ECE 255

29 August 2017

In this lecture, one would like to know what happens when one puts a p-type material next to an n-type material. This is known as a pn-junction. The p-type material could have been doped with phosphorous, and the n-type material could have been doped with boron. At the junction of these two materials, interesting physics happens.

1 pn junction

The first thing that happens when two junctions come into contact is the diffusion phenomenon. The preponderance of p-type carriers in the p-type material will cause their diffusion across the junction into the n-type material. The same for the n-type carries in the n-type material.

Therefore, an internal electric field is established that prevents further migration of the carriers across the junction. At this juncture, $J^{\text{drift}} = -J^{\text{diff}}$ for both kinds of carriers. As shown in the Appendix, the hole density and the potential ϕ are related to each other according to Boltzmann's law, namely

$$p(x) \sim e^{-q\phi(x)/(k_B T)} \tag{1.1}$$

Hence, the voltage drop across a pn junction at equilibrium is given as

$$V_0 = \phi_j = \frac{k_B T}{q} \ln \frac{p(x = -x_p)}{p(x = +x_n)} \tag{1.2}$$

where $p(x = -x_p)$ is the hole density in the p region (see Figure 2) or the left edge of the transition region (also called depletion region) of the junction, while similarly $p(x = +x_n)$ in the n region. In the p region, holes are majority carriers and its density is given by

$$p(-x_p) = N_A \tag{1.3}$$

while in the n region, holes are minority carrier, and the hole density is given by

$$p(+x_n) = n_i^2/N_A \tag{1.4}$$

Printed on August 31, 2017 at 10 : 45: W.C. Chew and Z.H. Chen.

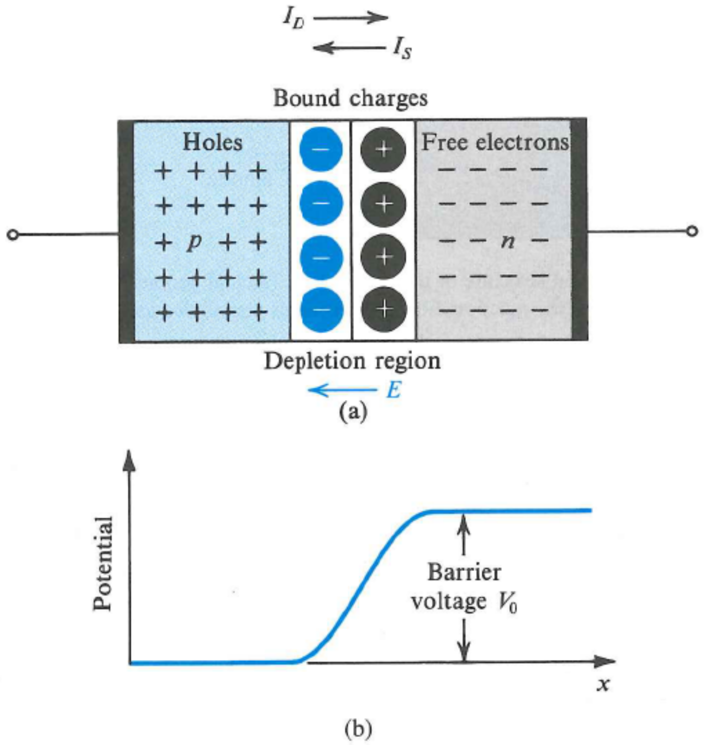

Figure 1: A pn junction of two different materials. The thicknesses of the depletion region need not be the same (Courtesy of Sedra and Smith).

Substituting the above back in (1.2), the following equation is arrived at yielding a formula for the voltage or potential difference between the two materials

$$V_0 = \phi_j = \frac{k_B T}{q} \ln \frac{N_D N_A}{n_i^2} \tag{1.5}$$

It is often defined that $V_T = k_B T/q$ which is 25 mV at room temperature. Moreover, because the diffusion current and the drift current cancel each other, it can be established that there is a relationship, called the Einstein relationship, between the diffusion coefficient D and mobility μ. Namely that (see Appendix)

$$\frac{D_n}{\mu_n} = \frac{D_p}{\mu_p} = \frac{k_B T}{q} = V_T \tag{1.6}$$

The above formulas also convey the notion that the higher the temperature T, the higher the diffusion coefficients. The higher the temperature, the more kinetic energy the electrons will acquire, and the faster they will diffuse.

If the length of the depletion region in the p region is x_p then the total charge contained in it is given by $qN_A x_p$. Similarly, if the length of the depletion region

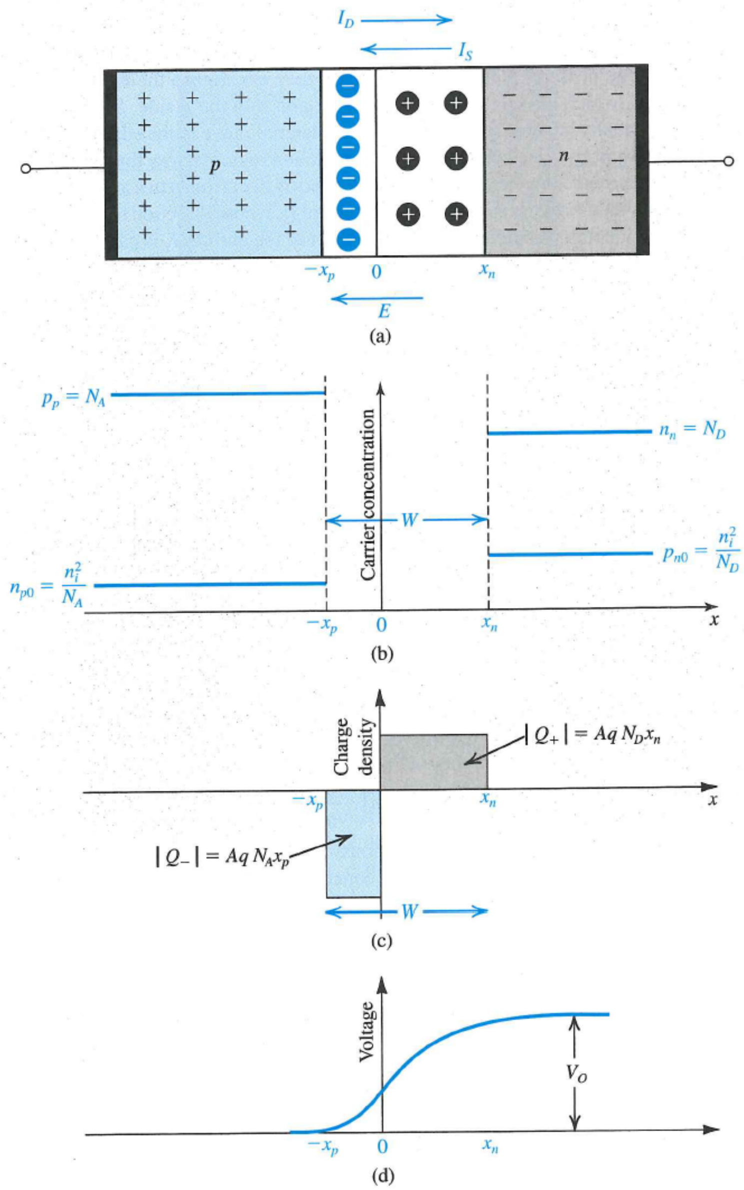

Figure 2: A pn junction of two different materials in the open circuit case. The thicknesses of the depletion region are not the same with x_n denoting the thickness of the n region, and x_p denoting the thickness of the p region (Courtesy of Sedra and Smith).

in n region is x_n, then the total charge contained in it is given by qN_Dx_n. By charge neutrality, then

$$qN_Dx_n = qN_Ax_p \tag{1.7}$$

$$\frac{x_n}{x_p} = \frac{N_A}{N_D} \tag{1.8}$$

As shown in the Appendix, the total width of the depletion region can be shown to be

$$W_D = x_p + x_n = \sqrt{\frac{2\epsilon_s}{q}\left(\frac{1}{N_A} + \frac{1}{N_D}\right) V_0} \tag{1.9}$$

Assuming that the p region is boron doped, and the n region is phosphorous doped, then as the carriers diffuse across the junction, the phosphorous region becomes positively charged due to the migration of electrons from the ions. The boron region becomes negatively charged due to the migration of electrons into their vacant bonds or holes. Hence, a depletion layer is formed that is not charge-neutral. This non-neutral region produces an electric field, producing a drift current that flows in opposition to the diffusion current. Or the electric field prevents the further diffusion of the carriers across the junction.

To see some an example, we consider a silicon pn-junction with $N_A = 10^{17}/\text{cm}^3$, and

$$N_D = 10^{20}/\text{cm}^3$$

Using the above formula for V_0, one obtains

$$V_0 = \frac{k_BT}{q} \ln \frac{N_DN_A}{n_i^2} \approx 0.979\,\text{V} \tag{1.10}$$

One can further calculate the width of the depletion region to be

$$W_D = x_p + x_n = \sqrt{\frac{2\epsilon_s}{q}\left(\frac{1}{N_A} + \frac{1}{N_D}\right) \phi_j} \approx 113\,\text{nm} \tag{1.11}$$

The ϵ_s of silicon is assumed to be 11.7 ϵ_0.

When the diode is in reverse bias, or $V_D < 0$ V, then the voltage drop ϕ between the two regions increases, because $\phi = V_0 + V_R$ where $V_R = -V_D$. The depletion region becomes wider and the current flow becomes smaller.

In the following, V_D implies the diode biasing voltage. When the diode is in forward bias, or $V_D > 0$ V, then the voltage drop ϕ between the two regions decreases, because $\phi = V_0 + V_R$. The depletion region becomes smaller and the current flow becomes larger across the pn junction.

2 Biasing of the *pn* Junction

First, it is worth describing the physical mechanism qualitatively of a diode under different biasing conditions, as shown in Figure 3. In the zero bias case, the diffusion currents are equal and opposite to the drift currents giving rise to zero current flow across the *pn* junction.

When the voltage is forward bias, it generates an electric field that cancels the original electric field in the transition region, reducing the voltage drop across the *pn* junction. In this case, the diffusion currents becomes larger than the drift currents, for both holes and electrons, giving rise to an increase current flow across the *pn* junction.

When the voltage is reverse bias, the electric field in the transition region is increased giving rise to greatly reduced diffusion currents for both holes and electrons. Hence, what remains is mainly the drift currents across the *pn* junction.

It is noteworthy that the drift currents remain the same for different biasing conditions. The reason is that there are no majority carriers in the transition region, and hence, even if an electric field exists there, there is no majority current flow. However, minority carriers, both from the p side (which are electrons in this case), and from the n side (which are holes in this case) can wander into the transition region. The moment they wonder into the transition region, because of their signs, they are swept across the transition region, contributing to a minuscule total drift current. These minority carriers are low in concentration after all. These minority carriers are generated by EHP generation due to thermal agitation, and they often recombine before they have a chance to drift to the transition region.

This drift current due to minority carriers is independent of the electric field: the stronger the electric field, the higher the velocity the carriers will be swept across the transition region, but the wider is the transition region they have to travel through. The width of the transition region is proportional the applied electric field or the potential across the *pn* junction. Therefore, the drift currents remain small, and is controlled by the EHP generation rate. These currents are also called generation currents.

2.1 Carrier Injection

Looking at the hole concentration, its equilibrium concentration on the p side of the transition region, p_{p0}, is related to its concentration no the n side of the same region, p_{n0}. Namely,

$$\frac{p_{p0}}{p_{n_0}} = e^{V_0/V_T} \tag{2.1}$$

according to Boltzmann's law, where V_0 is the voltage drop across the junction. When the junction is biased, assuming quasi-equilibrium and that Boltzmann's law still applies, then the ratio of the hole density is given by

$$\frac{p(-x_p)}{p(x_n)} = e^{(V_0-V)/V_T} \tag{2.2}$$

where V is the forward bias voltage. Then it is quite easy to show that

$$\frac{p(x_n)}{p_{n0}} = e^{V/V_T}, \qquad \text{assuming } p(-x_p) = p_{p0} \tag{2.3}$$

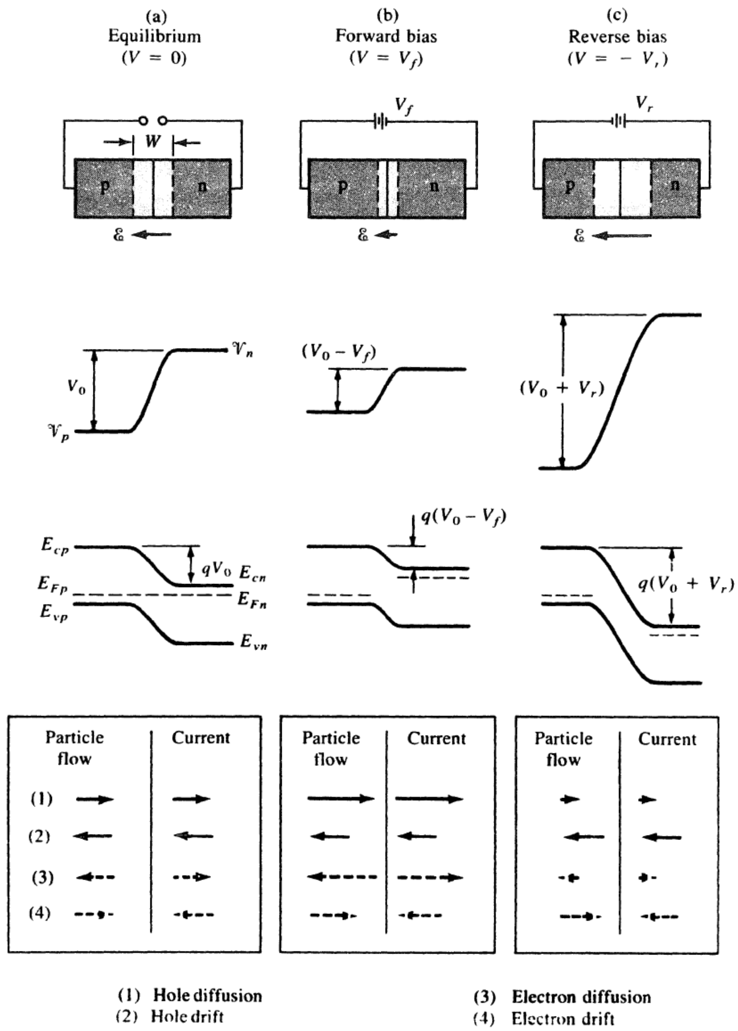

Figure 3: A *pn* junction under different bias conditions, showing transition region width, electric field, electrostatic potential, and energy band diagram, for (a) equilibrium, (b) forward bias, and (c) reverse bias (Courtesy of Streetman).

Then the excess carrier in the n region is given by

$$\Delta p_n = p(x_n) - p_{n0} = p_{n0}(e^{V/V_T} - 1) \tag{2.4}$$

Similarly, for the electrons,

$$\Delta n_p = n(-x_p) - n_{p0} = n_{p0}(e^{V/V_T} - 1) \tag{2.5}$$

As the deluge of minority carriers are injected into their respective regions, they will diffuse from the edges of the transition zones. Moreover, they will recombine with the majority carriers of their new host media, giving rise to exponential decay as they diffuse. Consequently,

$$p_n(x) = p_{n0} + p_{n0}\left(e^{V/V_T} - 1\right)e^{-(x-x_n)/L_p} \tag{2.6}$$

where L_p is the diffusion length very much dependent on the recombination rate of the EHP. And a gradient develops in the hole concentration giving rise to diffusion current

$$J_p(x) = -qD_p\frac{dp_n(x)}{dx} \tag{2.7}$$

Therefore,

$$J_p(x) = q\left(\frac{D_p}{L_p}\right)p_{n0}\left(e^{V/V_T} - 1\right)e^{-(x-x_n)/L_p} \tag{2.8}$$

The maximum current is at the edge of the transition region, $x = x_n$, giving the diffusion current to be

$$J_p(x_n) = q\left(\frac{D_p}{L_p}\right)p_{n0}\left(e^{V/V_T} - 1\right) \tag{2.9}$$

As this hole current diffuses into the host region on the right, it recombination with electrons has to be augmented by electron currents coming from the right. Hence, the total current remains the same throughout the host region.

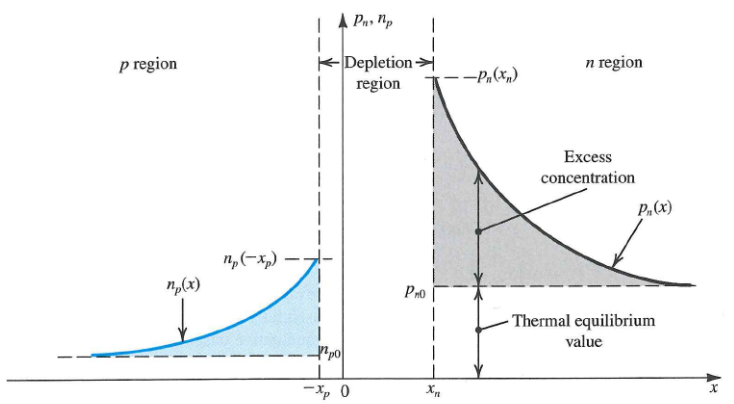

Figure 4: A *pn* junction with minority carrier distributions in the forward bias mode. The *p* region is more heavily doped than the *n* region (Courtesy of Sedra and Smith).

Similarly, for the electron injection into the *p* region, the resultant diffusion current is

$$J_n(-x_p) = q\left(\frac{D_n}{L_n}\right)n_{p0}\left(e^{V/V_T} - 1\right) \tag{2.10}$$

The total current is given by $I = AJ_{\text{total}} = A(J_p + J_n)$. With the use of the fact that $p_{n0} = n_i^2/N_D$ and $n_{p0} = n_i^2/N_A$, one gets

$$I = Aqn_i^2 \left(\frac{D_p}{L_p N_D} + \frac{D_n}{L_n N_A} \right) \left(e^{V/V_T} - 1 \right) = I_S \left(e^{V/V_T} - 1 \right) \tag{2.11}$$

where

$$I_S = Aqn_i^2 \left(\frac{D_p}{L_p N_D} + \frac{D_n}{L_n N_A} \right) \tag{2.12}$$

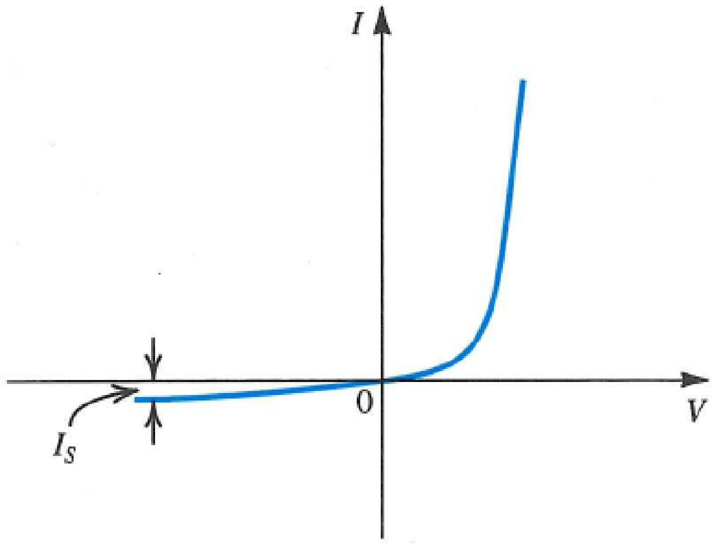

Figure 5: The $I - V$ characteristic of a *pn* junction diode. The current increases to exponentially large value under forward bias. Under reverse bias, the current is minuscule because it is due to minority carriers (Courtesy of Sedra and Smith).

Some Considerations

In the following, the forward bias diode voltage and current will be denoted as V_D and I_D, respectively. When a diode is being forward biased, $V_D > 0$ V. In this case, the E_{ext}, the external field, is opposite in sign to the internally built field, or E_{built}. Then I_D, or the diode current will be larger than zero.

When the diode is reverse biased, $V_D < 0$ V, and E_{ext} is in the same direction as the internally built field E_{built}. The voltage drop across the depletion layer will increase, and the width of the depletion layer, W_D will increase. Then the diode current $I_D \approx 0$ A.

$$W'_D = \sqrt{\frac{2\epsilon_s}{q} \left(\frac{1}{N_A} + \frac{1}{N_D} \right) (V_0 + V_R)} \tag{2.13}$$

When $V_D = 0$ V, then $I_D = 0$ A. The relation between I_D and V_D can be described by the ideal diode equation, which states that

$$I_D = I_S \left(e^{\frac{V_D}{V_T}} - 1 \right) \tag{2.14}$$

where V_T is about 25 mV at room temperature.

The reverse saturation current (also called generation current)

$$I_S = Aqn_i^2 \left(\frac{D_p}{L_p N_D} + \frac{D_n}{L_n N_A} \right) \tag{2.15}$$

Usually,

$$10^{-18}\text{A} < I_s < 10^{-12}\text{A}$$

Example 1

$$I_S = 0.1\,\text{fA} \tag{2.16}$$

$$I_D = 300\,\mu\text{A}, \quad \text{What is } V_D? \tag{2.17}$$

Answer:

$$I_S \left(e^{\frac{V_D}{V_T}} - 1 \right) \to V_D = V_T \ln \left(1 + \frac{I_D}{I_s} \right) \tag{2.18}$$

$$V_D \approx 0.718\text{V} \tag{2.19}$$

Example 2

$$I_D = 1\,\text{mA}, \quad \text{then using } V_D = V_T \ln \left(1 + \frac{I_D}{I_S} \right) \approx 0.748\,\text{V} \tag{2.20}$$

Usually, the turn-on voltage of a diode is $V_{\text{on}} \approx$ 0.6-0.7 V.

Example 3

$$I_S = 10\,\text{fA} \tag{2.21}$$

$$I_D = 300\,\mu\text{A}, \quad \text{What is } V_D? \tag{2.22}$$

Answer:

$$I_S \left(e^{\frac{V_D}{V_T}} - 1 \right) \to V_D = V_T \ln \left(1 + \frac{I_D}{I_S} \right) \tag{2.23}$$

$$V_D \approx 0.603\text{V} \tag{2.24}$$

For the I-V relation of a diode, given by

$$I_D = I_S \left(e^{\frac{V_D}{V_T}} - 1 \right) \tag{2.25}$$

it will be consider forward bias when $V_D \geq 4V_T$. In this case,

$$I_D \approx I_S e^{\frac{V_D}{V_T}} \tag{2.26}$$

It is reverse bias when $V_D \leq 4V_T$.

$$I_D \approx -I_S \tag{2.27}$$

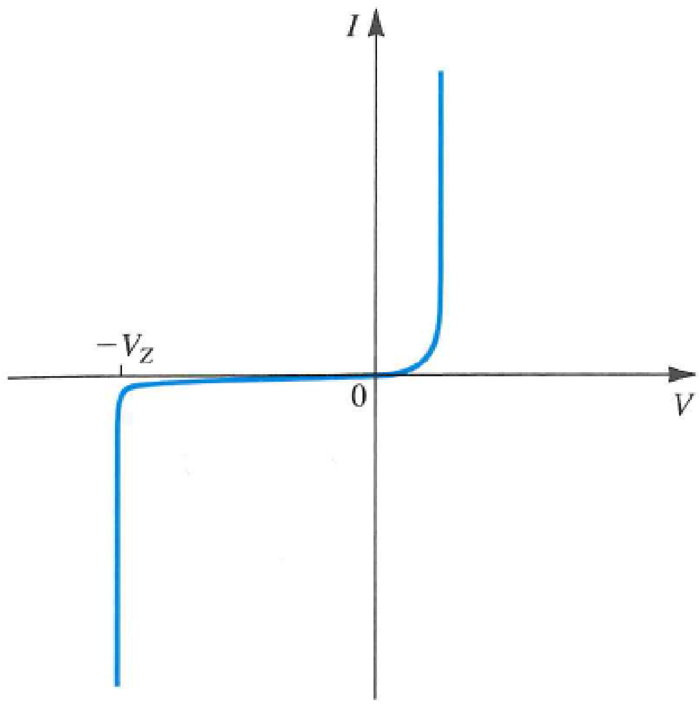

Figure 6: The $I - V$ characteristic of a *pn* junction diode under reverse bias breakdown (Courtesy of Sedra and Smith).

3 Reverse Breakdown

When $V_D \to -\infty$, the diode breaks down, and one can then pass a large current through the diode. When the reverse bias voltage is large, the width of the depletion region increases, and the electric field inside the depletion region also increases.

In an **avalanche breakdown**, usually larger than 7 V, the minority carriers in the transition region attain such a high velocity that they break the covalent bond of the silicon, giving rise to energetic minority carriers. This can be a cascading effect, giving rise to a huge reverse bias current.

When the reverse bias voltage is greater than 5 V, a **Zener breakdown** or **Zener effect** can occur. This breakdown is due to the tunneling of the electrons from the conduction band of the n region to the valence band of the p region. The breakdown voltage can be engineered precisely, and such diodes can be used for engineering designs.

Appendix A Derivation of the Einstein Relationship

The Einstein relationship is obtain by relatinng the equilibrium state of a pn junction to a state in thermal equilibrium, which should obey Boltzmann's law. When a junction is in equilibrium, the drift current cancels the diffusion current. In other words, looking at the hole current alone,

$$J_p(x) = q\left[\mu_p p(x)E(x) - D_p \frac{dp(x)}{dx}\right] = 0 \tag{A.1}$$

In the above, one can define $E(x) = -\frac{d\phi(x)}{dx}$ to obtain the equation

$$-\frac{\mu_p}{D_p}\frac{d\phi(x)}{dx} = \frac{1}{p(x)}\frac{dp(x)}{dx} \tag{A.2}$$

The right-hand side can be rewritten as

$$-\frac{\mu_p}{D_p}\frac{d\phi(x)}{dx} = \frac{d\ln p(x)}{dx} \tag{A.3}$$

After defining

$$\frac{q}{k_B T} = \frac{\mu_p}{D_p} \tag{A.4}$$

the above equation can be integrated to yield

$$p(x) = Ce^{-q\phi(x)/(k_B T)} \tag{A.5}$$

where C is an arbitrary constant independent of x. The above is just Boltzmann's law, and the derivation can be repeated for electron carriers. Moreover, (A.4) can be used to derive (1.6).

The above derivation indicates that the diffusion process is in fact a "thermalized" process. The fact that carriers diffuse is because they acquire kinetic energy from the thermal environment. The heat bath supplies energy to the material giving rise to lattice vibrations. These lattice vibrations in turn transfer kinetic energy to the electrons causing them to diffuse. The diffusion of the charged carriers then gives rise to uneven potential in the environment, yielding internal electric field that stops the diffusion process. The equilibrium of the diffusion current and the drift current is due to "thermal equilibrium".

Appendix B Width of the Depletion Region

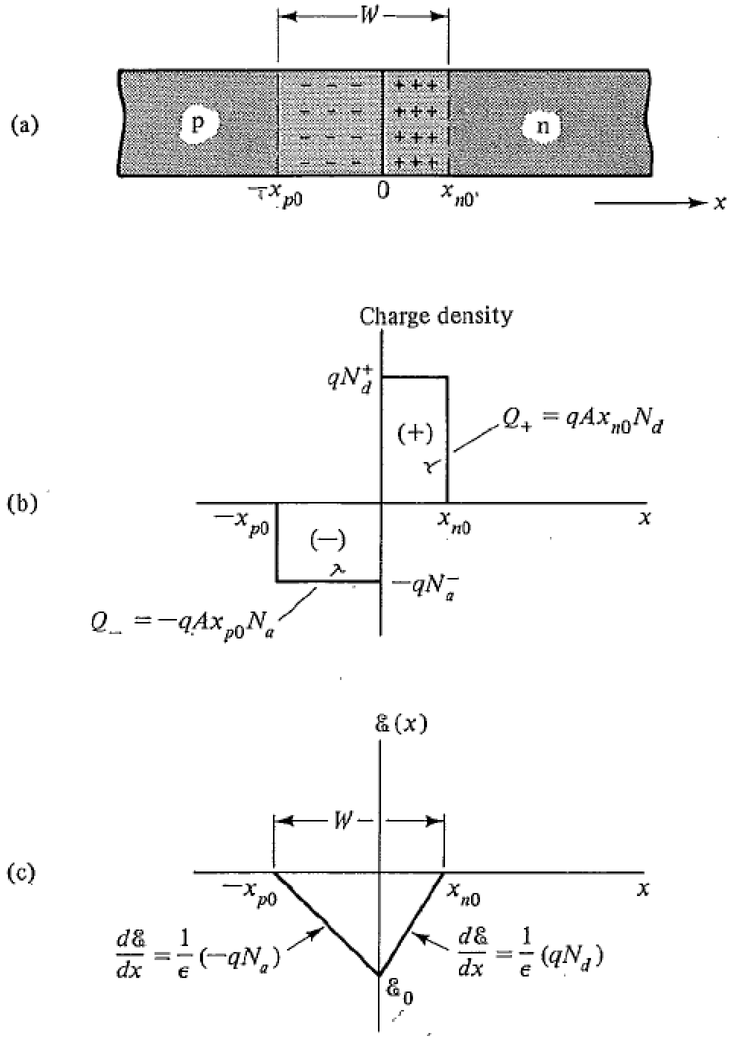

Figure 7: The depletion region and the approximate charge density, and electric field distribution in this region. (a) The transition region beginning at x_{po} and ending at x_{n0}, where $W = x_{p0} + x_{n0}$. (b) The approximate, simplified charge density within the transition region. (c) The approximate electric field distribution in the same region (Courtesy of Streetman).

Looking at Figure 7, it is quite clear that the peak field $\mathscr{E}_0$ is like the field between a parallel plate capacitor. This field is proportional to the surface charge density on the plate divided by ϵ, the permittivity of the silicon. The surface charge density in this case is[1] $qN_d x_{n0}$ or $qN_a x_{p0}$. These two quantities are equal to each other due to charge neutrality. Therefore,

$$\mathscr{E}_0 = -\frac{q}{\epsilon} N_d x_{n0} = -\frac{q}{\epsilon} N_a x_{p0} \tag{B.1}$$

Moreover, the voltage drop across the *pn* junction is given by

$$V_0 = -\int_{-x_{p0}}^{x_{n0}} \mathscr{E}(x) dx \tag{B.2}$$

Clearly, this is negative of the area under the curve in Figure 7(c). Then

$$V_0 = -\frac{1}{2} \mathscr{E}_0 W = \frac{1}{2} \frac{q}{\epsilon} N_d x_{n0} W \tag{B.3}$$

Since $N_d x_{n0} = N_a x_{p0}$, and that $W = x_{n0} + x_{p0}$, one gets $x_{n0} = W N_a / (N_a + N_d)$. Therefore, it follows that

$$V_0 = \frac{1}{2} \frac{q}{\epsilon} \frac{N_a N_d}{N_a + N_d} W^2 \tag{B.4}$$

The above equation can be used to derive W_D in (1.9).

[1] Here, the the math symbols used are commensurate with Figure 7 which is from Streetman's book.