

## Lecture 8: Semiconductor

Kit Matan

Mahidol University

In this chapter, we will use the band structures we calculate in the previous chapter to explain some properties of materials that cannot be explained by the free electron model. In the previous chapter, we have learned that electrons are allowed to occupy some range of energy and there is a region, which we call band gaps, at which there can be no electron. We have also learned how the picture of electron bands gives rise to metals (or conductors), semimetals, insulators and semiconductors. However, in this chapter we will only try to understand properties of semiconductors using the band structure.

The semiconductors are a group of materials with a full band, which we call the **valence band**, but the energy gap between the full band and the next empty band, which we call the **conduction band**, is quite small, roughly the same order as  $k_B T$ . The chemical composition of the semiconductors is  $AB$  where  $A$  are elements in Groups II, III and IV and  $B$  are in Groups VI, V and IV, respectively. For example, the III-V compounds are indium antimonide (InSb) and gallium arsenide (GaAs), those in the II-VI group are zinc sulfide (ZnS) and cadmium sulfide (CdS), and those in the IV-IV groups are silicon carbide (SiC), Si and Ge, which have the diamond structure.

In the semiconductors, the conductivity is due to two types of electrons. The first type of electrons is coming from the thermal excitation of electrons from the valence band to the conduction band, and the second type of electrons is from impurities. We will first only consider the contribution to the conductivity from the the first type, which we call the **intrinsic conductivity**.

## 8.1 Band gaps

Furthermore, in the semiconductors, there are two types of band gaps.

1. **Direct band gap:** This type of the band gap occurs when the lowest point of the conductor band is right on top of the highest point of the valence band, that is, they both are at the same  $\vec{k}$ -point. For example, InSb and  $\alpha$ Sn have the direct band gaps.
2. **Indirect band gap:** This band gap occurs when the lowest point of the conductor band are the highest point of the valence band are at different  $\vec{k}$ -points. The excitation between these two points requires an extra momentum transfer from phonons. Therefore, the indirect absorption process has to include the transfer of the phonon momentum besides the energy transfer and both momentum and energy have to be conserved, that is,

$$\vec{k} = \vec{k}_c + \vec{K}$$

and

$$\hbar\omega = E_g + \hbar\Omega,$$

where  $\vec{k}_c$  is the difference of the wave-vectors of the the lowest point of the conductor band are the highest point of the valence band,  $\vec{K}$  is the phonon wave-vector, and  $\hbar\Omega$  is the phonon energy, which is much smaller than  $\hbar\omega$  or the photon energy. Some examples of the indirect band gap semiconductors are Ge and Si.

We can determine the band-gap energy  $E_g$  by measuring the optical absorption. The absorption coefficient increase rapidly if the incident photon energy is larger than the energy gap and the on-set of the coefficient defines the value of the band gap.

## 8.2 Equation of motion for semiconductors

In the free electron model, in order to derive the electrical transport properties of metals we first derive the equation of motion for the free electron by applying Newton's law. We will have to do the same derivation for the electrons in the energy bands.

From the dispersion relation, we find that electrons in the energy band will travel with a group velocity  $v_g$ , which can be calculated from the dispersion relation:

$$v_g = \frac{d\omega}{dk} = \frac{1}{\hbar} \frac{dE}{dk} \quad \text{or} \quad \vec{v} = \frac{1}{\hbar} \nabla_k E(k),$$

where  $E(k)$  is the dispersion relation of electrons in the energy bands. The work done on the electron by an external force can be written as

$$\delta E = F v_g \delta t$$

or

$$\delta E = \left( \frac{dE}{dk} \right) \delta k$$

From these two equations, we obtain

$$F v_g \delta t = \left( \frac{dE}{dk} \right) \delta k \quad \Rightarrow \quad F \frac{1}{\hbar} \left( \frac{dE}{dk} \right) \delta t = \left( \frac{dE}{dk} \right) \delta k$$

Therefore, the equation of motion is

$$\hbar \frac{d\vec{k}}{dt} = \vec{F},$$

where in the last step, we rewritten the expression in a vector form. In this case,  $\vec{F}$  can be electromagnetic force and the calculations of the band electrons are similar to those of the free electrons. In Kittel, there is another derivation of the equation of motion. I encourage you to look at that derivation as well.

## 8.3 Holes and effective mass

In this section, we will consider physical properties of charge carriers given rise by the presence of energy bands. We will focus our attention only in semiconductors. You all know now that for semiconductors at the absolute zero temperature the valence band is filled and there is no electron in the conduction band. However, at a finite temperature some electrons can be excited by thermal energy from the valence band to the conduction band, leaving behind a vacancy in the valence band. This vacancy behaves like a charge carrier with a positive charge  $+e$ , that is, it can move around just like an electron but has a positive charge  $+e$ . We normally call this vacancy in the valence band a **hole**. Therefore, for ideal semiconductors (by an ideal semiconductor, I mean, a semiconductor without impurities) at a finite temperature, there can be charge carriers with both positive and negative charge. Which charge carriers dominate the transport properties depends on how easily they can move inside the lattice. This question is related to effective mass, which we will consider later. First, we will look at the properties of holes.

1.  $\vec{k}_h = -\vec{k}_e$

Since the total wavevector of electron in the filled valence band has to be zero, when one electron leaves the valence band taking with it the momentum of  $\hbar\vec{k}_e$  it must leave behind a negative momentum of  $-\hbar\vec{k}_e$ . This negative momentum is in fact the momentum of the hole. Therefore, we have the above mentioned relation.

2.  $E_h(\vec{k}_h) = -E_e(\vec{k}_e)$

Similar to the case of the wave vector, due to the conservation of energy when one electron leaves the valence band taking with it an energy of  $E_e(-\vec{k}_e)$ . Since we can think of the filled valence band as the ground state with the initial total energy of zero, the energy left behind by the electron must be negative and that is the energy of the hole.

3.  $\vec{v}_h = \vec{v}_e$

This is due to the fact that  $\nabla_k E_h(k_h) = \nabla_k E_e(k_e)$ .

4.  $m_h = -m_e$ , where  $m$ , in this case, is the effective mass, which we will discuss in detail later. The effective mass is inversely proportional to the curvature of energy band. Therefore, since the band of holes is an up-side-down version of the respective band of electrons, the curvature of hole bands must be opposite to that of electrons bands.

5. The equation of motion for electrons and holes are the same except for the fact that  $d\vec{k}_e/dt$  becomes  $d\vec{k}_h/dt$ ,  $-e$  becomes  $+e$ , and  $\vec{v}_e$  becomes  $\vec{v}_h$ .

**Effective mass** The effective mass indicates how difficult or easy it is for electrons or holes to move around. The definition of the effective mass comes from considering the free electron dispersion relation  $E(k) = \frac{\hbar^2 k^2}{2m}$ . We will immediately see that  $m$  in this case of equal to  $\hbar^2/(d^2E(k)/dk^2)$ . Therefore, we will adopt the same expression to define the effective mass for electrons and holes in the band structure, that is,

$$m^* = \frac{\hbar^2}{d^2E(k)/dk^2}.$$

Since  $\frac{d^2E(k)}{dk^2}$  is a curvature of the energy band, the effective mass  $m^*$  is inversely proportional to the curvature of the energy band. Therefore, if the curvature is large,  $m^*$  is small and electrons or holes can move easily around the lattice. In contrast, if the curvature of small,  $m^*$  is large and electrons or holes will have more difficulty moving around the lattice.

In three dimensions, we can generalize the effective mass and write it in terms of tensor.

$$\frac{1}{m^*} \Big|_{\mu\nu} = \frac{1}{\hbar^2} \frac{d^2E(\vec{k})}{dk_\mu dk_\nu},$$

and for the equation of motion

$$\frac{dv_\mu}{dt} = \frac{1}{\hbar} \frac{d^2E(\vec{k})}{dk_\mu dt} = \frac{1}{\hbar} \frac{d^2E(\vec{k})}{dk_\mu dk_\nu} \frac{dk_\nu}{dt} = \frac{1}{\hbar^2} \frac{d^2E(\vec{k})}{dk_\mu dk_\nu} F_\nu = \left( \frac{1}{m^*} \right)_{\mu\nu} F_\nu,$$

where  $\mu$  and  $\nu$  are indices in the Cartesian coordinate, equal to  $x, y, z$ .

We note that the effective mass is not the physical mass of electrons or holes. It is basically due to the fact that electrons and holes in the lattice can feel the potential energy from nucleuses, which influences its motion through the lattice as if its mass (inertial mass) has changed.

**Example: effective mass for one dimensional lattice** We will consider the one-dimensional lattice of which the dispersion has been calculated in the previous chapter. For the upper band or the conduction band, the dispersion relation near the zone boundary of electrons is

$$E_{+\bar{k}} = E_+ + \frac{\hbar^2 \bar{k}^2}{2m_e} \left[ 1 + \frac{2\lambda}{U} \right],$$

and therefore the effective mass is equal to

$$m_e^* = \frac{m_e}{2\lambda/U + 1},$$

where  $m_e$  is the mass of the electron. For the lower band or the valence band the dispersion relation near the zone boundary of holes is

$$E_{-\bar{k}} = E_- + \frac{\hbar^2 \bar{k}^2}{2m_e} \left[ 1 - \frac{2\lambda}{U} \right],$$

and therefore the effective mass of holes is equal to

$$m_h^* = -m_e^* = \frac{m_e}{2\lambda/U - 1}.$$

The curvature of the  $E_{-\bar{k}}$  band is negative but since  $m_h^* = -m_e^*$  for the hole in the valence band, the effective mass of the hole is positive. Note that  $U < \lambda$  ( $2\lambda/U > 1$ ) and hence both  $m_h^*$  in the valence band and  $m_e^*$  in the conduction band for the one-dimensional lattice are positive. However, in general, the effective mass can be negative.

Physically, we can think of the situation where electrons and holes exchange momentum with the lattice. A negative effective mass implies that the net momentum is transferred from electrons to the lattice, that is, electrons are losing its momentum and being decelerated. This effect happens at the top of the conduction band or the local minimum. The effective mass can become very larger if the curvature of the band is small. This is the case for  $4f$  electrons, where the band width is narrow and hence the curvature of the bands is small. We, therefore, collectively call the compounds with  $4f$  electrons the **heavy fermions**.

**Example of band structure in a real system: silicon and germanium** Fig. 8.1 shows the calculated band structure of germanium derived from the combine experimental and theoretical results. The shaded bands are the valence bands. We will only consider the top three bands of the valence bands and the lowest three bands of the conduction bands. Those six bands are the result of the hybridization of the  $p$ -orbitals, making the outer energy bands of the germanium in the lattice different from the energy level of the germanium atom. Electrons fill up three of these six  $p$ -orbitals, which becomes the valence bands. The rest becomes the conduction band.

We can see that for the three valence bands two of them are degenerate at  $k = 0$  with different curvature. One with less curvature gives rise to the larger effective mass hence called **heavy holes** (holes for the valence band), whereas the other with more curvature corresponds to the smaller effective mass hence called **light holes**. The third band splitting from the other two at  $k = 0$ . This splitting is due to the spin-orbit interaction. We, therefore, call it **split-off holes**. Germanium has an indirect gap since the top of the valence band is at  $k = 0$  whereas the bottom of the conduction band is located at  $k = \frac{2\pi}{a}(\frac{1}{2} \frac{1}{2} \frac{1}{2})$ .

The overall feature of the band structure of silicon is very similar to germanium with the exception that the bottom of the conduction band is along the direction (100).

Note that for simplicity we assume that the dispersion near the  $k = 0$  point is spherically, that is, the coefficients in front of  $k_x^2$ ,  $k_y^2$ , and  $k_z^2$  are all the same. However, in a real system, these coefficients are normally different and the dispersion forms an elliptical shape. In that case, the effective mass has to be more appropriately expressed in a form of tensor.

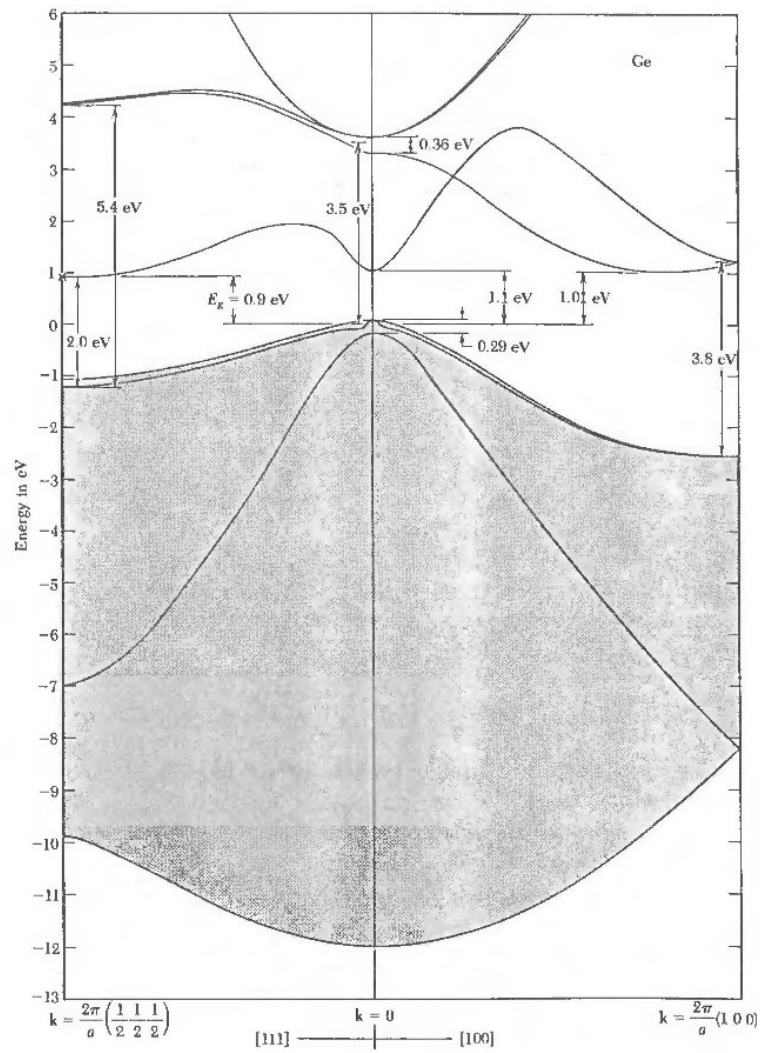


Figure 8.1: The calculated band structure for germanium (Ge) by C. Y. Fong shows the indirect band gap and the spin-orbit splitting at  $k = 0$ . The figure is taken from Kittel.

**Measurement of effective mass** We can measure the effective mass in semiconductors by means of cyclotron resonance. We are all familiar with the cyclotron motion of a free charged particle in a magnetic field. In this case, we can think of an electron in the lattice as a ‘free’ electron with different mass (effective mass). We will consider the electron near the bottom of the conduction band and a hole near the top of the valence band, where the dispersion varies as  $k^2$ . If we applied the external magnetic field  $\vec{B}$ , the equation of motion of this electron or hole become

$$m^* \frac{d\vec{v}}{dt} = \mp e\vec{v} \times \vec{B},$$

where the  $-$  and  $+$  signs indicate the electron and hole, respectively. You all know the solution to this equation of motion from the electromagnetic class that the trajectory of the motion of the charge particle is circular with the frequency of the circular motion or the **cyclotron frequency** is equal to

$$\omega_c = \frac{eB}{m^*}.$$

Note that in order to be able to measure this cyclotron frequency a sample must be pure of impurity to the level that the mean free length is long enough to allow the charge carrier to transverse around a circular path. We can measure this resonance cyclotron frequency by absorption of the electromagnetic wave whose frequency matches the cyclotron frequency.

As we have noted above that in general the dispersion can have an elliptical shape and therefore the cyclotron frequency also depends on the orientation of the applied magnetic field with respect to the ellipsoid. Therefore, by varying the orientation of the magnetic field we can measure the effective mass tensor of the semiconductors as well as shown in Fig. 8.2.

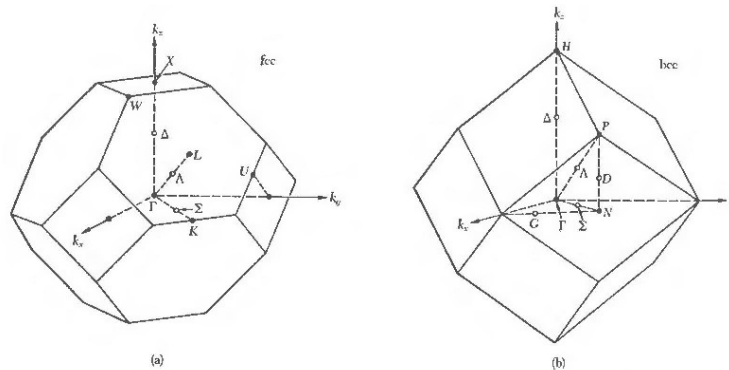


Figure 8.2: Standard labels of the symmetry point of the Brillouin zone of the fcc and bcc lattice. The figure is taken from Kittel.

## 8.4 Intrinsic charge carrier concentration

In this section, we will consider the charge carrier concentration of a pure semiconductor, that is, a semiconductor without any impurities. Let  $\mu$  be the Fermi level or the chemical potential. Our task is to calculate  $\mu$ . We will assume that the interest energy is near the Fermi level, that is,  $E - \mu \gg k_B T$ . Therefore, the Fermi-Dirac distribution function for electrons becomes

$$f_e = \frac{1}{e^{(E-\mu)/k_B T} + 1} \approx e^{(\mu-E)/k_B T}.$$

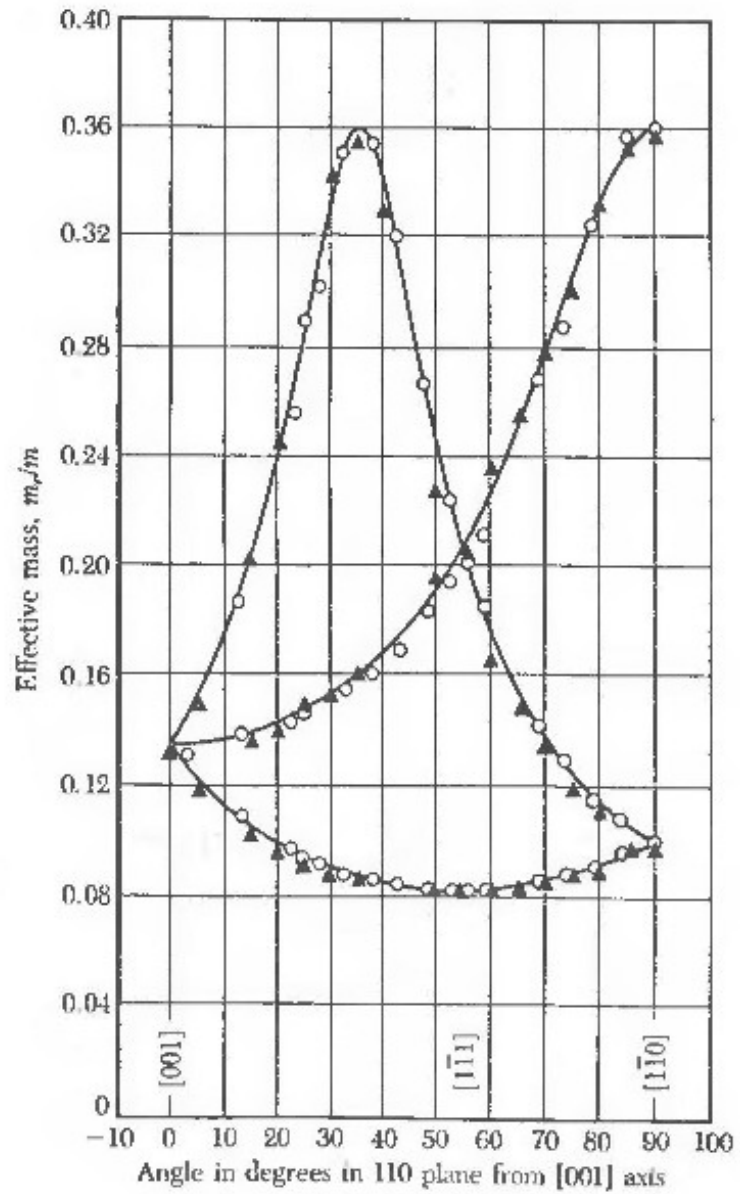


Figure 8.3: Effective mass of germanium as measured by the cyclotron resonance as a function of orientation of the applied magnetic field with respect to the ellipsoid. The figure is taken from Kittel.

The energy of an electron near the bottom of the conduction band is

$$E_{+k} = E_+ + \frac{\hbar^2 k^2}{2m_e},$$

which we have derived in the previous chapter. We have also derived the density of states for electrons:

$$D_e(E) = \frac{1}{2\pi^2} \left( \frac{2m_e}{\hbar^2} \right)^{3/2} (E - E_+)^{1/2},$$

where we ignore a factor of the volume  $V$  and replace  $\varepsilon$  with  $E - E_c$ . Now we are ready to calculate the concentration of electrons in the conduction band at temperature  $T$ ,

$$\begin{aligned} n &= \int_{E_+}^{\infty} D_e(E) f_e(E) dE \\ &= \frac{1}{2\pi^2} \left( \frac{2m_e}{\hbar^2} \right)^{3/2} e^{\mu/k_B T} \int_{E_+}^{\infty} (E - E_+)^{1/2} e^{-E/k_B T} dE \\ n &= 2 \left( \frac{m_e k_B T}{2\pi\hbar^2} \right)^{3/2} e^{(\mu - E_+)/k_B T}. \end{aligned} \quad (8.1)$$

We can do the same calculation to find the concentration of holes at temperature  $T$ . The distribution of holes at temperature is  $1 - f_e$ , which is

$$f_h = \frac{1}{e^{(\mu - E)/k_B T} + 1} \approx e^{(E - \mu)/k_B T}.$$

The dispersion relation of the hole is also different.

$$E_{-k} = E_- - \frac{\hbar^2 k^2}{2m_h}.$$

Therefore, the density of states for holes is

$$D_e(E) = \frac{1}{2\pi^2} \left( \frac{2m_h}{\hbar^2} \right)^{3/2} (E_- - E)^{1/2}.$$

The same calculation leads to the concentration of holes at temperature  $T$ ,

$$p = 2 \left( \frac{m_h k_B T}{2\pi\hbar^2} \right)^{3/2} e^{(E_- - \mu)/k_B T}. \quad (8.2)$$

We can see that the product of  $n$  and  $p$  is independent of the Fermi level  $\mu$ , that is,

$$np = 4 \left( \frac{k_B T}{2\pi\hbar^2} \right)^{3/2} (m_e m_h)^{3/2} e^{-E_g/k_B T} \equiv n_i^2,$$

where  $E_g = E_+ - E_-$ . This relation is called the law of mass action, which holds not only for the intrinsic case but also for the extrinsic case that we will consider later. However, at the equilibrium the concentration of holes and electrons must be equal, that is,

$$n = p = 2 \left( \frac{k_B T}{2\pi\hbar^2} \right)^{3/2} (m_e m_h)^{3/2} e^{E_g/2k_B T}.$$

We can also calculate the Fermi level by letting Eqs. 8.1 and 8.2 be equal to obtain

$$\mu = \frac{1}{2} E_g + \frac{3}{4} k_B T \log \left( \frac{m_h}{m_e} \right),$$

where  $\mu$  is measured from the top of the valence band. If  $m_h = m_e$ , then the Fermi level will lie at the middle of the gap, that is,  $\mu = \frac{1}{2} E_g$ .



**Doped semiconductors and extrinsic charge carrier** We sometimes call doped semiconductors **deficit semiconductors**. We can dope semiconductor without altering much its structure by replacing the quadrivalent atom of the semiconductors with its trivalent or pentavalent counterparts. Therefore, there are two types of doped semiconductors.

1. ***n*-type semiconductor**: In this case, the quadrivalent atom is replaced by the pentavalent atom, which will contribute an extra electron to the semiconductor. This excess electron can be easily ionized and can move around more or less freely around the lattice, increasing the conductivity of the semiconductor.
2. ***p*-type semiconductor**: In this case, the quadrivalent atoms is replaced by the trivalent atom, which will contribute an extra hole to the semiconductor.

The extra charge will feel the screened Coulomb potential from the impurity ion. The ionization energy of this extra charge can be estimated by assuming that the impurity ion introduced the Coulomb potential resembling that of the hydrogen atom. However, we have to replace the mass of an electron with the effective mass and include the dielectric constant due to the screening effect. You are asked to calculate this energy in the homework.

For an example of the doped semiconductors, let us consider a pure-germanium material doped with arsenic atoms. Germanium is a Group IV semiconductor, while arsenic is in Group V with five valence electrons. To first approximation, we will ignore the difference in size between the germanium and arsenic atom and assume that the structure of the lattice does not change. Since arsenic has five valence electrons, one of its electron will not form a bond with the nearby germanium atoms and become free, leaving behind a positively-charge ion center with charge  $+e$ . In this case, we call the arsenic atoms **donors**, since they give electrons to the semiconductors. On the other hand, if we dope germanium with atoms in Group III, those atoms will give holes or accept electrons from the valence band and we will call those impurities **acceptors**. The attractive force between this free electrons and the positive ion center is similar to the attraction between a electron and proton in the hydrogen atom. However, while the electron in the hydrogen is in the free space, the electron in the doped semiconductor is inside the lattice of germanium and it feels other forces from the lattice. The lattice will govern the motion of the electron in two ways

1. Due to the dielectric property of the material, the force that the electron feels is smaller by a factor of  $1/\epsilon$ , where  $\epsilon$  is the dielectric constant. In germanium,  $\epsilon \approx 16$  but in other semiconductors  $\epsilon$  can be as large as 100 or more. It turns out that  $\epsilon$  is large in those semiconductors with a small energy gap. In one limit, we know that for a metal the energy gap is zero and the dielectric constant becomes infinite. In the other limit, in the insulator, the energy gap is very large and the dielectric constant goes to one for the vacuum, which is the ideal insulator.
2. As we have already know, the mass of the electron inside the lattice is different from the free-space mass. Therefore, in calculating the binding energy between the electron from the donors and the ion center, we need to use the effective mass, which is smaller than the free-space electron mass by a factor of 0.1 or less.

Knowing these facts, we can calculate the ionization energy of the donor electrons  $E_d$  using the formulation of the ground state energy of the hydrogen atom and replace  $m$  by  $m^*$  and scale the Coulomb force by  $1/\epsilon$  to obtain

$$E_d = 13.6 \times \frac{1}{\epsilon^2} \frac{m^*}{m} \text{ eV}.$$

For a semiconductor that often has a small band gap and hence large  $\epsilon$ ,  $E_d$  is much smaller than the band gap. Therefore, the electrons from the donors (or holes from the acceptors) are much more likely to be excited into the conduction band (or into the valence band). These electrons, which we call **extrinsic charge carriers** hence dominate the electrical transport of the semiconductors as oppose to the intrinsic charge carriers, which are due to the thermal excitations of electrons from the valence band to the conduction band. The donor level  $E_d$  is near the bottom of the conduction band  $E_+$  and the acceptor level  $E_a$  is near the top of the valence band  $E_-$ , where  $E_+ - E_d$  and  $E_a - E_-$  are much smaller than  $E_g = E_+ - E_-$ .

Next we will calculate the the density of charge carriers for the case of doped semiconductors. Unlike for the intrinsic case, the density of electrons in the conduction band and that of holes in the valence can be different for the extrinsic case, that is,

$$n - p = \Delta n \neq 0.$$

However, like for the intrinsic case, the law of mass action still holds, that is,

$$np = n_i^2.$$

From these two relations, we can write  $n$  and  $p$  in term of  $n_i$  and  $\Delta n$  as

$$\begin{cases} n \\ p \end{cases} = \frac{1}{2} [(\Delta n)^2 + 4n_i^2]^{1/2} \pm \frac{1}{2}\Delta n.$$

We can also write the ration  $\frac{\Delta n}{n_i}$  in terms of the intrinsic chemical potential  $\mu_i$  and the new chemical potential for the doped semiconductors  $\mu$  as

$$\frac{\Delta n}{n_i} = 2 \sinh \beta(\mu - \mu_i),$$

where we note that

$$n = n_i e^{\beta(\mu - \mu_i)} \quad \text{and} \quad p = n_i e^{\beta(\mu_i - \mu)}.$$

We will next calculate the number of occupied donor level using what we have learned in statistical mechanics. The average number of occupied donor levels is equal to

$$\langle n_d \rangle = \frac{\sum N_j e^{-\beta(E_j - \mu N_j)}}{\sum e^{-\beta(E_j - \mu N_j)}}.$$

We will only consider two states, that is, the empty state and the singly occupied state. Note that even though two electrons with opposite spins can occupy the same state, their Coulomb repulsive interaction gives rise to a much higher energy level and hence the two-electron state is excluded. Therefore, we obtain

$$n_d = \langle n_d \rangle N_d = N_d \frac{2e^{-\beta(E_d - \mu)}}{1 + 2e^{-\beta(E_d - \mu)}} = \frac{N_d}{\frac{1}{2}e^{\beta(E_d - \mu)} + 1},$$

where a factor of 2 indicates the two spins of electrons.

For the acceptor levels, there can be two state; one with a single electron in the acceptor level and the other with two electrons in the acceptor level. Note that if there is no electron in the acceptor level, it will equivalently mean that there are two holes in the acceptor level, which requires quite a lot of energy. We, therefore, will exclude this state. Therefore, we obtain the average number of the acceptor levels equal to

$$p_a = (2 - \langle n_d \rangle) N_a = N_a \left( 2 - \frac{2e^{\beta\mu} + 2e^{-\beta(E_a - 2\mu)}}{2e^{\beta\mu} + e^{-\beta(E_a - 2\mu)}} \right) = \frac{N_a}{\frac{1}{2}e^{\beta(\mu - E_a)} + 1},$$

Now we will suppose that

$$E_d - \mu \gg k_B T$$

and

$$\mu - E_a \gg k_B T.$$

This assumption is quite similar to the assumption that  $E_g \gg k_B T$  for the intrinsic case. From the relations for  $n_d$  and  $p_a$  above, we can estimate that  $n_d \ll N_d$  and  $p_a \ll N_a$ . That means, the number of bound states for both electrons and holes are negligible and

$$\Delta n = n - p = N_d - N_a.$$

Therefore, we can write

$$\left\{ \begin{array}{l} n \\ p \end{array} \right\} = \frac{1}{2} [(N_d - N_a)^2 + 4n_i^2]^{1/2} \pm \frac{1}{2} (N_d - N_a).$$

and

$$\frac{N_d - N_a}{n_i} = 2 \sinh \beta (\mu - \mu_i),$$

In the limit of low impurity concentration  $(N_d - N_a)^2 \rightarrow 0$ , we have

$$\left\{ \begin{array}{l} n \\ p \end{array} \right\} = n_i \pm \frac{1}{2} (N_d - N_a).$$

On the other hand, in the limit of high impurity concentration, that is,  $\frac{N_d - N_a}{n_i} \gg 1$ , we obtain

$$\left. \begin{array}{l} n \approx N_d - N_a \\ p \approx \frac{n_i^2}{N_d - N_a} \end{array} \right\} N_d > N_a$$

and

$$\left. \begin{array}{l} n \approx \frac{n_i^2}{N_a - N_d} \\ p \approx N_a - N_d \end{array} \right\} N_a > N_d.$$

This series of equations show that the majority of charge carriers is due to the impurity donors and acceptors, which contribute electrons to the conduction band and holes to the valence band, respectively.

## 8.5 $p - n$ junction

Now we will consider a device that we can make from the doped semiconductors. We can combine the  $p$ -type semiconductor and the  $n$ -type semiconductor to create a diode. Some of you might be familiar with how the diode works. The diode will allow current to go through one direction but prohibit the current from passing through the opposite direction. In this section, we will try to understand how the diode works.

When we connect the  $p$ -type and  $n$ -type semiconductors, the electric potential from the ion centers will shift the chemical potential  $\mu$ . On the  $n$ -type semiconductor, the ion centers have positive charge while on the  $p$ -type semiconductor, the ion centers have negative charge. Therefore, the electric potential is higher on the  $n$ -type semiconductor than on the  $p$ -type semiconductor and the electric potential will point from the former to the latter. Next we will consider the electron and hole distribution. The majority of electrons will be on the  $n$ -type side while the majority of holes will be on the  $p$ -type side. There is repulsion from the negative ion centers on the  $p$ -type side on electrons preventing electrons to move to the  $p$ -type side and from the positive ion center on the  $n$ -type side on holes preventing holes to enter the  $n$ -type side. Therefore, the electrons on the  $n$ -type side cannot recombine with the holes in the  $p$ -type side and hence electrons and holes are separated. The small region where electrons and holes are separated due to the electric potential from the ion center is called the **depletion layer**, which extends for about  $10^2$  to  $10^4$  Å. Normally, some electrons from the  $n$ -type side can diffuse to the  $p$ -type side and vice versa for holes. The region that holes and electrons can diffuse to is called the **transition region**. However, for simplicity, we will assume that the boundary between the  $n$ -type and  $p$ -type semiconductors is abrupt and the charge carrier concentration can be expressed by

$$N_d(x) = \left\{ \begin{array}{ll} N_d, & x > 0 \\ 0, & x < 0 \end{array} \right\},$$

and

$$N_a(x) = \left\{ \begin{array}{ll} 0, & x > 0 \\ N_a, & x < 0 \end{array} \right\},$$

where  $x$  is the distance along the long axis and  $x = 0$  indicates the boundary. This abrupt approximation only works if the transition region is much smaller than the depletion region.

**$p-n$  junction in equilibrium** We will first consider the case where there is no external electric field. We now know that there is the electric potential  $\phi(x)$  between the  $n$ -type and  $p$ -type sides. From a perspective of electrons, this potential will shift both conduction and valence bands by a constant  $-e\phi(x)$ . Therefore, the carrier densities of electrons and holes at equilibrium, which now depends on  $x$  as well, become

$$n(x) = n_0 e^{-\frac{E_+ - e\phi(x) - \mu}{k_B T}}$$

for electrons, and

$$p(x) = p_0 e^{-\frac{\mu - E_- + e\phi(x)}{k_B T}}$$

for holes, where  $n_0 = 2 \left( \frac{m_e k_B T}{2\pi\hbar^2} \right)^{3/2}$  and  $p_0 = 2 \left( \frac{m_h k_B T}{2\pi\hbar^2} \right)^{3/2}$  as we have derive before. For simplicity, we will consider the carrier concentration in the limit where  $x$  is far from the boundary, that is,  $x \rightarrow \infty$  or  $x \rightarrow -\infty$ . We will assume that the  $p$ -type in on the left and the  $n$ -type side is on the right. Therefore, we obtain

$$\begin{aligned} n(\infty) &= N_d = n_0 e^{-\frac{E_+ - e\phi(\infty) - \mu}{k_B T}} \\ p(\infty) &= N_a = p_0 e^{-\frac{\mu - E_- + e\phi(-\infty)}{k_B T}}, \end{aligned}$$

since far to the left we should find that the concentration of electrons on the conduction band must be equal to the concentration of donors and far to the right we should find that the concentration of holes on the valence band must be equal to the concentration of acceptors. Now we will multiply the top two equations and solve from the change in the potential going from one side of the  $p-n$  junction to the other side, that is,

$$e\Delta\phi = e\phi(\infty) - e\phi(-\infty) = E_+ - E_- + k_B T \log \left[ \frac{N_d N_a}{n_0 p_0} \right]$$

or

$$e\Delta\phi = E_g + k_B T \log \left[ \frac{N_d N_a}{n_0 p_0} \right]$$

From this equation, we can see that the potential difference is large if the charge imbalance due to impurity is large, that is, there are a lot of donors and acceptors and hence a lot of ion centers on both  $n$ -type and  $p$ -type sides.

In the above discussion, we consider the  $p-n$  junction in such a way that the chemical potential  $\mu$  does not change but the energy bands shift due to the electric potential. Alternatively, we can consider the case where the energy bands remain the same but the chemical potential change coming from the  $p$ -type side to the  $n$ -type side. In this case, the equations for the carrier density become

$$\begin{aligned} n(x) &= n_0 e^{-\frac{E_+ - \mu_e(x)}{k_B T}} \\ p(x) &= p_0 e^{-\frac{\mu_e(x) - E_-}{k_B T}}, \end{aligned}$$

where  $\mu_e(x) = \mu + e\phi(x)$  and the subscript  $x$  indicates that the potential is from the perspective of electrons and not holes. We can then redefine the potential difference between the two sides by

$$e\Delta\phi = \mu_e(\infty) - \mu_e(-\infty).$$

Looking at this equation, we can see that the potential difference exists to bring the chemical potential on the side of the junction to the same level, which is consistent with the first case where we assume that the chemical potential on both sides of the junction is the same.

The next questions that we can ask are what is the electric potential and how large is the depletion region. To answer the first question, we will have to apply Poisson's equation from the classical theory of electromagnetism, that is,

$$-\nabla^2 \phi = -\frac{d^2 \phi}{dx^2} = \frac{\rho(x)}{\epsilon \epsilon_0}$$

Now we will have to approximate  $\rho(x)$ . First, we know that far away from the boundary the total charge is neutral since the number of charge carriers and ion centers with opposite charge are the same. Second, we know that inside the depletion region electrons and holes recombine. Therefore, the total charge is due to the ion center, which is no longer neutral. On the  $n$ -type side,  $\rho(x)$  is approximately equal to  $-eN_d$  and on the  $p$ -type side to  $eN_a$ . Therefore, we obtain

$$\frac{d^2\phi}{dx^2} = \begin{cases} 0, & x > d_n \\ -\frac{eN_d}{\epsilon\epsilon_0}, & 0 < x < d_n \\ \frac{eN_a}{\epsilon\epsilon_0}, & d_p < x < 0 \\ 0, & x < -d_p \end{cases} ,$$

where  $d_n$  and  $d_p$  is the width of the depletion region on the  $n$ -type side and  $n$ -type side, respectively. We can integrate these equations to obtain  $\phi(x)$

$$\phi(x) = \begin{cases} \phi(\infty), & x > d_n \\ \phi(\infty) - \frac{eN_d}{2\epsilon\epsilon_0}(x-d_n)^2, & 0 < x < d_n \\ \phi(-\infty) + \frac{eN_a}{2\epsilon\epsilon_0}(x+d_p)^2, & d_p < x < 0 \\ \phi(-\infty), & x < -d_p \end{cases}$$

Next we will apply the boundary condition. At  $x = 0$ ,  $\phi$  and  $d\phi/dx$  is continuous. The continuity of  $d\phi/dx$  at  $x = 0$  gives

$$N_d d_n = N_a d_p.$$

That is, the width of the depletion region is inversely proportional to the number of donors or acceptors. The continuity of  $\phi$  at  $x = 0$  implies that

$$\left(\frac{e}{2\epsilon\epsilon_0}\right) (N_d d_n^2 + N_a d_p^2) = \phi(\infty) - \phi(-\infty) = \Delta\phi.$$

We can solve for  $d_{n,p}$  to obtain

$$d_{n,p} = \left[ \frac{(N_a/N_d)^{\pm 1} 2\epsilon\epsilon_0 \Delta\phi}{(N_d + N_a) e} \right]^{1/2}$$

**Making of rectifier using the  $p-n$  junction** Now we will consider the case we apply electric field to the  $p-n$  junction. There are two ways we can connect the power supply to the  $p-n$  junction.

1. **Forward bias:** Current can pass through the  $p-n$  junction.
2. **Back or reverse bias:** Current cannot go through the  $p-n$  junction.

The principle behind the rectifier can be best explained using Figure 8.4.

## References

- [1] Kittel, C.: Introduction to Solid State Physics 7<sup>th</sup> Edition (Chapter 3), John Wiley & Sons, Inc. (1996).
- [2] Ashcroft, N. W. and Mermin, N. D.: Solid State Physics (Chapter 20), Thomson Learning, Inc. (1976).

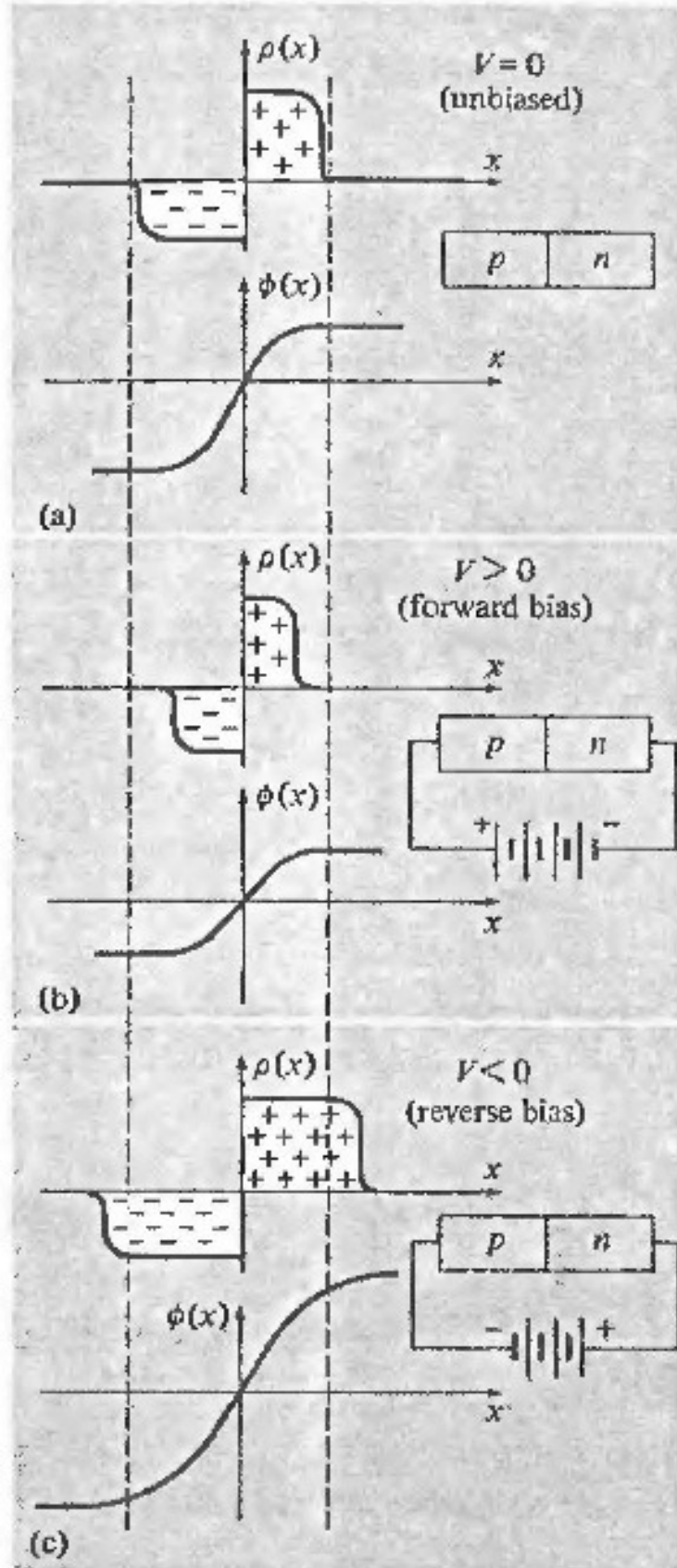


Figure 8.4: The charge density and electric potential in the depletion zone without the external voltage (a), with the forward bias (b) and with the back or reverse bias (c). The figure is taken from Ascroft&Mermin.