Semiconductor Physics

<div style="text-align:right">

17

</div>

17.1 Electrons and Holes

Suppose we start with an insulator or semiconductor and we excite one electron from the valence band to the conduction band, as shown in the left of Fig. 17.1. This excitation may be due to absorbing a photon, or it might be a thermal excitation. (For simplicity in the figure we have shown a direct band gap. For generality we have not assumed that the curvature of the two bands are the same.) When the electron has been moved up to the conduction band, there is an absence of an electron in the valence band known as a *hole*. Since a completely filled band is inert, it is very convenient to only keep track of the few holes in the valence band (assuming there are only a few) and to treat these holes as individual elementary particles. The electron can fall back into the empty state that is the hole, emitting energy (a photon, say) and "annihilating" both the electron from the conduction band and the hole from the valence band.[1] Note that while the electrical charge of an electron is negative the electrical charge of a hole (the absence of an electron) is positive—equal and opposite to that of the electron.[2]

[1] This is equivalent to pair annihilation of an electron with a positron. In fact, the analogy between electron–hole and electron–positron is fairly precise. As soon as Dirac constructed his equation (in 1928) describing the relativistic motion of electrons, and predicting positrons, it was understood that the positron could be thought of as an absence of an electron in a filled sea of states. The filled sea of electron states with a gap to exciting electron–positron pairs is the inert vacuum, which is analogous to an inert filled valence band.

[2] If this does not make intuitive sense consider the process of creating an electron–hole pair as described in Fig. 17.1. Initially (without the excited electron–hole pair) the system is charge neutral. We excite the system with a photon to create the pair, and we have not moved any additional net charge into the system. Thus if the electron is negative, the hole must be positive to preserve overall charge neutrality.
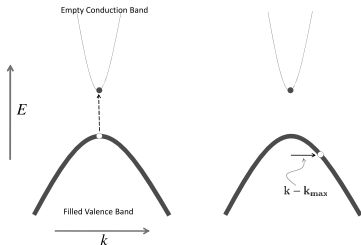

Fig. 17.1 Electrons and holes in a semiconductor. **Left:** A single hole in the valence band and a single electron in the conduction band. **Right:** Moving the hole to a momentum away from the top of the valence band costs *positive* energy—like pushing a balloon under water. As such, the effective mass of the hole is defined to be positive. The energy of the configuration on the right is greater than that on the left by $E = \hbar^2|\mathbf{k} - \mathbf{k_{max}}|^2/(2m^*)$.

[3]It is an important principle that near a minimum or a maximum one can always expand and get something quadratic plus higher order corrections.

[4]For simplicity we have assumed the system to be isotropic. In the more general case we would have

$$E = E_{min} + \alpha_x (k_x - k_x^{min})^2$$
$$+ \alpha_y (k_y - k_y^{min})^2$$
$$+ \alpha_z (k_z - k_z^{min})^2 + \ldots$$

for some orthogonal set of axes (the "principal axes") x, y, z. In this case we would have an effective mass which can be different in the three different principal directions.

[5]For simplicity we also neglect the spin of the electron here. In general, spin–orbit coupling can make the dispersion depend on the spin state of the electron. Among other things, this can modify the effective mass *g*-factor.

[6]It often occurs that the bottom of conduction band has more than one minimum at different points $\mathbf{k}_{min}^{(n)}$ in the Brillouin zone with exactly the same energy. We then say that there are multiple "valleys" in the band structure. Such a situation occurs due to the symmetry of the crystal. For example, in silicon (an fcc structure with a basis, see Fig. 12.21), six conduction band minima with the same energy occur approximately at the k-points $(\pm 5.3/a, 0, 0)$, $(0, \pm 5.3/a, 0)$, and $(0, 0, \pm 5.3/a)$.

[7]More accurately, $\mathbf{v} = \nabla_\mathbf{k} E(\mathbf{k})/\hbar + \mathbf{K}$ where the additional term $\mathbf{K}$ is known as the "Karplus–Luttinger" anomalous velocity and is proportional to applied electric field. This correction, resulting from subtle quantum-mechanical effects, is almost always neglected in solid state texts and rarely causes trouble (this is related to footnote 9 in Chapter 11). Only recently has research focused more on systems where such terms do matter. Proper treatment of this effect is beyond the scope of this book.

[8]Be warned: a few books define the mass of holes to be negative. This is a bit annoying but not inconsistent as long as the negative sign shows up somewhere else as well!

Effective Mass of Electrons

As mentioned in Sections 11.2 and 15.1.1, it is useful to describe the curvature at the bottom of a band in terms of an effective mass. Let us assume that near the bottom of the conduction band (assumed to be at $\mathbf{k} = \mathbf{k}_{min}$) the energy is given by[3,4,5,6]

$$E = E_{min} + \alpha |\mathbf{k} - \mathbf{k}_{min}|^2 + \ldots$$

with $\alpha > 0$, where the dots mean higher-order term in the deviation from $\mathbf{k}_{min}$. We then *define the effective mass* to be given by

$$\frac{\hbar^2}{m^*} = \frac{\partial^2 E}{\partial k^2} = 2\alpha \quad (17.1)$$

at the bottom of the band (with the derivative being taken in any direction for an isotropic system). Correspondingly, the (group) velocity is given by[7]

$$\mathbf{v} = \frac{\nabla_\mathbf{k} E}{\hbar} = \frac{\hbar(\mathbf{k} - \mathbf{k}_{min})}{m^*}. \quad (17.2)$$

This definition is chosen to be in analogy with the free electron behavior $E = \hbar^2 |\mathbf{k}|^2/(2m)$ with corresponding velocity $\mathbf{v} = \nabla_\mathbf{k} E/\hbar = \hbar \mathbf{k}/m$.

Effective Mass of Holes

Analogously we can define an effective mass for holes. Here things get a bit more complicated. For the top of the valence band, the energy dispersion for electrons would be

$$E = E_{max} - \alpha |\mathbf{k} - \mathbf{k}_{max}|^2 + \ldots \quad (17.3)$$

with $\alpha > 0$. The modern convention is to *define the effective mass for holes at the top of a valence band to be always positive*[8]

$$\frac{\hbar^2}{m_{hole}^*} = -\frac{\partial^2 E}{\partial k^2} = 2\alpha. \quad (17.4)$$

The convention of the effective mass being positive makes sense because the energy to boost the hole from zero velocity ($\mathbf{k} = \mathbf{k}_{max}$ at the top of the valence band) to finite velocity is positive. This energy is naturally given by

$$E_{hole} = \text{constant} + \frac{\hbar^2 |\mathbf{k} - \mathbf{k}_{max}|^2}{2m_{hole}^*}.$$

The fact that boosting the hole away from the top of the valence band is positive energy may seem a bit counter-intuitive being that the dispersion of the hole band is an upside-down parabola. However, one should think of this as being like pushing a balloon under water. The lowest energy configuration has the *electrons* at the lowest energy possible and the hole at the highest energy possible. So pushing the hole under the electrons costs positive energy. (This is depicted in the right-hand side of Fig. 17.1.) A good way to handle this bookkeeping is to remember

$$E(\text{absence of electron in state } \mathbf{k}) = -E(\text{electron in state } \mathbf{k}). \quad (17.5)$$

The momentum and velocity of a hole

There is a bit of complication with signs in keeping track of the momentum of a hole. If an electron is added to a band in a state **k** then the crystal momentum contained in the band increases by $\hbar\mathbf{k}$. Likewise, if an electron in state **k** is *removed* from an otherwise filled band, then the crystal momentum in the band must decrease by $\hbar\mathbf{k}$. Then, since a fully filled band carries no net crystal momentum the absence of an electron in state **k** should be a hole whose crystal momentum is $-\hbar\mathbf{k}$. It is thus convenient to *define* the wavevector $\mathbf{k}_{\text{hole}}$ of a hole to be the negative of the wavevector $\mathbf{k}_{\text{electron}}$ of the corresponding absent electron.[9]

This definition of wavevector is quite sensible when we try to calculate the group velocity of a hole. Analogous to the electron, we write the hole group velocity as the derivative of the hole energy

$$\mathbf{v}_{\text{hole}} = \frac{\nabla_{\mathbf{k}_{\text{hole}}} E_{\text{hole}}}{\hbar}. \tag{17.6}$$

Now, using Eq. 17.5, and also the fact that that the wavevector of the hole is minus the wavevector of the missing electron, we get two canceling minus signs and we find that

$$\mathbf{v}_{\text{hole}} = \mathbf{v}_{\text{missing electron}}.$$

This is a rather fundamental principle. The time evolution of a quantum state is independent of whether that state is occupied with a particle or not!

Effective Mass and Equations of Motion

We have defined the effective masses above in analogy with that of free electrons, by looking at the curvature of the dispersion. An equivalent definition (equivalent at least at the top or bottom of the band) is to define the effective mass m^* as being the quantity that satisfies Newton's second law, $F = m^*a$ for the particle in question. To demonstrate this, our strategy is to imagine applying a force to an electron in the system and then equate the work done on the electron to its change in energy. Let us start with an electron in momentum state **k**. Its group velocity is $\mathbf{v} = \nabla_k E(\mathbf{k})/\hbar$. If we apply a force,[10] the work done per unit time is

$$dW/dt = \mathbf{F} \cdot \mathbf{v} = \mathbf{F} \cdot \nabla_k E(\mathbf{k})/\hbar.$$

On the other hand, the change in energy per unit time must also be (by the chain rule)

$$dE/dt = d\mathbf{k}/dt \cdot \nabla_k E(\mathbf{k}).$$

Setting these two expressions equal to each other we (unsurprisingly) obtain Newton's equation

$$\mathbf{F} = \hbar \frac{d\mathbf{k}}{dt} = \frac{d\mathbf{p}}{dt} \tag{17.7}$$

where we have used $\mathbf{p} = \hbar\mathbf{k}$.

[9]Other conventions are possible but this is probably the simplest.

[10]For example, if we apply an electric field **E** and it acts on an electron of charge $-e$, the force is $\mathbf{F} = -e\mathbf{E}$.

If we now consider electrons near the bottom of a band, we can plug in the expression Eq. 17.2 for the velocity, and this becomes

$$\mathbf{F} = m^* \frac{d\mathbf{v}}{dt}$$

exactly as Newton would have expected.   In deriving this result recall that we have assumed that we are considering an electron near the bottom of a band so that we can expand the dispersion quadratically (or similarly we assumed that holes are near the top of a band). One might wonder how we should understand electrons when they are neither near the top nor the bottom of a band. More generally Eq. 17.7 always holds, as does the fact that the group velocity is $\mathbf{v} = \nabla_k E/\hbar$. It is then sometimes convenient to define an effective mass for an electron as a function of momentum to be given by[11]

$$\frac{\hbar^2}{m^*(k)} = \frac{\partial^2 E}{\partial k^2}$$

which agrees with our definition (Eq. 17.1) near the bottom of band. However, near the top of a band it is the *negative* of the corresponding hole mass (note the sign in Eq. 17.4). Note also that somewhere in the middle of the band the dispersion must reach an inflection point ($\partial^2 E/\partial k^2 = 0$), whereupon the effective mass actually becomes infinite as it changes sign.

Aside: It is useful to compare the time evolution of electrons and holes near the top of bands. If we think in terms of holes (the natural thing to do near the top of a band) we have $\mathbf{F} = +e\mathbf{E}$ and the holes have a positive mass. However, if we think in terms of electrons, we have $\mathbf{F} = -e\mathbf{E}$ but the mass is negative. Either way, the acceleration of the **k**-state is the same, whether we are describing the dynamics in terms of an electron in the state or in terms of a hole in the state. As mentioned below Eq. 17.6, this equivalence is expected, since the time evolution of an eigenstate is independent of whether that eigenstate is filled with an electron or not.

[11]For simplicity we write this in its one-dimensional form.

17.1.1 Drude Transport: Redux

Back in Chapter 3 we studied Drude theory—a simple kinetic theory of electron motion. The main failure of Drude theory was that it did not treat the Pauli exclusion principle properly: it neglected the fact that in metals the high density of electrons makes the Fermi energy extremely high. However, in semiconductors or band insulators, when only a few electrons are in the conduction band and/or only a few holes are in the valence band, then we can consider this to be a low-density situation, and to a very good approximation, we can ignore Fermi statistics. (For example, if only a single electron is excited into the conduction band, then we can completely ignore the Pauli principle, since it is the only electron around—there is no chance that any state it wants to sit in will already be filled!) As a result, when there is a low density of conduction electrons or valence holes, it turns out that Drude theory works

extremely well! We will come back to this issue later in Section 17.3, and make this statement much more precise.

At any rate, in the semiclassical picture, we can write a simple Drude transport equation (really Newton's equations!) for electrons in the conduction band

$$m_e^* \, d\mathbf{v}/dt = -e(\mathbf{E} + \mathbf{v} \times \mathbf{B}) - m_e^* \, \mathbf{v}/\tau$$

with m_e^* the electron effective mass. Here the first term on the right-hand side is the Lorentz force on the electron, and the second term is a drag force with an appropriate scattering time τ. The scattering time determines the so-called *mobility* μ which measures the ease with which the particle moves. The mobility is generally defined as the ratio of the velocity to the electric field.[12] In this Drude approach we then obtain

$$\mu = |\mathbf{v}|/|\mathbf{E}| = |e\tau/m^*|.$$

Similarly, we can write equations of motion for holes in the valence band

$$m_h^* \, d\mathbf{v}/dt = e(\mathbf{E} + \mathbf{v} \times \mathbf{B}) - m_h^* \, \mathbf{v}/\tau$$

where m_h^* is the hole effective mass. Note again that here the charge on the hole is *positive*. This should make sense—the electric field pulls on an electron in a direction opposite to the direction that it pulls on the absence of an electron!

If we think back all the way to Chapters 3 and 4, one of the physical puzzles that we could not understand was why the Hall coefficient sometimes changes sign (see Table 3.1). In some cases it looked as if the charge carrier had positive charge. Now we understand why this is true. In some materials the main charge carrier is the hole!

17.2   Adding Electrons or Holes with Impurities: Doping

In a pure band insulator or semiconductor, if we excite electrons from the valence to the conduction band (either with photons or thermally) we can be assured that the density of electrons in the conduction band (typically called n, which stands for "negative" charges) is precisely equal to the density of holes left behind in the valence band (typically called p, which stands for "positive" charges). However, in an impure semiconductor or band insulator this is not the case.

Without impurities, a semiconductor is known as *intrinsic*. The opposite of intrinsic, the case where there are impurities present, is sometimes known as *extrinsic*.

Let us now examine the extrinsic case more carefully. Consider for example, silicon (Si), which is a semiconductor with a band gap of about 1.1 eV. Now imagine that a phosphorus (P) atom replaces one of the Si atoms in the lattice as shown on the top of Fig. 17.2. This P atom, being directly to the right of Si on the periodic table, can be thought
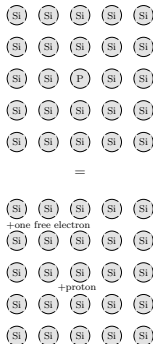
Fig. 17.2 Cartoon of doping a semiconductor. Doping Si with P adds one free electron to wander freely in the conduction band and leaves behind a positive charge on the nucleus.

[13]There are extra neutrons as well, but they don't do much in this context.

[14]"Dopant" generally means a chemical inserted into an object to alter its properties. This definition is true more broadly than the field of physics (e.g. Lance Armstrong, Jerry Garcia).

[15]Yes, it is annoying that the common dopant phosphorus has the chemical symbol P, but it is not a *p*-dopant in Si, it is an *n*-dopant.
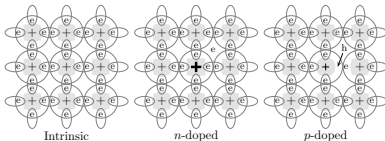
[16]More frequently than Al, boron (B) is used as a *p*-dopant in Si. Since B lies just below Al in the periodic table, it plays the same chemical role.

Fig.17.3 Cartoon of doping a semiconductor. **Left:** In the intrinsic case, all of the electrons are tied up in covalent bonds of two electrons. **Middle:** In the *n*-dopant case there is an extra unbound electron, and the dopant carries an extra nuclear charge. **Right:** In the *p*-dopant case there is one electron too few to complete all the bonds so there is an extra hole (denoted h) and the nuclear charge has one less positive charge than in the intrinsic case (the + sign is supposed to look slightly less large).

of as nothing more than a Si atom plus an extra proton and an extra electron,[13] as shown in the bottom of Fig. 17.2. Since the valence band is already filled this additional electron must go into the conduction band. The P atom is known as a *donor* (or *electron donor*) in silicon since it donates an electron to the conduction band. It is also sometimes known as an *n-dopant*,[14] since *n* is the symbol for the density of electrons in the conduction band.

Analogously, we can consider aluminum, the element directly to the left of Si on the periodic table. In this case, the aluminum dopant provides one fewer electron than Si, so there will be one electron missing from the valence band. In this case Al is known as an electron *acceptor*, or equivalently as a *p-dopant*, since *p* is the symbol for the density of holes.[15,16]

In a more chemistry-oriented language, we can depict the donors and acceptors as shown in Fig. 17.3. In the intrinsic case, all of the electrons are tied up in covalent bonds of two electrons. With the *n*-dopant, there is an extra unbound electron, whereas with the *p*-dopant there is an extra unbound hole (one electron too few).

Intrinsic   *n*-doped   *p*-doped

17.2.1 Impurity States

Let us consider even more carefully what happens when we add dopants. For definiteness let us consider adding an *n*-dopant such as P to a semiconductor such as Si. Once we add a single *n*-dopant to an otherwise intrinsic sample of Si, we get a single electron above the gap in the conduction band. This electron behaves like a free particle with mass m_e^*. However, in addition, we have a single extra positive charge $+e$ at some point in the crystal due to the P nucleus. The free electron is attracted back to this positive charge and forms a bound state that is similar to a hydrogen atom. There are two main differences between a real hydrogen atom and this bound state of an electron in the conduction band and the impurity nucleus. First of all, the electron has effective mass m_e^* which can be very different from the real (bare) mass of the electron (and is typically smaller than the bare mass of the electron). Secondly, instead of the two charges attracting each other with a potential $V = e^2/(4\pi\epsilon_0 r)$ they attract each other with a potential $V = e^2/(4\pi\epsilon_r\epsilon_0 r)$, where ϵ_r is the relative permittivity (or relative dielectric constant) of the material.

With these two small differences we can calculate the energies of the hydrogenic bound states exactly as we do for genuine hydrogen in our quantum mechanics courses.

We recall that the energy eigenstates of the hydrogen atom are given by $E_n^{H-atom} = -\text{Ry}/n^2$ where Ry is the Rydberg constant given by

$$\text{Ry} = \frac{me^4}{8\epsilon_0^2 h^2} \approx 13.6\text{eV}$$

with m the electron mass. The corresponding radius of this hydrogen atom wavefunction is $r_n \approx n^2 a_0$ with the Bohr radius given by

$$a_0 = \frac{4\pi\epsilon_0 \hbar^2}{me^2} \approx .51 \times 10^{-10}\text{m}.$$

The analogous calculation for a hydrogenic impurity state in a semiconductor gives precisely the same expression, only ϵ_0 is replaced by $\epsilon_0 \epsilon_r$ and m is replaced by m_e^*. One obtains

$$\text{Ry}^{eff} = \text{Ry}\left(\frac{m_e^*}{m} \frac{1}{\epsilon_r^2}\right)$$

and

$$a_0^{eff} = a_0\left(\epsilon_r \frac{m}{m_e^*}\right).$$

Because the dielectric constant of semiconductors is typically high (on the order of 10 for most common semiconductors) and because the effective mass is frequently low (a third of m or even smaller), the effective Rydberg Ry^{eff} can be tiny compared to the real Rydberg, and the effective Bohr radius a_0^{eff} can be huge compared to the real Bohr radius.[17] For example, in silicon[18] the effective Rydberg, Ry^{eff}, is much less than .1 eV and a_0^{eff} is above 30 Ångstroms. Thus this donor impurity forms an energy eigenstate with energy just below the bottom of the conduction band (Ry^{eff} below the band bottom only). At zero temperature this eigenstate will be filled, but it takes only a small temperature to excite a bound electron out of a hydrogenic orbital and into the conduction band.
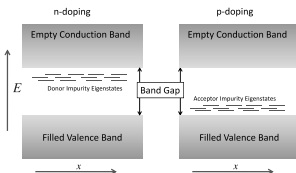
A depiction of this physics is shown in Fig. 17.4 where we have plotted an energy diagram for a semiconductor with donor or acceptor impurities. Here the energy eigenstates are plotted as a function of position. Between the valence and conduction band (which are uniform in position), there are many localized hydrogen-atom-like eigenstates. The energies of these states are not all exactly the same, since each impurity atom is perturbed by other impurity atoms in its environment. If the density of impurities is high enough, electrons (or holes) can hop from one impurity to the next, forming an *impurity band*.

Note that because the effective Rydberg is very small, the impurity eigenstates are only slightly below the conduction band or above the valence band respectively. With a small temperature, these donors or acceptors can be thermally excited into the band. Thus, except at low

[17]Note that the large Bohr Radius justifies *post facto* our use of a continuum approximation for the dielectric constant ϵ_r. On small length scales, the electric field is extremely inhomogeneous due to the microscopic structure of the atoms, but on large enough length scales we can use classical electromagnetism and simply model the material as a medium with a dielectric constant.

[18]Because silicon has an anisotropic band, and therefore an anisotropic mass, the actual formula is more complicated.

Fig. 17.4 Energy diagram of a doped semiconductor (left) with donor impurities, or (right) with acceptor impurities. The energy eigenstates of the hydrogenic orbitals tied to the impurities are not all the same because each impurity is perturbed by neighbor impurities. At low temperature, the donor impurity eigenstates are filled and the acceptor eigenstates are empty. But with increasing temperature, the electrons in the donor eigenstates are excited into the conduction band, and similarly the holes in the acceptor eigenstates are excited into the valence band.

enough temperature that the impurities bind the carrier, we can think of the impurities as simply adding carriers to the band. So the donor impurities donate free electrons to the conduction band, whereas the acceptor impurities give free holes to the valence band. However, at very low temperature these carriers get bound back to their respective nuclei so that they can no longer carry electricity—a phenomenon known as *carrier freeze out*. We will typically assume that we are at temperatures high enough (such as room temperature) such that freeze-out does not occur.

Note that in the absence of impurities, the Fermi energy (the chemical potential at zero temperature) is in the middle of the band gap. When donor impurities are added, at zero temperature, impurity states near the top of the band gap are filled. Thus the Fermi energy is moved up to the top of the band gap. On the other hand, when acceptors are added, the acceptor states near the bottom of the band gap are empty (remember it is a bound state of a hole to a nucleus!). Thus, the Fermi energy is moved down to the bottom of the band gap.

Optical Effects of Impurities (Redux)

As mentioned previously in Section 16.5.4, the presence of impurities in a material can have dramatic effects on its optical properties. There are two main optical effects of impurities. The first effect is that the impurities add charge carriers to an otherwise insulating material—turning an insulator into something that conducts at least somewhat. This obviously can have some important effects on the interaction with light. The second important effect is the introduction of new energy levels within the gap. Whereas before the introduction of impurities, the lowest energy transition that can be made is the full energy of the gap, now one can have optical transitions between impurity states, or from the bands to the impurity states.

17.3 Statistical Mechanics of Semiconductors

We now use our knowledge of statistical physics to analyze the occupation of the bands at finite temperature.

Imagine a band structure as shown in Fig. 17.5. The minimum energy of the conduction band is defined to be ϵ_c and the maximum energy of the valence band is defined to be ϵ_v. The band gap is correspondingly $E_{gap} = \epsilon_c - \epsilon_v$.

Recall from way back in Eq. 4.10 that the density of states per unit volume for free electrons (in three dimensions with two spin states) is given by

$$g(\epsilon \geqslant 0) = \frac{(2m)^{3/2}}{2\pi^2\hbar^3}\sqrt{\epsilon}.$$

The electrons in our conduction band are exactly like these free electrons, except that (a) the bottom of the band is at energy ϵ_c and (b) they have an effective mass m_e^*. Thus the density of states for these electrons near the bottom of the conduction band is given by

$$g_c(\epsilon \geqslant \epsilon_c) = \frac{(2m_e^*)^{3/2}}{2\pi^2\hbar^3}\sqrt{\epsilon - \epsilon_c}.$$

Similarly, the density of states for holes near the top of the valence band is given by

$$g_v(\epsilon \leqslant \epsilon_v) = \frac{(2m_h^*)^{3/2}}{2\pi^2\hbar^3}\sqrt{\epsilon_v - \epsilon}.$$

At fixed chemical potential μ the total density of electrons n in the conduction band, as a function of temperature T, is thus given by

$$n(T) = \int_{\epsilon_c}^{\infty} d\epsilon\, g_c(\epsilon)\, n_F(\beta(\epsilon-\mu)) = \int_{\epsilon_c}^{\infty} d\epsilon\, \frac{g_c(\epsilon)}{e^{\beta(\epsilon-\mu)}+1}$$

where n_F is the Fermi occupation factor, and $\beta^{-1} = k_B T$ as usual. If the chemical potential is "well below" the conduction band (i.e., if $\beta(\epsilon - \mu) \gg 1$), then we can approximate

$$\frac{1}{e^{\beta(\epsilon-\mu)}+1} \approx e^{-\beta(\epsilon-\mu)}\,.$$

In other words, Fermi statistics can be replaced by Boltzmann statistics when the temperature is low enough that the density of electrons in the band is very low. (We have already run into this principle in Section 17.1.1 when we discussed that Drude theory, a classical approach that neglects Fermi statistics, actually works very well for electrons above the band gap in semiconductors!) We thus obtain

$$n(T) \approx \int_{\epsilon_c}^{\infty} d\epsilon\, g_c(\epsilon) e^{-\beta(\epsilon-\mu)} = \frac{(2m_e^*)^{3/2}}{2\pi^2\hbar^3} \int_{\epsilon_c}^{\infty} d\epsilon\, (\epsilon-\epsilon_c)^{1/2} e^{-\beta(\epsilon-\mu)}$$

$$= \frac{(2m_e^*)^{3/2}}{2\pi^2\hbar^3} e^{\beta(\mu-\epsilon_c)} \int_{\epsilon_c}^{\infty} d\epsilon\, (\epsilon-\epsilon_c)^{1/2} e^{-\beta(\epsilon-\epsilon_c)}\,.$$
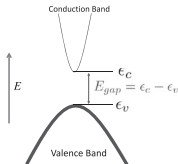
Fig. 17.5 A band diagram of a semiconductor near the top of the valence band (mostly filled) and the bottom of the conduction band (mostly empty). This diagram shows a direct band gap, but the considerations of this section apply to indirect gaps as well.

The last integral is (using $y^2 = x = \epsilon - \epsilon_c$).

$$\int_0^\infty dx\, x^{1/2} e^{-\beta x} = 2 \int_0^\infty dy\, y^2 e^{-\beta y^2} = -2\frac{d}{d\beta}\int_0^\infty dy\, e^{-\beta y^2}$$

$$= -\frac{d}{d\beta}\sqrt{\frac{\pi}{\beta}} = \frac{1}{2}\,\beta^{-3/2}\sqrt{\pi}.$$

Thus we obtain the standard expression for the density of electrons in the conduction band:

$$n(T) = \frac{1}{4}\left(\frac{2m_e^* k_B T}{\pi\hbar^2}\right)^{3/2} e^{-\beta(\epsilon_c - \mu)}. \qquad (17.8)$$

Note that this is mainly just exponential activation from the chemical potential to the bottom of the conduction band, with a prefactor which doesn't change too quickly as a function of temperature (obviously the exponential changes very quickly with temperature!).

Quite similarly, we can write the density of holes in the valence band p as a function of temperature:[19]

$$p(T) = \int_{-\infty}^{\epsilon_v} d\epsilon\, g_v(\epsilon)\left[1 - \frac{1}{e^{\beta(\epsilon-\mu)}+1}\right] = \int_{-\infty}^{\epsilon_v} d\epsilon\, \frac{g_v(\epsilon)e^{\beta(\epsilon-\mu)}}{e^{\beta(\epsilon-\mu)}+1}$$

Again, if μ is substantially above the top of the valence band, we have $e^{\beta(\epsilon-\mu)} \ll 1$ so we can replace this by

$$p(T) = \int_{-\infty}^{\epsilon_v} d\epsilon\, g_v(\epsilon)e^{\beta(\epsilon-\mu)}$$

and the same type of calculation then gives

$$p(T) = \frac{1}{4}\left(\frac{2m_h^* k_B T}{\pi\hbar^2}\right)^{3/2} e^{-\beta(\mu-\epsilon_v)} \qquad (17.9)$$

again showing that the holes are activated from the chemical potential *down* into the valence band (recall that pushing a hole down into the valence band costs energy!).

Law of Mass Action

A rather crucial relation is formed by combining Eqs. 17.8 and 17.9,

$$n(T)p(T) = \frac{1}{2}\left(\frac{k_B T}{\pi\hbar^2}\right)^3 (m_e^*\, m_h^*)^{3/2} e^{-\beta(\epsilon_c - \epsilon_v)}$$

$$= \frac{1}{2}\left(\frac{k_B T}{\pi\hbar^2}\right)^3 (m_e^*\, m_h^*)^{3/2} e^{-\beta E_{gap}} \qquad (17.10)$$

where we have used the fact that the gap energy $E_{gap} = \epsilon_c - \epsilon_v$. Eq. 17.10 is sometimes known as the *law of mass action*,[20] and it is true independent of doping of the material.

[19] If the Fermi factor gives the probability n_F that a state is occupied by an electron, then $1 - n_F$ gives the probability that the state is occupied by a hole.

[20] The nomenclature here, "law of mass action", is a reference to an analog in chemistry. In chemical reactions we may have an equilibrium between two objects A and B and their compound AB. This is frequently expressed as

$$A + B \rightleftharpoons AB$$

There is some chemical equilibrium constant K which gives the ratio of concentrations

$$K = \frac{[AB]}{[A][B]}$$

where $[X]$ is the concentration of species X. The law of mass action states that this constant K remains fixed, independent of the individual concentrations. In semiconductor physics it is quite similar, only the "reaction" is

$$e + h = 0,$$

the annihilation of an electron and a hole, so that the product of $[e] = n$ and $[h] = p$ is fixed.

Intrinsic Semiconductors

For an intrinsic (i.e., undoped) semiconductor the number of electrons excited into the conduction band must be equal to the number of holes left behind in the valence band, so $p = n$. We can then divide Eq. 17.8 by 17.9 to get

$$1 = \left(\frac{m_e^*}{m_h^*}\right)^{3/2} e^{-\beta(\epsilon_v + \epsilon_c - 2\mu)}.$$

Taking the log of both sides gives the useful relation

$$\mu = \frac{1}{2}(\epsilon_c + \epsilon_v) + \frac{3}{4}(k_B T)\log(m_h^*/m_e^*). \tag{17.11}$$

Note that at zero temperature, the chemical potential is precisely mid-gap.

Using either this expression, or by using the law of mass action along with the constraint $n = p$, we can obtain an expression for the *intrinsic* density of carriers in the semiconductor

$$n_{intrinsic} = p_{intrinsic} = \sqrt{np} = \frac{1}{\sqrt{2}}\left(\frac{k_B T}{\pi \hbar^2}\right)^{3/2}(m_e^* m_h^*)^{3/4} e^{-\beta E_{gap}/2}.$$

Doped Semiconductors

For doped semiconductors, the law of mass action still holds. If we further assume that the temperature is high enough so that there is no carrier freeze-out (i.e., carriers are not bound to impurities) then we have

$$n - p = (\text{density of donors}) - (\text{density of acceptors}).$$

This, along with the law of mass action, gives us two equations with two unknowns which can be solved.[21] The result is that if we are at a temperature where the undoped intrinsic carrier density is much greater than the dopant density, then the dopants do not matter much, and the chemical potential is roughly midgap as in Eq. 17.11 (this is the *intrinsic* regime). On the other hand, if we are at a temperature where the intrinsic undoped density is much smaller than the dopant density, then we can think of this as a low-temperature situation where the carrier concentration is mainly set by the dopant density (this is the *extrinsic* regime). In the n-doped case, the bottom of the conduction band gets filled with the density of electrons from the donors, and the chemical potential gets shifted up towards the conduction band. Correspondingly, in the p-doped case, holes fill the top of the valence band, and the chemical potential gets shifted down towards the valence band. Note that in this case of strong doping, the majority carrier concentration is obtained just from the doping, whereas the minority carrier concentration—which might be very small—is obtained via law of mass action. The ability to add carriers of either charge to semiconductors by doping is absolutely crucial to being able to construct semiconductor devices, as we will see in the next chapter.

[21] Here is how to solve these two equations. Let

$$D = \text{doping} = n - p.$$

Let us further assume that $n > p$ so $D > 0$ (we can do the calculation again making the opposite assumption, at the end). Also let

$$I = n_{intrinsic} = p_{intrinsic}$$

so that

$$I^2 = \frac{1}{2}\left(\frac{k_B T}{\pi \hbar^2}\right)^3 (m_e^* m_h^*)^{3/2} e^{-\beta E_{gap}}$$

from the law of mass action. Using $np = I^2$, we can then construct

$$D^2 + 4I^2 = (n - p)^2 + 4np = (n + p)^2$$

So we obtain

$$n = \frac{1}{2}\left(\sqrt{D^2 + 4I^2} + D\right)$$
$$p = \frac{1}{2}\left(\sqrt{D^2 + 4I^2} - D\right).$$

As stated in the main text, if $I \gg D$ then the doping D is not important. On the other hand, if $I \ll D$ then the majority carrier density is determined by the doping only, the thermal factor I is unimportant, and the minority carrier density is fixed by the law of mass action.

Chapter Summary

- Holes are the absence of an electron in the valence band. These have positive charge (electrons have negative charge), and positive effective mass. The energy of a hole gets larger at larger momentum (away from the maximum of the band) as they get pushed down into the valence band. The positive charge of the hole as a charge carrier explains the puzzle of the sign of the Hall coefficient.

- Effective mass of electrons is determined by the curvature at the bottom of the conduction band. Effective mass of holes is determined by the curvature at top of conduction band.

- Mobility of a carrier is $\mu = |e\tau/m^*|$ in Drude theory.

- When very few electrons are excited into the conduction band, or very few holes into the valence band, Boltzmann statistics is a good approximation for Fermi statistics, and Drude theory is accurate.

- Electrons or holes can be excited thermally, or can be added to a system by *doping* and can greatly change the optical and electrical properties of the material. The law of mass action assures that the product np is fixed independent of the amount of doping (it only depends on the temperature, the effective masses, and the band gap).

- Know how to derive the law of mass action!

- At very low temperature carriers may freeze out, binding to the impurity atoms that they came from. However, because the effective Rydberg is very small, carriers are easily ionized into the bands.

References

- Ashcroft and Mermin, chapter 28. (A very good discussion of holes and their effective mass is given in chapter 12. In particular see page 225 and thereafter.)
- Rosenberg, chapter 9
- Hook and Hall, sections 5.1–5.5
- Kittel, chapter 8
- Burns, chapter 10 not including 10.17 and after
- Singleton, chapters 5–6
- Ibach and Luth, sections 12–12.5
- Sze, chapter 2

Exercises

(17.1) **Holes**

(a) In semiconductor physics, what is meant by a hole and why is it useful?

(b) An electron near the top of the valence band in a semiconductor has energy

$$E = -10^{-37} |\mathbf{k}|^2$$

where E is in Joules and k is in m^{-1}. An electron is removed from a state $\mathbf{k} = 2 \times 10^8 \text{m}^{-1} \hat{x}$, where $\hat{x}$ is the unit vector in the x-direction. For a hole, calculate (and give the sign of!)

    (i) the effective mass

    (ii) the energy

    (iii) the momentum

    (iv) the velocity.

▷ If there is a density $p = 10^5 \text{m}^{-3}$ of such holes all having almost exactly this same momentum, calculate the current density and its sign.

(17.2) **Law of Mass Action and Doping of Semiconductors**

(a) Assume that the band-gap energy E_g is much greater than the temperature $k_B T$. Show that in a pure semiconductor at a fixed T, the product of the number of electrons (n) and the number of holes (p) depends only on the density of states in the conduction band and the density of states in the valence band (through their effective masses), and on the band-gap energy.

▷ Derive expressions for n for p and for the product np. You may need to use the integral $\int_0^\infty dx\, x^{1/2} e^{-x} = \sqrt{\pi}/2$.

(b) The band gaps of silicon and germanium are 1.1 eV and 0.75 eV respectively. You may assume the effective masses for silicon and germanium are isotropic, roughly the same, and are roughly .5 of the bare electron mass for both electrons and holes. (Actually the effective masses are not quite the same, and furthermore the effective masses are both rather anisotropic, but we are just making a rough estimate here.)

▷ Estimate the conduction electron concentration for intrinsic (undoped) silicon at room tempera-

▷ Make a rough estimate of the maximum concentration of ionized impurities that will still allow for this "intrinsic" behavior.

▷ Estimate the conduction electron concentration for germanium at room temperature.

(c) The graph in Fig. 17.6 shows the relationship between charge-carrier concentration for a certain n-doped semiconductor.

▷ Estimate the band gap for the semiconductor and the concentration of donor ions.

▷ Describe in detail an experimental method by which these data could have been measured, and suggest possible sources of experimental error.
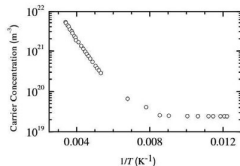

Fig. 17.6 Figure for Exercise 17.2.

(17.3) **Chemical Potential**

(a) Show that the chemical potential in an intrinsic semiconductor lies in the middle of the gap at low temperature.

(b) Explain how the chemical potential varies with temperature if the semiconductor is doped with (i) donors (ii) acceptors.

(c) A direct-gap semiconductor is doped to produce a density of 10^{23} electrons/m^3. Calculate the hole density at room temperature given that the gap is 1.0 eV, and the effective mass of carriers in the conduction and valence band are 0.25 and 0.4 electron masses respectively. Hint: use the result of Exercise 17.2.a.

(17.4) **Energy Density**

Show that the energy density of electrons in the valence band of a semiconductor is

$$\left(\epsilon_c + \frac{3}{2} k_B T\right) n$$

where n is the density of these electrons and ϵ_c is the energy of the bottom of the conduction band.

(17.5) **Semiconductors**

Describe experiments to determine the following properties of a semiconductor sample: (i) sign of the majority carrier (ii) carrier concentration (assume that one carrier type is dominant) (iii) band gap (iv) effective mass (v) mobility of the majority carrier.

(17.6) **More Semiconductors**

Outline the absorption properties of a semiconductor and how these are related to the band gap. Explain the significance of the distinction between a direct and an indirect semiconductor. What region of the optical spectrum would be interesting to study for a typical semiconducting crystal?

(17.7) **Yet More Semiconductors**

Outline a model with which you could estimate the energy of electron states introduced by donor atoms

into an n-type semiconductor. Write down an expression for this energy, explaining why the energy levels are very close to the conduction band edge.

(17.8) **Maximum Conductivity***

Suppose holes in a particular semiconductor have mobility μ_h and electrons in this semiconductor have mobility μ_e. The total conductivity of the semiconductor will be

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

with n and p the densities of electrons in the conduction band and holes in the valence band. Show that, independent of doping, the maximum conductivity that can be achieved is

$$\sigma = 2e\, n_{intrinsic} \sqrt{\mu_e \mu_h}$$

with $n_{intrinsic}$ the intrinsic carrier density. For what value of $n - p$ is this conductivity achieved?

(17.9) **Hall Effect with Both n- and p-Dopants***

Suppose a semiconductor has a density p of holes in the valence band with mobility μ_h and a density n of electrons in the conduction band with mobility μ_n. Use Drude theory to calculate the Hall resistivity of this sample.