Chapter 4

THE PN JUNCTION DIODE

4.1. Introduction

A PN junction is formed when a P-type and an N-type semiconductor are in contact. If the N-and P-type regions are made out of the same semiconductor material (e.g. N-type silicon and P-type silicon), the junction is a homojunction. If the semiconductor materials are different (e.g. N-type silicon and P-type germanium), the junction is a heterojunction. Heterojunctions are dealt with in Chapter 9.

A diode is a semiconductor device consisting of a single PN junction (Figure 4.1). Unlike a resistor, it has a highly non-linear current-voltage characteristic and is often used as a rectifying element. Some diodes can emit light (light-emitting diodes), and others can emit laser light (laser diodes). The proper combination of two PN junctions produces a bipolar transistor, a device capable of amplifying electric signals.

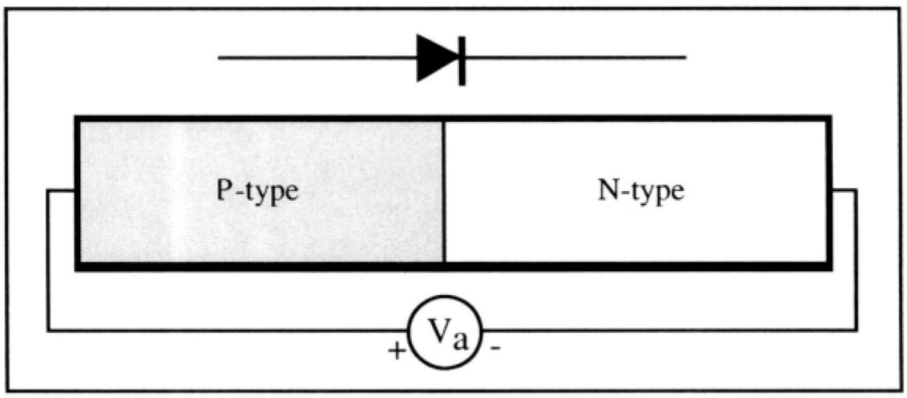

Figure 4.1: PN junction and symbol representing a diode.

The PN junction presents the following property: It allows current flow in one bias direction, but not in the other bias direction. Hence it rectifies the current. The sign convention used in this chapter is shown in Figure

4.1. The applied voltage, V_a, is positive if the potential applied to the P-side is higher than that on the N-side. As illustrated in Figure 4.2 current flows through the diode if V_a is positive, and does not if V_a is negative. If $V_a > 0$ the junction is said to be forward biased, and if $V_a < 0$ it is reverse biased.

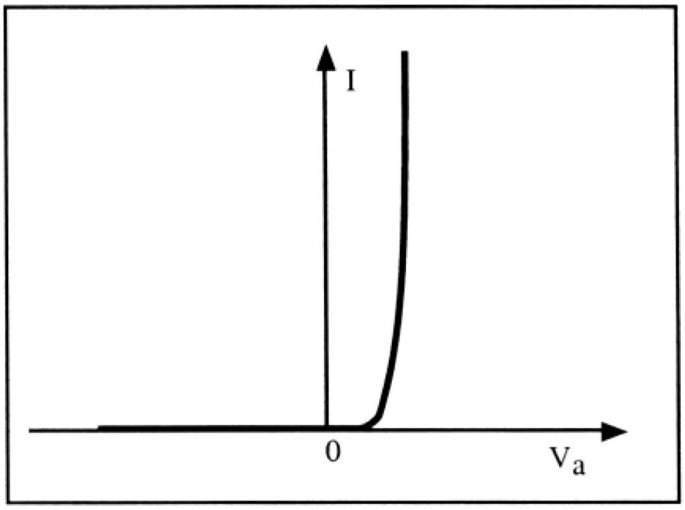

Figure 4.2: Current-voltage characteristics of a PN junction.

Experimental measurements show that the current in a PN junction, I, obeys the following equation:

$$I = I_s\left(exp\left[\frac{qV_a}{kT}\right] - 1\right) \qquad (4.1.1)$$

where I_s is a constant and V_a is the voltage applied to the diode.

An analogy of the diode is a valve which controls liquid flow (Figure 4.3). When a pressure differential is applied in the forward direction, the valve opens and allows the liquid flow. If the pressure differential is applied in the reverse direction, the valve closes, and no liquid flows, except for a few drops if the valve is imperfect and somewhat "leaky".

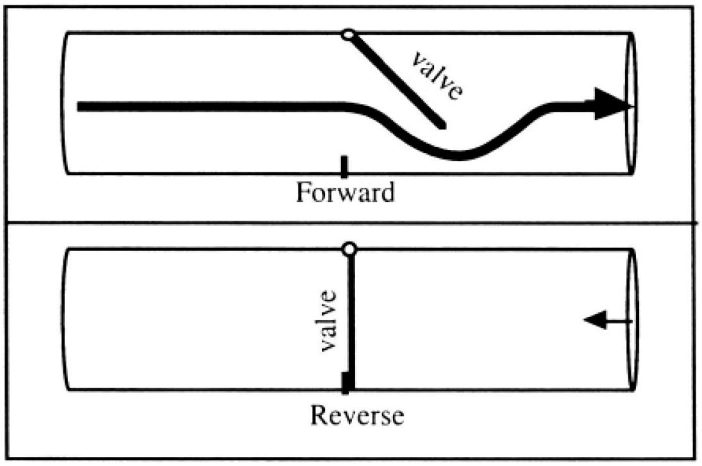

Figure 4.3: Fluid mechanics analogy of a pn junction to a valve.

4.2. Unbiased PN junction

We now consider a PN junction at thermodynamic equilibrium, *i.e.* in the absence of an applied bias ($V_a=0$). Let us first focus on the P-type and the N-type region taken separately, as if there were two separate pieces of semiconductor material. For simplicity, doping concentrations in both pieces are constant, and equal to N_d (cm^{-3}) in the N-type region, and N_a (cm^{-3}) in the P-type region. The energy band diagram of the two pieces of semiconductor are shown in Figure 4.4.

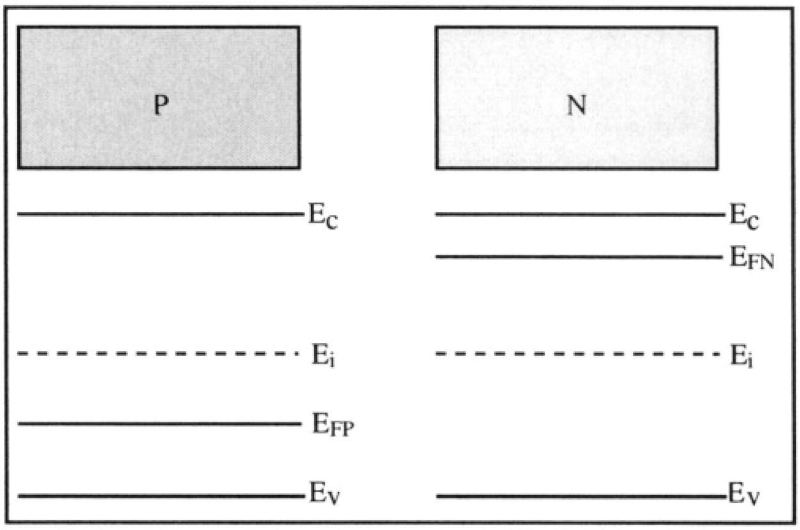

Figure 4.4: Energy band diagram in the N- and P-type regions taken separately.

Using Expressions 1.3.15a and 1.3.15b one can write:

$$E_{FN} - E_i = kT\ ln\left(\frac{N_d}{n_i}\right) \quad \text{in the N-type region, and}$$

$$E_i - E_{FP} = kT\ ln\left(\frac{N_a}{n_i}\right) \quad \text{in the P-type region.}$$

Let us now build the PN junction by connecting the P-type region to the N-type region. The surface where the contact is made is called the "metallurgical junction". A junction where the doping concentration "abruptly" switches from P-type to N-type (at the metallurgical junction) is called a step junction. We already know from Section 1.4 that the Fermi level is unique and constant in a structure under equilibrium: electrons instantly diffuse from the electron-rich N-type region into the electron-poor P-type region, and holes from the P-type material diffuse into the N-type region. As a result of the charge displacement an internal built-in potential called junction potential, Φ_o, is formed at the junction, as shown in Figure 4.5.

Within a multiplication factor *-q* the junction potential is equal to the curvature of the energy bands:

$$E_{FN} - E_{FP} = q\Phi_o = kT\ ln\left(\frac{N_d}{n_i}\right) + kT\ ln\left(\frac{N_a}{n_i}\right) = kT\ ln\left(\frac{N_a N_d}{n_i^2}\right) \quad (4.2.1)$$

and thus:

$$\Phi_o = \frac{kT}{q}\ ln\left(\frac{N_a N_d}{n_i^2}\right) \quad (4.2.2)$$

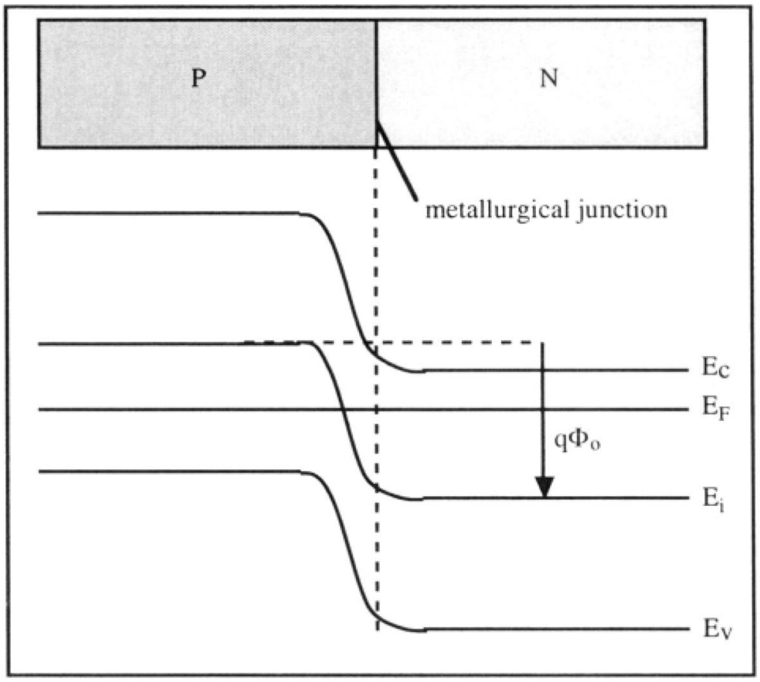

Figure 4.5: PN junction and corresponding energy band diagram.[1]

When electrons diffuse from the N-type region into the P-type material, they "leave behind" the ionized donor atoms they originated from. These atoms occupy substitutional sites in the crystal lattice and cannot move within the crystal. The region where these positively charged ions are located constitutes a space-charge region called a "depletion region" because it is depleted of electrons (Figure 4.6).

The positive charge in the depletion region attracts electrons such that at equilibrium, the force of diffusion pushing electrons into the P-type region is exactly balanced by the force of the built-in electric field that "recalls" the electrons back into the N-type region. Similarly, the diffusion of holes from the P-type into the N-type region gives rise to a depletion region in the P-type material. This region is depleted of holes and bears a negative charge because of the presence of negatively charged acceptor ionized atoms. There are several names for the depletion region located around the metallurgical junction; it can be called the "depletion region", the "space-charge region" or the "transition region".

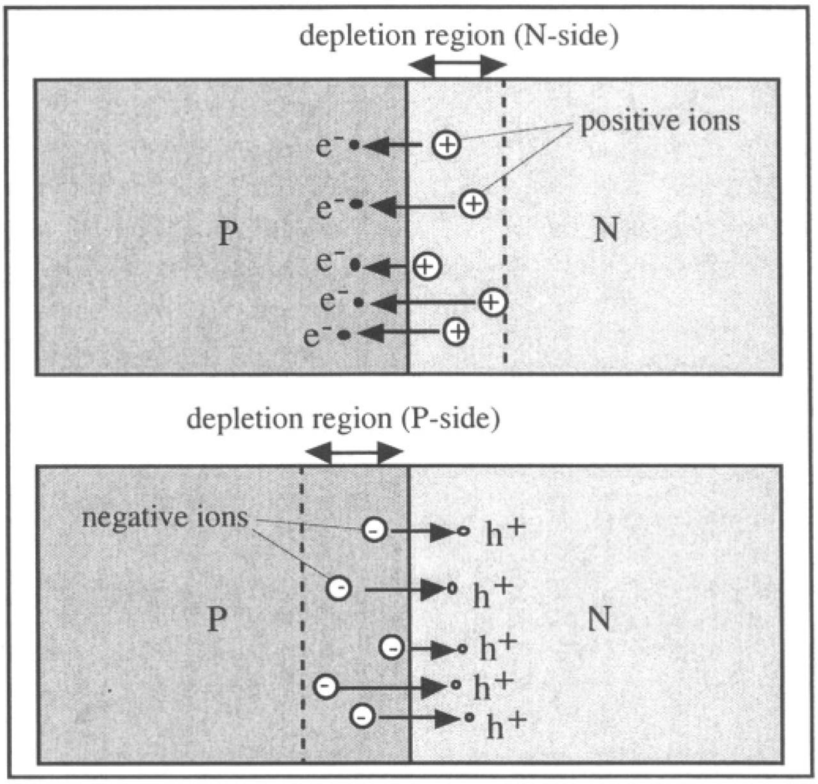

Figure 4.6: Creation of depletion regions by the diffusion of electrons and holes.

The electric field and the potential variation in the space-charge region can be calculated using the Poisson equation (Expression 2.6.2). For a one-dimensional junction the problem simplifies to:

$$\frac{d^2\Phi(x)}{dx^2} = -\frac{q}{\varepsilon_s}\left(p - n + N_d^+ - N_a^-\right) \qquad (4.2.3a)$$

Using the Boltzmann Relationships 1.3.20a and 1.3.20b we obtain:

$$\frac{d^2\Phi(x)}{dx^2} = -\frac{q}{\varepsilon_s}\left\{p_o\, exp\left[\frac{-q\Phi(x)}{kT}\right] - n_o\, exp\left[\frac{q\Phi(x)}{kT}\right] + N_d^+ - N_a^-\right\} \qquad (4.2.3b)$$

with $N_d^+ = N_d$ and $N_a^+ = N_a$.

Equation 4.2.3b cannot be solved analytically and a close-form solution for the potential cannot be found. It can, however, be simplified by using the "depletion approximation". The depletion approximation assumes that the space charge is composed only of ionized doping impurities, and that the contribution of free carriers to the local charge is negligible.

Furthermore, the carrier depletion in the space-charge regions is assumed to be complete. In other words, there are no free electrons in the depletion region on the N-type side, and no free holes in the depletion region on the P-type side. As a result, the charge densities in the depletion regions are equal to qN_d in the N-type material, and $-qN_a$ in the P-type material. The depletion regions extent to a distance l_{no} on the N-type side, and a distance $-l_{po}$ on the P-type side, where the metallurgical junction is taken as the origin (Figure 4.7). Additionally, the electric field and potential are shown in Figure 4.7, which can also be derived from Poisson's equation with the appropriate boundary conditions.

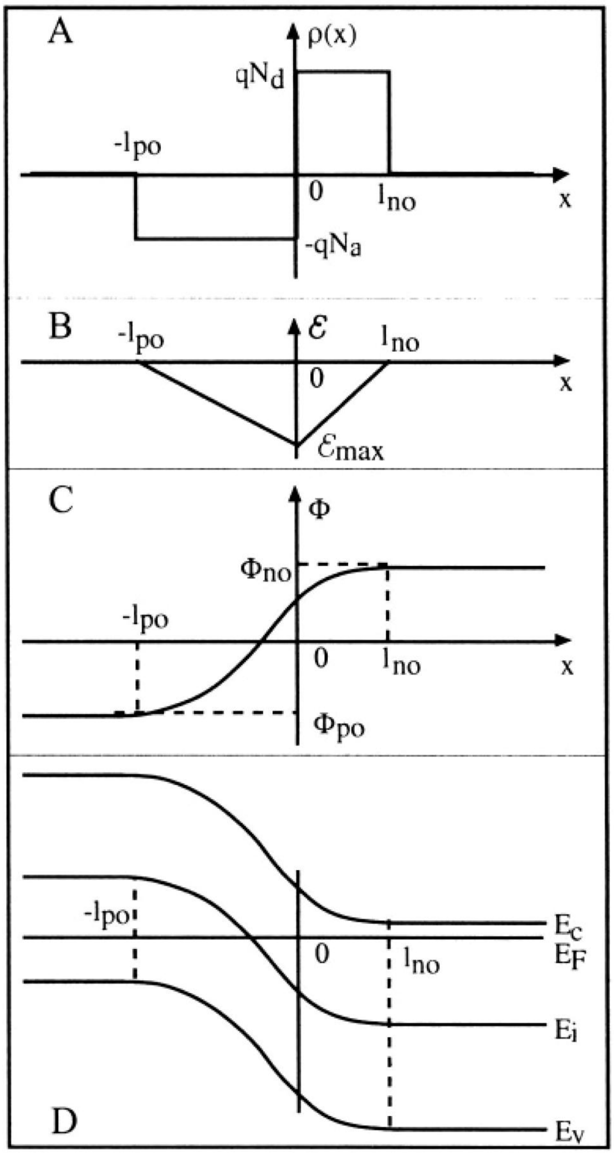

Figure 4.7: Charges (A), electric field (B) potential (C) and energy bands (D) in a PN junction.[2]

With the depletion approximation, a closed-form analytical expression can be found for the electric field $\mathcal{E}(x)$, the potential $\Phi(x)$, as well as for l_{po} and l_{no} by utilizing Poisson's equation and Gauss' law. The value of the charge density $\rho(x)$ can be expressed for four separate regions and are given by:

$$\rho(x) = \begin{cases} 0 & \text{for } -\infty < x < -l_{po} & \text{(quasi-neutral region)} \\ -qN_a & \text{for } -l_{po} < x < 0 & \text{(space-charge region)} \\ qN_d & \text{for } 0 < x < l_{no} & \text{(space-charge region)} \\ 0 & \text{for } l_{no} < x < \infty & \text{(quasi-neutral region)} \end{cases}$$

We will assume that charge neutrality exists in the quasi-neutral regions. Therefore, the electric field is zero in these regions. Using all the above assumptions the Poisson equation can be integrated a first time to yield the electric field:

for $-\infty < x < -l_{po}$: $\mathcal{E}(x)=0$

for $-l_{po} < x < 0$: $$\frac{d^2\Phi(x)}{dx^2} = -\frac{d\mathcal{E}(x)}{dx} = \frac{q}{\varepsilon_s}N_a \quad with \quad \mathcal{E}(-l_{po})=0$$

$$\Downarrow$$

$$\mathcal{E}(x) = -\frac{qN_a}{\varepsilon_s}(x + l_{po}) \tag{4.2.4}$$

for $0 < x < l_{no}$: $$\frac{d^2\Phi(x)}{dx^2} = -\frac{d\mathcal{E}(x)}{dx} = -\frac{q}{\varepsilon_s}N_d \quad with \quad \mathcal{E}(l_{no})=0$$

$$\Downarrow$$

$$\mathcal{E}(x) = -\frac{qN_d}{\varepsilon_s}(l_{no} - x) \tag{4.2.5}$$

and, for $l_{no} < x < \infty$ one obtains: $\mathcal{E}(x)=0$

The electric field is continuous at $x=0$ by imposing Gauss' law, which yields:

$$-\frac{qN_a}{\varepsilon_s}l_{po} = -\frac{qN_d}{\varepsilon_s}l_{no} \Rightarrow N_a l_{po} = N_d l_{no} \tag{4.2.6}$$

Relationship 4.2.6 reiterates charge neutrality in the device, since it states that the total negative charge in the depletion region on the N-side of the junction, $-qN_dl_{no}$, is equal, in absolute value, to the total positive charge on the P-side, qN_al_{po}. The potential distribution is obtained by integrating the Poisson equation a second time. In the P-type and N-type quasi-neutral regions the potentials are Φ_{po} and Φ_{no}, respectively. Using these as boundary conditions yields:

for $-\infty < x < -l_{po}$: $\Phi_o(x)=\Phi_{po}$

for $-l_{po} < x < 0$:
$$-\mathcal{E}(x) = \frac{d\Phi(x)}{dx} = \frac{qN_a}{\varepsilon_s}(x + l_{po})$$
$$\Downarrow$$
$$\Phi_o(x) = \frac{qN_a}{2\varepsilon_s}(x + l_{po})^2 + \Phi_{po} \qquad (4.2.7)$$

for $0 < x < l_{no}$:
$$-\mathcal{E}(x) = \frac{d\Phi(x)}{dx} = \frac{qN_d}{\varepsilon_s}(l_{no} - x)$$
$$\Downarrow$$
$$\Phi_o(x) = \Phi_{no} - \frac{qN_d}{2\varepsilon_s}(l_{no} - x)^2 \qquad (4.2.8)$$

for $l_{no} < x < \infty$:
$$\Phi_o(x) = \Phi_{no}$$

The potential is a continuous function at $x=0$. Combined with 4.2.2 this condition gives an alternate expression for the junction potential, Φ_o:

$$\frac{qN_a}{2\varepsilon_s} l_{po}^2 + \Phi_{po} = \Phi_{no} - \frac{qN_d}{2\varepsilon_s} l_{no}^2$$
$$\Downarrow$$

Junction Potential

$$\Phi_o = \Phi_{no} - \Phi_{po} = \frac{qN_a}{2\varepsilon_s} l_{po}^2 + \frac{qN_d}{2\varepsilon_s} l_{no}^2 = \frac{kT}{q} \ln\left(\frac{N_a N_d}{n_i^2}\right) \qquad (4.2.9)$$

The electric field has a single maximum value at $x=0$. Its expression can be obtained using 4.2.4 or 4.2.5:

Maximum electric field

$$\mathcal{E}_{max} = -\frac{qN_a}{\varepsilon_s} l_{po} = -\frac{qN_d}{\varepsilon_s} l_{no} \qquad (4.2.10)$$

Using Expressions 4.2.6 and 4.2.9 the width of the depletion regions, l_{po} and l_{no}, can be expressed as a function of the junction potential:

Width of Depletion Regions

$$l_{po} = \sqrt{\frac{2\varepsilon_s}{q} \frac{\Phi_o N_d}{N_a (N_a + N_d)}} \qquad (4.2.11a)$$

and

$$l_{no} = \sqrt{\frac{2\varepsilon_s}{q} \frac{\Phi_o N_a}{N_d (N_a + N_d)}} \qquad (4.2.11b)$$

The sum of the depletion regions is called the "transition region" which contains both ionized acceptor and donor impurities. The width of the transition region is given by:

$$l_{no} + l_{po} = \sqrt{\frac{2\varepsilon_s}{q} \frac{\Phi_o\,(N_a+N_d)}{N_a N_d}} \tag{4.2.12}$$

Actual PN junctions are strongly asymmetrical, which means that one side is doped much more heavily than the other. Consider the example of a PN^+ junction, with $N_a=10^{15}$ cm^{-3} and $N_d=10^{20}$ cm^{-3}. Since $N_d \gg N_a$, one obtains:

$$l_{po} = \sqrt{\frac{2\varepsilon_s}{q} \frac{\Phi_o}{N_a}} \gg l_{no} = \sqrt{\frac{2\varepsilon_s}{q} \frac{\Phi_o\,N_a}{N_d^2}} \tag{4.2.13}$$

and, therefore,

$$l_{no} + l_{po} \cong l_{po} \tag{4.2.14}$$

Comment: In a strongly asymmetrical junction, the width of the transition region is virtually equal to the width of the depletion region with the lowest doping concentration.

Example:

Calculate Φ_o, l_{no} and l_{po} in a silicon PN junction with $N_a=10^{15}$ cm^{-3} and $N_d=10^{19}$ cm^{-3}.

$$\varepsilon_{si} = \kappa_{si}\,\varepsilon_o = 11.7 \times 8.854\times10^{-14} \text{ F/cm}$$

$$q = 1.6\times10^{-19} \text{ C}$$

$\frac{kT}{q} = 26$ mV at room temperature (T = 300 K)

$n_i = 1.45\times10^{10}$ cm^{-3} at room temperature

$$\Phi_o = \frac{kT}{q} \ln\left(\frac{N_a N_d}{n_i^2}\right) = 0.82 \text{ V}$$

$$l_{po} = \sqrt{\frac{2\varepsilon_{si}}{q} \frac{\Phi_o\,N_d}{N_a\,(N_a+N_d)}} = 1.03\ \mu\text{m}$$

$$l_{no} = \sqrt{\frac{2\varepsilon_{si}}{q} \frac{\Phi_o\,N_a}{N_d\,(N_a+N_d)}} = 0.000103\ \mu\text{m} = 0.103 \text{ nm}$$

It can easily be seen that $l_{no} \ll l_{po}$

4.3. Biased PN junction

If no bias is applied to a PN junction the built-in junction potential is equal to Φ_o, as we have seen in the previous Section. The drift current generated by this potential variation is exactly equal and of opposite sign to the diffusion current caused by the carrier concentration gradients, such that the net current flow (drift + diffusion) is equal to zero. The potential variation $\Phi(x)$ actually acts as a barrier which prevents further diffusion of electrons into the P-type region and holes into the N-type region, once equilibrium has been established. That is why Φ_o is sometimes referred to as a "potential barrier" which the carriers must overcome in order to diffuse.

Consider the case when an external bias, V_a, is applied to the junction. V_a is considered positive if the potential of the P-type region is higher (more positive) than that of the N-type region. We will assume that the current flowing through the device is small enough such that the potential drops across the quasi-neutral regions are negligible. As a consequence, the external applied potential, V_a, is supported entirely by the transition region, and the internal potential, Φ, is equal to:

$$\Phi = \Phi_n - \Phi_p = \Phi_o - V_a \tag{4.3.1}$$

Noting that $-l_p$ and l_n are the edges of the transition region (Figure 4.8), the distribution of charges in the structure are:

$\rho(x) =$	0	for	$-\infty < x < -l_p$	(quasi-neutral region)
	$-qN_a$	for	$-l_p < x < 0$	(space-charge region)
	qN_d	for	$0 < x < l_n$	(space-charge region)
	0	for	$l_n < x < \infty$	(quasi-neutral region)

The Poisson equation can be solved just as it was in Equations 4.2.4 to 4.2.12, by replacing l_{no}, l_{po} and Φ_o by l_n, l_p and $(\Phi_o - V_a)$, respectively. The result is:

$$l_p = \sqrt{\frac{2\varepsilon_s}{q}\frac{(\Phi_o - V_a)\,N_d}{N_a\,(N_a+N_d)}} \tag{4.3.2}$$

and

$$l_n = \sqrt{\frac{2\varepsilon_s}{q}\frac{(\Phi_o - V_a)\,N_a}{N_d\,(N_a+N_d)}} \tag{4.3.3}$$

The total width of the transition region is equal to:

$$l_n + l_p = \sqrt{\frac{2\varepsilon_s}{q}\frac{(\Phi_o - V_a)\,(N_a+N_d)}{N_a N_d}} \tag{4.3.4}$$

It is worth noting that the width of the transition region increases when a reverse bias is applied $(V_a < 0)$ and that it decreases when a forward bias $(V_a > 0)$ is applied (Figure 4.8).

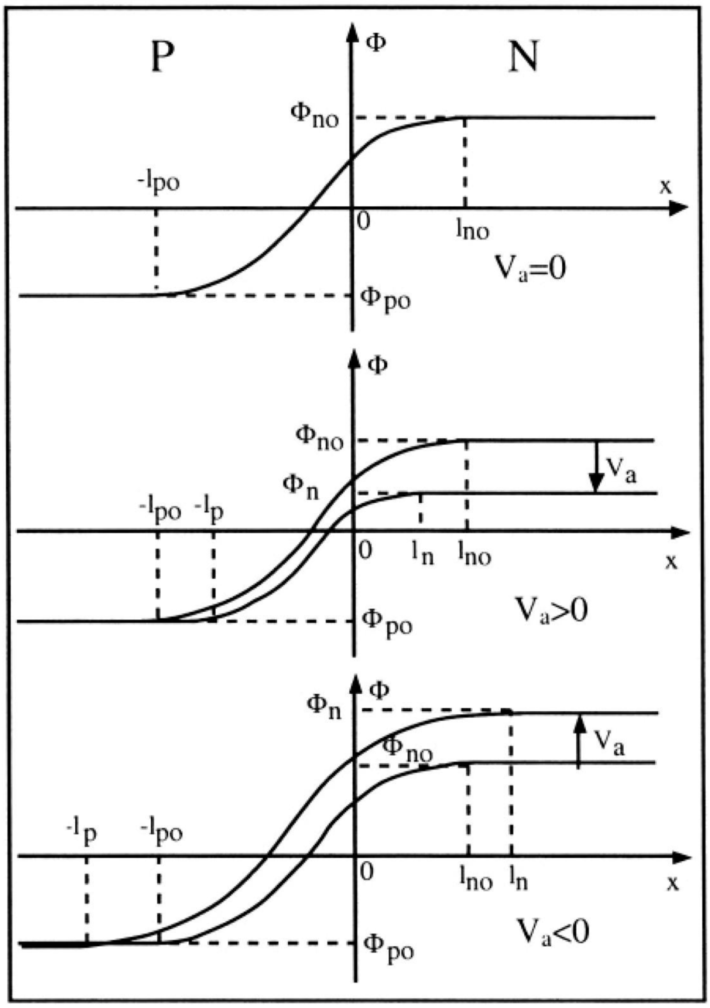

Figure 4.8: Potential in a PN junction for V_a=0, V_a>0 and V_a<0.

4.4. Current-voltage characteristics

As we have seen in the previous Section the potential drop across the transition region is equal to $\Phi_0 - V_a$, where V_a is the applied voltage. Therefore, if V_a is positive, the potential barrier in the junction is lower than its equilibrium value, Φ_0. As a result the diffusion and electric field forces are no longer equal and of opposite sign. Diffusion acting on the carriers is only partially compensated by the force resulting from the junction potential variation, and therefore, holes can flow from the P-type region into the N-type semiconductor and electrons can flow from the N-type region into the P-type semiconductor. The resulting currents are shown in Figure 4.9. The holes injected into the N-type region are excess minority carriers (current "1" in Figure 4.9). These carriers diffuse into the N-type quasi-neutral region an average distance called the "diffusion length" before recombining with the majority carriers (electrons). Since each recombination event consumes an electron, a resulting electron current appears in the N-type region where electrons are continuously supplied by the external contact (current "2" in Figure 4.9). Similarly, the electrons injected into the P-type region (current "3" in Figure 4.9) are excess minority carriers which recombine with holes in the P-type region. Since each recombination event consumes a hole, a resulting hole current appears in the P-type region (current "4" in Figure 4.9). It is worth noting that current "1" is equal to current "2" and that current "3" is equal to current "4", in Figure 4.9.

If the junction is reverse-biased ($V_a<0$) the amplitude of the potential barrier is increased beyond its equilibrium value, Φ_0. Diffusion of holes in the N-type region and diffusion of electrons in the P-type region are reduced and net current, resulting from the drift of holes from the N-type region into the P-type region and the drift of electrons from the P-type region into the N-type region, is observed. The magnitude of this current, however, is extremely small since it involves only *minority* carriers in the vicinity of the edges of the transition region.

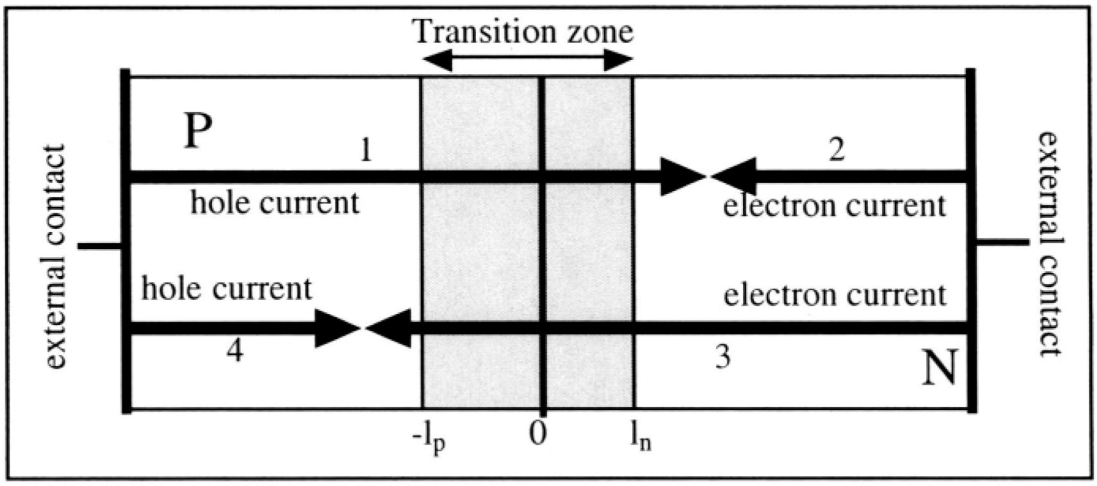

Figure 4.9: Forward-biased PN junction; 1: holes injected from the P-type region into the N-type region; 2: electrons recombining with the holes injected in the N-type region; 3: electrons injected from the N-type region into the P-type region; 4: holes recombining with the electrons injected in the N-type region.

A derivation of the current-voltage characteristics of the PN junction based on the currents of majority carriers would prove quite difficult. These are the hole current in the P-type material and the electron current in the N-type region, noted currents "4" and "2" in Figure 4.9, respectively. We know, however, that current "2" is equal to current "1" and current "4" is equal to current "3". Currents "1" and "3" are a result of minority carrier injection (holes in the N-type material and electrons in the P-type material) and the sum of these two components is equal to the total current in the device. The derivation of the modeling equations for the PN junction will, therefore, make use of currents "1" and "3".

Ultimately we want to have an equation for the PN junction which describes the current as a function of applied voltage.

4.4.1. Derivation of the ideal diode model

The notations used in this section are shown in Table 4.1:

Table 4.1: Notations used in this section

Symbol	Physical meaning	Unit
D_n	electron diffusion coefficient	$cm^2\ s^{-1}$
D_p	hole diffusion coefficient	$cm^2\ s^{-1}$
L_n	electron diffusion length	cm
L_p	hole diffusion length	cm
n_n	electron concentration in the N-type region	cm^{-3}
n_{no}	equilibrium electron concentration in the N-type region	cm^{-3}
n_p	electron concentration in the P-type region	cm^{-3}
n_{po}	equilibrium electron concentration in the P-type region	cm^{-3}
p_n	hole concentration in the N-type region	cm^{-3}
p_{no}	equilibrium hole concentration in the N-type region	cm^{-3}
p_p	hole concentration in the P-type region	cm^{-3}
p_{po}	equilibrium hole concentration in the P-type region	cm^{-3}
n'_p	excess electron concentration in the P-type region	cm^{-3}
p'_n	excess hole concentration in the N-type region	cm^{-3}
τ_n	electron lifetime	s
τ_p	hole lifetime	s

To simplify the PN junction model we will use the following starting assumptions:

1. Low-level injection assumption (or "weak injection"): the concentration of minority carriers injected in a quasi-neutral region is low compared to the majority carrier concentration.

2. The Boltzmann relationships 2.7.1 and 2.7.2 are valid in the quasi-neutral regions as well as in the transition region.

3. Current flow in the quasi-neutral regions is due to a diffusion mechanism (no potential drop, and therefore, no electric field is assumed in those regions).

4. The quasi-neutral regions of the diode are infinitely long.

5. Generation/recombination phenomena are neglected in the transition region.

Starting assumption #1- Low-level injection assumption (or "weak injection"). The concentration of minority carriers, p'_n and n'_p injected in a quasi-neutral region is low compared to the majority carrier concentration:

$$p'_n(x) \equiv p_n(x) - p_{no} << n_{no} \quad \text{in the N-type quasi-neutral region} \quad (4.4.1)$$
$$n'_p(x) \equiv n_p(x) - n_{po} << p_{po} \quad \text{in the P-type quasi-neutral region} \quad (4.4.2)$$

As a result of the low-level injection condition the concentration of majority carriers is not modified by the injection of minority carriers:

$$n_n(x) = n_{no} = N_d \quad \text{in the N-type quasi-neutral region} \quad (4.4.3)$$
$$p_p(x) = p_{po} = N_a \quad \text{in the P-type quasi-neutral region} \quad (4.4.4)$$

Starting assumption #2- The Boltzmann relationships 2.7.1 and 2.7.2 are valid in the quasi-neutral regions as well as in the transition region.

Considering the depletion region on the N-type side $(0 \leq x \leq l_n)$ one can write:

$$n(x) = n_i \, exp\left[\frac{E_{FN}-E_{io}}{kT}\right] exp\left[\frac{q\Phi(x)}{kT}\right] \quad (4.4.5)$$

From Relationship 4.4.3 we know that $n_n(x) = n_{no} = N_d \; for \; l_n \leq x \leq \infty$. Since the potential in the N-type quasi-neutral region is equal to Φ_n, we can write:

$$n_{no} = n_i \, exp\left[\frac{E_{FN}-E_{io}}{kT}\right] exp\left[\frac{q\Phi_n}{kT}\right] \quad (4.4.6)$$

Expression 4.4.6 can be substituted into 4.4.5 to give:

$$n(x) = n_{no} \, exp\left[\frac{q\Phi(x) - \Phi_n}{kT}\right]$$

Under the assumption that the Boltzmann relationships are valid in the transition region, the latter equation can be evaluated at $x = -l_p$ where $\Phi(x) = -\Phi_p$:

$$n_p(-l_p) = n_{no} \, exp\left[\frac{q(\Phi_p-\Phi_n)}{kT}\right]$$

Since $\Phi_p - \Phi_n = V_a - \Phi_o$ we can write:

$$n_p(-l_p) = n_{no} \, exp\left[\frac{-q\Phi_o}{kT}\right] exp\left[\frac{qV_a}{kT}\right]$$

From this we now have an expression for the minority carrier concentration at the edge of the transition region which is a function of the applied voltage. The equilibrium junction potential is defined by:

$$\Phi_o = \Phi_p - \Phi_n = \frac{kT}{q}\ln\left(\frac{N_a N_d}{n_i^2}\right) = \frac{kT}{q}\ln\left(\frac{p_{po}\, n_{no}}{n_i^2}\right)$$

Combining the two latter equations yields:

$$n_p(-l_p) = n_{no}\frac{n_i^2}{p_{po}\, n_{no}}\exp\left[\frac{qV_a}{kT}\right] = \frac{n_i^2}{p_{po}}\exp\left[\frac{qV_a}{kT}\right]$$

Since, by definition, $n_{po} = \dfrac{n_i^2}{p_{po}} = \dfrac{n_i^2}{N_a}$ we finally obtain:

$$n_p(-l_p) = n_{po}\exp\left[\frac{qV_a}{kT}\right] \qquad (4.4.7)$$

A similar calculation, carried out for holes at the N-side edge of the transition region, would yield:

$$p_n(l_n) = p_{no}\exp\left[\frac{qV_a}{kT}\right] \qquad (4.4.8)$$

As a result of Expressions 4.4.7 the concentration of *excess* electrons at the P-side edge of the transition region is equal to:

$$n'_p(-l_p) = n_p(-l_p) - n_{po} = n_{po}\left[\exp\left(\frac{qV_a}{kT}\right) - 1\right] \qquad (4.4.9)$$

Similarly the concentration of *excess* holes at the N-side edge of the transition region is given by:

$$p'_n(l_n) = p_n(l_n) - p_{no} = p_{no}\left[\exp\left(\frac{qV_a}{kT}\right) - 1\right] \qquad (4.4.10)$$

Starting assumption #3- Current flow in the quasi-neutral regions is due to a diffusion mechanism (no potential drop, and therefore, no electric field is assumed in those regions).

$$J_p = -qD_p\frac{dp}{dx} \quad \text{in the N-type quasi-neutral region} \qquad (4.4.11)$$

and

$$J_n = qD_n\frac{dn}{dx} \quad \text{in the P-type quasi-neutral region} \qquad (4.4.12)$$

Let us now write the Continuity Equation 2.6.6b for holes in the N-type quasi-neutral region, with the assumption that there is no generation from an external source:

$$\frac{\partial p_n}{\partial t} = -\frac{1}{q}\frac{\partial J_p}{\partial x} - \frac{p_n-p_{no}}{\tau_p} \tag{4.4.13}$$

and, replacing J_p by its value in Equation 4.4.11:

$$\frac{\partial p_n}{\partial t} = D_p\frac{\partial^2 p_n}{\partial x^2} - \frac{p_n-p_{no}}{\tau_p} \tag{4.4.14}$$

Assuming steady-state conditions ($\partial p/\partial t = 0$) the following differential equation is obtained:

$$D_p\frac{d^2 p_n}{dx^2} = \frac{p_n-p_{no}}{\tau_p} \tag{4.4.15}$$

which admits the solution:

$$p_n(x) = p_{no} + A\ exp(-x/L_p) + B\ exp\ (x/L_p) \tag{4.4.16}$$

where A and B are integration constants, and L_p is called the *diffusion length* of holes, defined by:

$$L_p = \sqrt{D_p\tau_p} \tag{4.4.17}$$

A similar calculation, made for electrons in the P-type quasi-neutral region, would yield:

$$n_p(x) = n_{po} + C\ exp(-x/L_n) + D\ exp\ (x/L_n) \tag{4.4.18}$$

where C and D are integration constants, and L_n is called the *diffusion length* of electrons, defined by:

$$L_n = \sqrt{D_n\tau_n} \tag{4.4.19}$$

Starting assumption #4- Consider a "long-base diode", *i.e.* a diode where the length of the quasi-neutral regions is much larger than the diffusion length of the minority carriers, L_n and L_p. From a mathematical point of view this condition is equivalent to assuming that the length of the quasi-neutral regions is infinite.

Using Expression 4.4.8 and $p_n(\infty) = p_{no}$ (thermodynamic equilibrium far from the junction) as boundary conditions for Equation 4.4.16 one obtains:

$$p_n(\infty) = p_{no} \Rightarrow B = 0 \tag{4.4.20}$$

$$p_n(l_n) = p_{no}\ exp\left[\frac{qV_a}{kT}\right] = p_{no} + A\ exp\left(\frac{-l_n}{L_p}\right)$$

which yields:

$$A = p_{no}\left[exp\left(\frac{qV_a}{kT}\right) - 1\right] exp\left(\frac{l_n}{L_p}\right) \tag{4.4.21}$$

Once the integration constants A and B are known the concentration of holes in the quasi-neutral N-type region can be derived from Equation 4.4.16:

$$p_n(x) = p_{no} + p_{no}\left[exp\left(\frac{qV_a}{kT}\right) - 1\right] exp\left(\frac{-(x-l_n)}{L_p}\right) \qquad (4.4.22)$$

The hole diffusion current in the quasi-neutral N-type region is, therefore, equal to:

$$J_p = -qD_p\frac{dp_n}{dx} = \frac{qD_p p_{no}}{L_p}\, exp\left(\frac{-(x-l_n)}{L_p}\right)\left[exp\left(\frac{qV_a}{kT}\right) - 1\right] \qquad (4.4.23)$$

Similarly, the electron diffusion current in the quasi-neutral P-type region is given by:

$$J_n = qD_n\frac{dn_p}{dx} = \frac{qD_n n_{po}}{L_n}\, exp\left(\frac{x+l_p}{L_n}\right)\left[exp\left(\frac{qV_a}{kT}\right) - 1\right] \qquad (4.4.24)$$

Since the diode considered here is a one-dimensional device with two access terminals the current flowing through it is constant and independent of the position x. One can, however, observe that the hole current density given by Expression 4.4.23 decreases when the value of x is increased (with $l_n < x < \infty$). This occurs because the holes, which are minority carriers in the N-type region, recombine with electrons, which are majority carriers. Since an electron must be supplied for every recombination event in which a hole disappears the current steadily transforms from a hole current into an electron current as x is increased. Similarly the electron current in the P-type region disappears to the benefit of a hole current as x (with $-\infty < x < -l_p$) is decreased. The net current density in the device is given by:

$$J = constant = J_n(x) + J_p(x), \text{ for any value of } x \qquad (4.4.25)$$

The minority carrier concentrations in the quasi-neutral regions and the hole and electron current densities are shown as a function of position, x, for $V_a > 0$, in Figure 4.10.

Since we have assumed no generation/recombination in the transition zone **(Starting assumption #5**) we can write:

$$J_p(-l_p) = J_p(l_n) \quad \text{and} \quad J_n(-l_p) = J_n(l_n) \qquad (4.4.26)$$

The current at the boundaries of the space-charge region is entirely due to the minority carriers which have been injected. As a result, the total current in the device will be the sum of these two components, *i.e.* the sum of expressions 4.4.23 and 4.4.24 evaluated at l_n and $-l_p$, respectively.

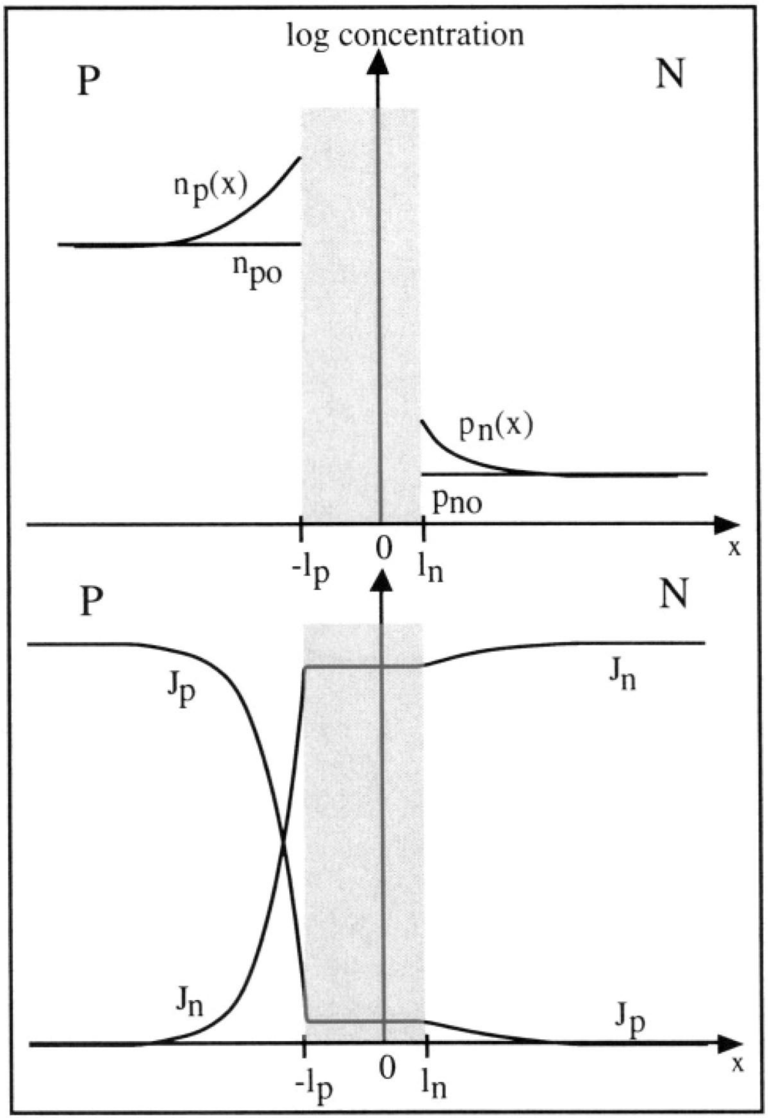

Figure 4.10: Minority carrier concentrations (top) and electron and hole current densities (bottom). [3]

Using the two latter Relationships we can write:

$$J_{TOTAL} = J = J_n(-l_p) + J_p(l_n)$$

$$\Downarrow$$

$$J = \frac{qD_n n_{po}}{L_n} exp\left(\frac{-l_p+l_p}{L_n}\right)\left[exp\left(\frac{qV_a}{kT}\right) - 1\right] + \frac{qD_p p_{no}}{L_p} exp\left(\frac{-(l_n-l_n)}{L_p}\right)\left[exp\left(\frac{qV_a}{kT}\right) - 1\right]$$

$$\Downarrow$$

$$J = \left\{\frac{qD_n n_{po}}{L_n} + \frac{qD_p p_{no}}{L_p}\right\}\left[exp\left(\frac{qV_a}{kT}\right) - 1\right]$$

$$\Downarrow$$

Current Density in the ideal PN junction

$$J = J_s \left[exp\left(\frac{qV_a}{kT}\right) - 1 \right] \qquad (4.4.27)$$

where J_s is called the "saturation current density" and is equal to:

Saturation Current Density

$$J_s = \frac{qD_n n_{po}}{L_n} + \frac{qD_p p_{no}}{L_p} = q\, n_i^2 \left(\frac{1}{N_a}\sqrt{\frac{D_n}{\tau_n}} + \frac{1}{N_d}\sqrt{\frac{D_p}{\tau_p}} \right) \qquad (4.4.28)$$

It is worthwhile noting that the magnitude of the current flowing in a reverse-biased PN junction $(V_a<0)$ is equal to J_s. J_s is independent of the applied bias and of the magnitude of the electric field in the structure. It is, however, quite dependent on temperature.

The current in the device can readily be obtained by multiplying the current density, J, of expression 4.4.27 by the cross-sectional area of the junction, A such that $I = AJ$ (amperes). The current expression obtained in Relationship 4.4.27 is in good agreement with experimental current-voltage characteristics, since Expression 4.4.27 is equivalent to Expression 4.1.1, where $I_s = A\, J_s$. Note that the reverse-bias current of the diode, $-I_s$, is sometimes called a "leakage current".

4.4.2. Generation/recombination current

We have so far calculated the current-voltage characteristics of an "ideal diode" and neglected generation/recombination mechanisms in the transition region. Actual diodes are, unfortunately, non-ideal and the effects of generation/recombination have to be taken into account to accurately model experimental device characteristics.

When an external bias, V_a, is applied, the *pn* product in the transition region is different from its equilibrium value, n_i^2, since excess carriers are injected into $(V_a>0)$ or extracted from $(V_a<0)$ the transition region. As a result the Fermi level splits into two quasi Fermi levels (E_{Fn} for electrons and E_{Fp} for holes). The difference between the two quasi Fermi levels is the applied voltage, V_a. According to Expression 2.7.6 the *pn* product in the transition region is equal to:

$$pn = n_i^2\, exp\left[\frac{E_{Fn} - E_{Fp}}{kT}\right] = n_i^2\, exp\left[\frac{qV_a}{kT}\right] \qquad (4.4.29)$$

Therefore, the SRH generation/recombination rate is equal to (Expression 3.5.14):

$$U = \frac{pn - n_i^2}{\tau_o\left(p + n + 2n_i \cosh\left[\frac{E_t - E_i}{kT}\right]\right)} = \frac{n_i^2\left[\exp\left(\frac{qV_a}{kT}\right) - 1\right]}{\tau_o\left(p + n + 2n_i \cosh\left[\frac{E_t - E_i}{kT}\right]\right)} \quad (4.4.30)$$

or, considering that the recombination centers are located at midgap ($E_t = E_i$), where recombination is the most effective (see Expression 3.5.14):

$$U = \frac{n_i^2\left[\exp\left(\frac{qV_a}{kT}\right) - 1\right]}{\tau_o\left(p + n + 2n_i\right)} \quad (4.4.31)$$

Using the continuity equations 2.6.7a and 2.6.7b in steady state, which assumes $U_p = U_n \equiv U$, we can write:

$$\frac{dJ_n}{dx} = qU = -\frac{dJ_p}{dx} \quad (4.4.32)$$

and, integrating over the transition region one obtains:

$$J_n(l_n) = J_n(-l_p) + q\int_{-l_p}^{l_n} U(x)\,dx \quad (4.4.33)$$

The net current density is given by:

$$J = J_p(l_n) + J_n(l_n) = J_p(l_n) + J_n(-l_p) + q\int_{-l_p}^{l_n} U(x)\,dx$$

$$= J_s\left[\exp\left(\frac{qV_a}{kT}\right) - 1\right] + q\int_{-l_p}^{l_n} U(x)\,dx$$

which can be rewritten:

$$J = J_s\left[\exp\left(\frac{qV_a}{kT}\right) - 1\right] + J_{rg} \quad (4.4.34)$$

For a given forward bias, V_a, the generation/recombination rate will have a maximum value at that location in the transition region where the sum of the electron and hole concentration, $p+n$, is at a minimum value, based on Expression 4.4.31.[4] Since the product of the electron and hole concentrations, pn, is a constant, the conditions $d(p+n)=0$ and $d(pn)=0$ lead to:

$$dp = -dn \quad \text{and} \quad n\,dp + p\,dn = 0 \quad \Rightarrow \quad p = n$$

This condition exists at a location within the transition region where the intrinsic Fermi level, E_i, is half-way between the quasi-Fermi level for electrons, E_{Fn}, and for holes, E_{Fp}. There, the carrier concentrations are given by 4.4.29:

$$n = p = n_i \exp(qV_a/2kT)$$

and, the recombination rate, U, can be found using Expression 4.4.31:

$$U = \frac{n_i^2\left[\exp\left(\frac{qV_a}{kT}\right) - 1\right]}{\tau_o\left(p + n + 2n_i\right)} = \frac{n_i^2\left[\exp\left(\frac{qV_a}{kT}\right) - 1\right]}{2n_i\left[\exp\left(\frac{qV_a}{2kT}\right) + 1\right]\tau_o} = \frac{n_i\left[\exp\left(\frac{qV_a}{2kT}\right) - 1\right]}{2\tau_o}$$

Assuming the latter expression is valid (*i.e.* generation/recombination is maximum) over the entire transition region Equation 4.4.33 can be solved analytically. The assumption of maximum generation/ recombination over the entire transition region will slightly overestimate the current Jrg, but it accurately reproduces its exponential dependence on $qV_a/2kT$. Using Relationship 4.3.4 for calculating the width of the transition region we find:

$$J_{rg} = q\int_{-l_p}^{l_n} U(x)\,dx = q\,U\,(l_n+l_p)$$

$$= \frac{qn_i}{2\tau_o}\sqrt{\frac{2\varepsilon_s}{q}(\Phi_o - V_a)\frac{N_a+N_d}{N_aN_d}}\left[\exp\left(\frac{qV_a}{2kT}\right) - 1\right] \quad (4.4.35)$$

For a silicon diode the generation/recombination current is larger than the diffusion current for small forward bias and adds to the reverse current when $V_a<0$. At small forward biases, therefore, the current dependence on the applied voltage follows an $\exp\left(\frac{qV_a}{2kT}\right)$ law, which is characteristic of a recombination-dominated current. At higher bias values, however, the $\exp\left(\frac{qV_a}{kT}\right)$ variation due to the diffusion current takes over (Figure 4.11) and completely overshadows the recombination current.

Generation current can be observed in the reverse-bias current-voltage characteristics. The physical origin of that current is the following: when the junction is reverse biased ($V_a<0$), the *pn* product, given by Equation 4.4.29, is smaller than n_i^2. Therefore, the SRH generation mechanism forces an increase in the *pn* product towards its equilibrium value. The generated carriers are separated by the electric field in the transition region. The generated holes are swept into the P-type quasi-neutral region, and the generated electrons into the N-type region. The motion of these carriers constitutes the generation current.

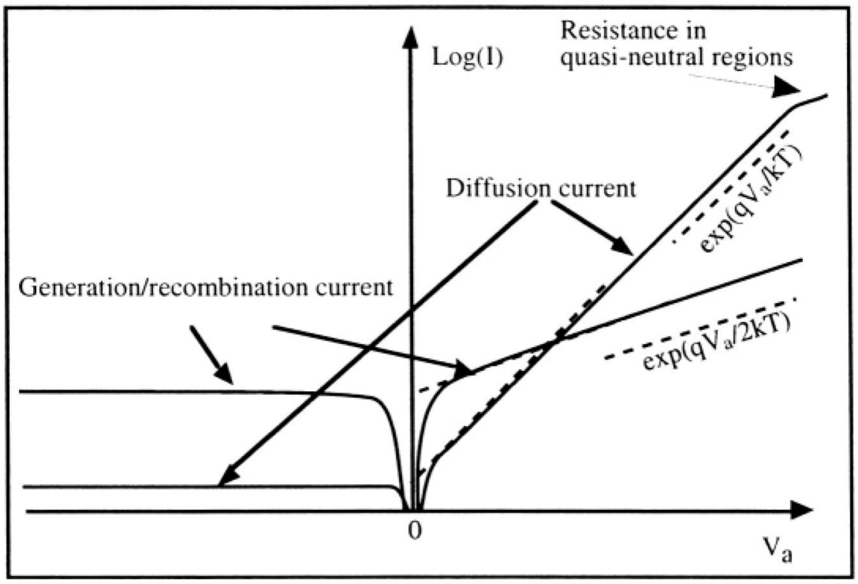

Figure 4.11: Diffusion, generation and recombination currents in the junction.

Often diffusion and generation/recombination currents are regrouped into a single current expression:

$$I = I_s \left[exp\left(\frac{qV_a}{nkT}\right) - 1 \right] \tag{4.4.36}$$

where n is called the "ideality factor". The ideality factor ranges between 1 and 2. It is equal to 1 in a diode where the current is completely dominated by diffusion mechanisms (ideal diode), and it is equal to 2 when the current is completely dominated by generation/recombination mechanisms.

Another divergence from ideality exists. At high forward-bias current levels the resistance in the quasi-neutral regions can no longer be neglected. If R is the sum of the resistances in the P and N neutral regions, then the potential difference at the edges of the transition region is not V_a, but rather V_a - IR. This causes a reduction of the current with applied voltage at high current levels (Figure 4.11).

4.4.3. Junction breakdown

When a PN junction is strongly reversed biased the electric field near the metallurgical junction can reach high values. The value of that field is given by Expression 4.2.10, where l_{no} and l_{po} are replaced by l_n and l_p, respectively. Carriers accelerated in that field can accumulate enough kinetic energy that they can, through a collision process, generate electron-hole pairs through impact ionization (see end of Section 3.3). The generated carriers can in turn be accelerated, and again through impact ionization, generate additional carriers. This carrier multiplication