

Semiconductor materials: pn junction

Contents

1 Semiconductor Materials	1
1.1 Charge Carriers in Semiconductors	2
1.2 Intrinsic Semiconductors	4
1.3 Extrinsic Semiconductors	4
2 $p - n$ Diode Junction	6
3 Polarisation of a $p - n$ Junction	8

1 Semiconductor Materials

Semiconductors are crystalline materials whose outer shell atomic levels exhibit an energy band structure with an energy gap between valence and conduction bands of the order of $\sim 1\text{eV}$. Figure 1 schematically illustrates this basic structure consisting of a valence band, a forbidden energy gap and a conduction band. The band configuration for conductors and insulators is also shown for comparison.

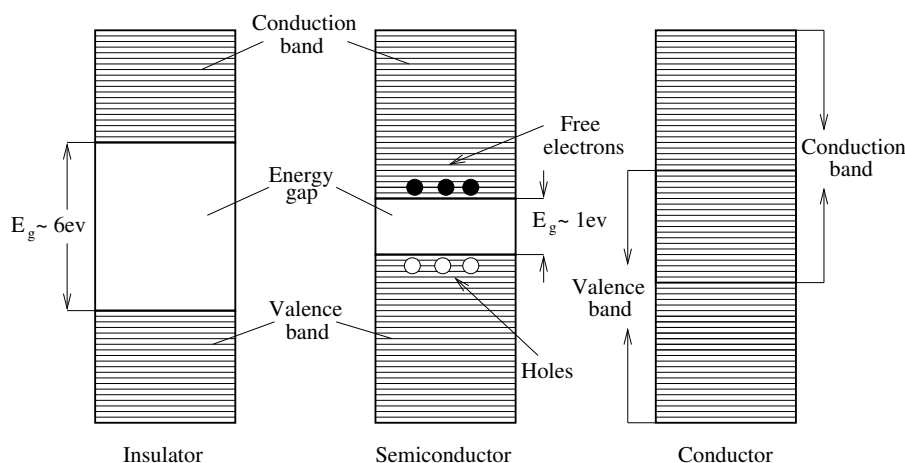


Figure 1: Energy band structure of conductors, insulators and semiconductors. The valence band is the highest energy band completely filled by electrons at 0K. The conduction band is the next higher possible energy band above the valence band, separated by a forbidden band gap in the case of insulators and semiconductors.

The energy bands are regions of many discrete levels so closely spaced that they may be considered as a continuum, while the forbidden energy gap is a region in which there are no available energy levels at all. This band structure arises because of the close, periodic arrangement of the atoms in the crystal which causes an overlapping of the electron wave-functions. Since the Pauli principle forbids more than one electron in the same state, the

degeneracy in the outer atomic shell energy levels breaks to form many discrete levels only slightly separated from each other. As two electrons of opposite spin may reside in the same level, there are as many levels as there are pairs of electrons in the crystal. This degeneracy breaking does not affect the inner atomic levels, however, which are more tightly bound.

The highest energy band is the conduction band. Electrons in this region are detached from their parent atoms and are free to roam about the entire crystal. The electrons in the valence band levels, however, are more tightly bound and remain associated to their respective lattice atoms.

The width of the gap and the bands is determined by the lattice spacing between the atoms. These parameters are thus dependent on the temperature and the pressure. In conductors, the energy gap is nonexistent, while in insulators the gap is large. At normal temperatures, the electrons in an insulator are all in the valence band, the thermal energy being insufficient to excite electrons across this gap. When an external electric field is applied, there is no movement of electrons through the crystal and thus no current. For a conductor, on the other hand, the absence of a band gap makes it very easy for thermally excited electrons to jump into the conduction band where they are free to move around the crystal. A current will therefore flow when an electric field is applied. In a semiconductor, the energy gap is intermediate in size such that only few electrons are excited into the conduction band through thermal excitation. When an electric field is applied, a small current is observed. However, if the semiconductor is cooled, almost all electrons will fall into the valence band and the conductivity of the semiconductor will decrease.¹

1.1 Charge Carriers in Semiconductors

At 0 K, all the electrons in the valence band participate in covalent bonding between the lattice atoms. At normal temperatures, however, the action of thermal energy can excite a valence electron into the conduction band leaving a hole in its original position. In this state, it is easy for a neighbouring valence electron to jump from its bond to fill the hole. This now leaves a hole in the neighbouring position. If the next neighbouring electron repeats the sequence and so on, the hole appears to move through the crystal. The hole acts therefore like a positive charge carrier, and its movement through the crystal constitutes an electric current with direction opposite to the direction of the electrons. In a semiconductor the electric current thus arises from two sources: the movement of free electrons in the conduction band and the movement of holes in the valence band.

Thermal ionization results in free electrons and holes in equal numbers and hence equal concentrations. During the movement of the electrons in the semiconductor crystal, some of them recombine with holes and therefore there are pairs of carriers which disappear. The recombination rate is proportional to the number of free electrons and holes, which is determined by the temperature dependent ionization rate. In thermal equilibrium, the recombination rate is equal to the thermal agitation.

¹The conductivity depends critically on the energy gap, because the fraction of electrons excited across the gap at temperature T is proportional to $\exp^{-E_g/2kT}$. With an energy gap of 4 eV, this factor is $\exp^{-80} \sim 10^{-35}$ at room temperature ($kT \sim 0.025$ eV), and essentially no electrons are excited across the gap. This is the case of insulators. If, however, E_g is 0.25 eV, then the factor at room temperature is $e^{-5} \sim 10^{-2}$, and observable conduction will occur.

The most important properties of a semiconductor are the number of electrons per unit volume in the conduction band (n_c) and the number of holes per unit volume in the valence band (p_v), which are a function of temperature. Suppose that the density of levels is $g_c(E)$ in the conduction band and $g_v(E)$ in the valence band, then we can express:

$$n_c(T) = \int_{E_c}^{\infty} dE g_c(E) \frac{1}{\exp^{(E-E_F)/kT} + 1}$$

$$p_v(T) = \int_{-\infty}^{E_v} dE g_v(E) \left(1 - \frac{1}{\exp^{(E-E_F)/kT} + 1} \right) = \int_{-\infty}^{E_v} dE g_v(E) \frac{1}{\exp^{(E_F-E)/kT} + 1}$$

where E_F (Fermi Energy) is the energy level at which the occupation probability of a state is one half. In the case of semiconductors this energy usually lies within the band gap (see figure 2).

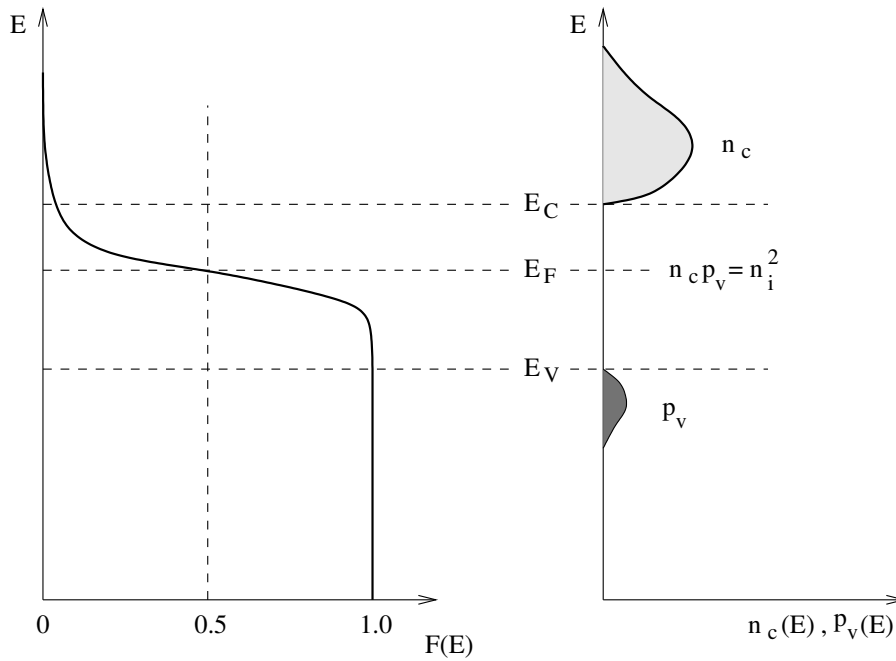


Figure 2: Occupation probability and carrier concentration for intrinsic semiconductors.

In the case $E_c - E_F \ll kT$ and $E_F - E_v \gg kT$, the Fermi-Dirac factors can be simplified:

$$\frac{1}{\exp^{(E-E_F)/kT} + 1} \sim \exp^{-(E-E_F)/kT}, \quad E > E_c$$

$$\frac{1}{\exp^{(E_F-E)/kT} + 1} \sim \exp^{-(E_F-E)/kT}, \quad E < E_v$$

reducing the number of carriers to

$$n_c(T) = N_c(T) \exp^{-(E_c-E_F)/kT}$$

$$p_v(T) = P_v(T) \exp^{-(E_F-E_v)/kT}$$

where

$$N_c(T) = \int_{E_c}^{\infty} dE g_c(E) \exp^{-(E-E_c)/kT}$$

$$P_v(T) = \int_{-\infty}^{E_v} dE g_v(E) \exp^{-(E_v-E)/kT}$$

1.2 Intrinsic Semiconductors

Intrinsic semiconductors contain negligible numbers of impurities compared to the number of thermally generated electrons and holes. In this case the concentration of free electrons $n_c(T)$ is equal to the concentration of free holes $p_v(T)$:

$$n_c(T) = p_v(T) = n_i(T)$$

The relation $n_i^2(T) = n_c(T)p_v(T)$ is called **mass-action law**. The product of $n_c(T)$ times $p_v(T)$ will always stay constant at a certain temperature, whereas their sum will change.

1.3 Extrinsic Semiconductors

An extrinsic semiconductor is a semiconductor with a small fraction of specific impurities that alter the properties of the material. The procedure of adding small proportions of selected impurities, which can be performed either during crystal growth or later in certain regions of the material (via diffusion), is called **doping**. Depending on the type of the added material, one obtains *n*-type semiconductors (excess of electrons in the conduction band) or *p*-type (additional holes in the valence band).

n-type semiconductors: If one of the atoms of the semiconductor crystal is replaced by a pentavalent atom (As, P, Bi, Sb), only four of the electrons are used for the formation of the covalent bonds with a neighbouring atom, while the fifth is not bonded to any specific atom. This additional electron is free for conduction.

p-type semiconductors: If the semiconductor atom is replaced by an trivalent atom (Al, B, Ga), then one electron is missing in the covalent bond and thus a hole is created. This hole may be filled by an electron from a neighbouring atom, this being equivalent to a movement of the hole, so the hole is free for conduction.

Figure 3 shows a two-dimensional schematic bond representation of a silicon crystal with both types of impurities.

In terms of the band model, adding an impurity in the crystal corresponds to the creation of localised energy levels in the band gap. In the case of donor impurities (*n*-type semiconductor), this energy level is close to the conduction band, and in the case of acceptor impurities (*p*-type semiconductor), it is close to the valence band. The energy levels introduced by impurities are in the range of few *meV* up to some tenths of *meV* wrt. the valence or conduction band, which means that at room temperature all extra electrons are in the conduction band and all holes have been filled by neighbouring atoms. Table 1 shows these energy levels introduced by some impurities for a few semiconductors.

In all cases the dopant concentration is small ($\sim 10^{16} - 10^{17} \text{ cm}^{-3}$), and depend on the type of semiconductor, the type of dopant and the temperature of the doping process, but

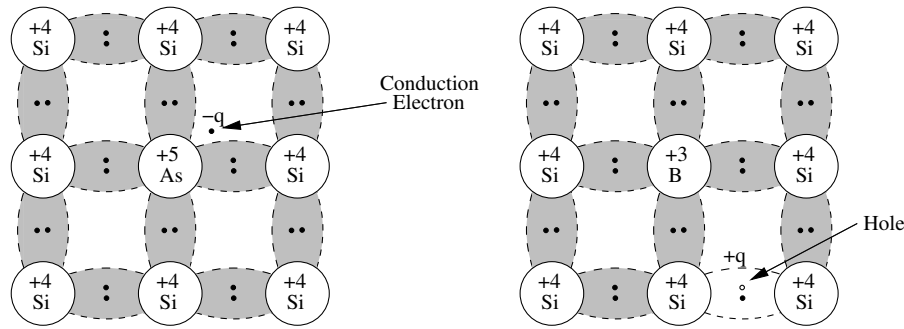


Figure 3: Bond representation of n-type (left) and p-type (right) semiconductors.

Semiconductor	Donor	$E_d[meV]$	Acceptor	$E_v[meV]$
GaAs	Si	5.8	Si	35
GaAs	Ge	6.0	Be	28
Si	As	54	B	45
Si	P	45	Ga	72
Ge	As	13	B	10
Ge	P	12	Ga	11

Table 1: Donor and acceptor energy levels for different semiconductors.

it is still much larger than the concentration of intrinsic charge carriers. As all dopants generate extra electrons or holes, one refers to two kinds of charge carriers depending on their abundance, the **majority** and **minority charge carriers**.

In the case of a *n*-type semiconductor, if N_d is the concentration of donor atoms, the electrons in the conduction band (n_c) will have two contributions

- electrons originating from thermal ionisation, that are equal to the number of holes in the valence band (p_v)
- N_d electrons coming from donor atoms

Thus $n_c = p_v + N_d$. As $N_d \gg p_v$ we have that $n_c \sim N_d$, therefore:

$$n_c = N_d \quad p_v = \frac{n_i^2}{N_d}$$

In this case the majority carriers are electrons and the minority carriers are holes. The charge carrier concentrations can then be expressed as:

$$n_c = N_d = N_c e^{-(E_c - E_F^{(n)})/kT}$$

where $E_F^{(n)}$ is the Fermi energy for a *n*-type semiconductor.

A similar reasoning can be applied for a p -type semiconductor, obtaining

$$p_v = N_a \quad n_c = \frac{n_i^2}{N_a}$$

In this case, the majority carriers are holes and the minority carriers electrons.

2 $p - n$ Diode Junction

One of the most important electronic structures is the $p - n$ junction, which is obtained by joining together extrinsic semiconductors of opposite doping. Such a structure shows diode characteristics, which means it will conduct current mainly in one direction. Figure 4-a shows the junction just after its formation; the p side has a large number of free holes and the n side an excess of free electrons. Due to their mutual repulsion, the free electrons of n side are diffused in all directions. Some of them can traverse the junction, creating an atom with a positive charge on the n side. When the electron enters the p side, it recombines with a hole (as holes are the majority carriers at the p side), generating a negative ion. This means that each electron that traverses the $p - n$ junction generates a couple of ions, one negative on the n side and one positive on the p side. After some time, around the junction almost all electrons and holes recombine, creating a negative charged zone on the p side and a positive charged zone on the n side. These two zones create a potential barrier that counteracts the diffusion of the electrons from the n side to the p side.

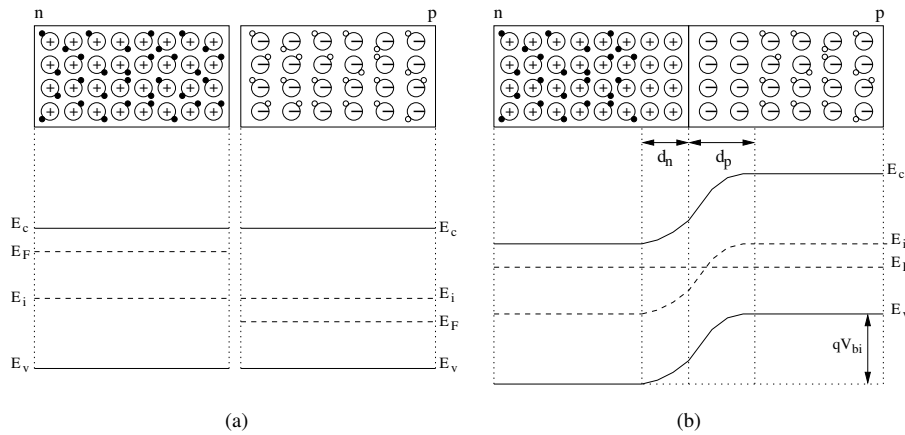


Figure 4: A $p - n$ junction in thermal equilibrium, with its parts separated (a) and brought together (b) forming a potential barrier.

This potential barrier, the so called built-in voltage (V_{bi}), is obtained using the band model and the requirement that the Fermi levels (E_F) have to line up in thermal equilibrium. It can be calculated from the intrinsic levels in the neutral p and n regions. Using the majority (in each case) carrier concentrations to be equal to the acceptor and donor concentrations respectively, we have:

$$N_d = n_i \exp \frac{E_F - E_c}{kT} \quad N_a = n_i \exp \frac{E_v - E_F}{kT}$$

$$N_a \cdot N_d = n_i^2 \exp \frac{E_v - E_c}{kT}$$

$$V_{bi} = \frac{1}{q}(E_v - E_c) = \frac{kT}{q} \ln \frac{N_a N_d}{n_i^2}$$

The value of this potential barrier is around 0.7 V in case of silicon and 0.3 V in case of germanium ($n_i = 1.45 \times 10^{10}$ for silicon and $n_i = 2.4 \times 10^{13}$ for germanium) around room temperature. It decreases with increasing temperature. The thermal equilibrium situation is shown in figure 4-b.

The width d_n , d_p of the depleted region as well as the maximum electric field E_{max} generated can be calculated requiring that the total charge in this depleted region corresponds to zero :

$$N_d d_n = N_a d_p$$

The electric field at the junction is then

$$E_{max} = \frac{1}{\epsilon \epsilon_0} q N_d d_n = \frac{1}{\epsilon \epsilon_0} q N_a d_p$$

and the voltage steps are:

$$V_n = \frac{E_{max} d_n}{2} = \frac{1}{\epsilon \epsilon_0} \frac{q N_a d_p d_n}{2}$$

$$V_p = \frac{E_{max} d_p}{2} = \frac{1}{\epsilon \epsilon_0} \frac{q N_d d_p d_n}{2}$$

This yields the voltage across the junction

$$V = V_n + V_p = \frac{q}{2\epsilon \epsilon_0} (N_a + N_d) d_n d_p = \frac{q}{2\epsilon \epsilon_0} \frac{(N_a + N_d) N_d}{N_a} d_n^2 = \frac{q}{2\epsilon \epsilon_0} \frac{(N_a + N_d) N_a}{N_d} d_p^2$$

As V is equal to the built-in voltage, we can easily obtain the depletion depths and the maximum electric field:

$$d_n = \sqrt{\frac{2\epsilon \epsilon_0}{q} \frac{N_a}{N_d(N_a + N_d)} V_{bi}}$$

$$d_p = \sqrt{\frac{2\epsilon \epsilon_0}{q} \frac{N_a}{N_d(N_a + N_d)} V_{bi}}$$

$$d = d_n + d_p = \sqrt{\frac{2\epsilon \epsilon_0 (N_a + N_d)}{q N_a N_d} V_{bi}}$$

$$E_{max} = \frac{1}{\epsilon \epsilon_0} q N_d d_n = \sqrt{\frac{2q}{\epsilon \epsilon_0} \frac{N_a N_d}{N_a + N_d} V_{bi}}$$

It is important to mention that a doped semiconductor has always an ohmic resistance called extrinsic resistance, that it is inversely proportional to the doping level.

3 Polarisation of a $p - n$ Junction

Diodes can be polarised in two ways (direct and reverse polarisation) that can be seen in figure 5. If we apply a voltage V and the potential of the p side increases with respect to the potential in the n side, this voltage is positive. Direct polarisation corresponds to the case $V > 0$, while reverse polarisation means $V < 0$.

Through biasing a diode in the correct way, it either blocks current flow in a circuit (under reverse bias) or lets current pass through (direct polarisation).

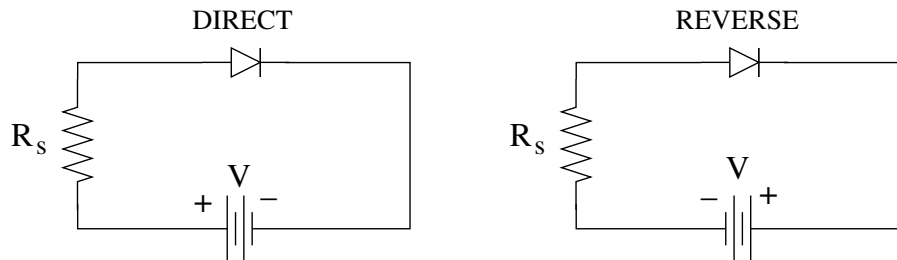


Figure 5: Diodes in direct and reverse polarisation.

In **direct polarisation**, the positive terminal of the power supply is connected to the p side and the negative to the n side. This pushes the holes and the electrons towards the junction. When they arrive at the junction, the external voltage delivers enough energy for them to cross the junction where they recombine. The electrons then become valence electrons. As valence electrons they move from hole to hole towards the positive terminal. Same for the holes moving in the opposite direction. In direct polarisation the potential barrier tries to oppose therefore the current flow following from the external voltage applied at the junction. This means that the external voltage has to be $> V_{bi}$ for the effect to show up.

In **reverse polarisation**, the positive terminal of the power supply is connected to the n side and the negative to the p side. The free electrons of the n side are therefore attracted towards the positive terminal of the supply and the holes on the p side to the negative one. The effect of this is that the depletion region gets enlarged, increasing the potential barrier until it equals the external potential. After reaching equilibrium, the current is very small, often negligible, and is only due to the thermal generation of electron-hole pairs². The width of the depleted region can be calculated for $V < 0$ as follows:

$$d = d_n + d_p = \sqrt{\frac{2\epsilon\epsilon_0(N_a + N_d)}{qN_aN_d}(V_{bi} - V)}$$

(Excellent reference: <http://jas2.eng.buffalo.edu>)

²Semiconductor detectors widely used in particle physics are simply reversed-biased diodes. Usually, they are operated fully depleted. They make use of the fact that their particle detection volume is free of charge carriers. When a charged particle traverses this volume, it creates electron-hole pairs through ionisation which are then collected at the 2 electrodes. The collected charge is proportional to the deposited energy of the particle to be detected.

References

- [1] Électronique, T.L. Floyd, Ed. Reynald Goulet
- [2] Principes d'Électronique, A.P. Malvino, Ed. Ediscience
- [3] The Art of Electronics, P.Horowitz, W.Hill, Ed. Cambridge Univ. Press
- [4] Techniques for Nuclear & Particle Physics Experiments, W.R. Leo, Ed. Springer-Verlag
- [5] Semiconductor Radiation Detectors, G. Lutz, Ed. Springer-Verlag
- [6] Solid State Physics, N.W. Ashcroft, N.D. Mermin.