

1. SEMICONDUCTORS

2. THE P-N JUNCTION

Introduction. A p-n junction consists of two semi-infinite semiconductors, which we imagine to fill the entire space. One of them has more donors than acceptors, and so it is negatively charged, (it is the n-part of the junction) and the other has more acceptors than donors, and so it is positively charged (and hence it is the p-member of the junction). These two semiconductors are attached to one another at $x = 0$, such that along the x -axis the combined system is negatively charged for $x > 0$, and is positively charged at $x < 0$, while along the other two directions y and z it is homogeneous (namely, its properties do not vary with position). A very crude model to describe the system will be to assume that

$$N_d(x) = \begin{cases} N_d, & x > 0 \\ 0, & x < 0 \end{cases}, \quad N_a(x) = \begin{cases} 0, & x > 0 \\ N_a, & x < 0 \end{cases}. \quad (2.1)$$

The non-uniformity in impurity concentrations induces non-uniformity in the density of the charge carriers (the electrons and the holes). It is plausible to expect that this non-uniformity is significant at and around $x = 0$ and decays far away as $|x| \rightarrow \infty$. Hence, there is formed a ‘layer’ around $x = 0$, which is called the ‘depletion layer’, where the electron and hole concentration depends on x . In order to determine the properties of the depletion layer, we can view the abrupt change in the impurity concentration, as described by Eq. (2.1), as causing some electrostatic potential, $\phi(x)$, which we need to determine in a self-consistent way, since the charge carriers which are moving around will tend to screen any electrostatic potential as best as they can.

Screening in a free electron gas. Suppose we have free electrons (that have only kinetic energy) moving against a positive background, so that the entire system is neutral. This describes a simple *metal*. Now let us introduce into this system a positively-charged particle, at a given position, and hold it there firmly. This charge will attract the electrons to it, and will thus create in its neighborhood a surplus of negative charge, which will reduce (i.e.,

‘screen’) its positive charge.

The relevant equations that govern this behavior are from electrostatics. Firstly, we have Poisson’s equation connecting the charge density of the positive ‘extra’ charge (which was introduced into our neutral system) and the potential it creates,

$$-\nabla^2\phi^{ext}(\mathbf{r}) = 4\pi\rho^{ext}(\mathbf{r}) . \quad (2.2)$$

Secondly, we have Poisson’s equation connecting the ‘true’ charge density, formed in the system after all the re-shuffling of the electrons, with the potential it creates

$$-\nabla^2\phi(\mathbf{r}) = 4\pi\rho(\mathbf{r}) . \quad (2.3)$$

Now, the true charge density $\rho(\mathbf{r})$, consists of the extra charge density inserted into the system, and the charge density it *induces* because it attracts electrons,

$$\rho(\mathbf{r}) = \rho^{ext}(\mathbf{r}) + \rho^{ind}(\mathbf{r}) . \quad (2.4)$$

To these (exact) electrostatic equations, we add a semi-classical approximation: we find a relation between the induced charge density and the (full) potential which the electrons ‘feel’. This approximation, known as the Thomas-Fermi approximation, is carried out as follows. At equilibrium, the chemical potential, μ , in the system fixes the electron density, n ,

$$n(\mu) = \sum_{\mathbf{k}} \frac{1}{e^{\beta(E_k - \mu)} + 1} , \quad (2.5)$$

where $\beta = 1/(k_B T)$ is the inverse of the temperature. When there is a *constant* electrostatic potential, ϕ , acting on the electrons, we may just add it to μ , i.e., $\mu \rightarrow \mu + e\phi$. When the electrostatic potential, $\phi(\mathbf{r})$, varies sufficiently smoothly, we nonetheless add it to the chemical potential. The result, according to Eq. (2.5), is that the electron density is changed, $n(\mu) \rightarrow n(\mu + e\phi(\mathbf{r}))$. The change in the electron density, times $-e$, is just the induced charge density. Thus we have

$$\rho^{ind}(\mathbf{r}) = -e[n(\mu + e\phi(\mathbf{r})) - n(\mu)] \simeq -e^2 \frac{\partial n}{\partial \mu} \phi(\mathbf{r}) . \quad (2.6)$$

Note that the derivative is to be found for $\phi = 0$.

* * * **exercise:** Calculate explicitly the quantity $e^2 \frac{\partial n}{\partial \mu}$ for the free electron gas in which $E_k = \hbar^2 k^2 / 2m$, estimate its numerical value, and in particular determine its dimensions.

Let us now write our equations in Fourier space. Equations (2.2) and (2.3) become

$$\phi^{ext}(\mathbf{q}) = \frac{4\pi}{q^2} \rho^{ext}(\mathbf{q}) , \quad \phi(\mathbf{q}) = \frac{4\pi}{q^2} \rho(\mathbf{q}) , \quad (2.7)$$

and Eqs. (2.4) and (2.6) are

$$\rho(\mathbf{q}) = \rho^{ext}(\mathbf{q}) + \rho^{ind}(\mathbf{q}) , \quad \rho^{ind}(\mathbf{q}) = \left[-e^2 \frac{\partial n}{\partial \mu} \right] \phi(\mathbf{q}) . \quad (2.8)$$

Simple algebraic manipulations give us the relation between the true potential in the system, and the extra charge or extra potential that have caused this potential (in the initially neutral system)

$$\phi(\mathbf{q}) = \frac{\phi^{ext}(\mathbf{q})}{1 + \frac{4\pi}{q^2} \left[e^2 \frac{\partial n}{\partial \mu} \right]} = \frac{4\pi \rho^{ext}(\mathbf{q})}{q^2 + 4\pi e^2 \frac{\partial n}{\partial \mu}} . \quad (2.9)$$

We deduce from this calculation the *dielectric function*, $\varepsilon(\mathbf{q})$, of the free electron gas in the Thomas-Fermi approximation,

$$\varepsilon(\mathbf{q}) = 1 + \frac{4\pi}{q^2} \left[e^2 \frac{\partial n}{\partial \mu} \right] . \quad (2.10)$$

For example, when the extra charge density is simply a (positive) point charge, $\rho^{ext}(\mathbf{r}) = e\delta(\mathbf{r})$ then $\rho^{ext}(\mathbf{q}) = e$, and the potential it creates is obviously $\phi^{ext}(\mathbf{r}) = e/r$, which in Fourier space becomes $\phi^{ext}(\mathbf{q}) = 4\pi e/q^2$. The true potential created in the system is ϕ , which according to Eq. (2.9) is given in this example by

$$\phi(\mathbf{q}) = \frac{4\pi e}{q^2 + \kappa^2} , \quad \kappa^2 \equiv 4\pi e^2 \frac{\partial n}{\partial \mu} . \quad (2.11)$$

Transforming back to real space,

$$\phi(\mathbf{r}) = \frac{e}{r} e^{-\kappa r} , \quad (2.12)$$

so that instead of the *long-range* Coulomb potential, that decays at long distances as $1/r$, the potential screened by the free electron gas decays *exponentially*. The exercise above will tell us how good this screening is, namely, how far (in space) extends the the potential (2.12).

***** exercise:** Work out explicitly the integral leading to the result (2.12).

Self-consistent determination of the electrostatic potential in a p-n junction.

Coming back to our p-n junction, we will calculate the electrostatic potential created by

the junction (or rather, by the depletion layer) employing the same type of approximation as the Thomas-Fermi approximation used above. Namely, we simply shift the chemical potential in Eqs. (1.17) and (1.18) by $e\phi(x)$, leading to

$$n_c(x) = N_c(T)e^{-\beta(\varepsilon_c - e\phi(x) - \mu)} , \quad p_v(x) = P_v(T)e^{-\beta(\mu + e\phi(x) - \varepsilon_v)} . \quad (2.13)$$

Confining ourselves to the completely *extrinsic* case, in which the charge carrier density is entirely determined by the impurity concentration, we write down the charge carrier densities far away from the junction

$$n_c(\infty) \equiv N_d = N_c(T)e^{-\beta(\varepsilon_c - e\phi(\infty) - \mu)} , \quad p_v(-\infty) \equiv N_a = P_v(T)e^{-\beta(\mu + e\phi(-\infty) - \varepsilon_v)} . \quad (2.14)$$

These two equations give

$$e\phi(\infty) + \mu = \varepsilon_c + k_B T \ln \frac{N_d}{N_c(T)} , \quad -e\phi(-\infty) - \mu = -\varepsilon_v + k_B T \ln \frac{N_a}{P_v(T)} , \quad (2.15)$$

so that we know the *total* electrostatic potential drop across the junction is

$$e\phi(\infty) - e\phi(-\infty) = E_{\text{gap}} + k_B T \ln \frac{N_d N_a}{N_c(T) P_v(T)} . \quad (2.16)$$

The *spatial variation* of the electrostatic potential is determined, as usual, by the Poisson equation

$$-\nabla^2 \phi = -\frac{d^2 \phi}{dx^2} = \frac{4\pi \rho(x)}{\epsilon} , \quad (2.17)$$

in which ϵ is the *static* dielectric constant of the semiconductor. (The static dielectric constant of the host material of the p-n junction takes into account all static screening effects which are there before the material turned into a p-n junction.) The charge density giving rise to the spatially-varying electrostatic potential is

$$\rho(x) = e \left[N_d(x) - N_a(x) + p_v(x) - n_c(x) \right] . \quad (2.18)$$

The meaning of the various factors here is simple. In the square brackets we have the number of completely ionized donors (positive charge density) + the number of completely ionized acceptors (negative charge density) + hole concentration (p_v) + electron concentration (negative charge, n_c).

In principle, we have now all that is needed to find the electrostatic potential, since we can just insert Eqs. (2.1) and (2.13) into Eq. (2.18), and then in Poisson's equation (2.17). However, the result will be an horrendous equation, so let us try to simplify it a bit. Combining Eqs. (2.13) with Eqs. (2.14) we find

$$n_c(x) = N_d e^{-\beta e(\phi(\infty) - \phi(x))} , \quad p_v(x) = N_a e^{-\beta e(\phi(x) - \phi(-\infty))} , \quad (2.19)$$

from which we see clearly that as $x \rightarrow \pm\infty$, and the respective exponentials approach unity, the electron and hole concentrations approach the concentration of the donors and the acceptors, respectively. We may thus say that at x larger than a certain distance, say d_n , $N_d(x) - n_c(x) \rightarrow 0$, while for x smaller than a certain distance, say $-d_p$, $N_a(x) - p_v(x) \rightarrow 0$. Therefore, a plausible approximate expression for the charge density $\rho(x)$ will be

$$\rho(x) = \left\{ \begin{array}{ll} 0 , & x > d_n , \\ eN_d , & d_n > x > 0 , \\ -eN_a , & 0 > x > -d_p , \\ 0 , & -d_p > x . \end{array} \right\} . \quad (2.20)$$

In other words, outside the depletion layer, which extends from $-d_p$ up to d_n , the electron density cancels the donor density (on the right side) and the hole density cancels the acceptor density (on the left side).

The solution of Poisson's equation with the approximate charge density, Eq. (2.20), is now simple. We have already found the values of ϕ at $\pm\infty$, [see Eq. (2.16)] and we know (from electrostatics) that $\phi(x)$ and its first derivative (which is minus the electric field) should be continuous. Therefore

$$\phi(x) = \left\{ \begin{array}{ll} \phi(\infty) , & x > d_n , \\ \phi(\infty) - \left(\frac{2\pi e N_d}{\epsilon} \right) (x - d_n)^2 , & d_n > x > 0 , \\ \phi(-\infty) + \left(\frac{2\pi e N_a}{\epsilon} \right) (x + d_p)^2 , & 0 > x > -d_p , \\ \phi(-\infty) , & -d_p > x . \end{array} \right\} . \quad (2.21)$$

It remains to apply the boundary conditions at $x = 0$. First, we look at the electric field at the origin and require that it will be continuous. This gives us

$$N_d d_n = N_a d_p . \quad (2.22)$$

Second, the continuity of the electrostatic potential itself gives

$$\phi(\infty) - \phi(-\infty) \equiv \Delta\phi = \left(\frac{2\pi e}{\epsilon}\right)(N_a d_p^2 + N_d d_n^2) . \quad (2.23)$$

We therefore find for the boundaries of the depletion layer the result

$$d_n = \sqrt{\frac{(N_a/N_d)\epsilon\Delta\phi}{(N_d + N_a)2\pi e}} , \quad d_p = \sqrt{\frac{(N_d/N_a)\epsilon\Delta\phi}{(N_d + N_a)2\pi e}} . \quad (2.24)$$

*** **exercise:** Plot the electrostatic potential across the depletion layer of a p-n junction, plot the charge carrier density, and estimate the width of the depletion layer as function of the temperature.

The biased p-n junction—simple considerations. So far, we have considered the p-n at equilibrium. Let us now examine what happens when the junction is biased by a potential drop, V . In that case the full change of the electrostatic potential is modified as compared to its equilibrium value. We may try to estimate the effect of the bias by using the approximation

$$\Delta\phi \equiv \phi(\infty) - \phi(-\infty) = (\Delta\phi)_0 - V , \quad (2.25)$$

where $(\Delta\phi)_0$ is the equilibrium value [given by Eq. (2.16)].

One result of the bias voltage is that [see Eq. (2.24)] the depletion layer thickness changes: the layer shrinks when $V > 0$ (i.e., V raises the the potential of the p-side), and expands when $V < 0$. The other result is that an electrical current will flow across the junction. Denoting by J the density current and by j the electrical current, we have

$$j_e = -eJ_e , \text{ and } j_h = eJ_h . \quad (2.26)$$

Let us consider first the flow of the holes. Firstly, there are the holes which are generated *in the n-side* of the junction. These are called “minority carriers”. There are not many of them, but once they are in the depletion layer, they will be immediately swallowed into the p-side by the strong electric field that exists in that layer. This current is called ‘the hole generation current’ and is not too sensitive to the bias V . Secondly, there is the current of holes going from the p-side of the junction to the n-side, which is called ‘the hole recombination current’ (because eventually these holes will combine with the electrons

in the n-side and will disappear). These holes have to overcome the potential barrier, and therefore we expect that

$$J_h^{\text{rec}} \propto e^{-e\beta((\Delta\phi)_0 - V)} . \quad (2.27)$$

Since at equilibrium, i.e., when $V = 0$, we expect the two currents to balance one another, $J_h^{\text{rec}} \Big|_{V=0} = J_h^{\text{gen}}$, we find that

$$J_h^{\text{rec}} = J_h^{\text{gen}} e^{\beta e V} , \quad (2.28)$$

so that the total current of holes flowing from the p-side of the junction to the n-side is given by

$$J_h = J_h^{\text{rec}} - J_h^{\text{gen}} = J_h^{\text{gen}} (e^{\beta e V} - 1) . \quad (2.29)$$

Carrying out the analogous calculation for the electron current, and combining the hole current with the electron current, the total electric current across the junction is

$$j = e(J_h^{\text{gen}} + J_e^{\text{gen}})(e^{\beta e V} - 1) . \quad (2.30)$$

In particular, when $V < 0$, the current saturates at a (small) negative value, $-e(J_h^{\text{gen}} + J_e^{\text{gen}})$, independent of the voltage. In order to obtain an estimate for that current, we need to carry out a more detailed calculation.

***** exercise:** Plot the current Eq. (2.30) as function of the bias voltage, and discuss the result.

The biased p-n junction—detailed calculation. Our more elaborate treatment of the p-n junction begins with the observation that in the presence of both an electric field, $E = -d\phi/dx$, and density gradients, the current densities obey

$$J_e = -\mu_n n_c E - D_n \frac{dn_c}{dx} , \quad J_h = \mu_p p_v E - D_p \frac{dp_v}{dx} . \quad (2.31)$$

Here, μ_n and μ_p are the electron and hole mobilities, respectively, and D_n and D_p are the diffusion coefficients. The latter are related to the mobilities by the Einstein relations

$$\mu_n = \frac{eD_n}{k_B T} , \quad \mu_p = \frac{eD_p}{k_B T} . \quad (2.32)$$

*** **exercise:** Derive the Einstein relation for a *degenerate* electron gas, and for the *non degenerate* one [which is just the one appearing in Eqs. (2.32)].

*** **exercise:** Prove that the Einstein relations (2.32) ensure that no current is flowing at thermal equilibrium.

We can re-write the mobilities in terms of the the mean free times between collisions and the effective masses. For example, if there are only electrons of uniform density in the system, then an electric field will give rise to electrical current, $j_e = -eJ_e = \sigma E$, where σ is the electrical conductivity. According to Drude theory, $\sigma = ne^2\tau^{\text{col}}/m$, and therefore in our case

$$\mu_n = \frac{e\tau_n^{\text{col}}}{m_n}, \quad \mu_p = \frac{e\tau_p^{\text{col}}}{m_p}. \quad (2.33)$$

We have distinguished here between the mean free time in between collisions of the electrons (τ_e^{col}) and of the holes (τ_p^{col}). We have also assigned different masses, m_n and m_p , to the charge carriers, but assumed the same mass for all three principal directions for each of the species.

At equilibrium, where J_e and J_h vanish, one can determine the carrier densities by using Eqs. (2.31) and (2.32). The solutions are just Eqs. (2.13).

*** **exercise:** Prove the statement above.

When the junction is biased, the system is no more at equilibrium. In this case we add two *phenomenological* equations,

$$\frac{\partial n_c}{\partial t} = \left(\frac{\partial n_c}{\partial t}\right)_{g-r} - \frac{\partial J_e}{\partial x}, \quad \frac{\partial p_v}{\partial t} = \left(\frac{\partial p_v}{\partial t}\right)_{g-r} - \frac{\partial J_h}{\partial x}, \quad (2.34)$$

which aim to reflect the fact that charge is conserved. In fact, charge is *not* conserved in our case, since electrons and holes are created by thermal fluctuations and disappear by recombining (namely, electrons go back to the valence band). These processes are described by the first terms on the right hand side of each of Eqs. (2.34). [In the absence of these two terms, each of Eqs. (2.34) is just the continuity equation for the electrons and the holes, respectively.] In order to determine the forms of the terms $(\partial n_c/\partial t)_{g-r}$ and $(\partial p_v/\partial t)_{g-r}$, we postulate that they act as to *restore* thermal equilibrium. Hence

$$\left(\frac{\partial n_c}{\partial t}\right)_{g-r} = -\frac{n_c - n_c^0}{\tau_n}, \quad \left(\frac{\partial p_v}{\partial t}\right)_{g-r} = -\frac{p_v - p_v^0}{\tau_p}, \quad (2.35)$$

where n_c^0 and p_v^0 are the equilibrium values of the densities.

Since our p-n junction is in a *steady state* (namely, it does not evolve in time), we set the ‘true’ time-derivatives in Eqs. (2.34) to be zero, so that

$$\frac{dJ_e}{dx} + \frac{n_c - n_c^0}{\tau_n} = 0, \quad \frac{dJ_h}{dx} + \frac{p_v - p_v^0}{\tau_p} = 0. \quad (2.36)$$

These equations have to be solved together with Eqs. (2.31). This is not a simple task in general; however, in the regions where the electric field is almost constant, we may ignore the terms including it, and combine Eqs. (2.31) and (2.36) to find

$$D_n \frac{d^2 n_c}{dx^2} = \frac{n_c - n_c^0}{\tau_n}, \quad D_p \frac{d^2 p_v}{dx^2} = \frac{p_v - p_v^0}{\tau_p}. \quad (2.37)$$

Inspection of these equations shows that each of them defines a length, the ‘diffusion length’,

$$L_n^2 = D_n \tau_n, \quad L_p^2 = D_p \tau_p. \quad (2.38)$$

For example, if the density of holes at some $x_0 > 0$ differs from its (equilibrium) value at $x \rightarrow +\infty$, then the second of Eqs. (2.38) tells us that

$$p_v(x) = p_v(\infty) + e^{-(x-x_0)/L_p} (p_v(x_0) - p_v(\infty)). \quad (2.39)$$

In other words, at $x > L_p$, the hole density (on the n-side of the junction) takes almost its equilibrium value (which is very small, see below). Since holes are created by thermal fluctuations at a rate p_v^0/τ_p , and are not recombined along a distance L_p , the flow of thermally generated holes per unit area into the depletion layer is about $L_p p_v^0/\tau_p \simeq (n_i^2/N_d)L_p/\tau_p$. Hence the saturation current is

$$J_h^{\text{gen}} + J_h^{\text{gen}} = \left(\frac{n_i^2}{N_d}\right) \frac{L_p}{\tau_p} + \left(\frac{n_i^2}{N_a}\right) \frac{L_n}{\tau_n}. \quad (2.40)$$

*** **exercise:** Discuss the temperature dependence of the saturation current, Eq. (2.40).

Summary.

- The diode is the simplest type of semiconductor device. In generally, it is a two-terminal electronic device that permits current flow predominantly in only one direction. The current they pass depends upon the voltage between the leads. They do not obey Ohm’s law! Most diodes are semiconductor devices; diode electron tubes, now uncommon, are also available.

- There are two main types of semiconductor materials: intrinsic - where the semiconducting properties of the material occur naturally i.e. they are intrinsic to the material's nature; extrinsic - the semiconducting properties of the material are manufactured, by us, to make the material behave in the manner which we require. Nearly all the semiconductors used in modern electronics are extrinsic. This means that they have been created by altering the electronic properties of the material. The two most common methods of modifying the electronic properties are: doping - the addition of 'foreign' atoms to the material; junction effects - the things that happen when we join differing materials together.
- The pn junction diode. Free electrons on the n-side and free holes on the p-side can initially wander across the junction. When a free electron meets a free hole it can 'drop into it'. So far as charge movements are concerned this means the hole and electron cancel each other and vanish. As a result, the free electrons near the junction tend to eat each other, producing a region depleted of any moving charges. This creates what is called the depletion zone. Now, any free charge which wanders into the depletion zone finds itself in a region with no other free charges. Locally it sees a lot of positive charges (the donor atoms) on the n-type side and a lot of negative charges (the acceptor atoms) on the p-type side. These exert a force on the free charge, driving it back to its 'own side' of the junction away from the depletion zone. The acceptor and donor atoms are 'nailed down' in the solid and cannot move around. However, the negative charge of the acceptor's extra electron and the positive charge of the donor's extra proton (exposed by its missing electron) tend to keep the depletion zone swept clean of free charges once the zone has formed. A free charge now requires some extra energy to overcome the forces from the donor/acceptor atoms to be able to cross the zone. The junction therefore acts like a barrier, blocking any charge flow (current) across the barrier. Usually, we represent this barrier by 'bending' the conduction and valence bands as they cross the depletion zone. Now we can imagine the electrons having to 'get uphill' to move from the n-type side to the p-type side.

The holes behave a bit like balloons bobbing up against a ceiling. In the bending-band diagram one requires energy to 'pull them down' before they can move from the p-type side to the n-type side. The energy required by the free holes and electrons

can be supplied by a suitable voltage applied between the two ends of the pn-junction diode. Notice that this voltage must be supplied the correct way around, this pushes the charges over the barrier. However, applying the voltage the 'wrong' way around makes things worse by pulling what free charges there are away from the junction! This is why diodes conduct in one direction but not the other.

- On the basis of the explanation given above we might expect no current to flow when the diode is reverse biased. In reality, the energies of the electrons and holes in the diode are not all the same. A small number will have enough energy to overcome the barrier. As a result, there will be a tiny current through the diode when we apply reverse bias. The polarity which produces a significant current is called forward bias. When we raise the forward bias voltage we reduce the diode's energy barrier. This essentially reduces the diode's resistance.