

## Semiconductor Basics

In solids, the electron energy levels are organized into bands. How does this come about? Imagine a regular crystal lattice composed of identical atoms. Imagine that the size of the crystal can be varied. So we can consider the crystal to have an atomic spacing anywhere from its actual spacing to very large separations between the atoms. What are the energy levels when the atoms are far apart? The energy levels are just the atomic levels but the degeneracy of each atomic level is multiplied by  $N$ , the total number of atoms in our crystal. Now imagine bringing the atoms closer together. Eventually, they will be close enough that the electrons in neighboring atoms can begin to have weak interactions. Each atomic energy level then gives rise to  $N$  very closely spaced levels corresponding to the original energy. As the atoms are brought still closer together, the interactions among neighboring atoms increase and the spreading out of the original atomic levels into a range of levels increases. In effect, each atomic level becomes a band of extremely closely spaced levels in the crystal. There can be gaps between bands and bands can overlap.

The bands are most important for the energy levels of the outer electrons. The wave functions of the inner electrons barely overlap from one atom to the next, so there is very little spreading out of the energy levels of the inner electrons. The inner electrons can be thought of as belonging to a specific atom. They are localized at the site of the atom. On the other hand the outer electrons are more properly thought of as bound to the crystal rather than any particular atom. Of the most interest for what we're about to discuss are the bands of the outermost electrons.

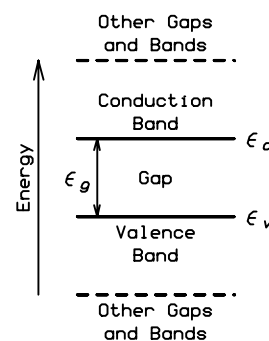
What's the difference between an insulator, a semi-conductor, and a conductor? Consider a crystal at  $\tau = 0$  (and suppose it doesn't become a superconductor!). The energy levels available to the outer electrons are in bands as we've discussed. The number of levels within a band is just the number of atoms in the crystal. At absolute 0, electrons fill up the energy levels until they reach the Fermi energy. Where the Fermi energy occurs with respect to a band edge is critical for determining whether the material is an insulator or a conductor. If a band is just filled, so there is an appreciable energy gap to the next available level, the material will be an insulator. To make any change to the electron distribution, for example, to produce a distribution of electrons with a non-zero average velocity, requires that electrons be given enough energy to overcome the gap between bands. If the gap is substantial (several electron volts, say) ordinary electric fields will not do it. (Several electron volts is the energy of a photon of visible light!) On the other hand, if a band is only partially filled, or if the gap from a filled band to the next higher empty band is very small, then it is easy to give electrons enough energy to change their distribution and in particular to put them in a distribution with a net average velocity in which case they are carrying an electric current. An insulator has a filled band with a big gap to the next band. A conductor has a partially filled band or else a filled band with a very small gap to the next band.

If the temperature is not zero, then the electron distribution does not end abruptly

at the Fermi energy. Instead, there will be electrons in states with energies above the Fermi level and there will be holes (missing electrons) in states at energies below the Fermi level. If at  $\tau = 0$  a band was filled with a gap to the next band, then if the gap is not too large, there may be an appreciable number of electrons thermally excited to the next band (leaving the same number of holes behind). In this case, there aren't as many charge carriers as in a good conductor, but there are more than in an insulator, so we have a *semiconductor*. Semiconductors are technologically useful because their electrical properties are relatively easy to control. We will investigate some of the properties of semiconductors starting with the electron distribution.

### Electron Distribution in Semiconductors

We will consider a situation where the *valence band* is completely full at  $\tau = 0$  but the *conduction band* is completely empty. This is in the absence of doping by impurities. With doping, there may be some excess electrons that must reside in the conduction band, even at  $\tau = 0$  or there may be some holes (a deficit of electrons) so the valence band is not completely full at  $\tau = 0$ . The energy of the top of the valence band is denoted by  $\epsilon_v$ , the bottom of the conduction band is  $\epsilon_c$ , and the energy gap is  $\epsilon_g = \epsilon_c - \epsilon_v$ .



The concentrations of electrons in the conduction band and holes in the valence band are denoted by  $n_e$  and  $n_h$ . With a pure semiconductor, they are equal so that the semiconductor is electrically neutral. But most of the fun with semiconductors comes from the doping. Impurity atoms which provide an extra electron are called donors. Silicon is the basis of much of the semiconductor industry. An atom from the periodic table column just to the right of silicon is a donor. Phosphorus is often used. An atom from the periodic table column just to the left of silicon is an acceptor. Boron is often used. Of course, not all donors and acceptors will be ionized, so the concentration of excess electrons will not be exactly the same as the  $n_d$ , the concentration of donor atoms. Instead, consider the concentration of donor and acceptor ions,  $n_d^+$  and  $n_a^-$ . Then

$$\Delta n = n_d^+ - n_a^- ,$$

is the net ionized donor concentration, or the net concentration of positive charges from the ions. If the sample is electrically neutral, this must be balanced by the net concentration of negative charges in electrons and holes in the conduction and valence bands,

$$n_e - n_h = \Delta n = n_d^+ - n_a^- .$$

The electron distribution (that is, the probability that a state of energy  $\epsilon$  is occupied) is the Fermi-Dirac distribution (see lectures 13 and 16),

$$f_e(\epsilon) = \frac{1}{e^{(\epsilon - \mu)/\tau} + 1} .$$

The chemical potential of the electrons is  $\mu$ . At this point, we have to mention a little jargon. Recall that when we discussed Fermi gases earlier, we used the term “Fermi energy,” and the symbol  $\epsilon_F$ , to denote the chemical potential of the electrons at  $\tau = 0$ ; that is, the energy of the highest filled state at  $\tau = 0$ . In the world of semiconductors, the term “Fermi level,” and the symbol  $\epsilon_F$  are used to denote the chemical potential of the electrons at non-zero temperature. We will follow K&K and continue to let  $\epsilon_F$  denote the chemical potential at  $\tau = 0$  and we will use  $\mu$  for the chemical potential at other temperatures.

The total number of conduction electrons is given by summing the electron distribution over all states in the conduction band,

$$N_e = \sum_{CB} f_e(\epsilon).$$

Note that this is the total number, not the concentration, since we’re summing over all states. A hole in the valence band is the absence of an electron. The probability that a state is occupied by a hole (that is, there’s not an electron in the state) plus the probability that the state is occupied by an electron must sum to one since these events are mutually exclusive and are the only things that can happen. The distribution function for holes is then

$$f_h(\epsilon) = 1 - f_e(\epsilon) = 1 - \frac{1}{e^{(\epsilon - \mu)/\tau} + 1} = \frac{1}{e^{(\mu - \epsilon)/\tau} + 1},$$

and curiously, a hole (I keep wanting to say anti-electron, but it’s not really!) looks as though it has energy and chemical potential opposite to that of the electron, just as it has charge opposite to that of the electron. In any case, the total number of holes in the valence band requires a sum over the states in the valence band

$$N_h = \sum_{VB} f_h(\epsilon).$$

These sums of the occupancies over states are similar in principle to the sums we did for the free electron gas. The big difference has to do with the number of states as a function of energy. In the free electron gas, we were able to use particle in a box states to get the number of states versus energy. With bands, the density of states are determined by the lattice structure, the doping, and so on. We will be making appropriate approximations!

K&K table 13.1 gives some data for popular semiconductors. The energy gap at room temperature is near an electron volt for most of the entries in the table. Recall that at room temperature  $\tau = kT = \frac{1}{40}$  eV and thermal energies are small compared to the gap energies. Even the small 0.18 eV gap of InSb is large compared to thermal energies. At least in pure (undoped) semiconductors, we expect to find the chemical potential somewhere in the middle of the gap. So let’s start by assuming that this is the case and we can check our results to see if they’re consistent with this assumption.

Consider the occupancy at the conduction band edge. It involves the exponential  $\exp((\epsilon_c - \mu)/\tau)$ . By the assumptions we've just made, the argument of the exponential is large and the exponential is even larger. The same argument applies but more so for higher energy states in the conduction band. Then the electron distribution is

$$f_e(\epsilon) = \frac{1}{e^{(\epsilon - \mu)/\tau} + 1} \approx e^{-(\epsilon - \mu)/\tau} \ll 1.$$

Similarly, the hole distribution is

$$f_h(\epsilon) = \frac{1}{e^{(\mu - \epsilon)/\tau} + 1} \approx e^{-(\mu - \epsilon)/\tau} \ll 1.$$

So, with our assumptions, both electrons and holes are distributed “classically.” Of course, they're not really, but we managed to get the distribution functions in a form that looks like the classical Boltzmann factor. What's actually going on is that the step from 100% occupancy to 0 occupancy is hidden in the band gap. There's a small exponential tail sticking into the conduction band and there's one minus a small exponential tail sticking into the valence band. By considering holes, not electrons, as the occupants of the valence band, we convert this into an exponential tail of holes sticking into the valence band.

Returning to the total number of electrons in the conduction band, we have

$$N_e = \sum_{CB} e^{-(\epsilon - \mu)/\tau} = e^{-(\epsilon_c - \mu)/\tau} \underbrace{\sum_{CB} e^{-(\epsilon - \epsilon_c)/\tau}}_{N_c}.$$

The expression for  $N_c$  looks very much like a single particle partition function. We're just summing Boltzmann factors over all the energy states in the conduction band. When we did this sum for a single particle in a box, we came up with  $Z_1 = n_Q V$  with  $n_Q = (m\tau/2\pi\hbar^2)^{3/2}$ . The  $\tau^{3/2}$  dependence comes about because the energy depends on momentum squared and the single particle states are uniformly distributed in 3 dimensions of momentum. (When we integrate, our volume element is  $dp_x dp_y dp_z$  and we need a  $\tau^{1/2}$  for each momentum.) Since we have a Boltzmann factor, only the states near  $\epsilon_c$  will be important in determining  $N_c$ . Assuming we knew the density of states, we might expect that the integral we would write would again have a volume element  $dp_x dp_y dp_z$ , so we might expect to get a  $\tau^{3/2}$  temperature dependence. Then we can account for the actual density of states by introducing an effective mass in place of the electron mass in the quantum concentration. So, we wind up with

$$N_c = 2 \left( \frac{m_e^* \tau}{2\pi\hbar^2} \right)^{3/2} V = n_c V,$$

where the factor of 2 accounts for the two spin states of the electron,  $n_c$  is the *conduction electron quantum concentration*, and  $m_e^*$  is called the *density of states effective mass*. This

means it's the fudge factor to be used in the density of states to get the overall number  $N_c$  to come out right. K&K table 13.1 lists these masses as well as the equivalent quantities for holes. Putting all this together, we have

$$n_e = n_c e^{-(\epsilon_c - \mu)/\tau},$$

and for our earlier assumptions concerning the size of the gap compared with thermal energies to be valid, we must have

$$n_e \ll n_c.$$

We do the same thing with the holes,

$$N_h = \sum_{VB} e^{-(\mu - \epsilon)/\tau} = e^{-(\mu - \epsilon_v)/\tau} \underbrace{\sum_{VB} e^{-(\epsilon_v - \epsilon)/\tau}}_{N_v},$$

The quantum concentration of the holes is

$$n_v = N_v/V = 2 \left( \frac{m_h^* \tau}{2\pi \hbar^2} \right)^{3/2},$$

where  $m_h^*$  is the density of states effective mass for the holes. Then the hole concentration is

$$n_h = n_v e^{-(\mu - \epsilon_v)/\tau}.$$

Note that the two effective masses are really just proportionality constants to get the number of electrons and holes to come out right. They need not be equal, even in the same semiconductor.

It should also be pointed out that effective masses occur in other contexts in condensed matter physics. In particular, as an electron or hole moves through a semiconductor, it interacts with the surrounding atoms and electrons, perhaps causing them to move as well, and the electron may have a “dynamical” effective mass that's different from the actual mass.

The two quantum concentrations  $n_c$  and  $n_v$  are also known as the *effective density of states* in the conduction and valence bands.

## Law of Mass Action for Semiconductors

We have expressions for the electron and hole concentrations in the conduction and valence bands. If we multiply them together, the chemical potential drops out,

$$n_e n_h = n_c e^{-(\epsilon_c - \mu)/\tau} n_v e^{-(\mu - \epsilon_v)/\tau} = n_c n_v e^{-(\epsilon_c - \epsilon_v)/\tau} = n_c n_v e^{-\epsilon_g/\tau} .$$

If we have a pure semiconductor, no donors nor acceptors, then the electron and hole concentrations must be equal. The common value is called the *intrinsic carrier concentration* and the semiconductor is said to be an intrinsic semiconductor. This concentration is

$$n_i = (n_c n_v)^{1/2} e^{-\epsilon_g/2\tau} .$$

So, we can write the previous expression as

$$n_e n_h = n_i^2(\tau) ,$$

where the dependence on temperature has been explicitly indicated as a reminder. There's a temperature in an exponential and three powers of temperature in the product of the quantum concentrations. Probably the earlier expression is to be preferred, since it hides less of what's going on. On the other hand, the expression in terms of  $n_i^2$  is certainly more elegant and also, I believe, it's possible to measure  $n_i$  more or less directly without having to measure the energy gap and density of states.

In any case, what we've just come up with is known as the law of mass action for semiconductors. The key assumption we made is that the chemical potential is far enough away from both the valence and conduction band edges that we can approximate a Fermi-Dirac distribution (in both bands) as a classical Maxwell-Boltzmann distribution. The law of mass action will continue to apply (with essentially the same  $n_i^2$ ) for doped semiconductors provided the chemical potential does not move too close to one of the band edges and our assumption that the electron and hole distributions are classical remains valid. Why would the chemical potential move around? Just as with the standard Fermi-Dirac ideal gas, the chemical potential determines the number of particles. For example, if a semiconductor is doped with donors, it will have more electrons and fewer holes, the chemical potential shifts towards the conduction band, the electrons are less strongly cut off by the Boltzmann factor, and the holes are more strongly cut off by the Boltzmann factor.

As a point of interest, we can use the data in table 13.1 of K&K to calculate the intrinsic carrier concentration for silicon at room temperature. The result is

$$n_{i,\text{Si}}(300 \text{ K}) = 4.8 \times 10^9 \text{ cm}^{-3} .$$

In the case of an intrinsic semiconductor, we can calculate the chemical potential in terms of the other quantities we've defined. We have

$$n_e = n_c e^{-(\epsilon_c - \mu)/\tau} = n_h = n_v e^{-(\mu - \epsilon_v)/\tau} .$$

Isolating the chemical potential gives

$$e^{2\mu/\tau} = \frac{n_v}{n_c} e^{(\epsilon_c + \epsilon_v)/\tau},$$

and taking the logarithm gives

$$\mu = \frac{1}{2}(\epsilon_c + \epsilon_v) + \frac{3\tau}{4} \log \frac{m_h^*}{m_e^*}.$$

For an intrinsic semiconductor, the chemical potential is at the center of the gap plus a small offset which comes from the differing densities of states at the two band edges. If for example,  $m_h^* > m_e^*$ , this means that there are a greater density of hole states than electron states. The correction increases  $\mu$  past the gap center to provide a greater occupancy of the electron states relative to the hole states so that there will be equal numbers of electrons and holes. The logarithmic term remains a small correction, even for rather large effective mass ratios because we're working in the regime where thermal energies ( $\sim \tau$ ) are small compared to  $\epsilon_g$ .

### Electron Distribution in Doped Semiconductors

As already mentioned, if we have a doped semiconductor the electron and hole concentrations are no longer equal but must adjust to maintain charge neutrality,

$$n_e - n_h = \Delta n = n_d^+ - n_a^-.$$

If the occupancies are small enough that the classical approximation remains valid, then the law of mass action provides another equation relating  $n_e$  and  $n_h$ . (We assume for the moment that  $\Delta n$  is known.) Solving for the concentrations, we have

$$n_e = \frac{1}{2} \left( \sqrt{\Delta n^2 + 4n_i^2} \pm \Delta n \right).$$

In many cases of practical interest, the doping concentration is much greater than the intrinsic concentration and we can expand the square root

$$\sqrt{\Delta n^2 + 4n_i^2} \approx |\Delta n| + \frac{2n_i^2}{|\Delta n|}.$$

For an  $n$ -type semiconductor (donors),  $\Delta n > 0$ , and

$$n_e \approx \Delta n, \quad n_h \approx \frac{n_i^2}{\Delta n}.$$

For a  $p$ -type semiconductor (acceptors),  $\Delta n < 0$ , and

$$n_e \approx -\frac{n_i^2}{\Delta n}, \quad n_h \approx -\Delta n.$$

In this limit, the semiconductor is called an *extrinsic* semiconductor and whichever carrier is the majority carrier has a concentration equal to the net concentration of dopants while the other carrier is inversely proportional to the dopant concentration. You can see that in an extrinsic  $n$ -type semiconductor, the chemical potential must move towards the conduction band edge compared to its location in the intrinsic case. Similarly, a  $p$ -type semiconductor has a lower chemical potential than the corresponding intrinsic semiconductor. In fact, if we assume that the doping doesn't change  $n_i$ , the effective masses, and the gap width, we can solve for the chemical potential, much as we did before, and we get

$$\mu_p^n = \frac{1}{2}(\epsilon_c + \epsilon_v) + \frac{3\tau}{4} \log \frac{m_h^*}{m_e^*} \pm \frac{\tau}{2} \log \frac{\Delta n^2}{n_i^2},$$

where we've assumed that  $|\Delta n| \gg n_i$ . The upper sign applies to  $n$ -type and the lower sign applies to  $p$ -type semiconductors. You might think that the additional term invalidates our classical treatment of electrons and holes when we get to high temperatures. Actually the problem is at low temperatures because  $n_i$  goes to 0 as the temperature goes to 0. As the temperature is lowered,  $\mu$  goes to one or the other band edge and we must use the full Fermi-Dirac distribution.