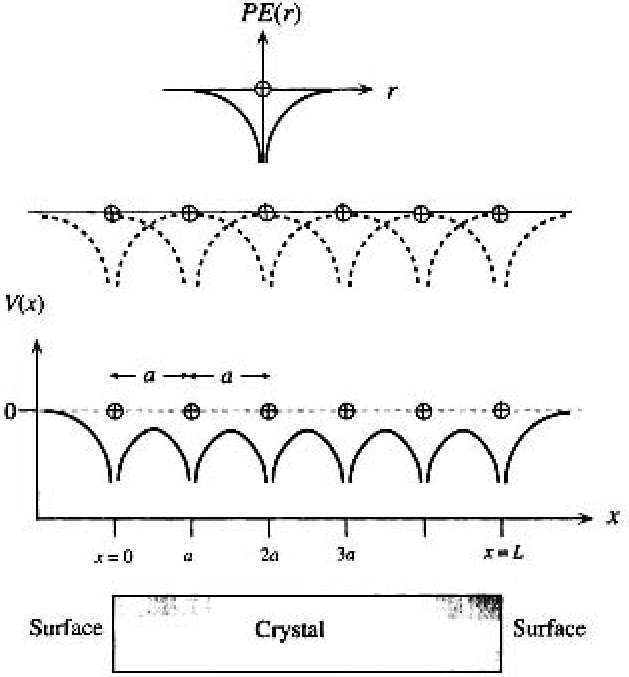
where $n = 1, 2, 3, \ldots$ The energy increases parabolically with the wavevector k_n. We also know that the electron momentum is given by $\hbar k_n$. This description can be used to represent the behavior of electrons in a metal within which their average potential energy can be taken to be roughly zero. In other words, we take $V(x) = 0$ within the metal crystal and $V(x)$ to be large [*e.g.*, $V(x) = V_o$] outside so that the electron is contained within the metal. This is the **nearly free electron model** of a metal that has been quite successful in interpreting many of the properties. Indeed, we were able to calculate the density of states $g(E)$ based on the three-dimensional potential well problem. It is quite obvious that this model is too simple since it does not take into account the actual variation of the electron potential energy in the crystal.

The potential energy of the electron depends on its location within the crystal and is periodic due to the regular arrangement of the atoms. How does a periodic potential energy affect the relationship between E and k? It will no longer simply be $E_n = (\hbar k_n)^2/2m_e$.

To find the energy of the electron in a crystal, we need to solve the Schrödinger equation for a periodic potential energy function in three dimensions. We first consider the hypothetical one-dimensional crystal shown in Figure 5.48. The electron potential energy functions for each atom add to give an overall potential energy function $V(x)$, which is clearly periodic in x with the periodicity of the crystal a. Thus,

Periodic potential energy

$$V(x) = V(x+a) = V(x+2a) = \cdots \qquad [5.71]$$



PE of the electron around an isolated atom.

When *N* atoms are arranged to form the crystal then there is an overlap of individual electron *PE* functions.

PE of the electron, *V*(*x*), inside the crystal is periodic with a period *a*.

Figure 5.48 The electron potential energy (*PE*), *V*(*x*), inside the crystal is periodic with the same periodicity *a* as that of the crystal. Far away outside the crystal, by choice, $V = 0$ (the electron is free and $PE = 0$).

and so on. Our task is therefore to solve the Schrödinger equation

Schrödinger equation

$$\frac{d^2\psi}{dx^2} + \frac{2m_e}{\hbar^2}[E - V(x)]\psi = 0 \tag{5.72}$$

subject to the condition that the potential energy $V(x)$ is periodic in a, that is,

Periodic potential

$$V(x) = V(x + ma) \qquad m = 1, 2, 3, \ldots \tag{5.73}$$

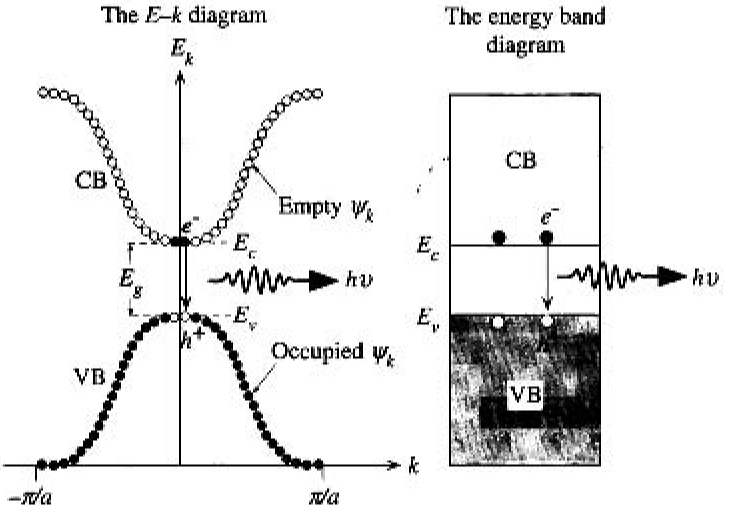
The solution of Equation 5.72 will give the electron wavefunction in the crystal and hence the electron energy. Since $V(x)$ is periodic, we should expect, by intuition at least, the solution $\psi(x)$ to be periodic. It turns out that the solutions to Equation 5.72, which are called **Bloch wavefunctions,** are of the form

Bloch wavefunction

$$\psi_k(x) = U_k(x)\exp(jkx) \tag{5.74}$$

where $U_k(x)$ is a periodic function that depends on $V(x)$ and has the same periodicity a as $V(x)$. The term $\exp(jkx)$, of course, represents a traveling wave. We should remember that we have to multiply this by $\exp(-jEt/\hbar)$, where E is the energy, to get the overall wavefunction $\Psi(x, t)$. Thus the electron wavefunction in the crystal is a traveling wave that is modulated by $U_k(x)$.

There are many such Bloch wavefunction solutions to the one-dimensional crystal, each identified with a particular k value, say k_n, which acts as a kind of quantum number. Each $\psi_k(x)$ solution corresponds to a particular k_n and represents a state with an energy E_k. The dependence of the energy E_k on the wavevector k is what we call the E–k diagram. Figure 5.49 shows a typical E–k diagram for the hypothetical one-dimensional solid for k values in the range $-\pi/a$ to $+\pi/a$. Just as $\hbar k$ is the momentum of a free electron, $\hbar k$ for the Bloch electron is the momentum involved in its interaction with external fields, for example, those involved in the photon absorption process. Indeed, the rate of change of $\hbar k$ is the externally applied force F_{ext} on the electron such as that due to an electric field ($F_{\text{ext}} = e\mathcal{E}$). Thus, for the electron within


The E–k diagram

The energy band diagram

Figure 5.49 The E–k diagram of a direct bandgap semiconductor such as GaAs.

The E–k curve consists of many discrete points, each corresponding to a possible state, wavefunction $\psi_k(x)$, that is allowed to exist in the crystal. The points are so close that we normally draw the E–k relationship as a continuous curve. In the energy range E_v to E_c, there are no points [$\psi_k(x)$ solutions].

the crystal,

$$\frac{d(\hbar k)}{dt} = F_{\text{ext}}$$

and consequently we call $\hbar k$ the **crystal momentum** of the electron.[7]

Inasmuch as the momentum of the electron in the x direction in the crystal is given by $\hbar k$, the E–k diagram is an **energy versus crystal momentum plot.** The states $\psi_k(x)$ in the lower E–k curve constitute the wavefunctions for the valence electrons and thus correspond to the states in the VB. Those in the upper E–k curve, on the other hand, correspond to the states in the conduction band (CB) since they have higher energies. All the valence electrons at absolute zero of temperature therefore fill the states, particular k_n values, in the lower E–k diagram.

It should be emphasized that an E–k curve consists of many discrete points, each corresponding to a possible state, wavefunction $\psi_k(x)$, that is allowed to exist in the crystal. The points are so close that we draw the E–k relationship as a continuous curve. It is clear from the E–k diagram that there is a range of energies, from E_v to E_c, for which there are no solutions to the Schrödinger equation and hence there are no $\psi_k(x)$ with energies in E_v to E_c. Furthermore, we also note that the E–k behavior is not a simple parabolic relationship except near the bottom of the CB and the top of the VB.

Above absolute zero of temperature, due to thermal excitation, however, some of the electrons from the top of the valence band will be excited to the bottom of the conduction band. According to the E–k diagram in Figure 5.49, when an electron and hole recombine, the electron simply drops from the bottom of the CB to the top of the VB without any change in its k value, so this transition is quite acceptable in terms of momentum conservation. We should recall that the momentum of the emitted photon is negligible compared with the momentum of the electron. The E–k diagram in Figure 5.49 is therefore for a **direct bandgap semiconductor.**

The simple E–k diagram sketched in Figure 5.49 is for the hypothetical one-dimensional crystal in which each atom simply bonds with two neighbors. In real crystals, we have a three-dimensional arrangement of atoms with $V(x, y, z)$ showing periodicity in more than one direction. The E–k curves are then not as simple as that in Figure 5.49 and often show unusual features. The E–k diagram for GaAs, which is shown in Figure 5.50a, as it turns out, has main features that are quite similar to that sketched in Figure 5.49. GaAs is therefore a direct bandgap semiconductor in which electron–hole pairs can recombine directly and emit a photon. It is quite apparent that light emitting devices use direct bandgap semiconductors to make use of direct recombination.

---

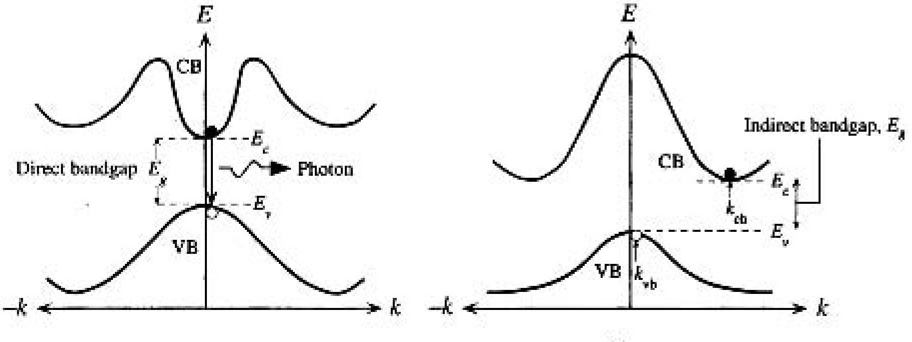
[7] The actual momentum of the electron, however, is not $\hbar k$ because

$$\frac{d(\hbar k)}{dt} \neq F_{\text{external}} + F_{\text{internal}}$$

where $F_{\text{external}} + F_{\text{internal}}$ are all forces acting on the electron. The true momentum p_e satisfies

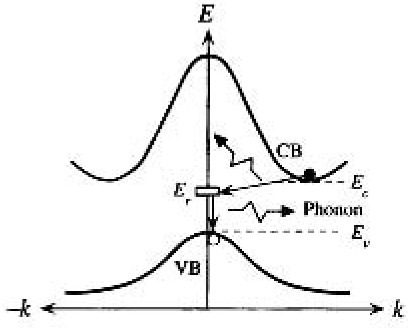
$$\frac{dp_e}{dt} = F_{\text{external}} + F_{\text{internal}}$$

However, as we are interested in interactions with external forces such as an applied field, we treat $\hbar k$ as if it were the momentum of the electron in the crystal and use the name **crystal momentum.**



(a) GaAs

(b) Si



(c) Si with a recombination center

Figure 5.50

(a) In GaAs the minimum of the CB is directly above the maximum of the VB. GaAs is therefore a direct bandgap semiconductor.

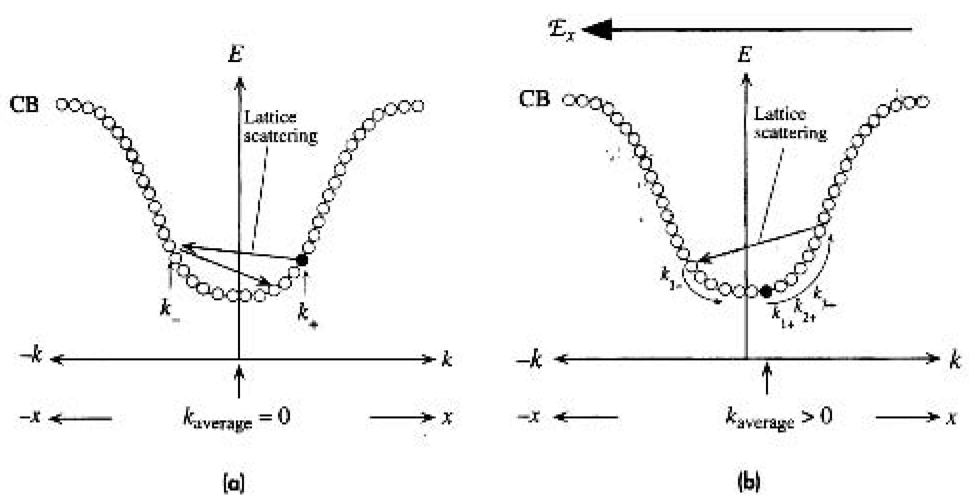
(b) In Si, the minimum of the CB is displaced from the maximum of the VB and Si is an indirect bandgap semiconductor.

(c) Recombination of an electron and a hole in Si involves a recombination center.

In the case of Si, the diamond crystal structure leads to an E–k diagram that has the essential features depicted in Figure 5.50b. We notice that the minimum of the CB is not directly above the maximum of the VB. An electron at the bottom of the CB therefore cannot recombine directly with a hole at the top of the VB because, for the electron to fall down to the top of the VB, its momentum must change from k_{cb} to k_{vb}, which is not allowed by the law of conservation of momentum. Thus direct electron–hole recombination does not take place in Si and Ge. The recombination process in these elemental semiconductors occurs via a recombination center at an energy level E_r. The electron is captured by the defect at E_r, from where it can fall down into the top of the VB. The indirect recombination process is illustrated in Figure 5.50c. The energy of the electron is lost by the emission of phonons, that is, lattice vibrations. The E–k diagram in Figure 5.50b for Si is an example of an **indirect bandgap semiconductor.**

In some indirect bandgap semiconductors such as GaP, the recombination of the electron with a hole at certain recombination centers results in photon emission. The E–k diagram is similar to that shown in Figure 5.50c except that the recombination centers at E_r are generated by the purposeful addition of nitrogen impurities to GaP. The electron transition from E_r to E_v involves photon emission.

Electron Motion and Drift We can understand the response of a conduction band electron to an applied external force, for example, an applied field, by examining the E–k diagram. Again, for simplicity, we consider the one-dimensional crystal. The



(a)

(b)

Figure 5.51

(a) In the absence of a field, over a long time, the average of all k values is zero; there is no net momentum in any one particular direction.

(b) In the presence of a field in the $-x$ direction, the electron accelerates in the $+x$ direction increasing its k value along x until it is scattered to a random k value. Over a long time, the average of all k values is along the $+x$ direction. Thus the electron drifts along $+x$.

electron is wandering around the crystal quite randomly due to scattering from lattice vibrations. Thus the electron moves with a certain k value in the $+x$ direction, say k_+, as illustrated in the E–k diagram of Figure 5.51a. When it is scattered by a lattice vibration, its k value changes, perhaps to k_-, which is also shown in Figure 5.51a. This process of k changing randomly from one scattering to another scattering process continues all the time, so over a long time the average value of k is zero; that is, average k_+ is the same as average k_-.

When an electric field is applied, say in the $-x$ direction, then the electron gains momentum in the $+x$ direction from the force of the field $e\mathcal{E}_x$. With time, while the electron is not scattered, it moves up in the E–k diagram from k_{1+} to k_{2+} to k_{3+} and so on until a lattice vibration randomly scatters the electron to say k_{1-} (or to some other random k value) as shown in Figure 5.51b. Over a long time, the average of all k_+ is no longer equal to the average of all k_- and there is a net momentum in the $+x$ direction, which is tantamount to a drift in the same direction.

Effective Mass The usual definition of inertial mass of a particle in classical physics is based on

$$\text{Force} = \text{Mass} \times \text{Acceleration}$$

$$F = ma$$

When we treat the electron as a wave within the semiconductor crystal, we have to determine whether we can still, in some way, use the convenient classical $F = ma$ relation to describe the motion of an electron under an applied force such as $e\mathcal{E}_x$ and, if so, what the apparent mass of the electron in the crystal should be.

We will evaluate the velocity and acceleration of the electron in the CB in response to an electric field $\mathcal{E}_x$ along $-x$ that imposes an external force $F_{\text{ext}} = e\mathcal{E}_x$ in the $+x$ direction, as shown in Figure 5.51b. Our treatment will make use of the quantum mechanical E–k diagram.

Since we are treating the electron as a wave, we have to evaluate the group velocity v_g, which, by definition, is $v_g = d\omega/dk$. We know that the time dependence of the wavefunction is $\exp(-jEt/\hbar)$ where the energy $E = \hbar\omega$ (ω is an "angular frequency" associated with the wave motion of the electron). Both E and ω depend on k. Thus, the group velocity is

Electron's group velocity

$$v_g = \frac{1}{\hbar}\frac{dE}{dk} \tag{5.75}$$

Thus the group velocity is determined by the **gradient** of the E–k curve. In the presence of an electric field, the electron experiences a force $F_{\text{ext}} = e\mathcal{E}_x$ from which it gains energy and moves up in the E–k diagram until, later on, it collides with a lattice vibration, as shown in Figure 5.51b. During a small time interval δt between collisions, the electron moves a distance $v_g\,\delta t$ and hence gains energy δE, which is

$$\delta E = F_{\text{ext}} v_g\,\delta t \tag{5.76}$$

To find the acceleration of the electron and the effective mass, we somehow have to put this equation into a form that looks like $F_{\text{ext}} = m_e a$, where a is the acceleration. From Equation 5.76, the relationship between the external force and energy is

$$F_{\text{ext}} = \frac{1}{v_g}\frac{dE}{dt} = \hbar\frac{dk}{dt} \tag{5.77}$$

where we used Equation 5.75 for v_g in Equation 5.76. Equation 5.77 is the reason for interpreting $\hbar k$ as the **crystal momentum** inasmuch as the rate of change of $\hbar k$ is the externally applied force.

The acceleration a is defined as dv_g/dt. We can use Equation 5.75,

$$a = \frac{dv_g}{dt} = \frac{d\left[\dfrac{1}{\hbar}\dfrac{dE}{dk}\right]}{dt} = \frac{1}{\hbar}\frac{d^2E}{dk^2}\frac{dk}{dt} \tag{5.78}$$

From Equation 5.78, we can substitute for dk/dt in Equation 5.77, which is then a relationship between F_{ext} and a of the form

External force and acceleration

$$F_{\text{ext}} = \frac{\hbar^2}{\left[\dfrac{d^2E}{dk^2}\right]}a \tag{5.79}$$

We know that the response of a free electron to the external force is $F_{\text{ext}} = m_e a$, where m_e is its mass in vacuum. Therefore it is quite clear from Equation 5.79 that the **effective mass** of the electron in the crystal is

Effective mass

$$m_e^* = \hbar^2\left[\frac{d^2E}{dk^2}\right]^{-1} \tag{5.80}$$

Thus, the electron responds to an external force and moves as if its mass were given by Equation 5.80. The effective mass obviously depends on the E–k relationship, which in turn depends on the crystal symmetry and the nature of bonding between the atoms. Its value is different for electrons in the CB and for those in the VB, and moreover, it depends on the energy of the electron since it is related to the curvature of the E–k behavior (d^2E/dk^2). Further, it is clear from Equation 5.80 that the effective mass is a quantum mechanical quantity inasmuch as the E–k behavior is a direct consequence of the application of quantum mechanics (the Schrödinger equation) to the electron in the crystal.

It is interesting that, according to Equation 5.80, when the E–k curve is a downward concave as at the top of a band (*e.g.*, Figure 5.49), the effective mass of an electron at these energies in a band is then negative. What does a negative effective mass mean? When the electron moves up on the E–k curve by gaining energy from the field, it actually decelerates, that is, moves more slowly. Its acceleration is therefore in the opposite direction to an electron at the bottom of the band. Electrons in the CB are at the bottom of a band, so their effective masses are positive quantities. At the top of a valence band, however, we have plenty of electrons. These electrons have negative effective masses and under the action of a field, they decelerate. Put differently, they accelerate in the opposite direction to the applied external force F_{ext}. It turns out that we can describe the collective motion of these electrons near the top of a band by considering the motion of a few holes with positive masses.

It should be mentioned that Equation 5.80 defines the meaning of the effective mass in quantum mechanical terms. Its usefulness as a concept lies in the fact that we can measure it experimentally, for example, by cyclotron resonance experiments, and have actual values for it. This means we can simply replace m_e by m_e^* in equations that describe the effect of an external force on electron transport in semiconductors.

Holes To understand the concept of a hole, we consider the E–k curve corresponding to energies in the VB, as shown in Figure 5.52a. If all the states are filled, then there are no empty states for the electrons to move into and consequently an electron cannot gain energy from the field. For each electron moving in the positive x direction with a momentum $\hbar k_+$, there is a corresponding electron with an equal and opposite momentum $\hbar k_-$, so there is no net motion. For example, the electron at b is moving toward the

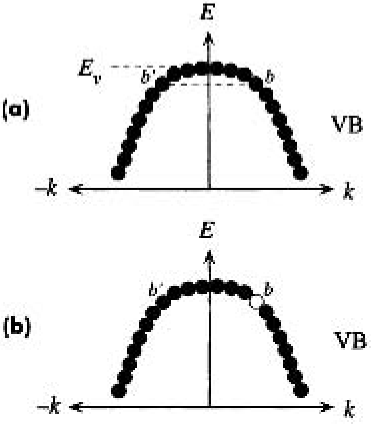


Figure 5.52

(a) In a full valence band, there is no net contribution to the current. There are equal numbers of electrons (*e.g.*, at b and b') with opposite momenta.

(b) If there is an empty state (*hole*) at b at the top of the band, then the electron at b' contributes to the current.

EXAMPLE 5.22

CURRENT DUE TO A MISSING ELECTRON IN THE VB First, let us consider a completely full valence band that contains, say, N electrons. $N/2$ of these are moving with momentum in the $+x$, and $N/2$ in the $-x$ direction. Suppose that the crystal is unit volume. An electron with charge $-e$ moving with a group velocity $\mathbf{v}_{gi}$ contributes to the current by an amount $-e\mathbf{v}_{gi}$. We can determine the current density $\mathbf{J}_N$ due to the motion of all the electrons (N of them) in the band,

$$\mathbf{J}_N = -e\sum_{i=1}^{N} \mathbf{v}_{gi} = 0$$

$\mathbf{J}_N$ is zero because for each value of $\mathbf{v}_{gi}$, there is a corresponding velocity equal in magnitude but opposite in direction (b and b' in Figure 5.52a). Our conclusion from this is that the contribution to the current density from a full valence band is nil, as we expect.

Suppose now that the jth electron is missing (b in Figure 5.52b). The net current density is due to $N - 1$ electrons in the band, so

$$\mathbf{J}_{N-1} = -e\sum_{i=1,i\neq j}^{N} \mathbf{v}_{gi} \qquad [5.81]$$

where the summation is for $i = 1$ to N and $i \neq j$ (jth electron is missing). We can write the sum as summation to N including the jth electron and minus the missing jth electron contribution,

$$\mathbf{J}_{N-1} = -e\sum_{i=1}^{N} \mathbf{v}_{gi} - (-e\mathbf{v}_{gj})$$

that is,

$$\mathbf{J}_{N-1} = +e\mathbf{v}_{gj} \qquad [5.82]$$

where we used $\mathbf{J}_N = 0$. We see that when there is a missing electron, there is a net current due to that empty state (jth). The current appears as the motion of a charge $+e$ with a velocity $\mathbf{v}_{gj}$, where $\mathbf{v}_{gj}$ is the group velocity of the missing electron. In other words, the current is due to the motion of a positive charge $+e$ at the site of the missing electron at k_j, which is what we call a hole. One should note that Equation 5.81 describes the current by considering the motions of *all* the $N - 1$ electrons, whereas Equation 5.82 describes the same current by simply considering the missing electron as if it were a positively charged particle ($+e$) moving with a velocity equal to that of the missing electron. Equation 5.82 is the convenient description universally adopted for a valence band containing missing electrons.

5.12 INDIRECT RECOMBINATION

We consider the recombination of minority carriers in an extrinsic indirect bandgap semiconductor such as Si or Ge. As an example, we consider the recombination of electrons in a p-type semiconductor. In an indirect bandgap semiconductor, the recombination mechanism involves a recombination center, a third body that may be a crystal defect or an impurity, in the recombination process to satisfy the requirements of conservation of momentum. We can view the recombination process as follows. Recombination occurs when an electron is captured by the recombination center at the energy level E_r. As soon as the electron is captured, it will recombine with a hole because holes are abundant in a p-type semiconductor. In other words, since there are many majority carriers, the limitation on the rate of recombination is the actual capture of the minority carrier by the center. Thus, if τ_e is the electron recombination time, since the electrons will have to be captured by the centers, τ_e is given by

$$\tau_e = \frac{1}{S_r N_r v_{\mathrm{th}}} \qquad [5.83]$$

where S_r is the capture (or recombination) cross section of the center, N_r is the concentration of centers, and v_{th} is the mean speed of the electron that you may take as its effective thermal velocity.

Equation 5.83 is valid under small injection conditions, that is, $p_{po} \gg n_p$. There is a more general treatment of indirect recombination called the Shockley–Read statistics of indirect recombination and generation, which is treated in more advanced semiconductor physics textbooks. That theory eventually arrives at Equation 5.83 for low-level injection conditions. We derived Equation 5.83 from a purely physical reasoning.

Gold is frequently added to silicon to aid recombination. It is found that the minority carrier recombination time is inversely proportional to the gold concentration, following Equation 5.83.

5.13 AMORPHOUS SEMICONDUCTORS

Up to now we have been dealing with crystalline semiconductors, those crystals that have perfect periodicity and are practically flawless unless purposefully doped for use in device applications. They are used in numerous solid-state devices including large-area solar cells. Today's microprocessor uses a single crystal of silicon that contains millions of transistors; indeed, we are heading for the 1-billion-transistor chip. There are, however, various applications in electronics that require inexpensive large-area devices to be fabricated and hence require a semiconductor material that can be prepared in a large area. In other applications, the semiconductor material is required to be deposited as a film on a flexible substrate for use as a sensor. Best known examples of large-area devices are flat panel displays based on thin-film transistors (TFTs), inexpensive solar cells, photoconductor drums (for printing and photocopying), image sensors, and newly developed X-ray image detectors. Many of these applications typically use hydrogenated amorphous silicon, a-Si:H.

A distinctive property of an electron in a crystalline solid is that its wavefunction is a traveling wave, a Bloch wave, ψ_k, as in Equation 5.74. The Bloch wavefunction is a consequence of the periodicity of an electron's potential energy PE, $V(x)$, within the crystal. One can view the electron's motion as tunneling through the periodic potential energy hills. The wavefunctions ψ_k form **extended states** because they *extend* throughout the whole crystal. The electron belongs to the whole crystal, and there is an equal probability of finding an electron in any unit cell. The wavevector k in this traveling wave ψ_k acts as a quantum number. There are many discrete k_n values, which form a nearly continuous set of k values (see Figure 5.49). We can describe the interaction of the electron with an external force, or with photons and phonons, by assigning a momentum $\hbar k$ to the electron, which is called the electron's crystal momentum.

The electron's wavefunction ψ_k is frequently scattered by lattice vibrations (or by defects or impurities) from one k-value to another, *e.g.*, from ψ_k to $\psi_{k'}$. The scattering of the wavefunction imposes a mean free path ℓ on the electron's motion, that is, a mean distance over which a wave can travel without being scattering. Over the distance ℓ, the wavefunction is coherent, that is, well defined and predictable as a traveling Bloch wave; ℓ is also known as the coherence length of the wavefunction. The mobility is determined by the mean free path ℓ, which at room temperature is typically of the order of several hundreds of mean interatomic separations. The crystal periodicity and the unit cell atomic structure control the types of Bloch wave solutions one can obtain to the Schrödinger equation. The solutions allow the electron energy E to be examined as a function of k (or momentum $\hbar k$) and these $E - k$ diagrams categorize crystalline semiconductors into two classes: direct bandgap (GaAs type) and indirect bandgap (Si type) semiconductors.

Hydrogenated amorphous silicon (a-Si:H) is the noncrystalline form of silicon in which the structure has no long-range order but only short-range order; that is, we can only identify the nearest neighbors of a given atom. Each Si atom has four neighbors as in the crystal, but there is no periodicity or long-range order as illustrated in Figure 1.59. Without the hydrogen, pure a-Si would have dangling bonds. In such a structure sometimes a Si atom would not be able to find a fourth neighboring Si atom to bond with and will be left with a dangling bond as in Figure 1.59b. The hydrogen in the structure (~10 percent) passivates (*i.e.*, neutralizes) the unsatisfied ("dangling") bonds inherent in a noncrystalline structure and so reduces the density of dangling bonds or defects. a-Si:H belongs to a class of solids called **amorphous semiconductors** that do not follow typical crystalline concepts such as Bloch wavefunctions. First, due to the lack of periodicity, we cannot describe the electron as a Bloch wave. Consequently, we cannot use a wavevector k, and hence $\hbar k$, to describe the electron's motion. These semiconductors however do have a short-range order and also possess an energy bandgap that separates a conduction band and a valence band. A window glass has a noncrystalline structure but also has a bandgap, which makes it transparent. Photons with energies less than the bandgap energy can pass through the window glass.

The examination of the structure of a-Si:H in Figure 1.59c should make it apparent that the potential energy $V(x)$ of the electron in this noncrystalline structure fluctuates randomly from site to site. In some cases, the local changes in $V(x)$ can be quite strong, forming effective local PE wells (obviously finite wells). Such fluctuations in the PE within the solid can capture or trap electrons, that is, localize electrons at certain spatial locations. A localized electron will have a wavefunction that resembles the wavefunction in the hydrogen atom, so the probability of finding the electron is localized to the site. Such locations that can trap electrons, give them localized wavefunctions, are called **localized states.** The amorphous structure also has electrons that possess extended wavefunctions; that is, they belong to the whole solid. These extended wavefunctions are distinctly different than those in the crystal because they have very short coherence lengths due to the random potential fluctuations; the electron is scattered from site to site and hence the mean free path is of the order of a few atomic spacings. The extended wavefunction has random phase fluctuations. Figure 5.53 compares localized and extended wavefunctions in an amorphous semiconductor.

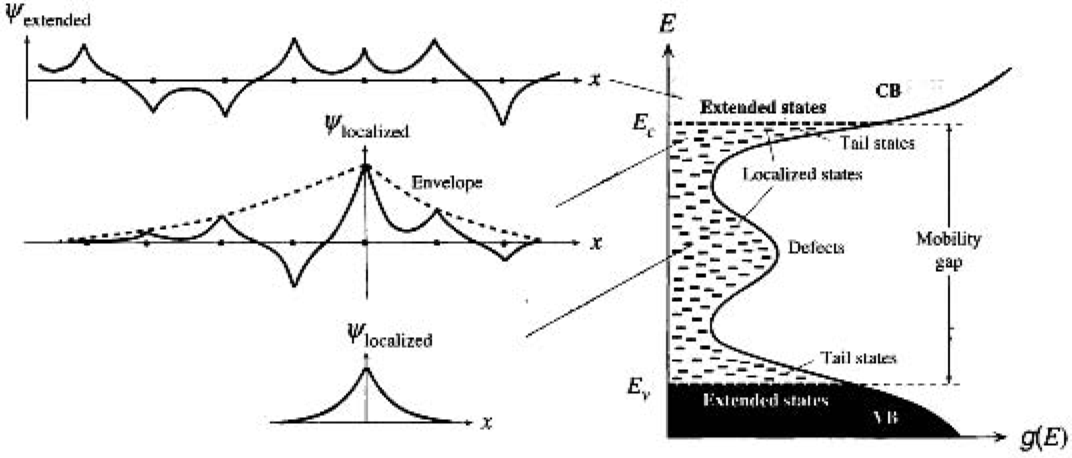


Figure 5.53 Schematic representation of the density of states $g(E)$ versus energy E for an amorphous semiconductor and the associated electron wavefunctions for an electron in the extended and localized states.

Electronic properties of all amorphous semiconductors can be explained in terms of the energy distribution of their density of states (DOS) function, $g(E)$. The DOS function has well-defined energies E_v and E_c that separate extended states from localized states as in Figure 5.53. There is a distribution of localized states, called **tail states** below E_c and above E_v. The usual **bandgap** $E_c - E_v$ is called the **mobility gap.** The reason is that there is a change in the character of charge transport, and hence in the carrier mobility, in going from extended states above E_c to localized states below E_c.

Electron transport above E_c in the conduction band is dominated by scattering from random potential fluctuations arising from the disordered nature of the structure. The electrons are scattered so frequently that their effective mobility is much less than what it is in crystalline Si: μ_e in a-Si:H is typically 5–10 cm^2 V^{-1} s^{-1} whereas it is 1400 cm^2 V^{-1} s^{-1} in a single crystal Si. Electron transport below E_c, on the other hand, requires an electron to jump, or hop, from one localized state to another, aided by thermal vibrations of the lattice, in an analogous way to the diffusion of an interstitial impurity in a crystal. We know from Chapter 1 that the jump or diffusion of the impurity is a thermally activated process because it relies on the thermal vibrations of all the crystal atoms to occasionally give the impurity enough energy to make that jump. The electron's mobility associated with this type of hopping motion among localized states is thermally activated, and its value is small. Thus, there is a change in the electron mobility across E_c, which is called the conduction band **mobility edge.**

The localized states (frequently simply called *traps*) between E_v and E_c have a profound effect on the overall electronic properties. The tail localized states are a direct result of the structural disorder that is inherent in noncrystalline solids, variations in the bond angles and length. Various prominent peaks and features in the DOS within the mobility gap have been associated with possible structural defects, such as under- and overcoordinated atoms in the structure, dangling bonds, and dopants. Electrons that drift in the conduction band can fall into localized states and become immobilized (trapped) for a while. Thus, electron transport in a-Si:H occurs by multiple trapping in shallow localized states. The effective electron drift mobility in a-Si:H is therefore reduced to ~1 cm^2 V^{-1} s^{-1}. Low drift mobilities obviously prevent the use of amorphous semiconductor materials in high-speed or high-gain electronic applications. Nonetheless, low-speed electronics is just as important as high-speed electronics in the electronics market in such applications as flat panel displays, solar cells, and image sensors. A low-speed flat panel display made from hydrogenated amorphous silicon (a-Si:H) TFTs costs very roughly the same as a high-speed crystalline Si microchip that runs the CPU.

CD Selected Topics and Solved Problems

Selected Topics

Hall Effect in Semiconductors
Transferred Electron Devices: Gunn Effect
Elements of Photoconductivity
Thermoelectric Effects in Semiconductors: Voltage Drift in Semiconductor Devices

Solved Problems

Piezoresistance: Pressure Sensors and Strain Gauges
Hall Effect
Ionization Region in Doped Semiconductors
Compensation Doping of Semiconductors
Electron-Hole Recombination in Semiconductors and Photoconductivity

DEFINING TERMS

Acceptor atoms are dopants that have one less valency than the host atom. They therefore accept electrons from the VB and thereby create holes in the VB, which leads to a $p > n$ and hence to a p-type semiconductor.

Average energy of an electron in the CB is $\frac{3}{2}kT$ as if the electrons were obeying Maxwell–Boltzmann statistics. This is only true for a nondegenerate semiconductor.

Bloch wave refers to an electron wavefunction of the form $\psi_k = U_k(x)\exp(jkx)$, which is a traveling wave that is modulated by a function $U_k(x)$ that has the periodicity of the crystal. The Bloch wavefunction is a consequence of the periodicity of an electron's potential energy within the crystal.

Compensated semiconductor contains both donors and acceptors in the same crystal region that compensate for each other's effects. For example, if there are more donors than acceptors, $N_d > N_a$, then some of the electrons released by donors are captured by acceptors and the net effect is that $N_d - N_a$ number of electrons per unit volume are left in the CB.

Conduction band (CB) is a band of energies for the electron in a semiconductor where it can gain energy from an applied field and drift and thereby contribute to electrical conduction. The electron in the CB behaves as if it were a "free" particle with an effective mass m_e^*.

Degenerate semiconductor has so many dopants that the electron concentration in the CB, or hole concentration in the VB, is comparable with the density of states in the band. Consequently, the Pauli exclusion principle is significant and Fermi–Dirac statistics must be used. The Fermi level is either in the CB for a n^+-type degenerate or in the VB for a p^+-type degenerate semiconductor. The superscript + indicates a heavily doped semiconductor.

Diffusion is a random process by which particles move from high-concentration regions to low-concentration regions.

Donor atoms are dopants that have a valency one more than the host atom. They therefore donate electrons to the CB and thereby create electrons in the CB, which leads to $n > p$ and hence to an n-type semiconductor.

Effective density of states (N_c) at the CB edge is a quantity that represents all the states in the CB per unit volume as if they were all at E_c. Similarly, N_v at the

VB edge is quantity that represents all the states in the VB per unit volume as if they were all at E_v.

Effective mass (m_e^*) of an electron is a quantum mechanical quantity that behaves like the inertial mass in classical mechanics, $F = ma$, in that it measures the object's inertial resistance to acceleration. It relates the acceleration a of an electron in a crystal to the applied external force F_{ext} by $F_{ext} = m_e^* a$. The external force is most commonly the force of an electric field $e\mathcal{E}$ and excludes all internal forces within the crystal.

Einstein relation relates the diffusion coefficient D and the drift mobility μ of a given species of charge carriers through $(D/\mu) = (kT/e)$.

Electron affinity (χ) is the energy required to remove an electron from E_c to the vacuum level.

Energy of the electron in the crystal, whether in the CB or VB, depends on its momentum $\hbar k$ through the E–k behavior determined by the Schrödinger equation. E–k behavior is most conveniently represented graphically through E–k diagrams. For example, for an electron at the bottom of the CB, E increases as $(\hbar k)^2/m_e^*$ where $\hbar k$ is the momentum and m_e^* is the effective mass of the electron, which is determined from the E–k behavior.

Excess carrier concentration is the excess concentration *above* the thermal equilibrium value. Excess carriers are generated by an external excitation such as photogeneration.

Extended state refers to an electron wavefunction ψ_k whose magnitude does not decay with distance; that is, it is extended in the crystal. An extended wavefunction of an electron in a *crystal* is a **Bloch wave,** that is, $\psi_k = U_k(x)\exp(jkx)$, which is a traveling wave that is modulated by a function $U_k(x)$ that has the periodicity of the crystal. There is an equal probability of finding an electron in any unit cell of the crystal. Scattering of an electron in the crystal by lattice vibrations or impurities, etc., corresponds to the electron being scattered from one ψ_k to another $\psi_{k'}$, *i.e.* a change in the wavevector from $\mathbf{k}$ to $\mathbf{k}'$. Valence and conduction bands in a crystal have extended states.

Extrinsic semiconductor is a semiconductor that has been doped so that the concentration of one type of charge carrier far exceeds that of the other. Adding donor impurities releases electrons into the CB and n far exceeds p; thus, the semiconductor becomes n-type.

Fermi energy or level (E_F) may be defined in several equivalent ways. The Fermi level is the energy level corresponding to the energy required to remove an electron from the semiconductor; there need not be any actual electrons at this energy level. The energy needed to remove an electron defines the work function Φ. We can define the Fermi level to be Φ below the vacuum level. E_F can also be defined as that energy value below which all states are full and above which all states are empty at absolute zero of temperature. E_F can also be defined through a difference. A difference in the Fermi energy ΔE_F in a system is the external electrical work done per electron either on the system or by the system such as electrical work done when a charge e moves through an electrostatic PE difference is $e\Delta V$. It can be viewed as a fundamental material property.

Intrinsic carrier concentration (n_i) is the electron concentration in the CB of an intrinsic semiconductor. The hole concentration in the VB is equal to the electron concentration.

Intrinsic semiconductor has an equal number of electrons and holes due to thermal generation across the bandgap E_g. It corresponds to a pure semiconductor crystal in which there are no impurities or crystal defects.

Ionization energy is the energy required to ionize an atom, for example, to remove an electron.

Ionized impurity scattering limited mobility is the mobility of the electrons when their motion is limited by scattering from the ionized impurities in the semiconductor (*e.g.*, donors and acceptors).

$\boldsymbol{k}$ is the wavevector of the electron's wavefunction. In a crystal the electron wavefunction, $\psi_k(x)$ is a *modulated traveling wave* of the form

$$\psi_k(x) = U_k(x)\exp(jkx)$$

where k is the wavevector and $U_k(x)$ is a periodic function that depends on the PE of interaction between the electron and the lattice atoms. k identifies all possible states $\psi_k(x)$ that are allowed to exist in the crystal. $\hbar k$ is called the *crystal momentum* of the electron as its rate of change is the externally applied force to the electron, $d(\hbar k)/dt = F_{\text{external}}$.

Lattice-scattering-limited mobility is the mobility of the electrons when their motion is limited by scattering from thermal vibrations of the lattice atoms.

Localized state refers to an electron wavefunction $\psi_{\text{localized}}$ whose magnitude, or the envelope of the wavefunction, decays with distance, which localizes the electron to a spatial region in the semiconductor. For example, a 1*s*-type wavefunction of the form $\psi_{\text{localized}} \propto \exp(-\alpha r)$, where r is the distance measured *from some* center at $r = 0$, and α is a positive constant, would represent a localized state centered at $r = 0$.

Majority carriers are electrons in an *n*-type and holes in a *p*-type semiconductor.

Mass action law in semiconductor science refers to the law $np = n_i^2$, which is valid under thermal equilibrium conditions and in the absence of external biases and illumination.

Minority carrier diffusion length (L) is the mean distance a minority carrier diffuses before recombination, $L = \sqrt{D\tau}$, where D is the diffusion coefficient and τ is the minority carrier lifetime.

Minority carrier lifetime (τ) is the mean time for a minority carrier to disappear by recombination. $1/\tau$ is the mean probability per unit time that a minority carrier recombines with a majority carrier.

Minority carriers are electrons in a *p*-type and holes in an *n*-type semiconductor.

Nondegenerate semiconductor has electrons in the CB and holes in the VB that obey Boltzmann statistics. Put differently, the electron concentration n in the CB is much less than the effective density of states N_c and similarly $p \ll N_v$. It refers to a semiconductor that has not been heavily doped so that these conditions are maintained; typically, doping concentrations are less than 10^{18} cm^{-3}.

Ohmic contact is a contact that can supply charge carriers to a semiconductor at a rate determined by charge transport through the semiconductor and not by the contact properties itself. Thus the current is limited by the conductivity of the semiconductor and not by the contact.

Peltier effect is the phenomenon of heat absorption or liberation at the contact between two dissimilar materials as a result of a dc current passing through the junction. The rate of heat generation Q' is proportional to the dc current I passing through the contact so that $Q' = \pm\Pi I$, where Π is called the Peltier coefficient and the sign depends on whether heat is absorbed or released.

Phonon is a quantum of energy associated with the vibrations of the atoms in the crystal, analogous to the photon. A phonon has an energy $\hbar\omega$ where ω is the frequency of the lattice vibration.

Photoconductivity is the change in the conductivity from dark to light, $\sigma_{\text{light}} - \sigma_{\text{dark}}$.

Photogeneration is the excitation of an electron into the CB by the absorption of a photon. If the photon is absorbed by an electron in the VB, then its excitation to the CB will generate an EHP.

Photoinjection is the photogeneration of carriers in the semiconductor by illumination. Photogeneration may be VB to CB excitation, in which case electrons and holes are generated in pairs.

Piezoresistivity is the change in the resistivity of a semiconductor due to an applied mechanical stress σ_m. **Elastoresistivity** refers to the change in the resistivity due to an induced strain in the substance. Application of stress normally leads to strain, so piezoresistivity and elastoresistivity refer to the same phenomenon. In simple terms, the change in the resistivity may be due to a change in the concentration of carriers or due to a change in the drift mobility of the carriers. The fractional change in the resistivity $\delta\rho/\rho$ is proportional to the applied stress σ_m, and the proportionality constant is called the **piezoresistive coefficient** π (1/Pa units), which is a tensor quantity because a stress in one direction in a crystal can alter the resistivity in another direction.

Recombination of an electron–hole pair involves an electron in the CB falling down in energy into an empty state (hole) in the VB to occupy it. The result is the annihilation of an EHP. Recombination is direct when the electron falls directly down into an empty state in the VB as in GaAs. Recombination is indirect if the electron is first captured locally by a defect or an impurity, called a recombination center, and from there it falls down into an empty state (hole) in the VB as in Si and Ge.

Schottky junction is a contact between a metal and a semiconductor that has rectifying properties. For a metal/n-type semiconductor junction, electrons on the metal side have to overcome a potential energy barrier Φ_B to enter the conduction band of the semiconductor, whereas the conduction electrons in the semiconductor have to overcome a smaller barrier eV_o to enter the metal. Forward bias decreases eV_o and thereby greatly encourages electron emissions over the barrier $e(V_o - V)$. Under reverse bias, electrons have to overcome Φ_B and the current is very small.

Thermal equilibrium carrier concentrations are those electron and hole concentrations that are solely determined by the statistics of the carriers and the density of states in the band. Thermal equilibrium concentrations obey the mass action law, $np = n_i^2$.

Thermal velocity (v_{th}) of an electron in the CB is its mean (or effective) speed in the semiconductor as it moves around in the crystal. For a nondegenerate semiconductor, it can be obtained simply from $\frac{1}{2}m_e^* v_{th}^2 = \frac{3}{2}kT$

Vacuum level is the energy level where the *PE* of the electron and the *KE* of the electron are both zero. It defines the energy level where the electron is just free from the solid.

Valence band (VB) is a band of energies for the electrons in bonds in a semiconductor. The valence band is made of all those states (wavefunctions) that constitute the bonding between the atoms in the crystal. At absolute zero of temperature, the VB is full of all the bonding electrons of the atoms. When an electron is excited to the CB, this leaves behind an empty state, which is called a hole. It carries a positive charge and behaves as if it were a "free" positively charged entity with an effective mass of m_h^*. It moves around the VB by having a neighboring electron tunnel into the unoccupied state.

Work function (Φ) is the energy required to remove an electron from the solid to the vacuum level.

QUESTIONS AND PROBLEMS

5.1 **Bandgap and photodetection**

a. Determine the maximum value of the energy gap that a semiconductor, used as a photoconductor, can have if it is to be sensitive to yellow light (600 nm).

b. A photodetector whose area is 5×10^{-2} cm^2 is irradiated with yellow light whose intensity is 2 mW cm^{-2}. Assuming that each photon generates one electron–hole pair, calculate the number of pairs generated per second.

c. From the known energy gap of the semiconductor GaAs ($E_g = 1.42$ eV), calculate the primary wavelength of photons emitted from this crystal as a result of electron–hole recombination.

d. Is the above wavelength visible?

e. Will a silicon photodetector be sensitive to the radiation from a GaAs laser? Why?

5.2 **Intrinsic Ge** Using the values of the density of states effective masses m_e^* and m_h^* in Table 5.1, calculate the intrinsic concentration in Ge. What is n_i if you use N_c and N_v from Table 5.1? Calculate the intrinsic resistivity of Ge at 300 K.

5.3 **Fermi level in intrinsic semiconductors** Using the values of the density of states effective masses m_e^* and m_h^* in Table 5.1, find the position of the Fermi energy in intrinsic Si, Ge, and GaAs with respect to the middle of the bandgap ($E_g/2$).

5.4 **Extrinsic Si** A Si crystal has been doped with P. The donor concentration is 10^{15} cm^{-3}. Find the conductivity and resistivity of the crystal.

5.5 **Extrinsic Si** Find the concentration of acceptors required for an n-Si crystal to have a resistivity of 1 Ω cm.

5.6 **Minimum conductivity**

a. Consider the conductivity of a semiconductor, $\sigma = en\mu_e + ep\mu_h$. Will doping always increase the conductivity?

b. Show that the minimum conductivity for Si is obtained when it is p-type doped such that the hole concentration is

$$p_m = n_i\sqrt{\frac{\mu_e}{\mu_h}}$$

and the corresponding minimum conductivity (maximum resistivity) is

$$\sigma_{\min} = 2en_i\sqrt{\mu_e\mu_h}$$

c. Calculate p_m and $\sigma_{\min}$ for Si and compare with intrinsic values.

5.7 **Extrinsic p-Si** A Si crystal is to be doped p-type with B acceptors. The hole drift mobility μ_h depends on the total concentration of ionized dopants N_{dopant}, in this case acceptors only, as

$$\mu_h \approx 54.3 + \frac{407}{1 + 3.745 \times 10^{-18} N_{\text{dopant}}} \quad \text{cm}^2\ \text{V}^{-1}\ \text{s}^{-1}$$

where N_{dopant} is in cm^{-3}. Find the required concentration of B doping for the resistivity to be 0.1 Ω cm.

5.8 **Thermal velocity and mean free path in GaAs** Given that the electron effective mass m_e^* for the GaAs is $0.067m_e$, calculate the thermal velocity of the conduction band (CB) electrons. The electron drift mobility μ_e depends on the mean free time τ_e between electron scattering events (between electrons and lattice vibrations). Given $\mu_e = e\tau_e/m_e^*$, and $\mu_e = 8500$ cm^2 V^{-1} s^{-1} for GaAs, calculate τ_e, and hence the mean free path ℓ of CB electrons. How many unit cells is ℓ if the lattice constant a of GaAs is 0.565 nm? Calculate the drift velocity $v_d = \mu_e\mathcal{E}$ of the CB electrons in an applied field $\mathcal{E}$ of 10^4 V m^{-1}. What is your conclusion?

5.9 **Compensation doping in Si**

a. A Si wafer has been doped n-type with 10^{17} As atoms cm^{-3}.

1. Calculate the conductivity of the sample at 27 °C.
2. Where is the Fermi level in this sample at 27 °C with respect to the Fermi level (E_{Fi}) in intrinsic Si?
3. Calculate the conductivity of the sample at 127 °C.

b. The above n-type Si sample is further doped with 9×10^{16} boron atoms (p-type dopant) per centimeter cubed.

1. Calculate the conductivity of the sample at 27 °C.
2. Where is the Fermi level in this sample with respect to the Fermi level in the sample in (*a*) at 27 °C? Is this an n-type or p-type Si?

5.10 **Temperature dependence of conductivity** An n-type Si sample has been doped with 10^{15} phosphorus atoms cm^{-3}. The donor energy level for P in Si is 0.045 eV below the conduction band edge energy.

a. Calculate the room temperature conductivity of the sample.

b. Estimate the temperature above which the sample behaves as if intrinsic.

c. Estimate to within 20 percent the lowest temperature above which all the donors are ionized.

d. Sketch schematically the dependence of the electron concentration in the conduction band on the temperature as $\log(n)$ versus $1/T$, and mark the various important regions and critical temperatures. For each region draw an energy band diagram that clearly shows from where the electrons are excited into the conduction band.

e. Sketch schematically the dependence of the conductivity on the temperature as $\log(\sigma)$ versus $1/T$ and mark the various critical temperatures and other relevant information.

***5.11** **Ionization at low temperatures in doped semiconductors** Consider an n-type semiconductor. The probability that a donor level E_d is occupied by an electron is

$$f_d = \frac{1}{1 + \dfrac{1}{g}\exp\left(\dfrac{E_d - E_F}{kT}\right)} \qquad [5.84]$$

Probability of donor occupancy

where k is the Boltzmann constant, T is the temperature, E_F is the Fermi energy, and g is a constant called the degeneracy factor: in Si, $g = 2$ for donors, and for the occupation statistics of acceptors $g = 4$. Show that

Electron concentration in extrinsic semiconductors

$$n^2 + \frac{nN_c}{g\exp\left(\frac{\Delta E}{kT}\right)} - \frac{N_d N_c}{g\exp\left(\frac{\Delta E}{kT}\right)} = 0 \qquad [5.85]$$

where n is the electron concentration in the conduction band, N_c is the effective density of states at the conduction band edge, N_d is the donor concentration, and $\Delta E = E_c - E_d$ is the ionization energy of the donors. Show that Equation 5.85 at low temperatures is equivalent to Equation 5.19. Consider a p-type Si sample that has been doped with 10^{15} gallium (Ga) atoms cm^{-3}. The acceptor energy level for Ga in Si is 0.065 eV above the valence band edge energy, E_v. Estimate the lowest temperature (°C) above which 90 percent of the acceptors are ionized by assuming that the acceptor degeneracy factor $g = 4$.

5.12 **Compensation doping in n-type Si** An n-type Si sample has been doped with 1×10^{17} phosphorus (P) atoms cm^{-3}. The drift mobilities of holes and electrons in Si at 300 K depend on the total concentration of dopants N_{dopant} (cm^{-3}) as follows:

Electron drift mobility

$$\mu_e \approx 88 + \frac{1252}{1 + 6.984 \times 10^{-18} N_{\text{dopant}}} \quad \text{cm}^2\ \text{V}^{-1}\ \text{s}^{-1}$$

and

Hole drift mobility

$$\mu_h \approx 54.3 + \frac{407}{1 + 3.745 \times 10^{-18} N_{\text{dopant}}} \quad \text{cm}^2\ \text{V}^{-1}\ \text{s}^{-1}$$

a. Calculate the room temperature conductivity of the sample.

b. Calculate the necessary acceptor doping (*i.e.*, N_a) that is required to make this sample p-type with approximately the same conductivity.

5.13 **GaAs** Ga has a valency of III and As has V. When Ga and As atoms are brought together to form the GaAs crystal, as depicted in Figure 5.54, the three valence electrons in each Ga and the five valence electrons in each As are all shared to form four covalent bonds per atom. In the GaAs crystal with some 10^{23} or so equal numbers of Ga and As atoms, we have an average of four valence electrons per atom, whether Ga or As, so we would expect the bonding to be similar to that in the Si crystal: four bonds per atom. The crystal structure, however, is not that of diamond but rather that of zinc blende (Chapter 1).

a. What is the average number of valence electrons per atom for a pair of Ga and As atoms and in the GaAs crystal?

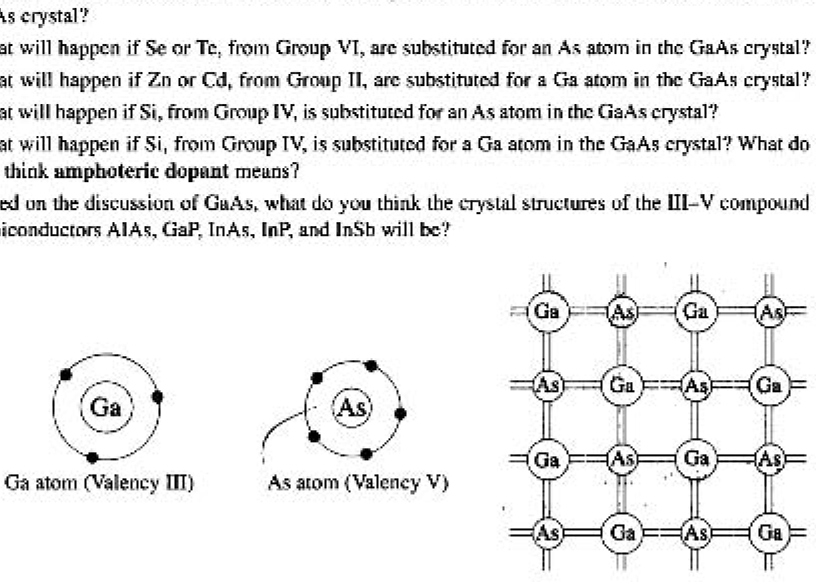
b. What will happen if Se or Te, from Group VI, are substituted for an As atom in the GaAs crystal?

c. What will happen if Zn or Cd, from Group II, are substituted for a Ga atom in the GaAs crystal?

d. What will happen if Si, from Group IV, is substituted for an As atom in the GaAs crystal?

e. What will happen if Si, from Group IV, is substituted for a Ga atom in the GaAs crystal? What do you think **amphoteric dopant** means?

f. Based on the discussion of GaAs, what do you think the crystal structures of the III–V compound semiconductors AlAs, GaP, InAs, InP, and InSb will be?


Ga atom (Valency III)

As atom (Valency V)

Figure 5.54 The GaAs crystal structure in two dimensions.

Average number of valence electrons per atom is four. Each Ga atom covalently bonds with four neighboring As atoms and vice versa.

5.14 Doped GaAs Consider the GaAs crystal at 300 K.

a. Calculate the intrinsic conductivity and resistivity.

b. In a sample containing only 10^{15} cm^{-3} ionized donors, where is the Fermi level? What is the conductivity of the sample?

c. In a sample containing 10^{15} cm^{-3} ionized donors and 9×10^{14} cm^{-3} ionized acceptors, what is the free hole concentration?

5.15 Varshni equation and the change in the bandgap with temperature The Varshni equation describes the change in the energy bandgap E_g of a semiconductor with temperature T in terms of

$$E_g = E_{go} - \frac{AT^2}{B+T}$$

Varshni equation

where E_{go} is the bandgap at $T = 0$ K, and A and B are material-specific constants. For example, for GaAs, $E_{go} = 1.519$ eV, $A = 5.405 \times 10^{-4}$ eV K^{-1}, $B = 204$ K, so that at $T = 300$ K, $E_g = 1.42$ eV. Show that

$$\frac{dE_g}{dT} = -\frac{AT(T+2B)}{(B+T)^2} = -\frac{(E_{go} - E_g)}{T}\left(\frac{T+2B}{T+B}\right)$$

Bandgap shift with temperature

What is dE_g/dT for GaAs? The Varshni equation can be used to calculate the shift in the peak emission wavelength of a light emitting diode (LED) with temperature or the cutoff wavelength of a detector. If the emitted photon energy from an electron and hole recombination is $h\nu \approx E_g + kT$, find the shift in the emitted wavelength from 27 °C down to −30 °C from a GaAs LED.

5.16 Degenerate semiconductor Consider the general exponential expression for the concentration of electrons in the CB,

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

and the mass action law, $np = n_i^2$. What happens when the doping level is such that n approaches N_c and exceeds it? Can you still use the above expressions for n and p?

Consider an n-type Si that has been heavily doped and the electron concentration in the CB is 10^{20} cm^{-3}. Where is the Fermi level? Can you use $np = n_i^2$ to find the hole concentration? What is its resistivity? How does this compare with a typical metal? What use is such a semiconductor?

5.17 Photoconductivity and speed Consider two p-type Si samples both doped with 10^{15} B atoms cm^{-3}. Both have identical dimensions of length L (1 mm), width W (1 mm), and depth (thickness) D (0.1 mm). One sample, labeled A, has an electron lifetime of 1 μs whereas the other, labeled B, has an electron lifetime of 5 μs.

a. At time $t = 0$, a laser light of wavelength 750 nm is switched on to illuminate the surface ($L \times W$) of both the samples. The incident laser light intensity on both samples is 10 mW cm^{-2}. At time $t = 50$ μs, the laser is switched off. Sketch the time evolution of the minority carrier concentration for both samples on the same axes.

b. What is the photocurrent (current due to illumination alone) if each sample is connected to a 1 V battery?

***5.18 Hall effect in semiconductors** The Hall effect in a semiconductor sample involves not only the electron and hole concentrations n and p, respectively, but also the electron and hole drift mobilities μ_e and μ_h. The Hall coefficient of a semiconductor is (see Chapter 2)

$$R_H = \frac{p - nb^2}{e(p+nb)^2} \qquad [5.86]$$

Hall coefficient of a semiconductor

where $b = \mu_e/\mu_h$.

a. Given the mass action law $np = n_i^2$, find n for maximum $|R_H|$ (negative and positive R_H). Assume that the drift mobilities remain relatively unaffected as n changes (due to doping). Given the electron and hole drift mobilities $\mu_e = 1350$ cm^2 V^{-1} s^{-1} and $\mu_h = 450$ cm^2 V^{-1} s^{-1} for silicon, determine n for maximum $|R_H|$ in terms of n_i.

b. Taking $b = 3$, plot R_H as a function of electron concentration n/n_i from 0.01 to 10.

c. Show that, when $n \gg n_i$, $R_H = -1/en$ and when $n \ll n_i$, $R_H = +1/ep$.

5.19 **Hall effect in semiconductors** Most Hall-effect high-sensitivity sensors typically use III–V semiconductors, such as GaAs, InAs, InSb. Hall-effect integrated circuits with integrated amplifiers, on the other hand, use Si. Consider nearly intrinsic samples in which $n \approx p \approx n_i$, and calculate R_H for each using the data in Table 5.4. What is your conclusion? Which sensor would exhibit the worst temperature drift? (Consider the bandgap, and drift in n_i.)

Table 5.4 Hall effect in selected semiconductors

	E_g(eV)	n_i(cm^{-3})	μ_e (cm^2 V^{-1}s^{-1})	μ_h (cm^2 V^{-1}s^{-1})	b	R_H (m^3 A^{-1}s^{-1})
Si	1.10	1×10^{10}	1,350	450	3	−312
GaAs	1.42	2×10^{6}	8,500	400	?	?
InAs	0.36	1×10^{15}	33,000	460	?	?
InSb	0.17	2×10^{16}	78,000	850	?	?

***5.20** **Compound semiconductor devices** Silicon and germanium crystalline semiconductors are what are called elemental Group IV semiconductors. It is possible to have compound semiconductors from atoms in Groups III and V. For example, GaAs is a compound semiconductor that has Ga from Group III and As from Group V, so in the crystalline structure we have an "effective" or "mean" valency of IV per atom and the solid behaves like a semiconductor. Similarly GaSb (gallium antimonide) would be a III–V type semiconductor. Provided we have a stoichiometric compound, the semiconductor will be ideally intrinsic. If, however, there is an excess of Sb atoms in the solid GaSb, then we will have nonstoichiometry and the semiconductor will be extrinsic. In this case, excess Sb atoms will act as donors in the GaSb structure. There are many useful compound semiconductors, the most important of which is GaAs. Some can be doped both n- and p-type, but many are one type only. For example, ZnO is a II–VI compound semiconductor with a direct bandgap of 3.2 eV, but unfortunately, due to the presence of excess Zn, it is naturally n-type and cannot be doped to p-type.

a. GaSb (gallium antimonide) is an interesting direct bandgap semiconductor with an energy bandgap $E_g = 0.67$ eV, almost equal to that of germanium. It can be used as an light emitting diode (LED) or laser diode material. What would be the wavelength of emission from a GaSb LED? Will this be visible?

b. Calculate the intrinsic conductivity of GaSb at 300 K taking $N_c = 2.3 \times 10^{19}$ cm^{-3}, $N_v = 6.1 \times 10^{19}$ cm^{-3}, $\mu_e = 5000$ cm^2 V^{-1} s^{-1}, and $\mu_h = 1000$ cm^2 V^{-1} s^{-1}. Compare with the intrinsic conductivity of Ge.

c. Excess Sb atoms will make gallium antimonide nonstoichiometric, that is, GaSb$_{1+\delta}$, which will result in an extrinsic semiconductor. Given that the density of GaSb is 5.4 g cm^{-3}, calculate δ (excess Sb) that will result in GaSb having a conductivity of 100 Ω^{-1} cm^{-1}. Will this be an n- or p-type semiconductor? You may assume that the drift mobilities are relatively unaffected by the doping.

5.21 **Excess minority carrier concentration** Consider an n-type semiconductor and weak injection conditions. Assume that the minority carrier recombination time τ_h is constant (independent of injection—hence the weak injection assumption). The rate of change of the instantaneous hole concentration $\partial p_n/\partial t$ due to recombination is given by

Recombination rate

$$\frac{\partial p_n}{\partial t} = -\frac{p_n}{\tau_h} \qquad [5.87]$$

The net rate of increase (change) in p_n is the sum of the total generation rate G and the rate of change due to recombination, that is,

Excess carriers under uniform photogeneration and recombination

$$\frac{dp_n}{dt} = G - \frac{p_n}{\tau_h} \qquad [5.88]$$

By separating the generation term G into thermal generation G_o and photogeneration G_{ph} and considering the dark condition as one possible solution, show that

$$\frac{d\Delta p_n}{dt} = G_{ph} - \frac{\Delta p_n}{\tau_h} \qquad [5.89]$$

How does your derivation compare with Equation 5.27? What are the assumptions inherent in Equation 5.89?

*5.22 **Direct recombination and GaAs** Consider recombination in a direct bandgap p-type semiconductor, *e.g.*, GaAs doped with an acceptor concentration N_a. The recombination involves a direct meeting of an electron–hole pair as depicted in Figure 5.22. Suppose that excess electrons and holes have been injected (*e.g.*, by photoexcitation), and that Δn_p is the excess electron concentration and Δp_p is the excess hole concentration. Assume Δn_p is controlled by recombination and thermal generation only; that is, recombination is the equilibrium storing mechanism. The recombination rate will be proportional to $n_p p_p$, and the thermal generation rate will be proportional to $n_{po} p_{po}$. In the dark, in equilibrium, thermal generation rate is equal to the recombination rate. The latter is proportional to $n_{po} p_{po}$. The rate of change of Δn_p is

$$\frac{\partial \Delta n_p}{\partial t} = -B[n_p p_p - n_{po} p_{po}] \qquad [5.90]$$

Recombination rate

where B is a proportionality constant, called the **direct recombination capture coefficient.** The **recombination lifetime** τ_r is *defined* by

$$\frac{\partial \Delta n_p}{\partial t} = -\frac{\Delta n_p}{\tau_r} \qquad [5.91]$$

Definition of recombination lifetime

a. Show that for *low-level injection*, $n_{po} \ll \Delta n_p \ll p_{po}$, τ_r is constant and given by

$$\tau_r = \frac{1}{Bp_{po}} = \frac{1}{BN_a} \qquad [5.92]$$

Low injection recombination time

b. Show that under *high-level injection*, $\Delta n_p \gg p_{po}$,

$$\frac{\partial \Delta n_p}{\partial t} \approx -B\Delta p_p\,\Delta n_p = -B(\Delta n_p)^2 \qquad [5.93]$$

High injection

so that the recombination lifetime τ_r is now given by

$$\tau_r = \frac{1}{B\Delta p_p} = \frac{1}{B\Delta n_p} \qquad [5.94]$$

High-injection recombination time

that is, the lifetime τ_r is inversely proportional to the injected carrier concentration.

c. Consider what happens in the presence of photogeneration at a rate G_{ph} (electron–hole pairs per unit volume per unit time). Steady state will be reached when the photogeneration rate and recombination rate become equal. That is,

$$G_{ph} = \left(\frac{\partial \Delta n_p}{\partial t}\right)_{\text{recombination}} = B[n_p p_p - n_{po} p_{po}]$$

Steady-state photogeneration rate

A photoconductive film of n-type GaAs doped with 10^{13} cm^{-3} donors is 2 mm long (L), 1 mm wide (W), and 5 μm thick (D). The sample has electrodes attached to its ends (electrode area is therefore 1 mm × 5 μm) which are connected to a 1 V supply through an ammeter. The GaAs photoconductor is uniformly illuminated over the surface area 2 mm × 1 mm with a 1 mW laser radiation of wavelength $\lambda = 840$ nm (infrared). The recombination coefficient B for GaAs is 7.21×10^{-16} m^3 s^{-1}. At $\lambda = 840$ nm, the absorption coefficient is about 5×10^3 cm^{-1}. Calculate the photocurrent I_{photo} and the electrical power dissipated as Joule heating in the sample. What will be the power dissipated as heat in the sample in an open circuit, where $I = 0$?

5.23 **Piezoresistivity application to deflection and force measurement** Consider the cantilever in Figure 5.38c. Suppose we apply a force F to the free end, which results in a deflection h of the tip of the cantilever from its horizontal equilibrium position. The maximum stress σ_m is induced at the support end of the cantilever, at its surface where the piezoresistor is embedded to measure the stress. When the cantilever is bent, there is a tensile or longitudinal stress σ_L on the surface because the top surface is extended and the bottom surface is contracted. If L, W, and D are respectively the length, width, and thickness of the cantilever, then the relationships between the force F and deflection h, and the maximum stress σ_L are

Cantilever equations

$$\sigma_L(\max) = \frac{3YDh}{2L^2} \qquad \text{and} \qquad F = \frac{WD^3Y}{4L^3}h$$

where Y is the elastic (Young's) modulus. A particular Si cantilever has a length (L) of 500 μm, width (W) of 100 μm, and thickness (D) of 10 μm. Given $Y = 170$ GPa, and that the piezoresistor embedded in the cantilever is along the [110] direction with $\pi_L \approx 72 \times 10^{-11}$ Pa^{-1}, find the percentage change in the resistance, $\Delta R/R$, of the piezoresistor when the deflection is 0.1 μm. What is the force that would give this deflection? (Neglect the transverse stresses on the piezoresistor.) How does the design choice for the length L of the cantilever depend on whether one is interested in measuring the deflection h or the force F? (Note: σ_L depends on the distance x from the support end; it decreases with x. Assume that the length of the piezoresistor is very short compared with L so that σ_L does not change significantly along its length.)

5.24 **Schottky junction**

a. Consider a Schottky junction diode between Au and n-Si, doped with 10^{16} donors cm^{-3}. The cross-sectional area is 1 mm^2. Given the work function of Au as 5.1 eV, what is the theoretical barrier height Φ_B from the metal to the semiconductor?

b. Given that the experimental barrier height Φ_B is about 0.8 eV, what is the reverse saturation current and the current when there is a forward bias of 0.3 V across the diode? (Use Equation 4.37.)

5.25 **Schottky junction** Consider a Schottky junction diode between Al and n-Si, doped with 5×10^{16} donors cm^{-3}. The cross-sectional area is 1 mm^2. Given that the electron affinity χ of Si is 4.01 eV and the work function of Al is 4.28 eV, what is the theoretical barrier height Φ_B from the metal to the semiconductor? What is the built-in voltage? If the experimental barrier height Φ_B is about 0.6 eV, what is the reverse saturation current and the current when there is a forward bias of 0.2 V across the diode? Take $B_e = 110$ A cm^{-2} K^{-2}.

5.26 **Schottky and ohmic contacts** Consider an n-type Si sample doped with 10^{16} donors cm^{-3}. The length L is 100 μm; the cross-sectional area A is 10 μm × 10 μm. The two ends of the sample are labeled as B and C. The electron affinity (χ) of Si is 4.01 eV and the work functions Φ of four potential metals for contacts at B and C are listed in Table 5.5.

Table 5.5 Work functions in eV

Cs	Li	Al	Au
1.8	2.5	4.25	5.0

a. Ideally, which metals will result in a Schottky contact?

b. Ideally, which metals will result in an ohmic contact?

c. Sketch the I–V characteristics when both B and C are ohmic contacts. What is the relationship between I and V?

d. Sketch the I–V characteristics when B is ohmic and C is a Schottky junction. What is the relationship between I and V?

e. Sketch the I–V characteristics when both B and C are Schottky contacts. What is the relationship between I and V?

5.27 **Peltier effect and electrical contacts** Consider the Schottky junction and the ohmic contact shown in Figures 5.39 and 5.43 between a metal and n-type semiconductor.

a. Is the Peltier effect similar in both contacts?

b. Is the sign in $Q' = \pm\Pi I$ the same for both contacts?

c. Which junction would you choose for a thermoelectric cooler? Give reasons.

***5.28** **Peltier coolers and figure of merit (FOM)** Consider the thermoelectric effect shown in Figure 5.45 in which a semiconductor has two contacts at its ends and is conducting an electric current I. We assume that the cold junction is at a temperature T_c and the hot junction is at T_h and that there is a temperature difference of $\Delta T = T_h - T_c$ between the two ends of the semiconductor. The current I flowing through the cold junction absorbs Peltier heat at a rate Q'_P, given by

$$Q'_P = \Pi I \qquad [5.95]$$

where Π is the Peltier coefficient for the junction between the metal and semiconductor. The current I flowing through the semiconductor generates heat due to the Joule heating of the semiconductor. The rate of Joule heat generated through the bulk of the semiconductor is

$$Q'_J = \left(\frac{L}{\sigma A}\right) I^2 \qquad [5.96]$$

We assume that half of this heat flows to the cold junction.

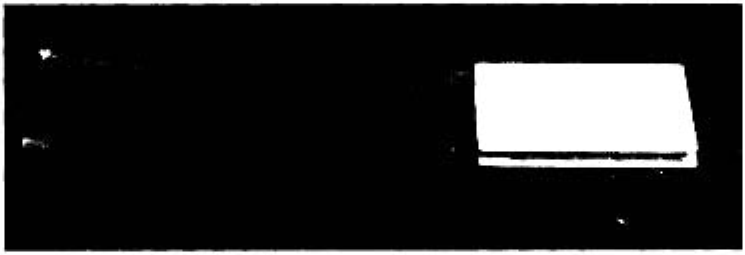
In addition there is heat flow from the hot to the cold junction through the semiconductor, given by the thermal conduction equation

$$Q'_{TC} = \left(\frac{A\kappa}{L}\right) \Delta T \qquad [5.97]$$

The net rate of heat absorption (cooling rate) at the cold junction is then

$$Q'_{\text{net cool}} = Q'_P - \frac{1}{2}Q'_J - Q'_{TC} \qquad [5.98]$$

By substituting from Equations 5.95 to 5.97 into Equation 5.98, obtain the net cooling rate in terms of the current I. Then by differentiating $Q'_{\text{net cool}}$ with respect to current, show that maximum cooling is



A commercial thermoelectric cooler (by Melcor); an example of the Peltier effect. The device area is 5.5 cm × 5.5 cm (approximately 2.2 inches × 2.2 inches). Its maximum current is 14 A; maximum heat pump ability is 67 W; maximum temperature difference between the hot and cold surfaces is 67 °C.

Table 5.6

Material	Π (V)	ρ (Ω m)	κ (W m^{-1}K^{-1})	FOM
n-Bi_2Te_3	6.0×10^{-2}	10^{-5}	1.70	
p-Bi_2Te_3	7.0×10^{-2}	10^{-5}	1.45	
Cu	5.5×10^{-4}	1.7×10^{-8}	390	
W	3.3×10^{-4}	5.5×10^{-8}	167	

obtained when the current is

$$I_m = \left(\frac{A}{L}\right)\Pi\sigma \qquad [5.99]$$

and the maximum cooling rate is

Maximum cooling rate

$$Q'_{\text{max cool}} = \frac{A}{L}\left[\frac{1}{2}\Pi^2\sigma - \kappa\,\Delta T\right] \qquad [5.100]$$

Under steady-state operating conditions, the temperature difference ΔT reaches a steady-state value and the net cooling rate at the junction is then zero (ΔT is constant). From Equation 5.100 show that the maximum temperature difference achievable is

Maximum temperature difference

$$\Delta T_{\max} = \frac{1}{2}\frac{\Pi^2\sigma}{\kappa} \qquad [5.101]$$

The quantity $\Pi^2\sigma/\kappa$ is defined as the **figure of merit** (FOM) for the semiconductor as it determines the maximum ΔT achievable. The same expression also applies to metals, though we will not derive it here.

Use Table 5.6 to determine the FOM for various materials listed therein and discuss the significance of your calculations. Would you recommend a thermoelectric cooler based on a metal-to-metal junction?

*5.29 **Seebeck coefficient of semiconductors and thermal drift in semiconductor devices** Consider an n-type semiconductor that has a temperature gradient across it. The right end is hot and the left end is cold, as depicted in Figure 5.55. There are more energetic electrons in the hot region than in the cold region. Consequently, electron diffusion occurs from hot to cold regions, which immediately exposes negatively charged donors in the hot region and therefore builds up an internal field and a built-in voltage, as shown in Figure 5.55. Eventually an equilibrium is reached when the diffusion of electrons is balanced by their drift driven by the built-in field. The net current must be zero. The Seebeck coefficient (or thermoelectric power) S measures

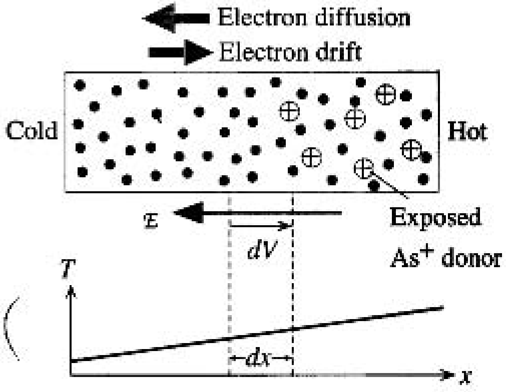


Figure 5.55 In the presence of a temperature gradient, there is an internal field and a voltage difference.

The Seebeck coefficient is defined as dV/dT, the potential difference per unit temperature difference.

this effect in terms of the voltage developed as a result of an applied temperature gradient as

$$S = \frac{dV}{dT} \qquad [5.102]$$

a. How is the Seebeck effect in a *p*-type semiconductor different than that for an *n*-type semiconductor when both are placed in the same temperature gradient in Figure 5.55? Recall that the sign of the Seebeck coefficient is the polarity of the voltage at the cold end with respect to the hot end (see Section 4.8.2).

b. Given that for an *n*-type semiconductor,

Seebeck coefficient n-type semiconductor

$$S_n = -\frac{k}{e}\left[2 + \frac{(E_c - E_F)}{kT}\right] \qquad [5.103]$$

what are typical magnitudes for S_n in Si doped with 10^{14} and 10^{16} donors cm^{-3}? What is the significance of S_n at the semiconductor device level?

c. Consider a *pn* junction Si device that has the *p*-side doped with 10^{18} acceptors cm^{-3} and the *n*-side doped with 10^{14} donors cm^{-3}. Suppose that this *pn* junction forms the input stage of an op amp with a large gain, say 100. What will be the output signal if a small thermal fluctuation gives rise to a 1 °C temperature difference across the *pn* junction?

5.30 **Photogeneration and carrier kinetic energies** Figure 5.35 shows what happens when a photon with energy $h\nu > E_g$ is absorbed in GaAs to photogenerate an electron and a hole. The figure shows that the electron has a higher kinetic energy (*KE*), which is the excess energy above E_c, than the hole, since the hole is almost at E_v. The reason is that the electron effective mass in GaAs is almost 10 times less than the hole effective mass, so the photogenerated electron has a much higher *KE*. When an electron and hole are photogenerated in a direct bandgap semiconductor, they have the same **k** vector. Energy conservation requires that the photon energy $h\nu$ divides according to

Photogeneration

$$h\nu = E_g + \frac{(\hbar k)^2}{2m_e^*} + \frac{(\hbar k)^2}{2m_h^*}$$

where k is the wavevector of the electron and hole and m_e^* and m_h^* are the effective masses of the electron and hole, respectively.

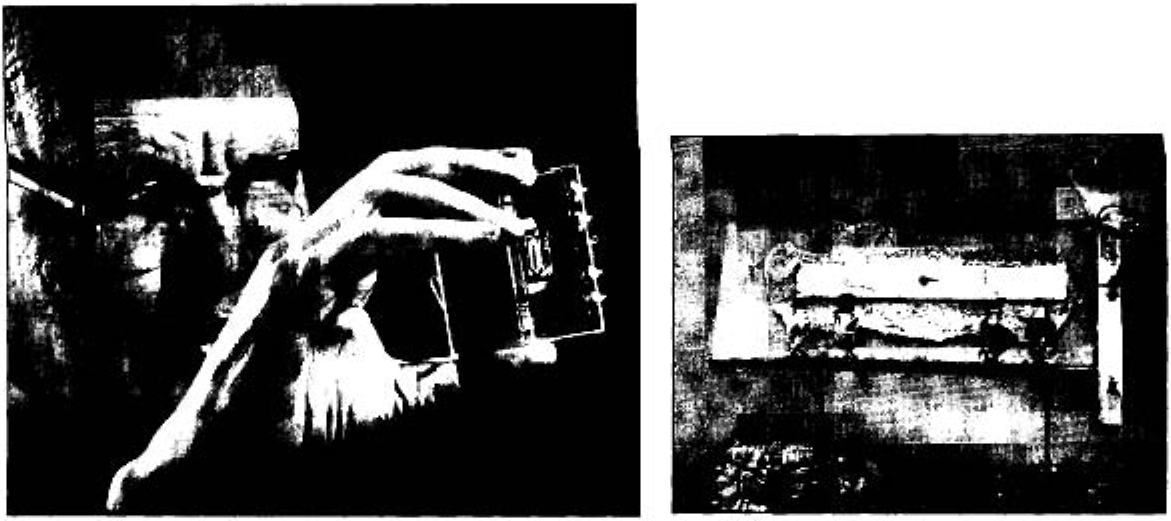
a. What is the ratio of the electron to hole *KE*s right after photogeneration?

b. If the incoming photon has an energy of 2.0 eV, and $E_g = 1.42$ eV for GaAs, calculate the *KE*s of the electron and the hole in eV, and calculate to which energy levels they have been excited with respect to their band edges.

c. Explain why the electron and hole wavevector k should be approximately the same right after photogeneration. Consider k_{photon} for the photon, and the momentum conservation.

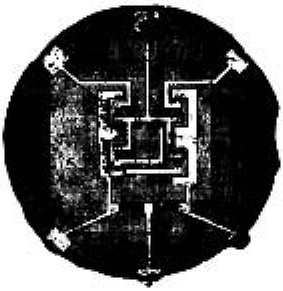
William Shockley and his group celebrate Shockley's Nobel prize in 1956. First left, sitting, is G. E. Moore (chairman emeritus of Intel), standing fourth from right is R. N. Noyce, inventor of the integrated circuit, and standing at the extreme right is J. T. Last.

SOURCE: P. K. Bondyopadhyay, "W = Shockley, the Transistor Pioneer—Portrait of an Inventive Genius," *Proceedings IEEE*, vol. 86, no. 1, January 1998, p. 202, figure 16 (Courtesy of IEEE.)



The first monolithic integrated circuit, about the size of a fingertip, was documented and developed at Texas Instruments by Jack Kilby in 1958; he won the 2000 Nobel prize in physics for his contribution to the development of the first integrated circuit. The IC was a chip of a single Ge crystal containing one transistor, one capacitor, and one resistor. Left: Jack Kilby holding his IC (photo, 1998). Right. The photo of the chip.

| SOURCE: Courtesy of Texas Instruments.



Robert Noyce and Jean Hoerni (a Swiss physicist) were responsible for the invention of the first planar IC at Fairchild (1961). The planar fabrication process was the key to the success of their IC. The photograph is that of the first logic chip at Fairchild.

| SOURCE: Courtesy of Fairchild Semiconductor.

Left to right: Andrew Grove, Robert Noyce (1927–1990), and Gordon Moore, who founded Intel in 1968. Andrew Grove's book *Physics and Technology of Semiconductor Devices* (Wiley, 1967) was one of the classic texts on devices in the sixties and seventies. "Moore's law" that started as a rough rule in 1965 states that the number of transistors in a chip will double every 18 months; Moore updated it in 1995 to every couple of years.

| SOURCE: Courtesy of Intel.