Chapter 6

PN Junction Diode

Outline

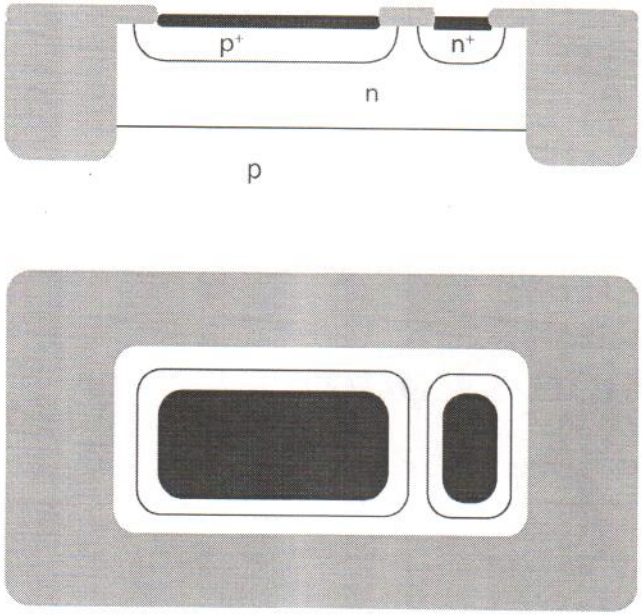


Figure 6.1: Cross section and top view of integrated PN junction diode.

The PN junction diode (or simply "PN diode") is the first microelectronics device that we study in detail in this book. The PN diode is a two-terminal device that exploits the *rectifying characteristics* of a PN junction. What this means is that under a certain voltage polarity, current readily flows across the device, while for the reverse polarity, the device is essentially open.

A top view and cross-sectional view of an integrated PN diode is shown in Fig. 6.1. It consists of a p^+ region created inside an n-type well fabricated on a p-type wafer. The contact to the n-type well is made through an n^+ region. The p region is often called the *anode* and the n region is referred to as the *cathode*.

PN diodes are very useful in many applications. They are, for example, used to provide electrostatic discharge (ESD) protection in input/output circuitry of integrated circuits. They also perform a variety of functions in analog circuits and power electronics. A number of other familiar semiconductor devices are in essence PN diodes. Examples are solar cells, light-emitting diodes and laser diodes.

In addition, PN junctions themselves are present in virtually every microelectronic device. For example, two PN junctions back-to-back constitute the essence of the bipolar junction transistor, a very important device that we will study in detail in Ch. 11. A Complementary Metal-Oxide-Semiconductor (CMOS) transistor pair also includes several PN junctions, as can be seen in Fig. 6.2. Finally, and this is perhaps their most widespread application, reverse biased PN junctions exhibit excellent isolation properties and are therefore employed in integrated circuits to isolate devices from each other. An example is seen in Fig. 6.1 where the n-well/p-substrate junction isolates the PN diode from the rest of the devices on the wafer.

This chapter deals with the physics of the PN junction and the PN junction diode. These are

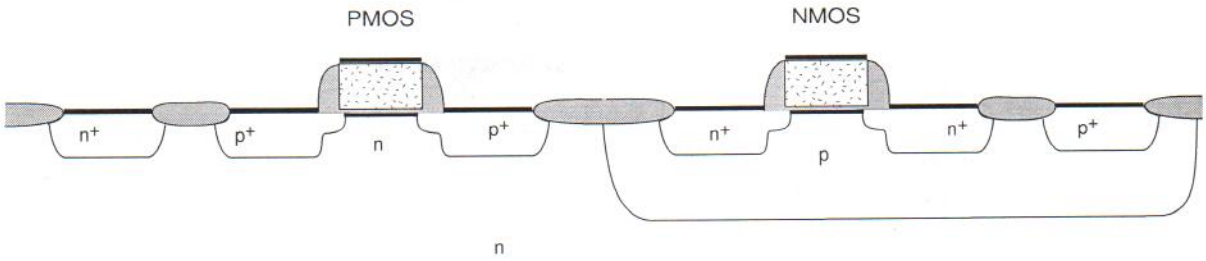


Figure 6.2: Cross section of a Complementary Metal-Oxide-Semiconductor (CMOS) transistor pair showing several PN junctions that naturally occur in this structure.

minority-carrier-type structures. In our analysis, we will heavily draw from our earlier study of minority-carrier type transport of Ch. 5. Since minority carrier behavior is also at the heart of bipolar transistor operation, some of the materials that we develop in this chapter will reappear in Ch. 11.

This chapter is organized as follows. We start by defining the notion of the ideal PN junction diode. We follow this with an analysis of the PN diode in thermal equilibrium. We then discuss the consequences of driving the PN junction out of equilibrium through the application of a bias. This culminates with the derivation of the rectifying current-voltage characteristics of the PN junction. The rest of the chapter is devoted to studying in detail the PN junction diode: equivalent circuit models, parasitics, second-order effects and design considerations. The large-signal dynamics of the PN diode receive special attention as they are distinctly different in nature from those of the other kind of diode, the Schottky diode that we study in Ch. 7. We will understand the reasons behind the relatively slow dynamic response of PN diodes that makes them unsuitable for many high-frequency applications.

This is a key chapter in this book. Important concepts presented earlier are encountered again here but now in a device context. New concepts are also introduced that will reappear later on in different chapters.

6.1 The ideal PN junction diode

It is useful to define the notion of an ideal PN junction diode. This is a concept device with a simple geometry, simplified physics and no parasitics that captures the essence of the PN diode. We define this model device in this section and then analyze it in the following sections. At the end of this chapter, we will study the most significant non idealities and the parasitics that affect real PN diodes.

Fig. 6.3 shows the circuit symbol of a PN diode and a sketch of an ideal diode. This consists of a p-region with doping level N_A on top of an n region with doping level N_D. The transition between p-type and n-type is perfectly abrupt. Each region is electrically accessed through an ohmic contact. This device stands alone and is completely isolated from the rest of the world.

In the analysis of the ideal PN diode, we will make a number of approximations. Some of

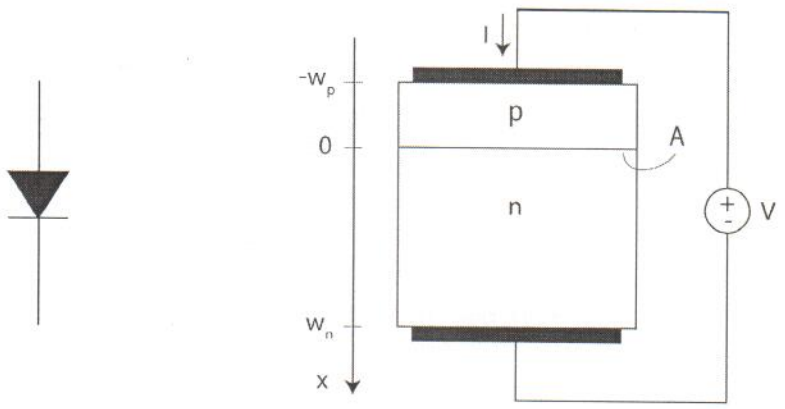


Figure 6.3: Left: circuit symbol for PN diode. Right: sketch of ideal PN diode.

these will be "undone" later on in this chapter. These are our simplifying assumptions:

- All carrier flow is one dimensional. There are no 2D or 3D effects.
- Doping levels are uniform throughout (we study the impact of non-uniform doping levels in Sec. 6.6.5).
- We treat the p-n transition region under the so-called depletion approximation. We consider the rest of the semiconductor on both sides as quasi-neutral regions. This is explained below.
- We assume that the quasi-neutral regions are much longer than the respective minority carrier diffusion lengths (we call this a "long" diode; a "short" diode in which the quasi-neutral regions are much shorter than the diffusion lengths is discussed in Sec. 6.6.1).
- We assume non-degenerate statistics for electrons and holes everywhere.
- Under conditions of excess carriers, low-level injection prevails at all times in all regions (some of the implications of high-level injection are examined in Sec. 6.6.6).
- We neglect generation and recombination of carriers in the depletion region (this is added in Sec. 6.6.2).
- We ignore any resistance effects from the semiconductor regions or the ohmic contacts (we study the impact of parasitic resistance in Sec. 6.6.3).
- We assume ideal ohmic contacts as defined in Sec. 5.2.2.
- We ignore any effects associated with the sidewalls of the device.

Fig. 6.3 defines the axis that we will use to analyze the PN diode. Since we assume this one to be a one-dimensional situation, there is only one axis to consider. We place its origin at the so-called *metallurgical junction* where the semiconductor changes from p-type to n-type. The extent of the p region is w_p. The extent of the n region is w_n. The area of the junction is A.

The voltage and current notation are also shown in Fig. 6.3. We always define the applied voltage with the p region positive with respect to the n region. When $V > 0$, the PN diode is under forward bias. When $V < 0$, the PN diode is under reverse bias. The current entering the p region is defined as positive. This notation is consistent with a naming convention for the two regions that is not widely used. The p region is at times called *anode* while the n-region is denoted as *cathode*.

6.2 Ideal PN junction in thermal equilibrium

A good place to start the analysis of the ideal PN diode is with an analysis of the structure in thermal equilibrium. We can envision creating such a situation by bringing together a p and an n region in intimate contact so that charge redistribution can take place. Fig. 6.4 sketches what happens. Isolated, the n-side contains lots of electrons and very few holes. The p-side, on the other hand, has many holes and very few electrons. A different way of looking at this is that the conduction band on the n-side is populated with lots of electrons, while that on the p-side is rather empty. Similarly, the valence band on the p-side has a sizable amount of holes while that on the n-side has very few.

Once intimate contact is established, electrons on the n-side will tend to flow towards the p-side since they are presented with empty states at the same energy. At the same time, holes on the p-side will flow towards the n-side where there are lots of available states at the same energy. We can alternatively view this intermingling of carriers as being driven by diffusion, since there are more carriers of one type on each side of the junction. While this view works well in this particular case, in general it is not a sure way to argue the sense of carrier redistribution. The energy view always provides a good understanding of the sense of carrier redistribution.

As this carrier intermingling proceeds, two things happen. First, electron-hole recombination takes place. Second and more important for us here, an electric field develops as the ionized acceptors and donors have their charge uncompensated by their respective majority carriers. The more carrier diffusion takes place, the higher the electric field. It is this electric field that eventually brings the diffusion process to a halt by introducing a drift force on the carriers that precisely compensates their diffusion tendency.

Once thermal equilibrium has been established and charge redistribution has stopped, a charge dipole has developed across the metallurgical junction. Sufficiently far away on either side of the junction, bulk conditions ought to prevail and the majority carrier concentrations equal the respective doping concentrations. In the vicinity of the junction, the n-side has less electrons and more holes than when it was isolated. The situation is reversed on the p-side. This results in net *positive charge* on the n-side and net *negative charge* on the p-side, as sketched in Fig. 6.5. This dipole of charge produces an electric field which results in a potential energy that must be added to the energy band diagram. The magnitude and spatial distribution of the electric field has to be such as to make the Fermi level flat everywhere, as sketched in Fig. 6.4.

The dipole of charge at the metallurgical interface of a PN junction results in a *built-in potential* ϕ_{bi}. The energy band diagram of Fig. 6.4 allows us to easily calculate ϕ_{bi}. Let us place

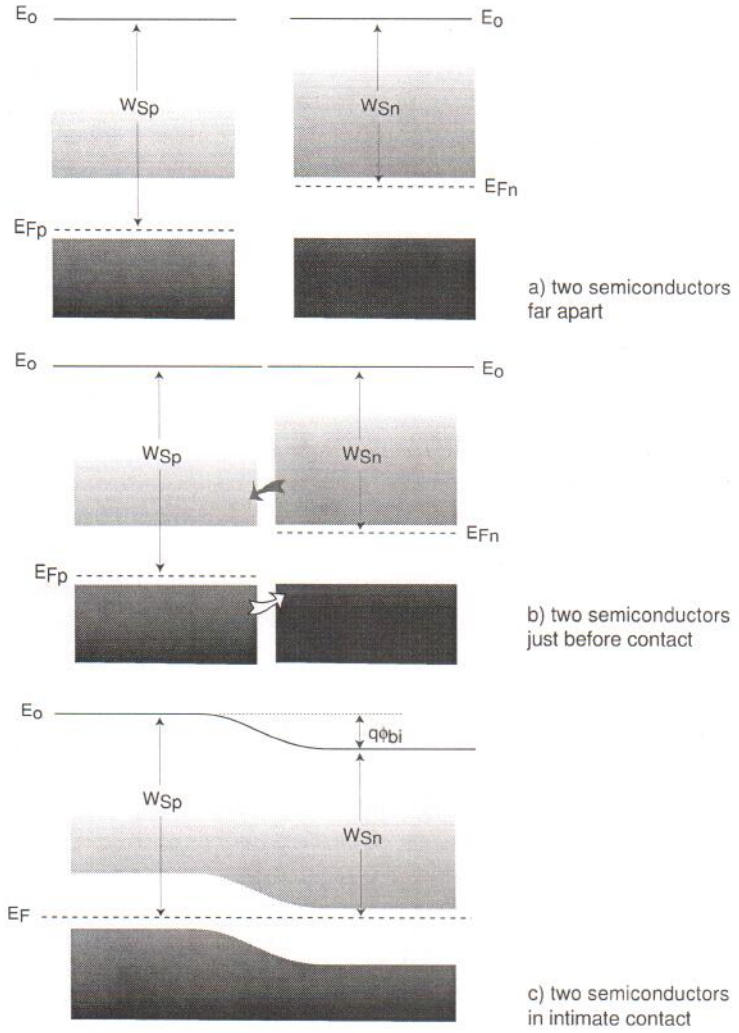


Figure 6.4: Formation of the energy band diagram of a PN junction in thermal equilibrium. Top: the p- and n-regions isolated from each other. Bottom: after contact has been established and equilibrium has been reached.

the coordinate origin $x = 0$ at the metallurgical junction and denote $x \gg 0$ and $x \ll 0$ as the n- and p-regions sufficiently far away from the junction so that bulk conditions prevail. Then:

$$\begin{aligned} q\phi_{bi} &= W_{Sp}' - W_{Sn} = (E_C - E_F)|_{x \ll 0} - (E_C - E_F)|_{x \gg 0} \\ &= kT \ln \frac{N_c}{n_o(x \ll 0)} - kT \ln \frac{N_c}{n_o(x \gg 0)} = kT \ln \frac{n_o(x \gg 0)}{n_o(x \ll 0)} \end{aligned} \tag{6.1}$$

Far away on the right where bulk conditions exist, electrons are majority carriers. Hence: $n_o(x \gg 0) = N_D$. Similarly far away on the left, $n_o(x \ll 0) = n_i^2/N_A$. Inserting this in Eq. 6.1, we get:

$$\phi_{bi} = \frac{kT}{q} \ln \frac{N_D N_A}{n_i^2} \tag{6.2}$$

It is important to recognize the limits to the validity of this expression. Three implicit assumptions were made in its derivation. First, it was assumed that both regions are extrinsic, that is, that $N_D \gg n_i$ and $N_A \gg n_i$. This is always the case near room temperature. Second, all dopants were assumed to be ionized, also a good approximation for Si at room temperature if common dopants are utilized. Finally, in Eq. 6.1 Maxwell-Boltzmann statistics were utilized. This actually is not a good assumption in many practical PN junctions in which often one side is heavily doped. In this case, Fermi-Dirac statistics and heavy-doping effects, as described in Advanced Topic AT2.6, must be taken into account. When heavy doping effects are important, ϕ_{bi} is slightly modified from the result given in Eq. 6.2. However, as a result of the presence of the logarithmic term, the difference is small and in general the error incurred in using Eq. 6.2 is acceptable for most applications.

Examining the volume charge density distribution in a PN junction in thermal equilibrium in Fig. 6.5, one can distinguish three different regions. Around the metallurgical junction, there is a *space-charge region* (SCR). On both sides of the SCR, there are two quasi-neutral regions (QNR), one p-type and the other n-type. Deep in the QNR's, $\rho \simeq 0$. As we advance towards the metallurgical junction from the n side, the electron concentration drops and the hole concentration increases. Since electrons are majority carriers, the drop of n_o is more significant from a charge point of view than the rise of p_o, and there is net positive charge due to the exposed donor atoms. Hence $\rho \simeq qN_D$. A similar situation occurs on the p side of the junction where $\rho \simeq -qN_A$ right next to the metallurgical junction. Because of the exponential dependence of the carrier concentration on the electrostatic potential, the transition between the QNR's and the SCR is fairly sharp.

This understanding suggests that an approximate model to the electrostatics of the PN junction in equilibrium can be obtained by performing what is called the *depletion approximation*. This approximation assumes that the two quasi-neutral regions are perfectly charge neutral and that the space-charge region is perfectly devoid of carriers, i.e., depleted. Mathematically, this can be expressed as:

$$\rho_o(x) \simeq 0 \qquad \text{in p-QNR: } x < -x_p \tag{6.3}$$

$$\rho_o(x) \simeq -qN_A \qquad \text{in SCR: } -x_p < x < 0 \tag{6.4}$$

$$\rho_o(x) \simeq qN_D \qquad \text{in SCR: } 0 < x < x_n \tag{6.5}$$

$$\rho_o(x) \simeq 0 \qquad \text{in n-QNR: } x_n < x \tag{6.6}$$

In these equations, the subindex "o" indicates thermal equilibrium. The location of the edges of the SCR, $-x_p$ and x_n, is determined below.

Within the depletion approximation, models for the electrostatics of the PN junction are easy to obtain. Integration of this volume charge yields the electric field profile:

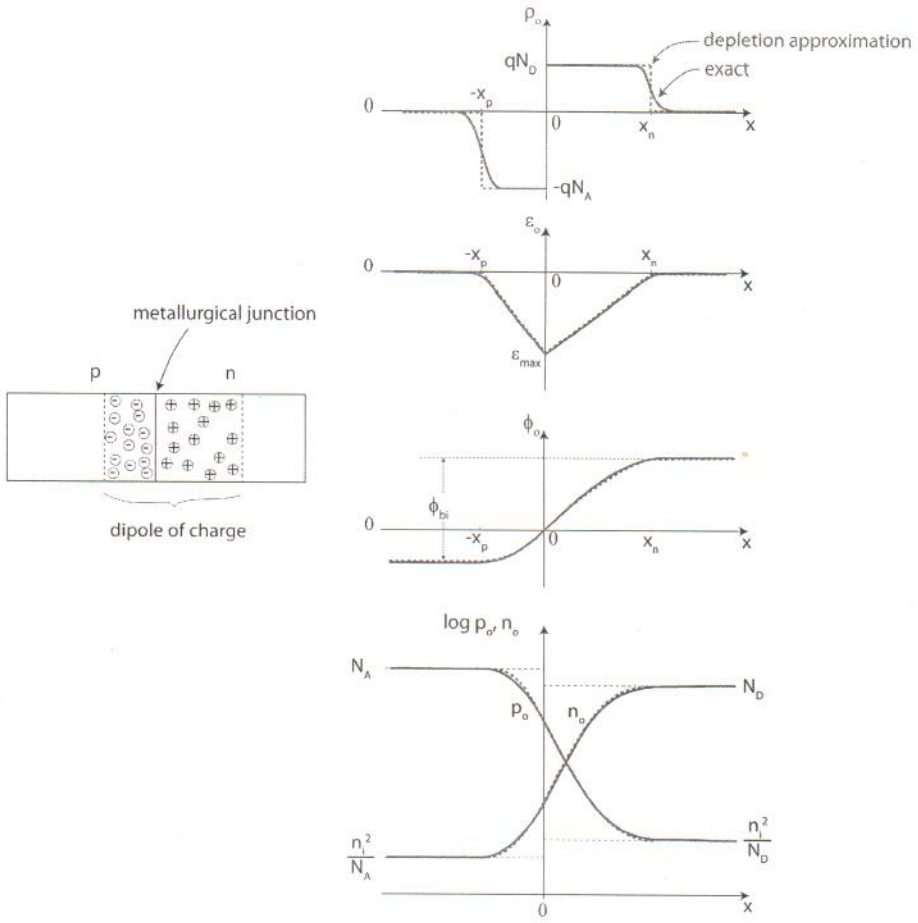


Figure 6.5: Electrostatics of a PN junction in thermal equilibrium. From top to bottom: volume charge density, electric field, electrostatic potential, and equilibrium carrier concentration. The continuous line is the exact solution. The discontinuous line is the calculation following the depletion approximation.

$$\mathcal{E}_o(x) \simeq 0 \qquad \text{in p-QNR: } x \leq -x_p \tag{6.7}$$

$$\mathcal{E}_o(x) \simeq -\frac{qN_A}{\epsilon}(x + x_p) \qquad \text{in SCR: } -x_p \leq x \leq 0 \tag{6.8}$$

$$\mathcal{E}_o(x) \simeq \frac{qN_D}{\epsilon}(x - x_n) \qquad \text{in SCR: } 0 \leq x \leq x_n \tag{6.9}$$

$$\mathcal{E}_o(x) \simeq 0 \qquad \text{in n-QNR: } x_n \leq x \tag{6.10}$$

One more integration yields the electrostatic potential. Selecting $\phi(x = 0) = 0$, as reference:

$$\phi_o(x) \simeq -\frac{qN_A x_p^2}{2\epsilon} \qquad \text{in p-QNR: } x \leq -x_p \tag{6.11}$$

$$\phi_o(x) \simeq \frac{qN_A}{2\epsilon}(x^2 + 2x_p x) \qquad \text{in SCR: } -x_p \leq x \leq 0 \tag{6.12}$$

$$\phi_o(x) \simeq -\frac{qN_D}{2\epsilon}(x^2 - 2x_n x) \qquad \text{in SCR: } 0 \leq x \leq x_n \tag{6.13}$$

$$\phi_o(x) \simeq \frac{qN_D x_n^2}{2\epsilon} \qquad \text{in n-QNR: } x_n \leq x \tag{6.14}$$

x_n and x_p are obtained by demanding that two conditions be met. First, overall charge neutrality must be satisfied since each of the starting pieces of semiconductor were charge neutral and we have not allowed any particles to escape. This implies that:

$$qN_A x_p = qN_D x_n \tag{6.15}$$

Second, the total electrostatic potential difference across the entire structure must equal the built-in potential derived above in Eq. 6.2. This results in:

$$\phi_o(x_n) - \phi_o(-x_p) = \frac{qN_D x_n^2}{2\epsilon} + \frac{qN_A x_p^2}{2\epsilon} = \phi_{bi} \tag{6.16}$$

Solving in these two equations for x_n and x_p, we get:

$$x_n = \sqrt{\frac{2\epsilon N_A \phi_{bi}}{qN_D(N_D + N_A)}} \tag{6.17}$$

$$x_p = \sqrt{\frac{2\epsilon N_D \phi_{bi}}{qN_A(N_D + N_A)}} \tag{6.18}$$

The total extent of the SCR, the sum $x_n + x_p$ is:

$$x_{SCR} = \sqrt{\frac{2\epsilon(N_D + N_A)\phi_{bi}}{qN_A N_D}} \tag{6.19}$$

The maximum electric field occurs at the metallurgical junction and is easily found from Eq. 6.8 or 6.9 at $x = 0$:

$$|\mathcal{E}_{o,max}| = \sqrt{\frac{2qN_A N_D \phi_{bi}}{\epsilon(N_D + N_A)}} \tag{6.20}$$

Exercise 6.1: *Consider a pn junction like that of Fig. 6.5 with* $N_A = 10^{18}\ cm^{-3}$ *and* $N_D = 10^{16}\ cm^{-3}$. *At room temperature, calculate a) the built-in potential of this junction, b) the extent of the depletion region on the p-side, c) the extent of the depletion region on the n-side, and d) the magnitude of the electric field at the metallurgical junction.*

a) To calculate the built-in potential, we use Eq. 6.2:

$$\phi_{bi} = \frac{kT}{q} \ln \frac{N_D N_A}{n_i^2} = 0.026\ V \ln \frac{10^{18}\ cm^{-3} \times 10^{16}\ cm^{-3}}{(1.1 \times 10^{10}\ cm^{-3})^2} = 0.84\ V$$

b) The extent of the depletion region on the p-side is given by Eq. 6.18:

$$x_p = \sqrt{\frac{2\epsilon N_D \phi_{bi}}{qN_A(N_D + N_A)}} = \sqrt{\frac{2 \times 11.7 \times 8.85 \times 10^{-14}\ F/cm \times 10^{16}\ cm^{-3} \times 0.84\ V}{1.6 \times 10^{-19}\ C \times 10^{18}\ cm^{-3}(10^{16} + 10^{18}\ cm^{-3})}} = 3.3\ nm$$

c) To obtain the extent of the depletion region on the n-side, the easiest is to use the charge neutrality condition of Eq. 6.15:

$$x_n = x_p \frac{N_A}{N_D} = 3.3\ nm \frac{10^{18}\ cm^{-3}}{10^{16}\ cm^{-3}} = 0.33\ \mu m$$

d) The electric field at the metallurgical junction can be obtained from Eq. 6.20 but is faster to use Eq. 6.8 or Eq. 6.9 at $x = 0$:

$$|\mathcal{E}_{max}| = \frac{qN_D}{\epsilon} x_n = \frac{1.6 \times 10^{-19}\ C \times 10^{16}\ cm^{-3} \times 3.3 \times 10^{-5}\ cm}{11.7 \times 8.85 \times 10^{-14}\ F/cm} = 5.1 \times 10^4\ V/cm$$

It is interesting to examine the electrostatics of the junction as a function of the relative doping level on both sides. If $N_A = N_D$, Eq. 6.15 implies that $x_p = x_n$. If $N_A > N_D$, then $x_p < x_n$, and vice versa. Notice that in this instance, Eqs. 6.11 and 6.14 indicate that $|\phi_o(-x_p)| < \phi_o(x_n)$. In other words, as the doping level becomes more asymmetric, the depletion region extends preferentially over the lowly-doped side and most of the electrostatic potential build up occurs there.

In practical devices, it is very common to have very asymmetric junctions in which one side is much more heavily doped than the other. For example, in a $p^+ - n$ junction, $N_A \gg N_D$. In a case like this, Eqs. 6.17-6.19 imply that:

$$x_n \simeq x_{SCR} \simeq \sqrt{\frac{2\epsilon\phi_{bi}}{qN_D}} \gg x_p \tag{6.21}$$

The space-charge region is essentially confined to the lowly doped region *and* its thickness is basically determined by the doping level on the lowly doped side. As a consequence of this, the maximum electric field is given by:

$$|\mathcal{E}_{o,max}| \simeq \sqrt{\frac{2qN_D\phi_{bi}}{\epsilon}} \tag{6.22}$$

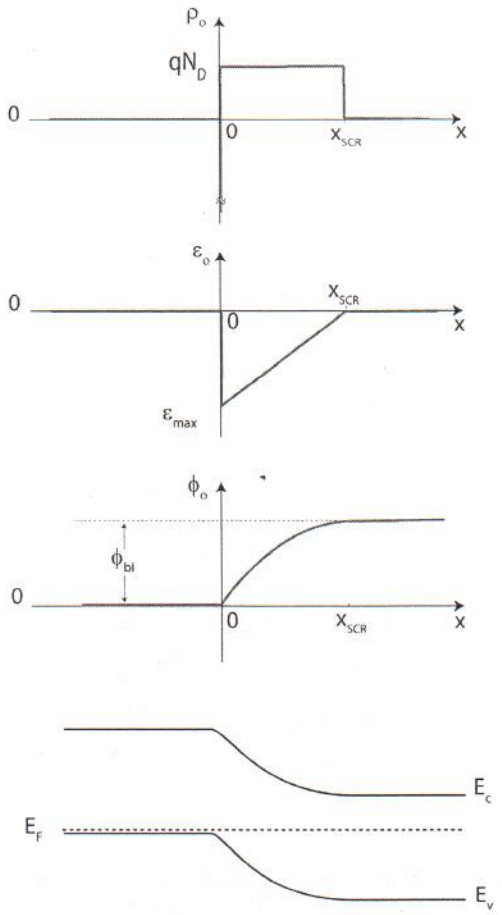


Figure 6.6: Electrostatics of a p^+-n junction in thermal equilibrium. From top to bottom: volume charge density, electric field, electrostatic potential, and energy band diagram.

and is also determined by the doping level on the lowly doped side alone. Additionally, the electrostatic potential drops mainly on the lowly doped side. From Eqs. 6.11 and 6.14, we find that:

$$\phi_o(x_n) \simeq \phi_{bi} \gg |\phi_o(-x_p)| \tag{6.23}$$

This situation is depicted in Fig. 6.6. In an asymmetric junction, the lowly-doped side dominates the electrostatics. As the doping level of the lowly-doped side increases, $x_{SCR} \downarrow$ but $|\mathcal{E}_{o,max}| \uparrow$. There are important consequences associated with this that we will discuss later on.

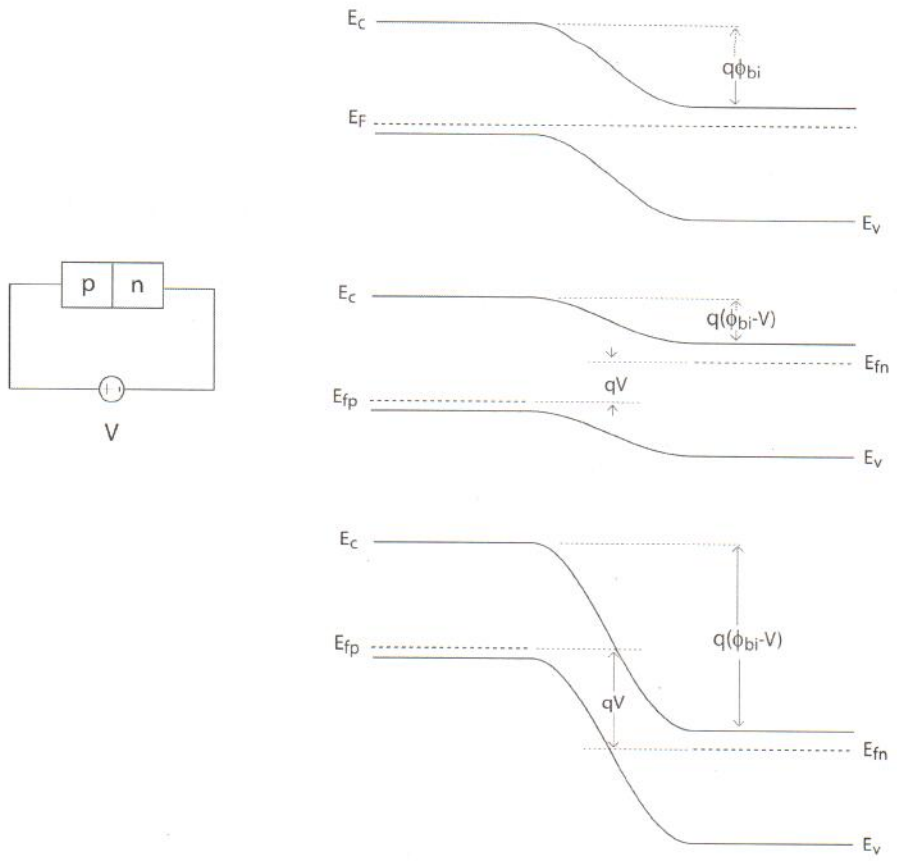


Figure 6.7: Energy band diagram of a PN junction in equilibrium (top), forward bias (middle) and reverse bias (bottom). Only the majority quasi-Fermi levels in the QNR's are drawn.

6.3 Current-voltage characteristics of the ideal PN diode

A PN diode displays rectifying behavior. With a forward bias across (p-side positive with respect to the n-side), current readily flows. Under reverse bias, only a trickle current flows and the device can be considered essentially an open circuit. This section explain this behavior of the PN diode and builds models for its current-voltage characteristics. We start by examining the modifications to the electrostatics that follows the application of a voltage.

6.3.1 Electrostatics under bias

Applying a voltage across the two terminals of a PN junction modifies the electrostatics under equilibrium that we described in the previous section. In a first pass analysis, it is easier to understand the impact of a voltage by initially neglecting any influence of the current that flows. We can later on apply appropriate corrections to this simple analysis.

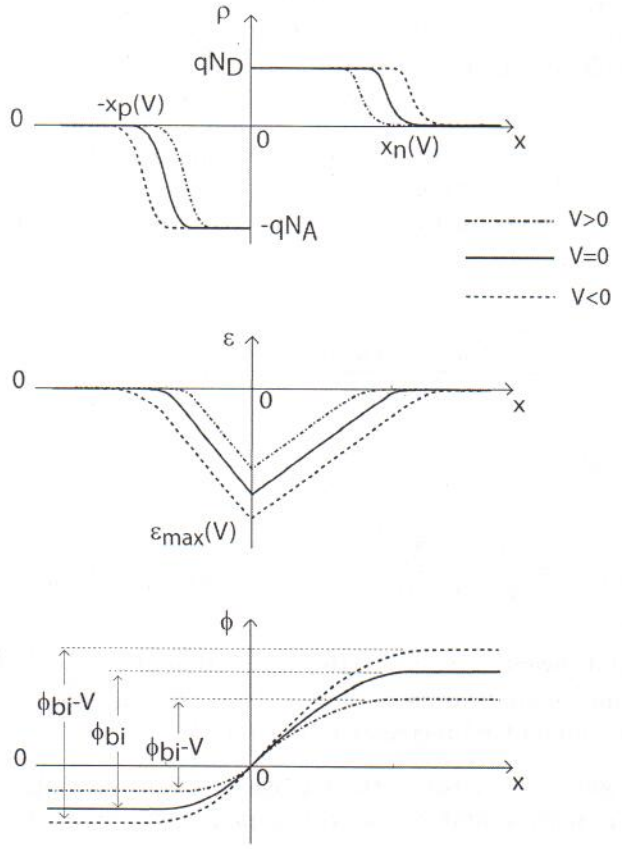


Figure 6.8: Volume charge density, electric field and electrostatic potential of a PN junction as a function of bias.

As we will see in Chapter 7, the ohmic contacts of the diode provide a way for the terminals of the battery to grab on the *majority carrier Fermi levels* on either side of the junction. This means that in forward bias, for example, the quasi-Fermi level for electrons on the n side is raised with respect to the quasi-Fermi level for holes on the p side by an energy difference qV, where V is the voltage of the battery. The contrary happens in reverse bias. Since the quasi-neutral regions contain lots of majority carriers, their concentrations are hard to upset in a significant way. In consequence, as the majority quasi-Fermi levels split, the entire band diagram on each side is dragged along with them. In each QNR, the bands remain flat reflecting our assumption that if no current flows, no ohmic drop occurs in the QNR's. This is illustrated in Fig. 6.7. In this figure, only the majority quasi-Fermi levels in the QNR's are drawn. The rest of the diagram has to wait until we have had a chance to discuss the physics of the current.

As a consequence of applying a voltage, the total potential difference across the junction is upset from its equilibrium value ϕ_{bi}. In forward bias, it is reduced to $\phi_{bi} - V$. In reverse bias, it is increased to $\phi_{bi} - V$ (in reverse bias V is negative). This in turn demands that the dipole

of charge that appears across the junction be modified too. In forward bias it has to be reduced, in reverse bias it has to be increased. Since the doping levels cannot be changed, the only way to accomplish this is to change the thickness of the SCR. The changes of the electrostatics of the PN junction upon the application of a bias are summarized in Fig. 6.8.

Qualitatively, the electrostatics of the PN junction under bias do not differ from the thermal equilibrium situation. All the equations that we have derived above basically apply in forward and reverse bias if we substitute ϕ_{bi} by $\phi_{bi} - V$. In particular, the extent of the space-charge region becomes:

$$x_{SCR}(V) = \sqrt{\frac{2\epsilon(N_D + N_A)(\phi_{bi} - V)}{qN_AN_D}} = x_{SCR}(V = 0)\sqrt{1 - \frac{V}{\phi_{bi}}} \tag{6.24}$$

The maximum electric field is:

$$|\mathcal{E}_{max}(V)| = \sqrt{\frac{2qN_AN_D(\phi_{bi} - V)}{\epsilon(N_D + N_A)}} = |\mathcal{E}_{max}(V = 0)|\sqrt{1 - \frac{V}{\phi_{bi}}} \tag{6.25}$$

both valid in forward and reverse bias. With the application of a forward bias, the space-charge region shrinks and the maximum field decreases. A reverse bias produces the contrary effect, the SCR widens and the maximum field increases in magnitude.

If the junction is asymmetric, most of the voltage that is applied drops on the lowly doped side of the junction. The space-charge region widens preferentially into the lowly doped side.

6.3.2 I-V characteristics: qualitative discussion

A PN junction is a *minority-carrier device*. It is the behavior of the minority carriers that represents the bottleneck to current flow. Let us first discuss this qualitatively.

In thermal equilibrium, the net current everywhere along the diode is zero. Actually, the principle of detailed balance demands that separately, the hole and electron currents be zero at all locations. In general, carrier current is composed of drift and diffusion. In the QNR's, the carrier concentration is uniform and there is no field, so both drift and diffusion are zero separately. Inside the SCR, there is an electric field and the gradients of carrier concentrations are very steep. Here, drift has to balance diffusion precisely at every point for thermal equilibrium to hold. That means that the diffusion tendency of carriers is precisely counterbalanced by their drift inside the electric field. It is important to notice that the sign of the electric field is the correct one for this to happen.

The application of a voltage breaks this dynamic balance inside the SCR. This can be seen in Fig. 6.9. A forward bias reduces the electric field which produces a decrease in the drift component of the carrier current. In consequence, diffusion prevails over drift and there is a net electron flow from the n-side (where they are majority carriers) to the p-side (where they are

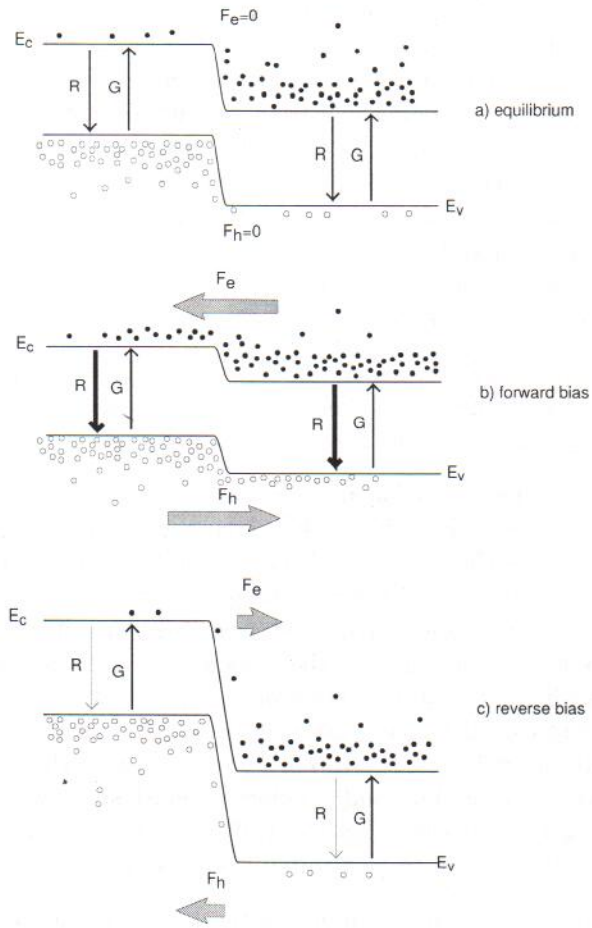


Figure 6.9: Current balance across SCR in thermal equilibrium (top), forward bias (middle), and reverse bias (bottom). In equilibrium, there is no net current crossing the SCR. In forward bias, minority carrier injection and recombination into the QNR's takes place. In reverse bias, minority carrier generation and extraction out of the QNR's dominates.

minority carriers). Similarly, holes flow from the p-side to the n-side. The term that is used to describe this process is *minority carrier injection*.

In reverse bias, the contrary happens. A negative voltage increases the electric field inside the SCR with the consequence that drift prevails over diffusion. As a result, electrons preferentially flow from the p-region (where they are minority carriers) to the n-region (where they are majority carriers). Similarly, holes flow from the n-region to the p-region. This receives the name of *minority carrier extraction*.

The energy view provides an intuitive picture of what happens under bias, as sketched in Fig. 6.9. In thermal equilibrium electrons are in equilibrium among themselves across the whole diode. There are many more electrons on the n-side than on the p-side. However, there is a steep energy barrier (of a height $q\phi_{bi}$) separating these two regions. Some electrons on the

n-side do have enough energy to get injected into the p-region. But this is precisely balanced out by electrons on the p-side "falling down" or getting extracted into the n-region. The net flow of electrons is precisely zero in thermal equilibrium. In forward bias, this energy barrier is reduced. In consequence, many more electrons on the n-region now have enough energy to overcome the reduced barrier and get injected to the p-side. Electron extraction, on the other hand, is diminished because the electric field inside the SCR has been reduced. Conversely, in reverse bias, the barrier is increased, and, relative to equilibrium, a smaller number of electrons have enough energy to overcome the barrier and get injected into the p-side. Electron extraction from the p-side to the n-side is also increased, since the electric field inside the SCR is of a higher magnitude. A similar series of events is experienced by the holes.

- But this cannot be the entire picture. What happens to the electrons injected into the p-side? How about the holes injected into the n-side? We see net current flowing through the SCR, but how is this current supported through the quasi-neutral regions?

Under forward bias, minority carrier injection raises the total minority carrier concentration above the equilibrium value in the QNRs. The presence of excess minority carriers drives up the recombination rate above the value that balances the thermal generation rate. The resulting net recombination extinguishes all the injected minority carriers and makes room for more to be injected. There is then a continuous carrier flow from the majority carrier side to the minority carrier side where they recombine. Most of the recombination takes place within a few diffusion lengths of the junction where the injection takes place. Hence, at sufficient distance from the edge of the SCR, the minority carrier concentration returns to the equilibrium value. The resulting minority carrier profile is sketched at the top of Fig. 6.10. Since the supply of majority carriers in the QNR is virtually inexhaustible (and is being replenished anyway by the ohmic contact) it is the dynamics of minority carrier flow that controls the magnitude of the current. It is because of this that the PN junction is said to be a minority-carrier device.

In reverse bias, minority carrier extraction is the net consequence of drift dominating over diffusion in the SCR. The minority carrier concentrations, consequently, drop in both QNRs below their equilibrium values. This breaks again the balance between thermal generation and recombination with net generation taking place to supply the minority carriers that are extracted. A few diffusion lengths away from the edge of the SCR, the minority carrier concentrations return to the equilibrium values. The resulting carrier profiles are sketched at the bottom of Fig. 6.10. A continuous current path is established with the generated minority carriers in the QNR's being extracted by the SCR and coming out the ohmic contact to the outside world.

Does our qualitative understanding explain the rectifying characteristics of the PN junction? Surely. Let us just focus on electrons. In forward bias, the minority carrier current is proportional to the number of electrons that have enough energy to overcome the energy barrier. Since this barrier is reduced as $\sim -V$, the number of electrons with enough energy to overcome it goes as $\sim e^{qV/kT}$, and so does the current. In reverse bias, once the barrier is made large enough, electron injection from the n-side to the p-side is completely suppressed. The current is then entirely supported by electron generation in the p-QNR. This reaches its maximum value when the electron concentration drops to zero. The electron generation rate cannot increase any further beyond that and the current saturates.

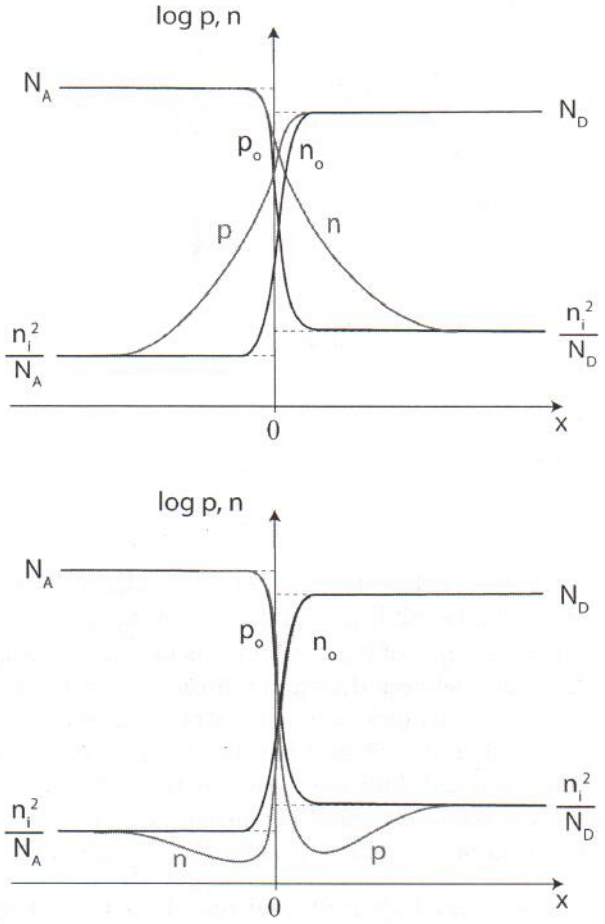


Figure 6.10: Sketch of carrier profiles across pn diode in forward bias (red lines, top) and reverse bias (blue lines, bottom). In both cases the black lines represent the equilibrium carrier concentrations.

6.3.3 I-V characteristics: quantitative models

Armed with the understanding produced in the previous section, we can now lay out a strategy for developing a model for the current through the PN junction. We must focus our work on the minority carriers in each quasi-neutral region (in the ideal PN diode, we neglect minority carrier generation and recombination in the SCR; we will add this later in Sec. 6.6.2). The calculation is basically a four-step process. First, we must calculate how many minority carriers are present at the respective edge of the QNR, $n(-x_p)$ and $p(x_n)$ (see Fig. 6.11). Second, we must derive expressions for the velocity at which minority carriers are injected into or extracted from the QNR's, $v_e(-x_p)$ and $v_h(x_n)$. Third, we compute the current that minority carriers support in their flow across the SCR, $J_e(-x_p)$ and $J_h(x_n)$. Finally, we add the contributions of electrons and holes and multiply by the area of the junction.

Let us start with the minority carrier boundary conditions across the SCR. These can be

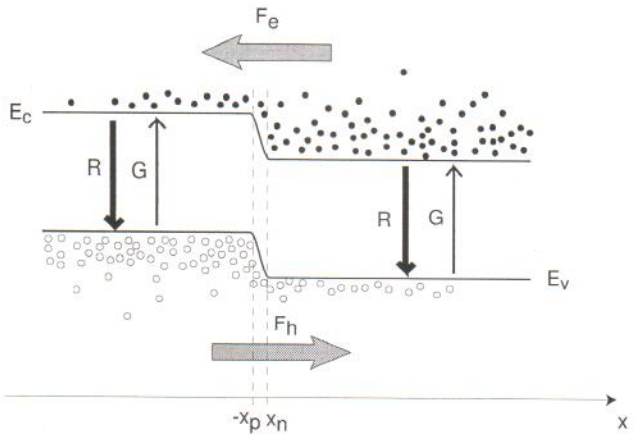


Figure 6.11: Sketch of carrier flux across PN junction in forward bias indicating the edges of the SCR where current calculations are performed.

derived in various ways. An approach is to focus on the balance between drift and diffusion currents inside the SCR. In thermal equilibrium, this balance is perfect at every point in the SCR and $J_e = J_h = 0$. As argued above, out of equilibrium this balance is broken inside the SCR and $J_e \neq 0$ and $J_h \neq 0$. The difference between drift and diffusion currents results in a net current for each type of carrier. If we were to estimate how far out of balance drift and diffusion get inside the SCR for typical diode operation, we will find (see Problem 6.13) that it is actually very little. Relative to their magnitude, drift and diffusion never get to differ very much inside the SCR of a PN diode under bias. Only a relatively small perturbation is required to support the current that originates on the generation or recombination of minority carriers in the QNR's.

The realization of this has a very important and practical consequence that we can exploit here. In Section 4.5.2 we learned that in thermal equilibrium, there is a relationship between the *ratio* of the carrier concentrations at two points and the *difference* of the electrostatic potential between those two same points. This is the Boltzmann relation. Eq. 4.58 captures this in its most general form for electrons. Across the SCR of a PN diode, the Boltzmann relation can be written, for electrons and holes, as:

$$\phi_o(x_n) - \phi_o(-x_p) = \phi_{bi} = \frac{kT}{q} \ln \frac{n_o(x_n)}{n_o(-x_p)}$$

$$\phi_o(x_n) - \phi_o(-x_p) = \phi_{bi} = -\frac{kT}{q} \ln \frac{p_o(x_n)}{p_o(-x_p)}$$

Under bias, strictly speaking these equations do not apply. However, since for each carrier type, the net carrier current is much smaller than its drift and diffusion components, then in the scale of these currents, $J_e \simeq 0$, and $J_h \simeq 0$. Hence, to first order, the Boltzmann relations should be reasonably accurate across the SCR. They can be written in the following form:

$$\phi_{bi} - V \simeq \frac{kT}{q} \ln \frac{n(x_n)}{n(-x_p)} \tag{6.26}$$

$$\phi_{bi} - V \simeq -\frac{kT}{q} \ln \frac{p(x_n)}{p(-x_p)} \tag{6.27}$$

If low-level injection conditions prevail on both sides, $n(x_n) \simeq N_D$ and $p(-x_p) \simeq N_A$ (see a discussion of the impact of high-level injection on Section 6.6.6). Using Eq. 6.2 and solving for the carrier concentrations on the minority carrier side, we get:

$$n(-x_p) \simeq \frac{n_i^2}{N_A} \exp \frac{qV}{kT} \tag{6.28}$$

$$p(x_n) \simeq \frac{n_i^2}{N_D} \exp \frac{qV}{kT} \tag{6.29}$$

The excess carrier concentrations (defined in Eq. 3.44 and 3.45) are:

$$n'(-x_p) \simeq \frac{n_i^2}{N_A} (\exp \frac{qV}{kT} - 1) \tag{6.30}$$

$$p'(x_n) \simeq \frac{n_i^2}{N_D} (\exp \frac{qV}{kT} - 1) \tag{6.31}$$

These boundary conditions have all the features that we have qualitatively discussed earlier. For electrons, for example, for $V = 0$, $n'(-x_p) = 0$. If $V \gg kT/q$, then $n'(-x_p) \gg n_i^2/N_A$ and grows exponentially with V. If, on the other hand, $V \ll -kT/q$, then $n'(-x_p)$ saturates to $\simeq -n_i^2/N_A$. With the derivation of the minority carrier boundary conditions, the first step of the diode current computation is complete.

Exercise 6.2: *Consider a pn junction identical to that of Exercise 6.1. At room temperature and with an applied forward bias of $V = 0.6\ V$, calculate: a) the potential difference between the n-side and the p-side; b) the excess minority carrier concentration at the edges of the depletion region.*

a) In Exercise 6.2, we computed the potential difference between the n-side and the p-side of this junction in thermal equilibrium. We found this to be $\phi_{bi} = 0.84\ V$. Under forward bias, this potential difference is reduced by an amount equal to the applied voltage. Hence, we have:

$$\phi(x_n) - \phi(-x_p) = \phi_{bi} - V = 0.84 - 0.6 = 0.24\ V$$

b) The excess electron concentration at the edge of the depletion region on the p-type side can be obtained from Eq. 6.30:

$$n'(-x_p) \simeq \frac{n_i^2}{N_A}(\exp\frac{qV}{kT} - 1) = \frac{1.1 \times 10^{20}\ cm^{-6}}{10^{18}\ cm^{-3}}(\exp\frac{0.6}{0.0259} - 1) = 1.3 \times 10^{12}\ cm^{-3}$$

The excess hole concentration at the edge of the depletion region on the n-tpe side can be obtained from the combination of Eqs. 6.30 and 6.31:

$$p'(x_n) = \frac{N_A}{N_D}n'(-x_p) = \frac{10^{18}\ cm^{-3}}{10^{16}\ cm^{-3}}1.3 \times 10^{12}\ cm^{-3} = 1.3 \times 10^{14}\ cm^{-3}$$

In both cases, the excess minority carrier concentrations are much smaller than the corresponding doping levels. We can then conclude that this diode, under these conditions, is in low-level injection.

Let us now focus on the second step, the calculation of the velocity at which minority carriers are injected or extracted into or out of the QNRs. Fortunately, we have already done all the work in Ch. 5. In that chapter, we studied minority carrier flow in several quasi-neutral low-level injection situations. In particular, we studied carrier flow in semiconductor bars illuminated by a thin beam of light that produced a sheet of carrier generation. Among other things, we calculated the velocity at which minority carriers diffuse away from the generation point.

Our problem here shares a lot with those already studied in Ch. 5. In a PN junction in forward bias, the reduction of potential barrier across the SCR results in minority carrier injection into the QNRs. Once inside the QNR, the behavior of the minority carriers is identical to that of the illuminated bar. In particular, if the boundary conditions are similar, the velocity at which minority carriers diffuse into the QNR should be identical to that of photogenerated carriers.

Here we are studying a "long" diode in which the QNRs are much longer than the respective minority carrier diffusion lengths. In Section 5.6.1 we found that the excess minority carrier profile in a long bar follows an exponential decay in space where the minority carrier diffusion length was the characteristic length of the exponential. We expect similar excess carrier profiles in the QNR's in the long diode, as sketched in Fig. 6.12. These profiles then follow the following analytical relations:

$$n'(x) = n'(-x_p)\exp\frac{x + x_p}{L_e} \qquad \text{in p-QNR: } x \leq -x_p \tag{6.32}$$

$$p'(x) = p'(x_n)\exp\frac{-x + x_n}{L_h} \qquad \text{in n-QNR: } x \geq x_n \tag{6.33}$$

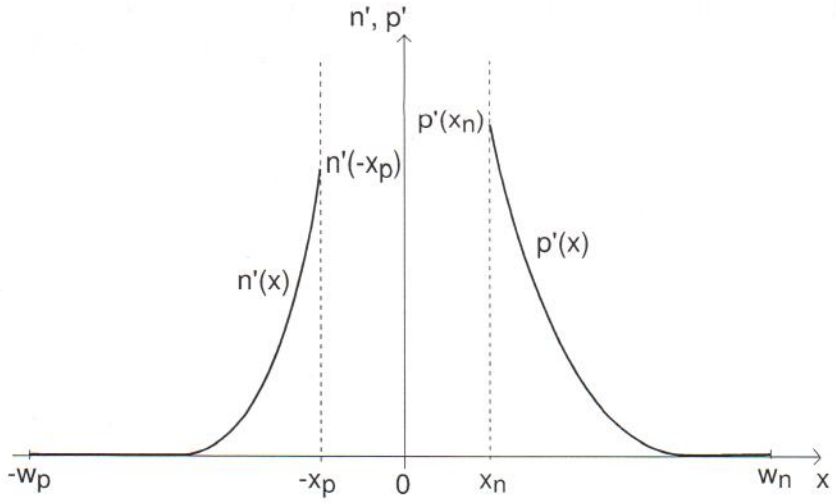


Figure 6.12: Excess minority carrier profiles in quasi-neutral regions of "long" diode under forward bias.

Also in 5.6.1 we found that the velocity at which photogenerated holes diffuse away from the generation point in a long n-type bar was given by $v_h^{diff} = D_h/L_h$ (Eq. 5.69). This result also applies to the PN junction. The velocity at which minority carriers diffuse into the QNRs from their respective edges is then given by:

$$v_e^{diff}(-x_p) = -\frac{D_e}{L_e} \tag{6.34}$$

$$v_h^{diff}(x_n) = \frac{D_h}{L_h} \tag{6.35}$$

where D_e and L_e pertain to minority carrier electrons in the p-type region, and D_h and L_h are associated with minority carrier holes in the n-type region. This is the result that we were looking for in step 2 of our calculation of the PN-diode current.

Step 3 consists of computing the minority carrier current injected into each QNR. Since we have both the minority carrier concentrations and the carrier velocities at the edges of the QNRs, we simply use the general results given by Eqs. 4.18 and 4.20. We get, respectively.

$$J_e(-x_p) \simeq -qv_e^{diff}(-x_p)n'(-x_p) = q\frac{n_i^2}{N_A}\frac{D_e}{L_e}(\exp\frac{qV}{kT} - 1) = J_{es}(\exp\frac{qV}{kT} - 1) \tag{6.36}$$

$$J_h(x_n) \simeq qv_h^{diff}(x_n)p'(x_n) = q\frac{n_i^2}{N_D}\frac{D_h}{L_h}(\exp\frac{qV}{kT} - 1) = J_{hs}(\exp\frac{qV}{kT} - 1) \tag{6.37}$$

where, for convenience, we have defined J_{es} and J_{hs} as the pre-exponential factors.

The total current density through the junction, step 4, is the sum of these two current densities. This is consistent with our assumption that no carriers recombine inside the SCR. Adding Eqs. 6.36 and 6.37, we get:

$$J = (J_{es} + J_{hs})(\exp \frac{qV}{kT} - 1) \tag{6.38}$$

Denoting the preexponential factor as the *saturation current density*, J_s, we can rewrite 6.38 as:

$$J = J_s(\exp \frac{qV}{kT} - 1) \tag{6.39}$$

which exhibits a classical rectifying behavior.

Exercise 6.3: *Consider a pn diode identical to that of Exercises 6.1 and 6.2. Now consider that the p- and n-regions are "long" from the minority carrier point of view. At room temperature and with an applied forward bias of $V = 0.6\ V$, estimate the current density flowing through this diode.*

To estimate the diode current density we need to separately compute the current density injected into each quasi-neutral region and add them up. In Exercise 6.2, we already computed the excess minority carrier concentrations at the edges of the depletion region under the same conditions as given here. The best way to proceed is then to estimate the velocity at which excess carriers are injected into each QNR using Eqs. 6.34 and 6.35. For this, we need to know the respective diffusion coefficients and diffusion lengths.

To estimate the diffusion coefficients, we can use the mobility data graphed in Fig. 4.3 or the corresponding fits given in Appendix E. For electrons in the p-type region, we obtain $\mu_e = 344\ cm^2/V.s$ and $D_e = 9.0\ cm^2/s$. For holes in the n-type region, we obtain $\mu_h = 476\ cm^2/V.s$ and $D_h = 12.4\ cm^2/s$.

To estimate the diffusion lengths, we can either use Fig. 5.17 or we can obtain the corresponding minority carrier lifetimes and then use Eq. 5.54 to obtain the diffusion length. We proceed this way. From Appendix E, we estimate the minority carrier lifetime in the p-type region as $\tau_e = 2.3 \times 10^{-6}\ s$ and in the n-type region as $\tau_h = 1.3 \times 10^{-4}\ s$. This yields diffusion lengths that are $L_e = 4.6 \times 10^{-3}\ cm$ and $L_h = 4.0 \times 10^{-2}\ cm$, respectively.

Now, using Eqs. 6.34 and 6.35 we estimate the velocity at which minority carriers are injected into each QNR. For electron injection into the p-QNR, we have:

$$|v_e^{diff}(-x_p)| = \frac{D_e}{L_e} = \frac{9.0\ cm^2/s}{4.6 \times 10^{-3}\ cm} = 2.9 \times 10^3\ cm/s$$

For hole injection into the n-QNR, we have:

$$v_h^{diff}(x_n) = \frac{D_h}{L_h} = \frac{12.4\ cm^2/s}{4.0 \times 10^{-2}\ cm} = 3.1 \times 10^2\ cm/s$$

The current density injected into each QNR can be obtained using Eqs. 6.36 and 6.37. For the p-QNR, we have:

$$\begin{aligned} J_e(-x_p) &= q\,|v_e^{diff}(-x_p)|\,n'(-x_p) = 1.6 \times 10^{-19}\ C \times 2.9 \times 10^3\ cm/s \times 1.3 \times 10^{12}\ cm^{-3} \\ &= 6.0 \times 10^{-4}\ A/cm^2 \end{aligned}$$

Similarly, for the n-QNR:

$$\begin{aligned} J_h(x_n) &= q\,v_h^{diff}(x_n)\,p'(x_n) = 1.6 \times 10^{-19}\ C \times 3.1 \times 10^2\ cm/s \times 1.3 \times 10^{14}\ cm^{-3} \\ &= 6.5 \times 10^{-3}\ A/cm^2 \end{aligned}$$

The total diode current density is the sum of these two components:

$$J = J_e(-x_p) + J_h(x_n) = 7.1 \times 10^{-3}\ A/cm^2$$

If the junction has an area A, the currents injected into the p- and n-regions are respectively given by:

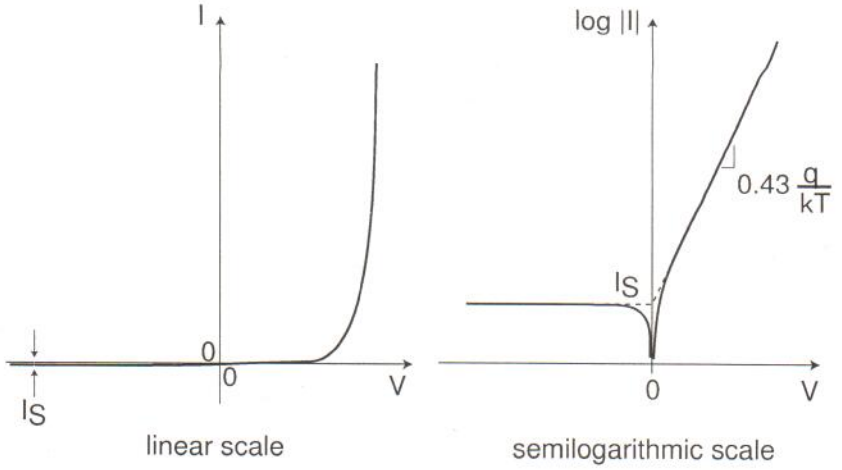


Figure 6.13: Sketch of I-V characteristics of ideal PN junction in linear and semilogarithmic scales.

$$I_p = AJ_e(-x_p) = AJ_{es}(\exp\frac{qV}{kT} - 1) \tag{6.40}$$

$$I_n = AJ_h(x_n) = AJ_{hs}(\exp\frac{qV}{kT} - 1) \tag{6.41}$$

and the total current is:

$$I = I_p + I_n = I_s(\exp\frac{qV}{kT} - 1) \tag{6.42}$$

with $I_s = AJ_s$. I_s is known as the diode *saturation curent.*

Eq. 6.42 is the result that we were looking for. It gives the current through the diode as a function of the applied voltage. It is all in terms of physically meaningful parameters and fundamental constants. In forward bias, the current grows exponentially, as we expected. In reverse bias, the current saturates to $-I_S$, also as expected. This behavior is plotted in Fig. 6.13. Since I_S tends to be small, it is common to graph the $I - V$ characteristics in a semilogarithmic scale, as shown on the right of Fig. 7.14. In a semilog scale, the forward bias current appears as a straight line with a slope of $(q/kT)\log\ e = 0.43q/kT$, which only depends on temperature. The higher the temperature, the lower the slope. It is common to refer to this slope in units of mV/dec, that is, the voltage required (in mV) to produce an increase in current of a factor of ten. At room temperature, this figure is 60 mV/dec, a handy number to remember.

Since the boundary conditions 6.30 and 6.31 capture both forward and reverse bias and the profiles 6.32 and 6.33 are valid in both forward and reverse bias, the equations that we have derived for the I-V characteristics, Eq. 6.38-6.42 apply in both forward and reverse bias.

Before we proceed further, it is of value to examine the relative magnitude of the two terms of

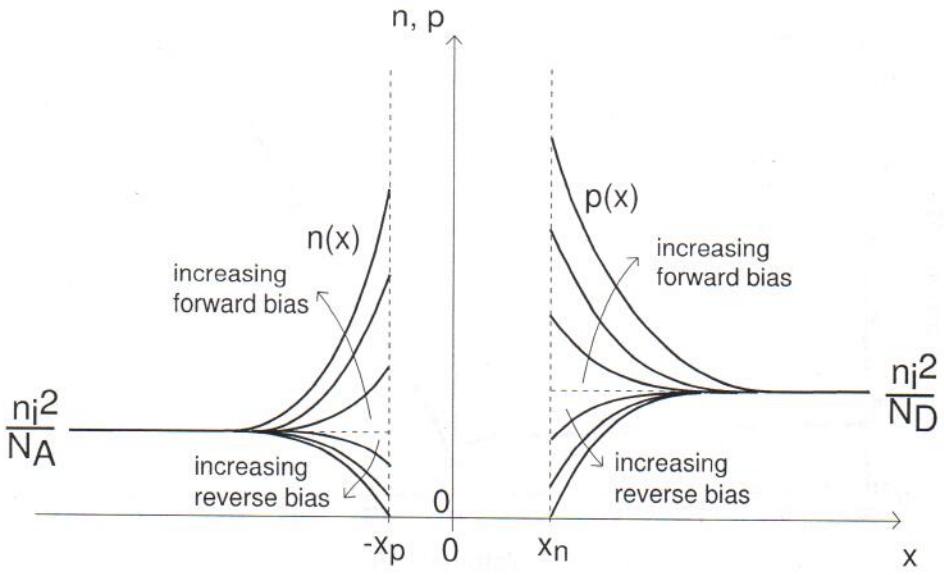


Figure 6.14: Evolution of minority carrier profiles in the quasi-neutral regions of a "long" diode as a function of bias.

the diode current that arise from each of the diode regions. If the doping levels are similar, these two terms are of comparable order of magnitude. The diffusion velocity for electrons is higher than that of holes so, for identical doping levels, the electron injection in the p-type region is somehow higher than hole injection into the n-type region. However, the current is inversely proportional to the doping level. So, in an asymmetric junction, current injection into the lowly-doped region tends to dominate.

It is important to emphasize here that the rectifying behavior of the diode is entirely due to the boundary condition at the edges of the SCR. This is clearly seen in the sketch of the minority carrier profiles as a function of bias shown in Fig. 6.14. In forward bias, the higher the voltage, the higher the concentration of minority carriers that are injected and contribute to current via their recombination with majority carriers. The relationship is of a Boltzmann type. In reverse bias, our model predicts that the minority carrier concentrations at the edges of the SCR drop to negligible values when $|V| \gg kT/q$. Since the minority carrier concentrations cannot be any smaller than zero, that means that when $|V| \gg kT/q$, the generation rate everywhere in the QNRs saturates and the reverse current also saturates. The reverse current is $-I_s$ and it is usually a very small value (typically $I_s \sim 1\ pA/cm^2$ in Si PN junctions at room temperature). [1]

Several key dependencies are observed in Eq. 6.42. First of all, in forward bias, the current increases exponentially with V. If plotted in a semilogarithmic scale, as sketched in Fig. 6.13, the forward characteristics appear as a straight line with a slope of $0.43q/kT$. It actually is common in the literature to refer to the inverse of this slope, whose ideal value at room temperature is

[1] Obviously, the carrier concentration cannot drop to zero as it would not be possible to support a finite current this way. This is a limit of this simplified formulation. In reality, the carrier concentrations at the edges of the SCR drop to very small values that are largely bias insensitive. This produces current saturation.

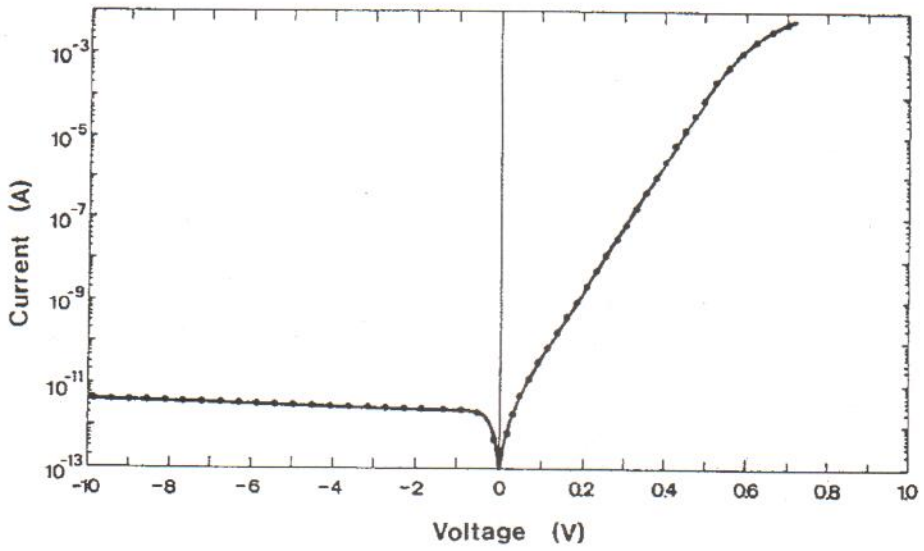


Figure 6.15: Experimental I-V characteristics of Si PN junction at room temperature [*Ref?*].

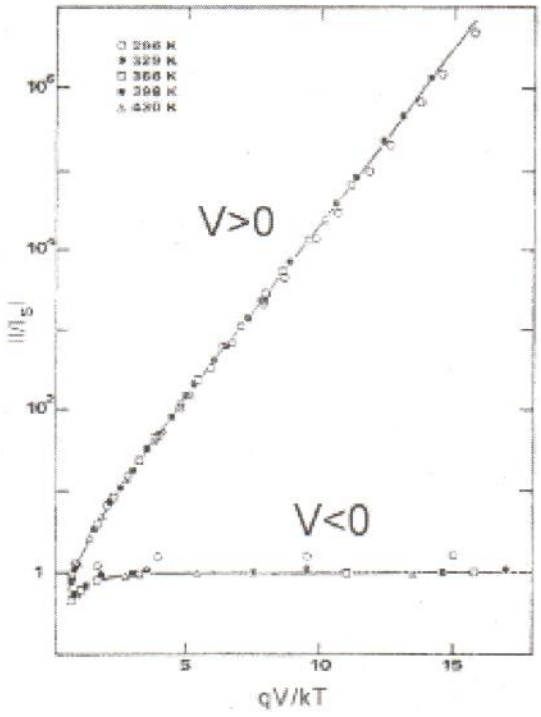


Figure 6.16: Normalized experimental I-V characteristics of Si PN junction at several temperatures [adapted from *P. Cappelletti et al., J. Appl. Phys. 57, 646, (1985)*].

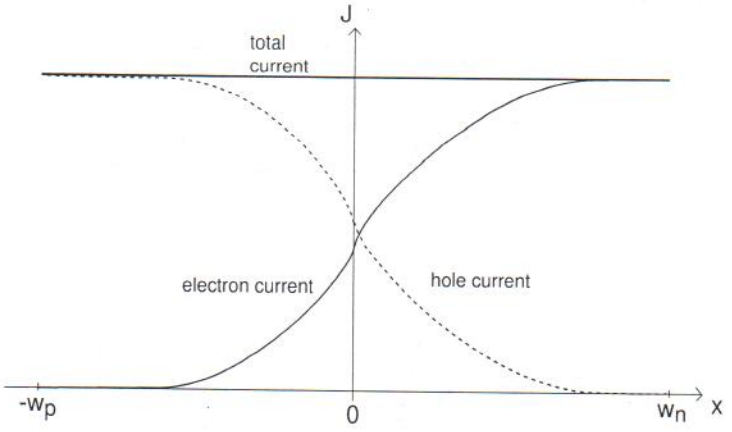


Figure 6.17: Sketch of current profiles across PN diode in forward bias.

60 $mV/decade$. The extrapolation of this straight line to $V = 0$ is I_s. This is also the magnitude of the current at a sufficient reverse bias.

These key features are actually seen in experimental PN junctions. Fig. 6.15 shows a semilogarithmic plot of the I-V characteristics of a PN diode at room temperature. These characteristics display a nearly ideal behavior. In forward bias, the current increases exponentially with a slope of about 62 mV/dec. The extrapolation of this current to $V = 0$ V gives $I_s \simeq 1$ pA. In reverse bias, the current is rather independent of bias and equal to about 4 pA, which is very close to I_s. The deviation from the perfect straight line at high forward bias is due to series resistance. The slight slope of the current in reverse bias is due to generation in the space-charge region. These non-idealities are discussed below.

Fig. 6.16 shows I/I_s vs. qV/kT for a PN junction at several temperatures. For all temperatures, the data follows Eq. 6.42 fairly well. Note in particular that the extrapolation of the forward and reverse characteristics meet at $V = 0$, as predicted by simple theory.

It is instructive to think about the electron and hole current profiles across the entire PN junction. They are shown in Fig. 6.17 for forward bias. A similar picture can be constructed in reverse bias. The current profiles are similar to those of the example of Section 5.6.1. The total current is the sum of the electron injected current to the p-region plus the hole-injected current to the n-region. Across the narrow space-charge region, we assume that there is negligible recombination and the currents do not change. In the quasi-neutral regions, the minority carrier currents fall off exponentially from the edge of the space-charge region. Far away from the junction, the total current is entirely supported by the majority carriers. At the end of the p region, the total current is entirely due to holes, while at the end of the n region, it is completely carried by electrons. The majority carrier currents flow through their respective ohmic contacts to the outside world.

Now that we have computed the carrier currents at all points in a PN junction, we are in a position to complete the energy band diagrams out of equilibrium by sketching the quasi-Fermi levels everywhere. This is done in Fig. 6.18 for a junction where both QNR's are much longer

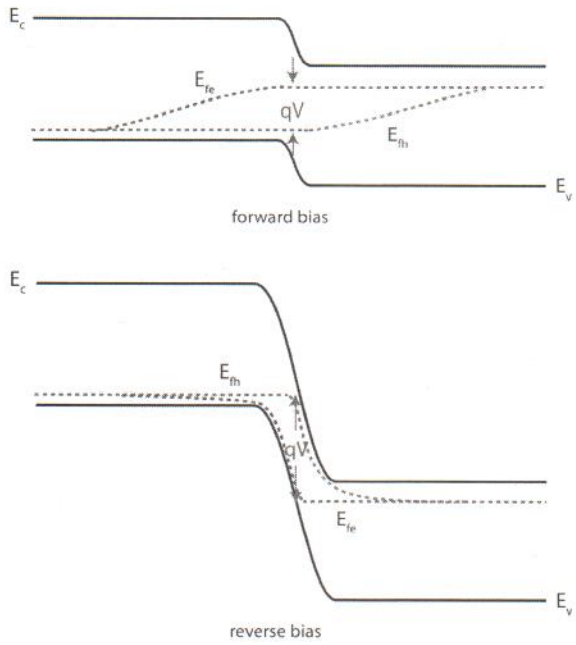


Figure 6.18: Schematic energy band diagram in forward and reverse bias showing the location of the quasi-Fermi levels.

than the respective minority carrier diffusion lengths. The assumption that we made inside the SCR that the carriers are in quasi-equilibrium implies that the quasi-Fermi levels are flat there. In the QNR's, the further away we move from their edge with the SCR, the closer we are to reestablishing equilibrium. A few diffusion lengths away, in fact, the carrier concentrations recover their equilibrium values. In consequence, the minority carrier quasi-Fermi levels approach the majority carrier ones and eventually merge into one as quasi-equilibrium is obtained sufficiently far away.

There is an interesting and relevant implication of quasi-equilibrium existing for electrons and holes separately and the quasi-Fermi levels being flat across the space-charge region. In Ch. 4, we saw that the np product was exponentially dependent on the difference of the quasi-Fermi levels (see Eq. 4.91). For the SCR of a PN junction, this difference is precisely qV. As a consequence, we can write:

$$np = n_i^2 \exp \frac{qV}{kT} \tag{6.43}$$

From this equation, it is trivial to derive the boundary conditions 6.28 and 6.29.

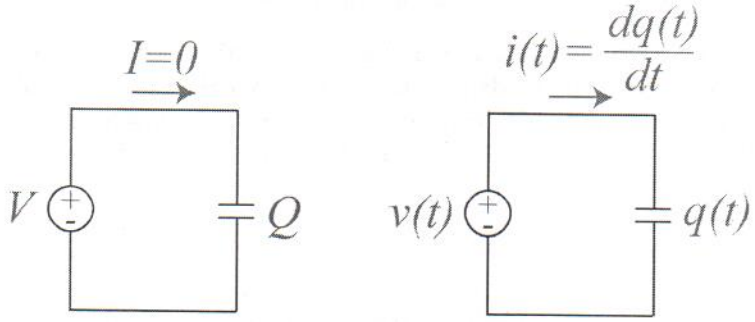


Figure 6.19: Behavior of capacitor to static voltage (left) and a dynamic voltage (right). In the static situation, the current is zero. In the dynamic situation, current flows as the plates of the capacitor charge or discharge.

6.4 Charge-voltage characteristics of ideal PN diode

At first sight, it would seem that a formulation of the current-voltage characteristics of a diode is all we could possibly need to analyze circuits containing diodes. However, the current-voltage characteristics only provide a simple *static* description of the device. In most applications, we also care about the *dynamic* operation of the diode. In power electronics, for example, the diode voltage often changes abruptly from a large reverse bias to a forward bias and back. In many analog applications, it is common to apply to the diode a rapidly varying small-signal voltage on top of a bias. The frequency of the small-signal can be quite high. To correctly describe the operation of the diode in situations like these, we need to keep track of the charge stored inside. What does this mean?

Let us think for a moment about about a parallel plate capacitor, as sketched in Fig. 6.19. The static current-voltage characteristics of an ideal parallel plate capacitor are very simple: for any voltage, the current is zero (left in Fig. 6.19). However, the dynamic behavior of the capacitor is much more interesting (right in Fig. 6.19). When the voltage changes, current flows to charge the plates of the capacitor. For example, if the voltage on one plate rises with respect to the other, positive charge is deposited to that plate and the same amount of positive charge is removed from the other plate. This produces a flow of charge through the circuit which results in current. [2] The current that flows through the circuit is given by the familiar expression $i(t) = dq(t)/dt$, where $q(t)$ is the charge stored in the capacitor (one plate is charged with $+q$ while the other plate is charged with $-q$.) If we were to describe the behavior of the capacitor by only its static current-voltage characteristics ($I = 0$ for any V), we would clearly miss its interesting dynamic features!

A similar situation occurs in virtually every semiconductor device. Just like in a parallel plate capacitor, in a semiconductor device there is stored charge with a magnitude that depends on the applied voltage. If the voltage changes, the stored charge has to change and that produces a current flow. This needs to be accounted for in the analysis of circuits that contain semiconductor

[2] In reality we know that current should be described by the retrograde movement of electrons. So, for the picture on the right of Fig. 6.19, when the voltage rises on one plate with respect to the other, electrons are taken away from the first plate and the same amount of electrons are added to the second plate.

devices. This is what the charge-voltage characteristics describe.

For the case of the PN diode, what charge is it that we need to account for? Two types, actually. One is associated with the depletion region across the metallurgical junction and the second one is associated with minority carrier injection into the quasi-neutral regions. The next two sections discuss this in more detail and derive models for these two types of stored charge.

6.4.1 Depletion charge

In Sec. 6.3.1 we saw that the extent of the depletion region in a PN diode is modulated by the applied voltage. A reverse bias enlarges the depletion layer and a forward bias shrinks it. The depletion layer consists of two distinct regions: on the n side of the junction there is a region that is positively charged while on the p-side of the junction there is another region that is negatively charged. Overall charge neutrality demands that the absolute amount of charge in both regions be the same. This clearly resembles the two plates of a capacitor.

This situation is depicted in Fig. 6.20. This graph shows the charge distribution in the depletion region of a diode at a certain bias V and the change in this charge when the bias increases by a small amount ΔV. For simplicity, let us assume that both V and ΔV are positive, though this is not necessary. At first sight, there seems like there is a sign discrepancy in the charge with respect to the situation in a parallel plate capacitor. The positive side of the diode (the p side) has negative charge in its share of the depletion region while the negative side of the diode (the n side) has positive charge in its portion of the SCR. This is not a problem because what matters is how this charge changes when V changes. So, if V increases by an amount ΔV, we know that the depletion region shrinks on both sides. That means that a sliver of positive charge must be deposited at the edge of the SCR on the p-side and the same amount of negative charge has to be deposited at the other edge of the SCR on the n side, as sketched in the bottom figure. The *change in charge* that occurs as a result of the *change in voltage* clearly has the right sign and current flows in the circuit with the expected sign.

From the discussion on the electrostatics of the PN diode of Sec. 6.3.1 we can immediately write the charge on either side of the depletion region as a function of the applied voltage. For the p-side, for example:

$$Q_j(V) = -AqN_A x_p = -A\sqrt{\frac{2q\epsilon N_D N_A(\phi_{bi} - V)}{N_D + N_A}} \tag{6.44}$$

The negative sign reflects the fact that the charge on the p side is due to exposed acceptors. We obtain an identical expression with a positive sign if we evaluate the charge on the n-side of the SCR. This expression includes the area of the junction so the units of Q_j are Coulombs.

This expression can be rewritten in the following form:

$$Q_j(V) = Q_j(V = 0)\sqrt{1 - \frac{V}{\phi_{bi}}} \tag{6.45}$$

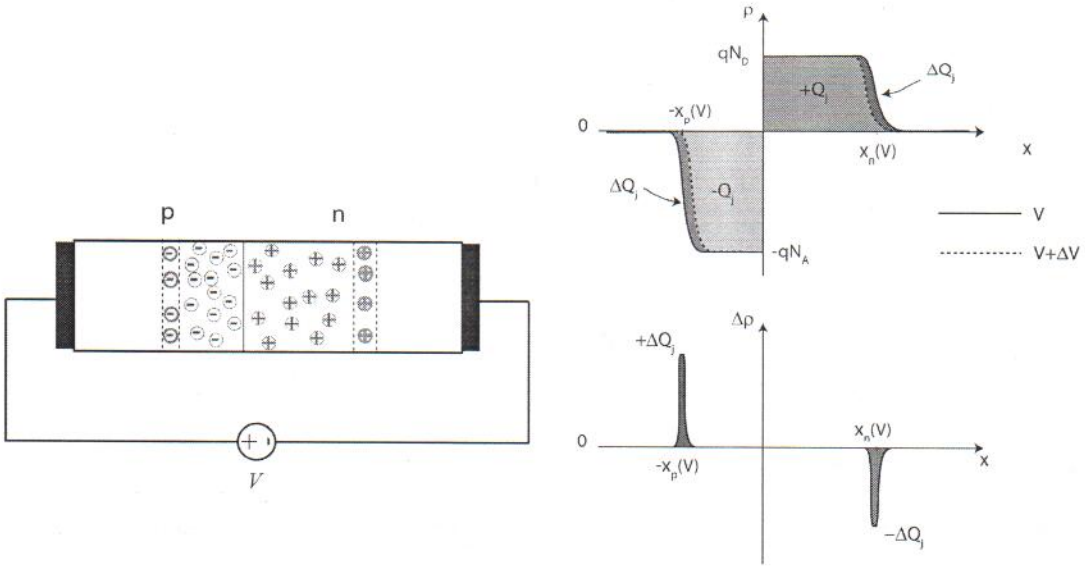


Figure 6.20: Sketch of volume charge density in the SCR of a PN junction at a certain bias V and its change upon the application of an additional small voltage ΔV.

As forward bias increases, the depletion region shrinks and Q_j is reduced in magnitude. As the reverse bias increases, on the other hand, Q_j increases in magnitude. The voltage dependence of Q_j is square root. This is sketched in Fig. 6.21. Notice how, in spite of the negative sign in Q_j, its derivative with V (the capacitance, discussed below) is positive, as it should be.

6.4.2 Minority carrier charge

The second type of stored charge in a PN diode has to do with minority carrier injection. In a forward biased PN junction, the two quasi-neutral regions are flooded with excess minority carriers. Quasi-neutrality is preserved by piling up a nearly identical concentration of majority carriers at every point in space, as sketched in Fig. 6.22. The required majority carriers are easily delivered by the battery through the respective ohmic contact. As studied in detail in Chapter 5, quasi-neutrality is a consequence of the relatively high conductivity of extrinsic semiconductor regions. Even a very small electric field can produce a substantial majority carrier current in a moderately doped region. The current circulates until the majority carrier distribution matches the minority carrier profile. This occurs in a time scale of the dielectric relaxation time, a very short time.

The modification of the majority carrier concentrations relative to the doping level is very small if low-level injection conditions are maintained. The assumption of low-level injection demands in a p-type region, for example, that $n' \ll p_o$. Since quasi-neutrality is satisfied when $n' \simeq p'$, then $p' \ll p_o$ and $p \simeq p_o$. Therefore, for most purposes, we can still assume that the majority carrier concentration equals the doping level everywhere.

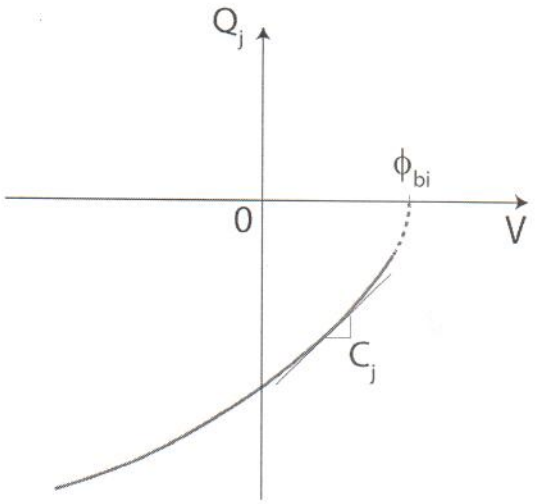


Figure 6.21: Sketch of the charge-voltage characteristics associated with the space-charge region of a PN diode.

Now let us consider what happens to the carrier profiles if we increase the forward voltage by a small amount, as sketched in Fig. 6.22. A slightly higher forward voltage injects more minority carriers at the edge of the SCR. This results in an increased overall minority carrier concentration everywhere. The majority carrier concentration must match up also this small increase in total charge in order to preserve quasi-neutrality. We again see that the change in voltage results in a change in stored charge, just like in a capacitor.

Let us look in more detail at the n-side. If the forward voltage increases by a small amount, the p-side injects holes to the n-QNR. This results in an overall increase ΔQ_{hn} in the positive charge stored in the n-region. In order to preserve quasi-neutrality, the ohmic contact of the n-type region delivers the same amount of electrons with a total charge $\Delta Q_{en} = -\Delta Q_{hn} \equiv -\Delta Q_n$. In this way, the net charge in the n-QNR is back to zero. To maintain quasi-neutrality point by point, the spatial charge distribution of additional holes matches very closely that of extra electrons, as indicated in Fig. 6.22. The additional holes injected into the n-QNR were provided by the positive side of the battery. The electrons were provided by the negative side of the battery consistent with the sign notation for a capacitor. The moment the minority carrier profile has reached its final distribution at the new forward voltage, the flow of charge bound to get stored ceases. The normal diode current continues to flow at a slightly higher value.

A similar sequence of events takes place on the p-QNR. If the forward voltage changes, the negative terminal of the battery delivers negative charge in the form of excess electrons, ΔQ_{ep}, which is matched point by point by positive charge delivered by the positive terminal of the battery in the form of excess holes through the ohmic contact to the p-QNR, that is, $\Delta Q_{ep} = -\Delta Q_{hp} \equiv -\Delta Q_p$. The p-QNR also displays the same capacitive effect as the n-QNR. Furthermore, both "capacitors" are located in parallel directly across the terminals of the PN junction.

We are now in a position to develop a model for the stored charge in the n- and p-QNRs. These can be obtained simply by integrating the respective excess minority carrier concentrations:

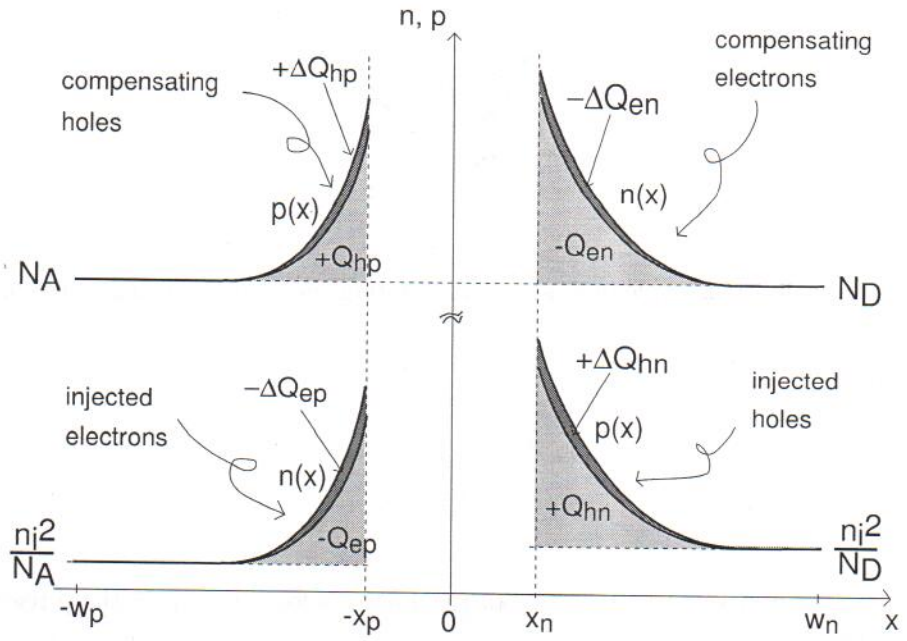


Figure 6.22: Sketch of minority and majority carrier profiles in a PN junction in forward bias. Both sides are long in comparison with the minority carrier diffusion lengths. In response to a small change in the voltage applied to the junction, both the minority and the majority carrier concentrations change by the same amount point-by-point in the quasi-neutral regions.

$$Q_p = qA\int_{-w_p}^{-x_p} n'(x)dx \tag{6.46}$$

$$Q_n = qA\int_{x_n}^{w_n} p'(x)dx \tag{6.47}$$

In these expressions, Q_p refers to the stored charge in the p-type QNR and and Q_n to that of the n-type QNR.

We derived expressions for the excess carrier concentrations in the two QNRs for our ideal PN diode (Eqs. 6.32 and 6.33). Plugging these expressions in and integrating, we get:

$$Q_p = qAL_e n'(-x_p) \tag{6.48}$$

$$Q_n = qAL_h p'(x_n) \tag{6.49}$$

We can rewrite these expressions in terms of the minority carrier current density injected into the respective regions by using Eqs. 6.36 and 6.37 on the one hand, and 6.34 and 6.35 on the other hand, to obtain:

$$Q_p = A\frac{L_e^2}{D_e}J_e(-x_p) \tag{6.50}$$

$$Q_n = A\frac{L_h^2}{D_h}J_h(x_n) \tag{6.51}$$

Finally, using the definitions made in Eqs. 6.40 and noting that L^2/D is equal to the minority carrier lifetime, we obtain:

$$Q_p = \tau_e I_p \tag{6.52}$$

$$Q_n = \tau_h I_n \tag{6.53}$$

where τ_e and τ_h are, respectively, the minority carrier lifetimes for electrons in the p-region and holes in the n-region.

The total stored charge is simply, then:

$$Q_d = Q_p + Q_n = \tau_e I_p + \tau_h I_n = Q_s(\exp\frac{qV}{kT} - 1) \tag{6.54}$$

This expression and the individual components of the current given by Eqs. 6.52 and 6.53 make good physical sense. If we think about this in forward bias, when Q_d can become sizable, the current flowing through the diode is entirely due to recombination. For each region of the diode, the characteristic time constant of this process is the recombination lifetime τ. Then the amount of stored charge has to be τI. Conversely, a charge Q_d recombining in a time span of τ gives rise to a current $I = Q_d/\tau$.

An important result obtained here is that the minority carrier charge stored in a diode goes as $Q_d \propto \exp(qV/kT) - 1$. This reflects the boundary conditions for excess carrier concentrations at the edges of the SCR. In forward bias, this charge increases exponentially. In reverse bias, this charge saturates to a very small value that is in most cases negligible.

Minority carrier storage in PN diodes is an issue that affects its switching characteristics in a very significant way. It we attempt to switch a diode from a forward bias state to a reverse state, the transitory is not complete until all the minority carrier charge that exists in forward bias is disposed of. This can take substantial time. Similarly, if we try to switch a PN diode from reverse bias to forward bias, a lot of current needs to be delivered to provide the minority carrier charge that needs to be stored. This also takes time. We analyze the situation in detail in Appendix AT6.3. The absence of minority carrier storage in Schottky diodes, as we will see in Ch. 7, makes these devices preferable for switching applications.

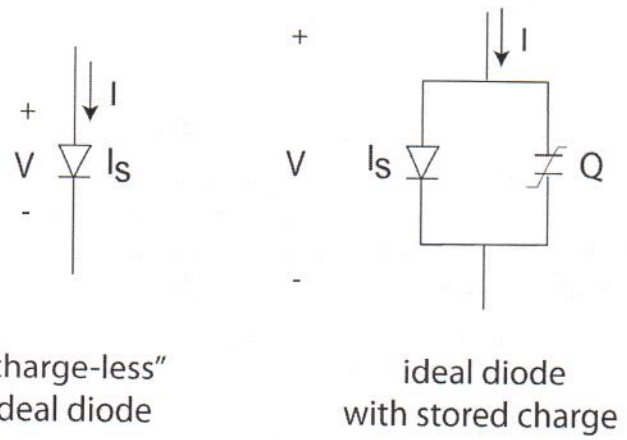


Figure 6.23: Left: symbol for charge-less ideal diode. I_s represents its saturation current. Right: large-signal equivalent circuit model of ideal diode.

6.5 Equivalent circuit models of the ideal PN diode

An equivalent circuit model for a semiconductor device is a circuit-like description of its behavior that uses elemental components such as resistors, capacitors, current sources, voltage sources, and others. One of these elemental components in the equivalent circuit models toolbox is actually a "charge-less" ideal diode. This is a two-terminal element that exhibits rectifying current-voltage characteristics as given by Eq. 6.42 but does not hold stored charge. It is then more akin to a non-linear resistor with rectifying characteristics. A symbol for a charge-less ideal diode is shown on the left of Fig. 6.23. A charge-less ideal diode is characterized by a single parameter, its saturation current I_s which is often written right next to the symbol. In the chapters that follow, we will use this element when constructing equivalent circuit models for the Schottky diode, the MOSFET and the bipolar junction transistor.

Our concept of ideal diode in this chapter does involve stored charge, as discussed in detail in the previous section. Therefore, to the charge-less ideal diode, we need to add a storage element in parallel. The reason for the parallel configuration is the fact that the same voltage that feeds the current through the diode also drives charge storage. This combination of elements is the most elemental equivalent circuit model for an ideal diode and is shown on the right of Fig. 6.23. The charge storage element is usually represented by a capacitor-like symbol with a characteristic broken line across, as indicated in the figure. This is to suggest that this is a non-linear element. As we now know, in an ideal diode, charge storage exhibits a complex voltage dependence given by the sum of the depletion region capacitance, Eq. 6.44, and the minority carrier charge, Eq. 6.54.

The equivalent circuit model on the right of Fig. 6.23 is referred to as a *large-signal equivalent circuit*. The two branches in parallel represent the direct diode current that flows upon the application of a voltage plus the current needed to charge or discharge the diode if the voltage changes. So, in a general dynamic situation, the terminal diode current is given by:

$$I = I_s(\exp\frac{qV}{kT} - 1) + \frac{dQ}{dt} \tag{6.55}$$

To keep the notation simple, I, V, and Q here denote instantaneous time dependent variables.

In many applications, it is also important to develop a good understanding of the small-signal behavior of semiconductor devices. This refers to their response to small excursions around its bias point. Devices are used this way in many analog and communications applications. The small-signal behavior of a device can be obtained by linearizing the current and charge equations.

For the diode, starting with the current-voltage characteristics, we apply a voltage across that consists of a bias V, plus a small signal v. The resulting current is:

$$I + i = I_s[\exp\frac{q(V+v)}{kT} - 1] \tag{6.56}$$

If v is small in the scale of kT/q, then we can linearize the exponential by selecting the first two terms of the Taylor series expansion:

$$I + i \simeq I_s[\exp\frac{qV}{kT}(1 + \frac{qv}{kT}) - 1] = I + \frac{q(I+I_s)}{kT}v \tag{6.57}$$

From the small-signal point of view, the ideal diode looks like a resistor of value:

$$r_d = \frac{kT}{q(I+I_s)} \tag{6.58}$$

r_d is referred to as the *dynamic resistance* of the diode. Its meaning is graphically illustrated in Fig. 6.24. The dynamic resistance is the inverse slope of the I-V characteristics of the diode at the selected bias point. Its value, therefore, depends on the bias that has been selected. In forward bias, if $V \gg kT/q$, then:

$$r_d \simeq \frac{kT}{qI} \tag{6.59}$$

This is an important result. The dynamic resistance of a diode is inversely proportional to the diode current. The dependence is completely fundamental, that is, it is completely specified by fundamental constants and the diode temperature. For a diode biased at $I = 1\ mA$ at room temperature, for example, $r_d = 26\ \Omega$. This fundamental relationship suggests that the value of r_d is independent of the design of the diode; a modern diode and a diode from the early days of microelectronics both present an identical dynamic resistance if biased at the same current level. Of course, this is to the extent that they both exhibit I-V characteristics that approach the ideal behavior.

We turn our attention now to the charge-voltage characteristics. It is interesting to realize that, using the chain rule, the second term in Eq. 6.55 can be written as:

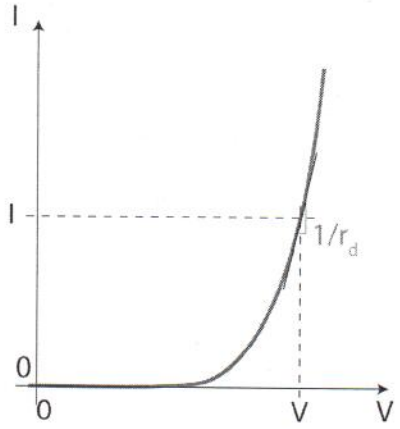


Figure 6.24: Illustration of dynamic resistance of a PN diode. The dynamic resistance characterizes the slope of the I-V characteristics of the diode at a given bias point. Its value depends on the selected bias point. For small excursions of voltage around this bias point, the dynamic resistance captures reasonably well the response of the diode.

$$\frac{dQ}{dt} = \frac{dQ}{dV}\frac{dV}{dt} = C\frac{dV}{dt} \tag{6.60}$$

where C has the units of capacitance. This suggests an alternative, but completely equivalent description of charge storage in a diode (and other semiconductor devices) in terms of a diode capacitance defined as:

$$C(V) = \frac{dQ}{dV} \tag{6.61}$$

This equation is written in a way to emphasize the fact that a device capacitance is in general bias dependent.

Describing charge storage in terms of capacitance has the advantage that only the variables I and V are needed to completely specify the situation. Also, Eq. 6.60 suggests that the absolute value of Q does not matter but it is dQ/dV, or capacitance, what really counts for the dynamics of a device.

Since in a pn diode there are two components of Q, the diode capacitance also contains two terms. Associated with charge storage in the SCR there is a capacitance that is called *depletion capacitance* or *junction capacitance*, C_j. We can easily evaluate an expression for it. If we go back to Eq. 6.44, for example, and use the capacitance definition of Eq. 6.61, we immediately obtain:

$$C_j(V) = A\sqrt{\frac{\epsilon q N_A N_D}{2(N_D + N_A)(\phi_{bi} - V)}} = \frac{C_j(V=0)}{\sqrt{1 - \frac{V}{\phi_{bi}}}} \tag{6.62}$$

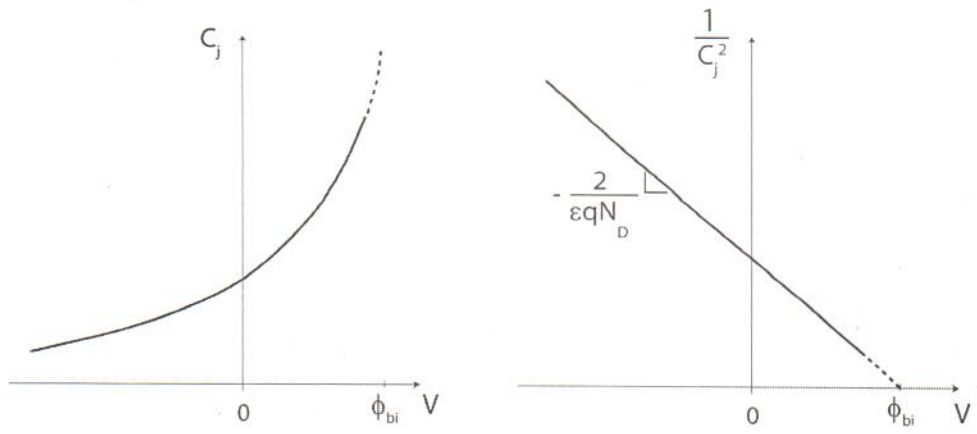


Figure 6.25: Sketch of C-V characteristics of a PN junction (left). Also plotted (right) is $1/C^2$ vs. V for an asymmetric p^+-n junction. The doping level on the lowly-doped side and the built potential can be determined from this plot.

This expression is sketched in Fig. 6.25. The shape of C_j with V makes good sense if we think of this as a parallel plate capacitor where the plates are separated by the thickness of the space-charge region, as Fig. 6.20 suggests. Starting from $V = 0$, as forward bias is applied, the space-charge region shrinks and the capacitance increases. If reverse bias is applied, the space-charge region widens and the capacitance decreases. In fact this way of thinking yields the expression of C_j very quickly if we write:

Alternatively:

$$C_j(V) = A\frac{\epsilon}{x_{SCR}(V)} \tag{6.63}$$

Substituting x_{SCR} from Eq. 6.24 into this expression immediately yields Eq. 6.62.

Measuring the capacitance-voltage characteristics often constitutes a powerful way to characterize semiconductor devices. We use an asymmetric $p^+ - n$ junction to illustrate this. In this case, Eq. 6.62 reduces to:

$$C_j(V) = A\sqrt{\frac{\epsilon q N_D}{2(\phi_{bi} - V)}} \tag{6.64}$$

Here we see that the capacitance is dominated by the doping on the lowly-doped side of the junction. In fact, $C_j(V)$ in Eq. 6.64 only depends on N_D and ϕ_{bi}. So C-V measurements can easily yield these two parameters. A particularly easy way to accomplish this is to plot $1/C_j^2$ as a function of V. From Eq. 6.64 for the p^+-n junction, we find:

$$\frac{1}{C_j^2} = \frac{2(\phi_{bi} - V)}{\epsilon q N_D A^2} \tag{6.65}$$

This expression shows that $1/C_j^2$ depends linearly on voltage with a negative slope that is a

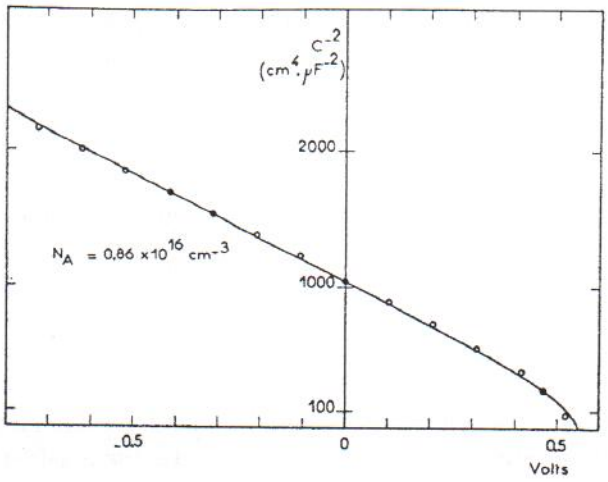


Figure 6.26: Experimental $1/C_j^2$ vs. V characteristics of a Si PN junction. From the slope of the straight line, the doping level on the lowly-doped side is determined [data from *Fortini et al., IEEE Trans. Electron Dev. ED-29, 1604 (1982)*].

function of N_D, the area of the diode and fundamental constants. Also, $1/C^2$ goes to zero as V goes to ϕ_{bi}. This behavior is sketched in Fig. 6.25. Fig. 6.26 provides an experimental demonstration of this. It is then clear that from C-V measurements, knowing the area of the diode we can determine the doping level on the lowly doped side of the diode and the built-in potential of the junction.

Exercise 6.4: *Consider a pn diode identical to that of Exercises 6.1, 6.2 and 6.3. The diode has a junction area of 10 μm^2. Estimate the junction capacitance associated with this diode at room temperature and under an applied forward bias of $V = 0.6$ V.*

This can be done in different ways. The fastest way is perhaps using the result of Exercise 6.1 where we estimated the extent of the SCR in this diode under zero bias. There we found that $x_{SCR} \simeq x_n = 0.33\ \mu m$. The junction capacitance at zero bias is then:

$$C_j(V=0) = A\frac{\epsilon}{x_{SCR}(V=0)} = 10 \times 10^{-8}\ cm^2 \times \frac{1.0 \times 10^{-12} F/cm}{0.33 \times 10^{-4}\ cm} = 3.1 \times 10^{-15}\ F = 3.1\ fF$$

The junction capacitance at a certain voltage can be obtained using Eq. 6.62:

$$C_j(V) = \frac{C_j(V=0)}{\sqrt{1-\frac{V}{\phi_{bi}}}} = \frac{3.1\ fF}{\sqrt{1-\frac{0.6}{0.84}}} = 5.8\ fF$$

The second source of stored charge in a diode is minority carrier charge. This can be described through a capacitance that is called the **diffusion capacitance**. We can obtain an expression for this by using Eq. 6.54 in Eq. 6.61:

$$C_d = \frac{dQ_d}{dV} = \frac{q}{kT} Q_s \exp\frac{qV}{kT} \tag{6.66}$$

The voltage dependence of C_d is a pure exponential. C_d goes to zero for reverse bias and increases exponentially with forward bias. If the forward bias voltage exceeds kT/q, then we can approximate C_d with:

$$C_d \simeq \frac{q}{kT} Q_d \tag{6.67}$$

or simply, q/kT times the total stored minority carrier charge in the diode.

The total capacitance of the diode is the sum of C_j and C_d:

$$C = C_j + C_d \tag{6.68}$$

Due to their unique voltage dependencies, in an ideal diode C_j dominates in reverse bias and small forward bias and C_d prevails in strong forward bias. This is illustrated in Fig. 6.27.

Exercise 6.5: *Consider a pn diode identical to that of Exercises 6.1-6.4. Note again that the p- and n-regions are "long" from the minority carrier point of view. At room temperature and with an applied forward bias of $V = 0.6\ V$, estimate a) the charge density stored in the two quasi-neutral regions, and b) the diffusion capacitance.*

a) In exercise 6.5 we obtained the magnitude of the current density injected into each QNR of this diode at this bias point. We also determined the minority carrier lifetimes. With these results, the fastest way to determine the charge stored in each QNR is using Eqs. 6.52 and 6.53. For the p-QNR, the stored charge per unit area is:

$$Q_p = \tau_e A J_e(-x_p) = 2.3 \times 10^{-6}\ s \times 10^{-7}\ cm^2 \times 6.0 \times 10^{-4}\ A/cm^2 = 1.4 \times 10^{-16}\ C$$

For the n-QNR, the stored charge per unit area is:

$$Q_n = \tau_h A J_h(x_n) = 1.3 \times 10^{-4}\ s \times 10^{-7}\ cm^2 \times 6.5 \times 10^{-3}\ A/cm^2 = 8.5 \times 10^{-14}\ C$$

b) The total carrier charge is the sum of these two and it is dominated by the n-QNR. The value is $Q_d = 8.5 \times 10^{-14}\ C$. The diffusion capacitance is then:

$$C_d \simeq \frac{q}{kT} Q_d = \frac{8.5 \times 10^{-14}\ C}{0.026\ V} = 3.3 \times 10^{-12}\ F = 3.3\ pF$$

When compared with the result obtained in Exercise 6.4, it is clear that the diffusion capacitance dominates at this bias point.

A small-signal equivalent circuit model of a device is a handy circuit-like description of its small-signal behavior. Small-signal equivalent circuits are a very intuitive way to grasp the behavior of a device from the small-signal point of view and are extensively used by circuit designers. By definition, a small-signal equivalent circuit model is a linearized version of the large-signal equivalent circuit model.

For an ideal PN diode, the small-signal equivalent circuit model looks like a pure RC parallel circuit as shown in Fig. 6.28. The resistance is the dynamic resistance that comes from the linearization of the ideal charge-less diode. The capacitance is the sum of the junction and diffusion capacitances and comes from the linearization of the charge storage element.

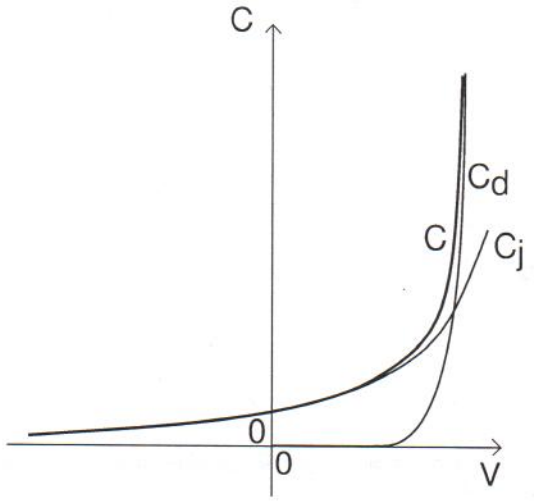


Figure 6.27: Sketch of diode capacitance and its components vs. diode voltage. Depletion capacitance dominates in reverse and small forward bias. For strong forward bias, diffusion capacitance dominates.

Figure 6.28: Small-signal equivalent circuit model of an ideal PN diode.

There is an additional kind of equivalent circuit model that is worth discussing. This is the model that is used to describe a device in the computer-aided design (CAD) of integrated circuits. These models tend to be based on physics but they have to pay attention to issues of computational efficiency and numerical convergence. These models also have to allow the parametrization of the device so that they can be matched to describe the characteristics of a given device. These issues are discussed in more detail in Appendix AT6.2 where a simple equivalent circuit model for circuit CAD for the PN diode is presented.

6.6 Non-ideal and second-order effects

In this section we proceed to discuss the most significant non-idealities and second-order effects in PN diodes. It is in this section that we undo several of the assumptions made in Sec. 6.1.

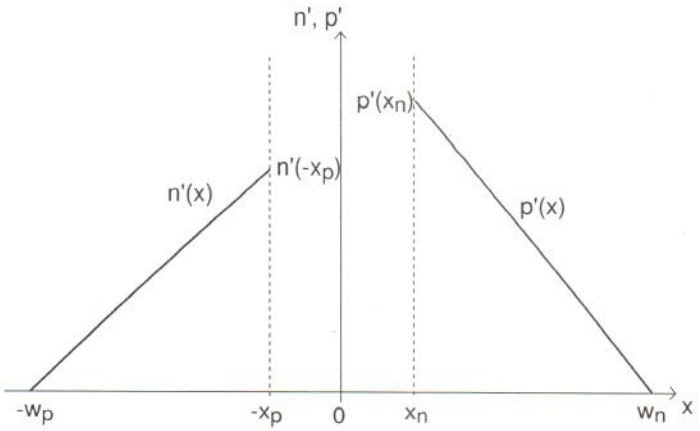


Figure 6.29: Schematic diagram of excess minority carrier profiles in quasi-neutral regions of a "short" diode with infinite surface recombination velocity at the ends.

6.6.1 Short diode

In our ideal PN diode we assumed that the quasi-neutral p- and n-regions were both much longer than their respective minority carrier diffusion lengths. Real diodes, depending on their design, can have QNRs that are long, short or comparable to the relevant diffusion length. Also, since there are two sides to a diode, several combinations are clearly possible. In this section we study the case of a diode in which both sides are much shorter than the respective diffusion lengths. We call this a "short diode." It is straightforward, though mathematically a bit more involved, to extend the analysis to more general situations.

The uniqueness of a short diode is that recombination and generation of minority carriers takes place at the surfaces of the QNRs (the ohmic contacts), as opposed to their bulk. The situation is similar to that of the "short" bar that we studied in Sec. 5.6.2 and the solutions that we obtained there can be readily used here. Other than accounting for this, the basic physics of the PN diode do not change.

If we think in terms of minority carrier injection, the excess carrier concentrations in a short diode are as depicted in Fig. 6.29. The ohmic contacts located at the surfaces of the diode (see Fig. 6.3) force the excess minority carrier concentrations there to zero. The boundary conditions at the edges of the SCR are set by the applied voltage, as determined above. In the absence of bulk recombination, the excess minority carrier profiles across the QNRs are straight lines and are therefore given by:

$$n'(x) = n'(-x_p)\frac{x + w_p}{-x_p + w_p} \qquad \text{in p-QNR: } x \leq -x_p \tag{6.69}$$

$$p'(x) = p'(x_n)\frac{-x + w_n}{w_n - x_n} \qquad \text{in n-QNR: } x \geq x_n \tag{6.70}$$

These linear minority carrier profiles lead to a minority carrier diffusion velocity at the edges of the QNRs that, in analogy with Eq. 5.77, are given by:

$$v_e^{diff}(-x_p) = -\frac{D_e}{w_p - x_p} \tag{6.71}$$

$$v_h^{diff}(x_n) = \frac{D_h}{w_n - x_n} \tag{6.72}$$

These are the velocities that should be used in the calculation of the diode current. Following a similar path as in Sec. 6.3.3, we obtain the two components of the diode current as:

$$I_p = qA\frac{n_i^2}{N_A}\frac{D_e}{w_p - x_p}(\exp\frac{qV}{kT} - 1) \tag{6.73}$$

$$I_n = qA\frac{n_i^2}{N_D}\frac{D_h}{w_n - x_n}(\exp\frac{qV}{kT} - 1) \tag{6.74}$$

Adding up I_p and I_n yields the expression of the diode I-V characteristics. The saturation current of the short diode is given by:

$$I_s \simeq qAn_i^2(\frac{1}{N_A}\frac{D_e}{w_p} + \frac{1}{N_D}\frac{D_h}{w_n}) \tag{6.75}$$

where we have further assumed that $w_p \gg x_p$ and $w_n \gg x_n$, as is commonly the case.

The ideal rectifying characteristics of the ideal PN diode are unchanged, whether the QNRs are long or short. This is because the peculiar voltage dependence of the diode emerges from the boundary conditions at the edges of the SCR and are independent of the design of the QNR.

The expression for the stored minority carrier charge in a short diode also needs to be revised. To be sure, the physics of stored charge is the same as in the long diode. The excess (or defect) of minority carrier concentrations imply a parallel modification in the majority carrier concentrations. All these charges need to be supplied by the contacts and remain stored in the diode.

With the excess minority carrier profiles being given by Eqs. 6.69 and 6.70, the minority carrier stored charge in each side of the diode is given by:

$$Q_p = qA\frac{1}{2}n'(-x_p)(w_p - x_p) = \tau_{tp}I_p \tag{6.76}$$

$$Q_n = qA\frac{1}{2}p'(x_n)(w_n - x_n) = \tau_{tn}I_n \tag{6.77}$$

where τ_{tp} and τ_{tn} are the transit times for minority carriers across the p- and n-QNRs, respectively. They follow the expressions:

$$\tau_{tp} = \frac{(w_p - x_p)^2}{2D_e} \tag{6.78}$$

$$\tau_{tn} = \frac{(w_n - x_n)^2}{2D_h} \tag{6.79}$$

This is an interesting result. It parallels what we obtained for the long diode in Eqs. 6.52 and 6.53. The charge stored in a QNR is given by the current injected into it times the relevant minority carrier time constant. For the long diode, this was the carrier lifetime. For the short diode, it obviously is the minority carrier transit time since carriers need to reach the ohmic contact before they can recombine. When the transport process across the QNR is diffusion and the profile is a straight line as in Fig. 6.29, we saw in Sec. 4.4 that the transit time is given by the square of the QNR length divided by two times the diffusion coefficient. This is exactly what we obtain here.

The total stored charge in a short diode is given by the sum of Eqs. 6.76 and 6.77:

$$Q_d = Q_p + Q_n = \tau_{tp} I_p + \tau_{tn} I_n = Q_s(\exp \frac{qV}{kT} - 1) \tag{6.80}$$

an expression that is functionally identical to that of the long diode.

Other than the need for using appropriate expressions for current and charge, the basic physics of the ideal diode are not modified in the short diode and the equivalent circuit models remain unchanged.

It is of interest, nevertheless, to sketch the energy band diagrams of a short diode. These are shown for forward and reverse bias in Fig. 6.30. Here we can see that the quasi-Fermi levels are split across the entire QNRs and only merge at the ohmic contacts. This reflects minority carrier concentrations that are out of equilibrium across the entire body of the semiconductor with the exception of the surfaces where the ohmic contacts force equilibrium to be maintained.

6.6.2 Space-charge generation and recombination

In the theory of the PN diode presented so far, we have neglected any generation and recombination in the space-charge region around the metallurgical junction. In our model, so far, quasi-neutral generation and recombination is all we need to explain the I-V characteristics experimentally observed in PN diodes.

It is not uncommon, in practice, to see PN diodes that do not quite follow such an ideal behavior. Fig. 6.31, for example, shows the I-V characteristics of a diode in which the ideal

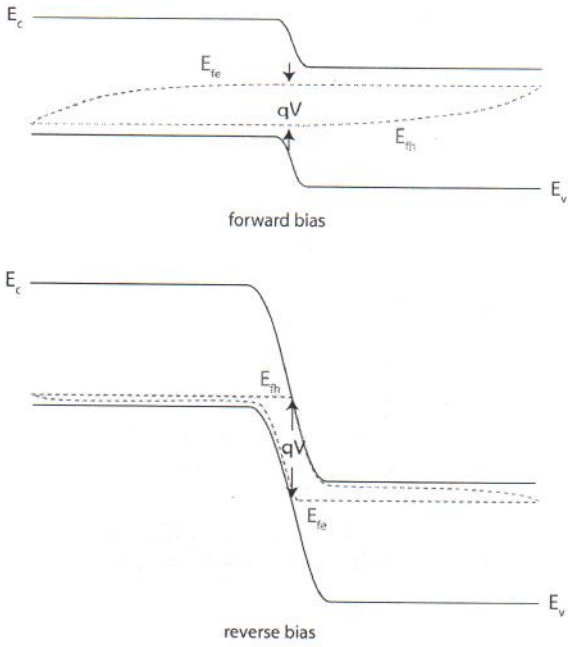


Figure 6.30: Schematic energy band diagram in forward and reverse bias showing the location of the quasi-Fermi levels in a short diode.

$e^{qV/kT}$ expected in forward bias does not extend all the way down to $V = 0$. In fact, for small forward bias, the current seems to follow a dependence more like $e^{qV/2kT}$. Similarly, the diode of Fig. 6.31 exhibits a reverse bias behavior in which the current is not saturated but is in excess of the ideal value extrapolated to $V = 0$ from forward bias.

These anomalies are due to recombination (in forward bias) and generation (in reverse bias) through traps located in the space-charge region. These mechanisms contribute additional current components to the pn diode.

A rigorous model for the diode current that includes SCR generation and recombination is mathematically involved. In most cases, it is also not useful to seek a very detailed description because the nature, concentration and electrical characteristics of the traps are not generally well known. What is most useful is to develop a simple model that captures the essential physics and provides the key dependencies.

The model that we develop in this section assumes that SCR generation and recombination represent a small perturbation to the ideal diode picture. In particular, we will assume that the carrier concentrations across the SCR are given by the expressions obtained under the ideal diode model.

We start by going back to Eq. 3.42 that provides the most general relation for the net recombination rate due to trap-assisted processes. If we use the expression for the pn product obtained for the ideal diode in Eq. 6.43, Eq. 3.42 becomes:

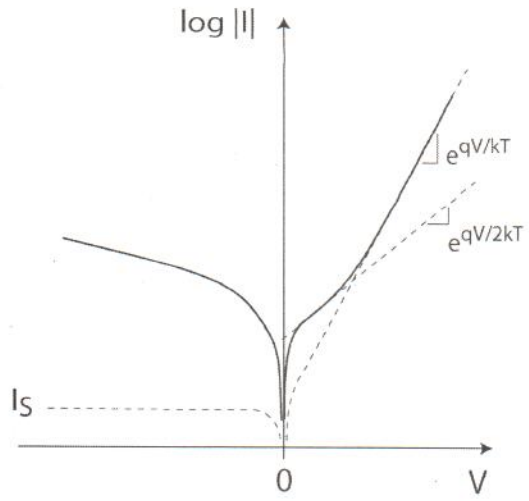


Figure 6.31: Sketch of I-V characteristics for typical Si PN junction. Notice the non-ideal I-V characteristics for small forward bias and for reverse bias associated with SCR recombination and generation, respectively.

$$U_{tr}|_{SCR} = \frac{n_i^2(\exp\frac{qV}{kT} - 1)}{\tau_{ho}n + \tau_{eo}p + (\tau_{ho} + \tau_{eo})n_i} \tag{6.81}$$

As discussed in Section 5.9, the SCR current can be computed by integrating the net recombination rate across the SCR and multiplying by the carrier charge:

$$J_{SCR} = q\int_{-x_p}^{x_n} U_{tr}dx \tag{6.82}$$

Inserting Eq. 6.81 into Eq. 6.82 does not lead to an analytical result. A relatively simple expression that displays all the dependencies observed experimentally can be obtained by identifying the highest value of U_{tr} anywhere in the SCR, and assuming that this value applies everywhere. This gives an upper limit to the desired result.

The denominator of Eq. 6.81 contains a term in n, and another one in p. Since the product of p and n is a constant, it is not difficult to prove that the highest value of U_{tr} is obtained when $\tau_{ho}n = \tau_{eo}p$. This yields a maximum net recombination rate given by:

$$U_{tr}|_{SCR,max} = \frac{n_i}{2\sqrt{\tau_{eo}\tau_{ho}}}(\exp\frac{qV}{2kT} - 1) \tag{6.83}$$

Inserting this into Eq. 6.82 and assuming it to be constant everywhere results in:

$$J_{SCR,max} = \frac{qn_i x_{SCR}}{2\sqrt{\tau_{eo}\tau_{ho}}}(\exp\frac{qV}{2kT} - 1) \tag{6.84}$$

This is an expression that is valid both in forward and reverse bias.

We can also define a saturation current for SCR generation/recombination. That is the preexponential factor in Eq. 6.84:

$$J_s(SCR) = \frac{qn_i x_{SCR}}{2\sqrt{\tau_{eo}\tau_{ho}}} \tag{6.85}$$

There are several features to notice in the result captured in Eq. 6.84.

- First, the forward current grows as $e^{qV/2kT}$, in contrast with recombination in a quasi-neutral region which goes as $e^{qV/kT}$. The coefficient in front of the kT factor in the denominator of the exponent is often called the *ideality factor*, n. $n = 1$ for QNR recombination, while $n = 2$ for SCR recombination.

  The reason for this rather different voltage behavior is easy to see if we set $\tau_{eo} = \tau_{ho}$. Inside the SCR and at its edges, we have that $np = n_i^2 \exp(qV/kT)$. At the edges of the SCR, since either n or p are fixed to the doping level, then the other carrier concentration goes as $n_i^2 \exp(qV/kT)$. The entire voltage is used to grow the minority carrier concentration and the recombination current exhibits this same dependence. Inside the SCR, however, at the point of highest recombination rate we have $n = p = n_i \exp(qV/2kT)$. That means that the voltage is used to grow the concentration of both types of carriers. Since the recombination rate is linear on n and p, then the recombination current exhibits a dependence of $n_i \exp(qV/2kT)$.

  In practice, SCR recombination is characterized by an experimental ideality factor $1 < n \leq 2$, the discrepancy arising from the simplifying approximations that we have made in the derivation of Eq. 6.84. A consequence of the higher ideality factor in the SCR recombination current is that it dominates for small forward bias, while QNR recombination dominates for large forward bias, as sketched in Fig. 6.31.

- Second, the space-charge region saturation current depends on n_i. This is in contrast with QNR current I_s that depends on n_i^2 (see Eqs. 6.36 and 6.37). Because of this, the temperature dependence of these two currents is quite different. The saturation current of QNR generation and recombination is thermally activated with an activation energy of about E_g, the bandgap of the semiconductor. In contrast, the saturation current of SCR-generation/recombination current displays an activation energy of $E_g/2$.

- Third, the magnitude of the SCR generation/recombination current depends on the identity and concentration of the trap or traps that are present inside the SCR. This can be seen in Eq. 6.84 in which the saturation current depends on $\sqrt{\tau_{eo}\tau_{ho}}$. In practice, this implies that the SCR generation/recombination currents are highly process sensitive. For high enough doping levels ($\sim 10^{16}\ cm^{-3}$ and above), recombination in the quasi-neutral regions, in contrast, is quite predictable, with the lifetime being essentially set by the doping level and the carrier transit time, for the case of the short diode, being set by geometry.

- Finally, in reverse bias, Eq. 6.85 exhibits a dependence on the SCR thickness. This has the important consequence that the reverse current due to SCR generation in a PN diode is not saturated but increases the higher the reverse bias. This is also sketched in Fig. 6.31.

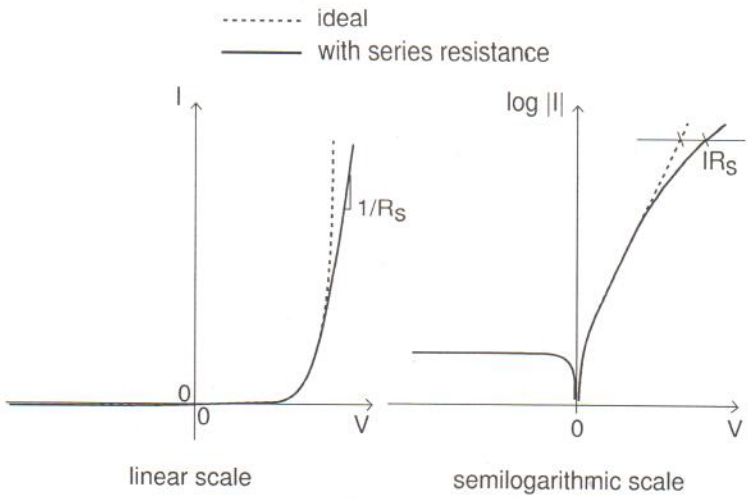


Figure 6.32: Sketch of I-V characteristics of an ideal PN diode and a diode with series resistance, both in a linear scale (left) and in a semilogarithmic scale (right).

The non-ideal currents described in this section are intimately associated with traps in the SCR. This can be the result of undesired impurities or crystallographic defects introduced by the process. Since it is eminently possible to obtain very ideal diodes, as we saw in Figs. 6.15 and 6.16, the type of non-ideal behavior discussed here is then a good indicator of the cleanliness and integrity of microelectronics processes and materials.

6.6.3 Series resistance

An important parasitic in a PN diode is series resistance. Series resistance arises from the finite resistance associated with the less than perfect ohmic contacts and the non-zero resistivity of the quasi-neutral regions. If current flows through a diode, series resistance produces an ohmic drop that adds on top of the junction voltage. In an alternative view, series resistance increases the voltage that needs to be applied to a diode to obtain a desirable level of current. Hence, series resistance "stretches" the I-V characteristics sideways. The effect is more pronounced the higher the current through the diode.

The impact of series resistance in the I-V characteristics can be seen in Fig. 6.32. On the right are the characteristics in a semilog scale. We see that at low current levels, the characteristics remain ideal. However, at high current, the voltage required to supply the current increases by an amount equal to the ohmic drop through the diode, or IR_s, where R_s is the series resistance. A consequence of this is that in a linear scale, for strong forward bias, the current no longer rises exponentially but tends to increase linearly with the voltage with a slope given by $1/R_s$.

An expression for the I-V characteristics of the diode that incorporates the impact of series resistance can be readily obtained by subtracting from V in the ideal diode current expression of Eq. 6.42 the ohmic drop produced by the series resistance, IR_s:

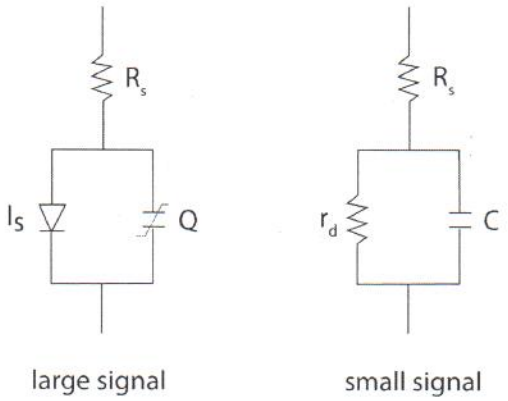


Figure 6.33: Equivalent circuit models of pn diode incorporating the effect of series resistance: large signal (left) and small signal (right).

$$I = I_s[\exp \frac{q(V - IR_s)}{kT} - 1] \tag{6.86}$$

It is clear in this expression that when IR_s is small in the scale of V, the impact of series resistance is negligible. However, for high values of I the impact can be very substantial. One unfortunate consequence of series resistance is that the current can no longer be solved in an explicit way in terms of the voltage (although the voltage can still be solved in terms of the current).

Series resistance can be readily incorporated into the equivalent circuit models of the diode. This only requires adding a resistance in series with the models for the ideal diode. This is shown in Fig. 6.33.

Given a certain diode structure, how does one compute its series resistance? This is not difficult for a diode with the structure shown in Fig. 6.3. There are two kinds of contributions to the series resistance. One is due to the contact resistances and the second one is due to the resistance of the QNRs. Both can be evaluated readily.

As we will discuss in Ch. 7, the resistance of an ohmic contact is characterized by its ohmic contact resistivity, ρ_c, given in $\Omega.cm^2$. If a contact has an area A, the ohmic contact resistance is then ρ_c/A. This has units of Ω and scales as $1/A$, that is, the larger the contact area, the smaller its resistance. For a diode like the one in Fig. 6.3 where both contacts have an area equal to the junction area, the total ohmic contact resistance is given by:

$$R_{cp} + R_{cn} = \frac{1}{A}(\rho_{cp} + \rho_{cn}) \tag{6.87}$$

where ρ_{cp} and ρ_{cn} refer to the contact resistivities of the ohmic contacts to the p region and n regions, respectively.

The second component of the series resistance is associated with the finite resistance of the

QNRs. This can be evaluated by computing their corresponding geometrical resistance as given by Eq. 5.39. In this case, we have:

$$R_p + R_n = \frac{1}{A}\left(\frac{w_p - x_p}{q\mu_h N_A} + \frac{w_n - x_n}{q\mu_e N_D}\right) \tag{6.88}$$

where the first term reflects the contribution of the p-QNR and the second one that of the n-QNR.

The total series resistance of the diode is the sum of these four terms:

$$R_s = R_{cp} + R_{cn} + R_p + R_n \tag{6.89}$$

In general, the series resistance of a diode is to be minimized. Eq. 6.89 suggests that achieving this requires paying attention to both contact resistances as well as the resistance of the QNRs which are affected by doping levels and geometrical design.

The resistance of the QNRs as given by Eq. 6.88 is simple and intuitive but it needs to be justified. It is not entirely obvious that in this complex situation in which we have excess minority carriers, a purely majority-carrier-like approach is appropriate.

As it turns out, it is not difficult to show that Eq. 6.88 is an excellent approximation for this situation. Solving this problem rigorously is a great opportunity to review key concepts introduced in Ch. 5. This is done in Appendix AT6.1.

6.6.4 Breakdown voltage

In reverse bias, the current in an ideal diode increases with voltage and quickly saturates to a very small value given by the saturation current. As we saw earlier, in actual diodes, the reverse bias current can be higher than expected as a result of carrier generation in the SCR. If this phenomenon is significant, the current increases with reverse voltage as the generation volume expands. Still, the reverse current is very small. However, for large enough reverse bias, a sudden increase in current is observed in all diodes. The current level can get substantial, just as high as in the forward regime, as sketched in Fig. 6.34. This condition is called *breakdown* and the voltage at which it happens is referred to as the *breakdown voltage*. The breakdown condition can be destructive since high current flows while a high voltage is being applied. The power dissipation and temperature rise that take place in the diode can create irreparable damage. The breakdown voltage is obviously a key consideration in PN diode design. In this section we study the physics of this effect and we discuss design guidelines for the breakdown voltage.

Breakdown can occur through several effects. One is band-to-band or Zener tunneling, briefly mentioned in Sec. 3.1. By far the most common process is impact ionization, whose basics physics were presented in Chs. 4 and 5. This is the mechanism that we discuss here.

As the reverse bias in a PN junction is increased, the electric field inside the SCR increases too. This makes it more likely that carriers that transit through the SCR gain enough energy

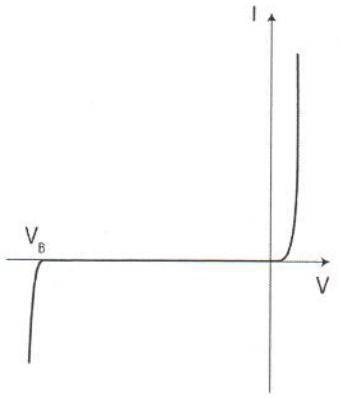


Figure 6.34: Sketch of reverse breakdown in a PN diode I-V characteristics.

from the electric field to undergo impact ionizing collisions with the lattice and generate electron-hole pairs. Even though this reverse bias current is very small, for high enough fields, substantial carrier multiplication might occur. At some critical field, avalanche breakdown eventually takes place and the reverse current rises abruptly (Fig. 6.34).

It is not possible to develop an analytical model for the breakdown voltage of a PN diode due to impact ionization. The electric field is non uniform in space and the expressions of the field-dependent impact ionization coefficient are algebraically too complicated to yield simple solutions. Through adequate simplifications, we can arrive to a model that captures the essential physics and illuminates the key dependencies.

The starting point for the model are the expressions for the multiplication factors derived in Sec. AT5.4 for a general non-uniform electric field case. Let us consider a simple situation in which the ionization coefficients $\alpha_e = \alpha_h = \alpha$. Notice from the data in Fig. 4.27 that this is not a bad assumption at high electric fields where breakdown occurs. In this instance, from either Eq. 5.159 or 5.160, the condition for breakdown becomes:

$$\int_{-x_p}^{x_n} \alpha dx = 1 \tag{6.90}$$

To further simplify the problem, we assume an asymmetric p^+-n junction where the electric field preferentially extends through the n-side of the SCR. In this case, $x_p \simeq 0$ and $x_n = x_{SCR}$ and the electric field distribution in the SCR can be written as $\mathcal{E}(x) = -|\mathcal{E}_{max}|(1 - x/x_{SCR})$. The negative sign in the electric field is a consequence of our choice of axis which is consistent with that of Sec. 6.2. With this and going back to Eq. 4.107, we can now write an expression for the impact ionization coefficient inside the PN junction SCR as:

$$\alpha = A \exp[-\frac{B}{|\mathcal{E}_{max}|(1 - \frac{x}{x_{SCR}})}] \tag{6.91}$$

If we use this expression, the integral in Eq. 6.90 does not yield an analytical result. However,

due to the exponential dependence of α on $\mathcal{E}$, the most significant contribution to the integral in Eq. 6.90 takes place around $x \simeq 0$ where $\mathcal{E} \simeq \mathcal{E}_{max}$. Around the metallurgical junction, we can approximate:

$$\frac{1}{1-\frac{x}{x_{SCR}}} \simeq 1 + \frac{x}{x_{SCR}} \tag{6.92}$$

We can now insert this into Eq. 6.91 and the result into Eq. 6.90 to yield:

$$1 = \int_0^{x_{SCR}} \alpha dx \simeq A \int_0^{x_{SCR}} \exp[-\frac{B}{|\mathcal{E}_{max}|}(1 + \frac{x}{x_{SCR}})]dx \tag{6.93}$$

This can now be readily integrated to yield:

$$1 \simeq \alpha(|\mathcal{E}_{max}|)\frac{|\mathcal{E}_{max}|x_{SCR}}{B}[1 - \frac{\alpha(|\mathcal{E}_{max}|)}{A}] \tag{6.94}$$

As we will see through some numbers below, at breakdown, $\alpha(|\mathcal{E}_{max}|) \ll A$. We also note that at breakdown, $|\mathcal{E}_{max}|x_{SCR} = 2V_B$, where V_B is the breakdown voltage (defined as a positive value). We can therefore approximate the breakdown condition as:

$$\alpha(|\mathcal{E}_{max}|)2V_B \simeq B \tag{6.95}$$

where $\mathcal{E}_{max}$ refers to its value at breakdown. From Eq. 6.25, this can be expressed in terms of V_B as:

$$|\mathcal{E}_{max}(V_B)| \simeq \sqrt{\frac{2qN_DV_B}{\epsilon}} \tag{6.96}$$

Because of the exponential dependence of α on $\mathcal{E}$, we cannot explicitly solve for V_B in Eq. 6.94. However, we can solve for N_D in terms of everything else. Plugging Eq. 6.96 into Eq. 4.107 and this into Eq. 6.94 and then solving for N_D yields:

$$N_D = \frac{\epsilon B^2}{2qV_B}\frac{1}{\ln^2(\frac{2AV_B}{B})} \tag{6.97}$$

This is now a relatively simple *implicit* model for V_B. The most important feature of this model is that V_B depends only on N_D and fundamental constants. The higher N_D, the lower V_B. This is understandable since the higher N_D, the higher the initial electric field inside the SCR and the easier it will be to reach the breakdown condition as the reverse bias increases. Note that because of the logarithmic term in the denominator, the dependence between V_B and N_D is not quite inverse linear but somehow weaker than that, particularly for small values of V_B.

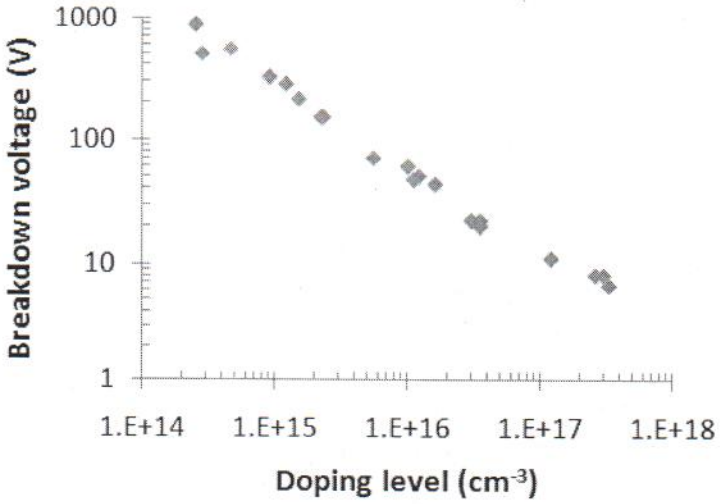


Figure 6.35: Experimental breakdown voltage of abrupt uniformly-doped PN diodes vs. doping level at 300 K.

A chart graphing experimental measurements of V_B vs. N_D for abrupt uniformly-doped Si PN diodes at room temperature is shown in Fig. 6.35. As the doping level increases, the breakdown voltage decreases. The decrease slows down for high doping levels.

An interesting feature of the breakdown voltage due to impact ionization is its temperature dependence. For Si, as the temperature increases, the impact ionization coefficients decrease. This is the result of the mean free path becoming smaller due to enhanced phonon scattering. A consequence of this is that the breakdown voltage increases as the temperature increases, that is, it exhibits a *positive temperature coefficient*. This is distinct from the other major breakdown mechanism that can be observed in PN diodes which is Zener breakdown. Zener breakdown takes place when direct tunneling occurs between the conduction band and the valence band across the SCR under strong reverse bias. Since as the temperature increases the bandgap gets smaller, the tunneling barrier decreases and the tunneling process becomes more likely. In consequence, the breakdown voltage decreases. Zener breakdown thus is characterized by a *negative temperature coefficient*.

6.6.5 Non-uniform doping distributions

The fabrication processes that are used to manufacture integrated microelectronic devices often result in doping distributions that are highly non-uniform. This does not change the basic physics of operation of a PN diode. However, the actual doping profile must be taken into account when computing the various diode figures of merit or when designing a diode for a specific application. Additionally, a non-uniform doping level changes the functional dependence of some important parameters, such as the capacitance vs. voltage characteristics.

A typical doping distribution of an integrated PN diode is sketched in Fig. 6.36. In this

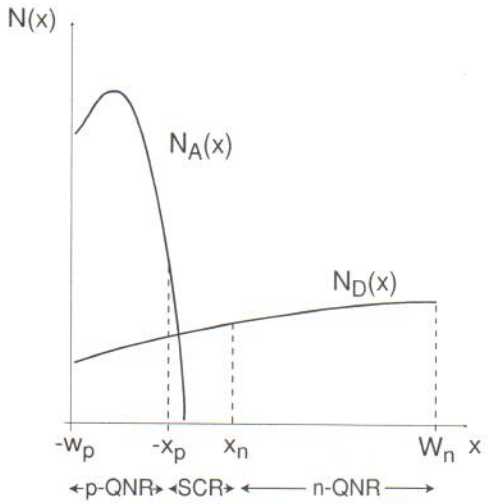


Figure 6.36: Sketch of a non-uniform doping level of an integrated diode. Typicaly the emitter is heavily doped and highly non-uniform. The base has a lower doping level and relatively more uniform. The back of the base is not shown.

particular case, a thin and heavily-doped p-type region is ion implanted or diffused into an n-type region. In analogy with bipolar transistor notation, the thin heavily doped region is often denoted as *emitter*, while the region with a lower doping level is known as *base*. As was the case in the ideal uniformly doped case studied above, from an electrical point of view, this diode can be divided into three regions: the n- and p-quasi-neutral regions and the space-charge region in between. The basic role played by these three regions is unchanged from what we have already studied. As in the ideal diode, the I-V characteristics are mainly determined by recombination in the quasi-neutral regions. The C-V characteristics are set by the electrostatics of the space-charge region. The dynamics are largely a function of the junction capacitance and the minority carrier storage in the quasi-neutral regions. For all three, the ideal diode treatment needs to be revised. Let us examine them in turn.

Depletion capacitance

When the doping distribution in both regions of a PN junction is non-uniform, the voltage dependence of the depletion capacitance does not follow any longer the simple square root law of Eq. 6.62. Arbitrary doping profiles, in general, do not give rise to simple functional laws for the C-V characteristics. Device simulators can accurately account for this and provide numerical calculations of the depletion capacitance that are acceptable. In a diode model for circuit CAD, however, a simple analytical relationship is required. Fortunately, in modern devices, the doping levels are high enough and consequently the space charge regions are thin enough that the doping level distribution across the SCR can often be described using a simple power law.

In order to illustrate the issues, let us solve a simple case. Consider a PN junction where the doping level changes linearly inside the SCR, that is:

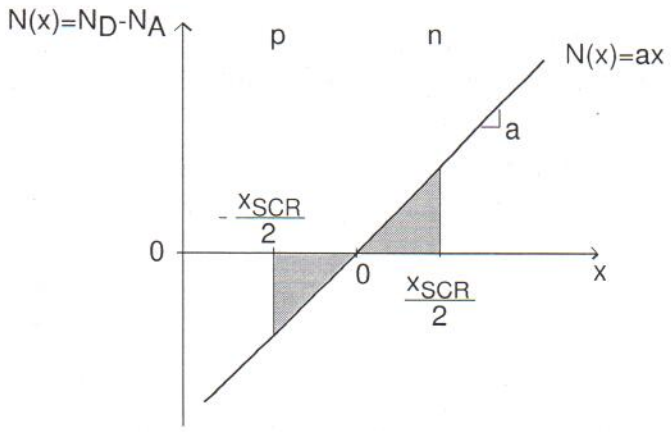


Figure 6.37: Doping distribution in PN junction with linearly graded impurity profile.

$$N_D - N_A = ax \tag{6.98}$$

a is called the *grading constant* and has units of cm^{-4}. This impurity profile is illustrated in Fig. 6.37.

We approach this problem in the context of the depletion approximation. In order to solve the electrostatics of this junction, we first make the depletion approximation. We assume that over a certain region of extent x_{SCR}, the carrier concentrations are much lower than the dopant concentrations. Outside, quasi-neutral conditions prevail. Due to the symmetry of the charge distribution, the space charge region extends equal amounts into the n-type region and the p-type region. Hence

$$x_n = x_p = \frac{x_{SCR}}{2} \tag{6.99}$$

We still do not know how much x_{SCR} is. In order to derive it, we must solve completely the electrostatics of this problem.

The volume charge density is given by:

$$\rho(x) = qax \qquad \text{for} -\frac{x_{SCR}}{2} \leq x \leq \frac{x_{SCR}}{2} \tag{6.100}$$

$$\rho(x) \simeq 0 \qquad \text{outside} \tag{6.101}$$

Using Gauss' law, the electric field is easily obtained:

$$\mathcal{E}(x) = \frac{qa}{2\epsilon}[x^2 - (\frac{x_{SCR}}{2})^2] \qquad \text{for} -\frac{x_{SCR}}{2} \leq x \leq \frac{x_{SCR}}{2} \tag{6.102}$$

$$\mathcal{E}(x) \simeq 0 \qquad \text{outside} \tag{6.103}$$

Strictly speaking, the electric field at the edges of the SCR is obviously not zero. If the doping level changes outside, a small field is developed. However, this field is likely to be much smaller than the field inside the SCR. For the purposes of solving the electrostatics of the SCR it can safely be neglected.

One further integration yields the electrostatic potential distribution. Selecting as reference $\phi(x=0)=0$,

$$\phi(x) = \frac{qa}{6\epsilon}[3(\frac{x_{SCR}}{2})^2 x - x^3] \qquad \text{for} -\frac{x_{SCR}}{2} \leq x \leq \frac{x_{SCR}}{2} \tag{6.104}$$

Outside, the potential distribution depends on the details of the impurity profile.

The electrostatic potential difference across the SCR is equal to the built-in potential minus the applied voltage, $\phi(x_{SCR}/2) - \phi(-x_{SCR}/2) = \phi_{bi} - V$. This allows us to derive an expression for x_{SCR}:

$$x_{SCR} = [\frac{12\epsilon(\phi_{bi} - V)}{qa}]^{1/3} \tag{6.105}$$

ϕ_{bi} is obtained from an equation similar to 6.1:

$$\phi_{bi} = \frac{kT}{q} \ln \frac{n_o(x_{SCR}/2)}{n_o(-x_{SCR}/2)} = 2\frac{kT}{q} \ln \frac{a x_{SCR}}{2n_i} \tag{6.106}$$

Since x_{SCR} in Eq. 6.105 depends on ϕ_{bi} and ϕ_{bi} in Eq. 6.106 depends on x_{SCR}, this pair of equations needs to be solved iteratively.

The depletion capacitance of the linearly graded junction can be most easily obtained using Eq. 6.63:

$$C_j = A[\frac{qa\epsilon^2}{12(\phi_{bi} - V)}]^{1/3} = \frac{C_j(V=0)}{(1 - \frac{V}{\phi_{bi}})^{1/3}} \tag{6.107}$$

This results suggests that $1/C_j^3$ has a linear dependence on V:

$$\frac{1}{C_j^3} = \frac{12(\phi_{bi} - V)}{qa\epsilon^2 A^3} \tag{6.108}$$

This is a simple relationship that can be used to extract ϕ_{bi} and the grading constant a.

Many actual PN junctions follow the linearly-graded behavior studied in this section. An example is shown in Fig. 6.38. In general, it is found in practice that the depletion capacitance of most PN diodes can be approximated by the functional form:

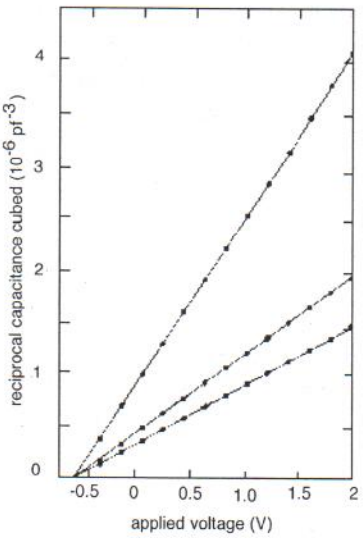


Figure 6.38: Experimental $1/C^3$ *vs.* V characteristics for three different Si PN junctions. As Eq. 6.108 predicts, $1/C^3$ has a linear dependence on voltage [data from *J. R. Lowney and W. R. Thurber, Electronics Lett. 20, 142 (1984)* (note: the voltage convention followed by these authors is contrary to that of this book)].

$$C_j = \frac{C_{jo}}{(1 - \frac{V}{\phi_j})^m} \tag{6.109}$$

with C_{jo} being the depletion capacitance that corresponds to $V = 0$, ϕ_j an effective junction built-in potential, and m a numerical exponent typically between 0.3 and 0.5.

Current-voltage characteristics

The forward I-V characteristics of the diode are dominated by recombination in the quasi-neutral regions. The total current is the sum of the current injected into the n-side plus the current injected into the p-side. The I-V characteristics of the diode are unchanged from those of Eq. 6.42. Only the individual expressions of the two components of the saturation current I_s are affected. In a non-uniformly doped diode, the problem of computing the saturation current is complicated by the fact that in addition to minority carrier diffusion and recombination, drift takes place. This is because, as studied in Chapter 4, a non-uniform doping level produces an electric field which acts on the carriers. The equation set that governs this family of problems was given in Table 5.5. The computation of the saturation current injected into a non-uniformly doped region with an arbitrary profile is not amenable to an analytical solution and must be done by computer. For the minority carrier holes, equation 5.51 needs to be solved subject to the proper boundary conditions at the edge of the space-charge region and at the surface of the diode. A parallel approach is followed for minority carrier electrons.

In many integrated PN diodes, the depth of the doped regions is shallow enough that minority carrier recombination in the bulk of the emitter or the base is negligible. Minority carrier recom-

bination takes place predominantly at the surfaces of the diode. This is what is called a short or *transparent region* [3]. In this case, a calculation of the saturation current is particularly easy. For the p-side, in the absence of external generation, under steady state, and neglecting volume recombination, the continuity equation equivalent to Eq. 5.50 becomes simply $dJ_e/dx = 0$, or a simple statement that the electron current does not change in space. This is obvious and not very useful in itself. Let us then use the minority carrier current equation (given in Table 5.5):

$$J_e = qn'\mu_e\mathcal{E}_o + qD_e\frac{dn'}{dx} \tag{6.110}$$

In a non-uniformly doped p-type region, the electric field in equilibrium is given by Eq. 4.56. In a minority carrier-type problem, we can assume that this field is negligibly modified out of equilibrium. Inserting this equation into Eq. 6.110, and using the Einstein relation for electrons, Eq. 4.38, we get:

$$J_e = q\frac{D_e}{N_A}\frac{d(n'N_A)}{dx} \tag{6.111}$$

After bringing the term qD_e/N_A to the left, this equation can be integrated across the non-uniformly doped region. For the emitter of the PN junction shown in Fig. 6.36, it becomes:

$$\int_{-w_p}^{-x_p} J_e\frac{N_A}{qD_e}dx = \int_{-w_p}^{-x_p}\frac{d(n'N_A)}{dx}dx = n'N_A|_{-x_p} - n'N_A|_{-w_p} \tag{6.112}$$

The boundary condition at $-x_p$, is simply as in Eq. 6.30. The boundary condition at $-w_p$ depends on the details of the surface. If the surface of the semiconductor is covered by an ohmic contact, then $n'(-w_p) = 0$. Getting J_e out of the integral in Eq. 6.112 and solving for it, we get:

$$J_e(-x_p) = \frac{qn_i^2}{\int_{-w_p}^{-x_p}\frac{N_A}{D_e}dx}(\exp\frac{qV}{kT} - 1) \tag{6.113}$$

If the doping level is uniform, this equation simplifies to that corresponding to a short diode (Eq. 6.73), as should be expected.

If the surface is characterized by a surface recombination S, then $n'(-w_p) = J_e(-w_p)/qS = J_e(-x_p)/qS$, and:

$$J_e(-x_p) = \frac{qn_i^2}{\int_{-w_p}^{-x_p}\frac{N_A}{D_e}dx + \frac{N_A(-w_p)}{S}}(\exp\frac{qV}{kT} - 1) \tag{6.114}$$

Eqs. 6.113 and 6.114 are very simple expressions. The integral in the denominator of both equations are easy to perform once the doping profile is known. In these integrals, the diffusion

[3] As opposed to a long or *opaque region* in which all minority carriers recombine before reaching the surface.

coefficient must remain inside the integral since, in general, it is a function of position if the doping level changes in space. Nevertheless, since D_e is not a strong function of N_A, sometimes D_e is taken out of the integral. In this case, the remaining integral $\int_{-w_p}^{-x_p} N_A dx$ is called the *Gummel number*, in honor of H. K. Gummel, a bipolar transistor pioneer. Gummel was first to realize that, to the first order, *the total amount of dopants* in a thin transparent region (that is when bulk recombination is negligible) determines the minority carrier injection current. Their particular spatial distribution is only of secondary importance. Interestingly, this implies that the magnitude of the minority carrier current is independent whether the electric field developed by the doping distribution aids or opposes minority carrier diffusion. This result, as you will see below, does not apply to the minority carrier storage, which is intimately related to the electric field distribution.

The current contribution arising from hole injection into the non-uniformly doped base is computed in the same way and a totally equivalent equation emerges.

Minority carrier storage

In a uniformly doped region, the dominant time constant for minority carrier behavior is the lifetime or the transit time, whichever is shortest. This does not change in a non-uniformly doped region, just the computation of these characteristic times is more complicated. In fact, in a general case, an analytical solution is not possible.

A case of importance for which an analytical solution is possible is the short or transparent case in which the carrier lifetime is much longer than the transit time. In this case, the transit time is the dominant time constant of the problem.

Let us compute the diffusion capacitance and the transit time for a p-region similar to the one examined just above. The starting point is the expression of the minority-carrier charge in the p-region given in Eq. 6.46. For this, we need the excess electron profile across the quasi-neutral p-region. This can be obtained by integrating Eq. 6.111 in a similar way as in Eq. 6.112 but from the surface at $x = -w_p$ to a point x inside:

$$\int_{-w_p}^{x} J_e \frac{N_A}{qD_e} dx = \int_{-w_p}^{x} \frac{d(n'N_A)}{dx} dx = n'N_A|_x - n'N_A|_{-w_p} \tag{6.115}$$

If the surface is covered by an ohmic contact, $n'(-w_p) = 0$ and we can then easily solve for $n'(x)$:

$$n'(x) = \frac{J_e(-x_p)}{qN_A(x)} \int_{-w_p}^{x} \frac{N_A}{D_e} dx \tag{6.116}$$

Inserting this result into Eq. 6.46 yields:

$$Q_p = J_e(-x_p) \int_{-w_p}^{-x_p} \frac{1}{N_A} \left(\int_{-w_p}^{x} \frac{N_A}{D_e} dx \right) dx \tag{6.117}$$

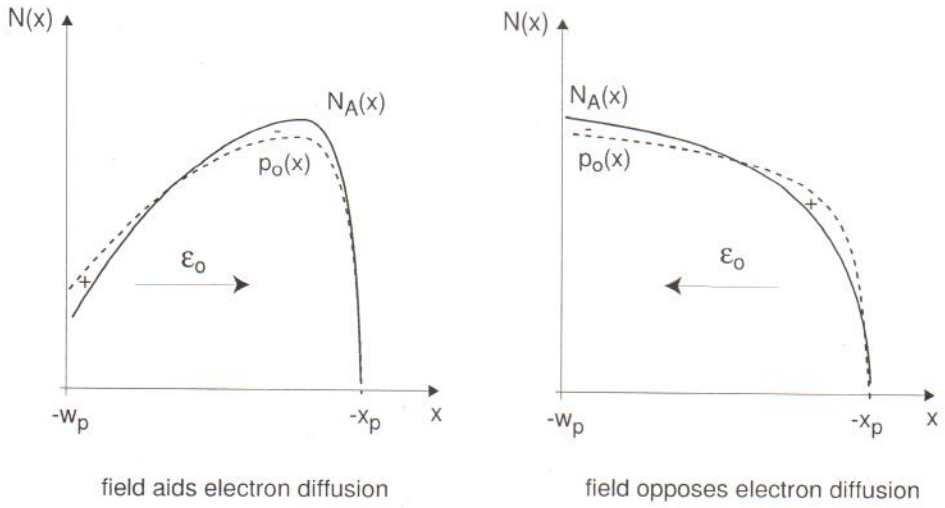


Figure 6.39: Sketch of majority carrier distributions and electric field in two non-uniformly doped p-type quasi-neutral regions. A "downgoing" gradient results in an electric field that helps electron diffusion. An "upgoing" gradient opposes electron diffusion.

The diffusion capacitance associated with the p-region is:

$$C_{dp} \simeq \frac{q}{kT} J_e(-x_p) \int_{-w_p}^{-x_p} \frac{1}{N_A} \left(\int_{-w_p}^{x} \frac{N_A}{D_e} dx \right) dx \tag{6.118}$$

From this expression, we can identify the transit time through the p-region as:

$$\tau_{tp} = \int_{-w_p}^{-x_p} \frac{1}{N_A} \left(\int_{-w_p}^{x} \frac{N_A}{D_e} dx \right) dx \tag{6.119}$$

You can easily verify that this expression converges to Eq. 6.78 for the uniformly-doped case.

It might not be obvious just by looking at Eq. 6.119, but the presence of an electric field in a quasi-neutral region can substantially affect the transit time of minority carriers. An electric field that makes minority carriers drift in the same direction as they diffuse will result in a shorter transit time. A field that opposes diffusion, will increase the transit time. Problem 6.4 shows an excellent example of this.

What type of an impurity gradient does one need in order to engineer an electric field that helps diffusion and results in a fast transit time? The sketches of Fig. 6.39 for a non-uniformly doped p-type region help us to determine this. By working out the equilibrium majority carrier profile in the quasi-neutral region, it is clear that a "downgoing" impurity gradient results in a field that aids electron diffusion. In contrast, an upgoing impurity gradient leads to a field that opposes electron diffusion. You can easily verify that the same happens for an n-type region.

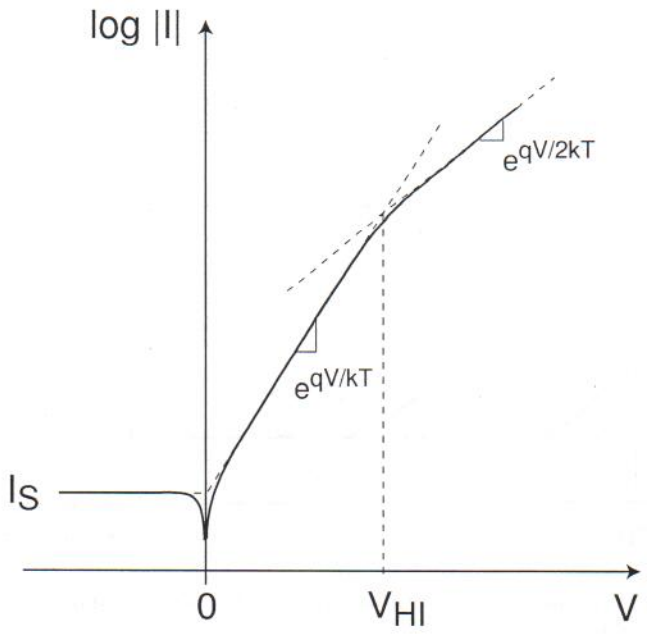


Figure 6.40: Sketch of forward bias I-V characteristics of a PN diode. For strong forward voltage, the diode can go into high-level injection and the current grows as $e^{qV/2kT}$.

6.6.6 High-injection effects

A basic assumption made in the derivation of the forward characteristics of the PN diode is low-level injection. Low-level injection means that the minority carrier concentration at the edges of the PN junction is always negligible in comparison with the majority carrier concentration. For high enough forward voltage, this assumption fails as the diode goes into "high-level injection." When this happens, our ideal theory does not apply any longer. You can see in Fig. 6.40 that, in fact, for strong forward voltage, the I-V characteristics deviate from their ideal $e^{qV/kT}$ behavior into what appears more to be $e^{qV/2kT}$. Let us discuss why this happens.

First, let us estimate the voltage at which high-level injection occurs. This is simple. In an asymmetric diode, high-level injection occurs first in the lowly-doped side of the diode. Let us consider here a p^+-n diode with a donor concentration N_D (an entirely equivalent result is obtained for a n^+-p diode). Fig. 6.41 shows the evolution of the minority and majority carrier concentrations (in a semilog scale) in the n-region as the diode forward voltage increases. This figure clearly shows that the location that is driven first into high-level injection is the edge of the SCR. For small forward voltages, the minority carrier concentration is much less than the doping level and the majority carrier concentration is hence unaffected by minority carrier injection. For a high-enough forward voltage, the minority carrier concentration at the edge of the SCR reaches a value equal to the doping level. At this voltage, the majority carrier concentration at the edge of the SCR is twice the doping level. Clearly, it is not possible to assume any longer that the majority carrier concentration is unaffected as this would violate quasi-neutrality. The n-side of the QNR has become high-level injected.

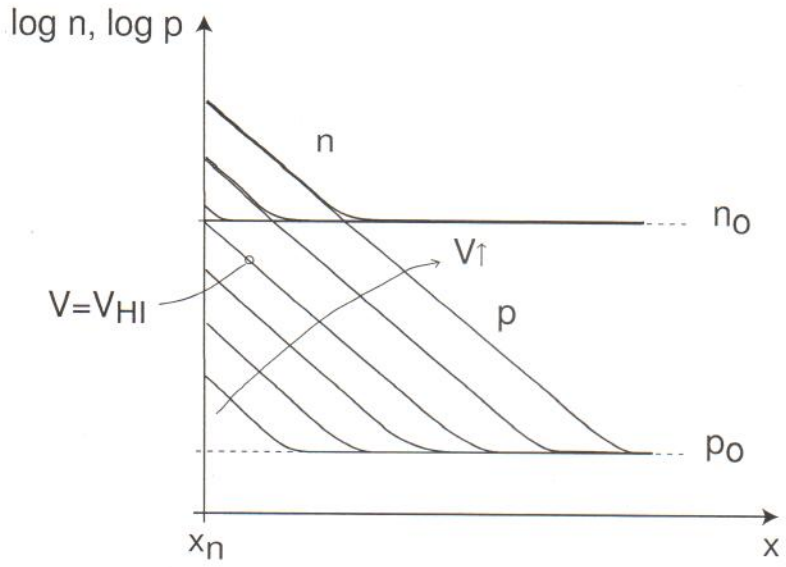


Figure 6.41: Minority and majority carrier concentrations in the n-side of a p^+-n diode as a function of forward bias. At V_{HI}, the n-region goes into high-level injection.

This is a reasonable definition for the onset of high-level injection. If we allow ourselves to "stretch" the low-level injection boundary condition, Eq. 6.29, up to this point, high-level injection occurs when:

$$p(x_n) = \frac{n_i^2}{N_D} \exp \frac{qV_{HI}}{kT} = N_D \tag{6.120}$$

where V_{HI} refers to the voltage at the onset of high-level injection. Solving for V_{HI} in this equation, we get:

$$V_{HI} = \frac{2kT}{q} \ln \frac{N_D}{n_i} \tag{6.121}$$

This result is reasonable. The higher the doping level (of the lowly-doped side), the higher the voltage that it takes to drive the diode into high-level injection.

In microelectronic device operation, we are not interested in general in the high-level injection regime. This is something to be avoided. In PN junctions, for example, the I-V characteristics grow more slowly, as sketched in Fig. 6.40, something that is undesirable for many applications. In bipolar transistors, for example, high-level injection of the base brings with it a drop in gain. Unless one is working with devices that have been designed to operate in high-level injection, such as power diodes or certain kinds of solar cells, high-level injection is a syndrome to avoid all together. For this reason, we will not treat it in detail here.

It is of value, nevertheless, to understand the origin of the $e^{qV/2kT}$ dependence of the current

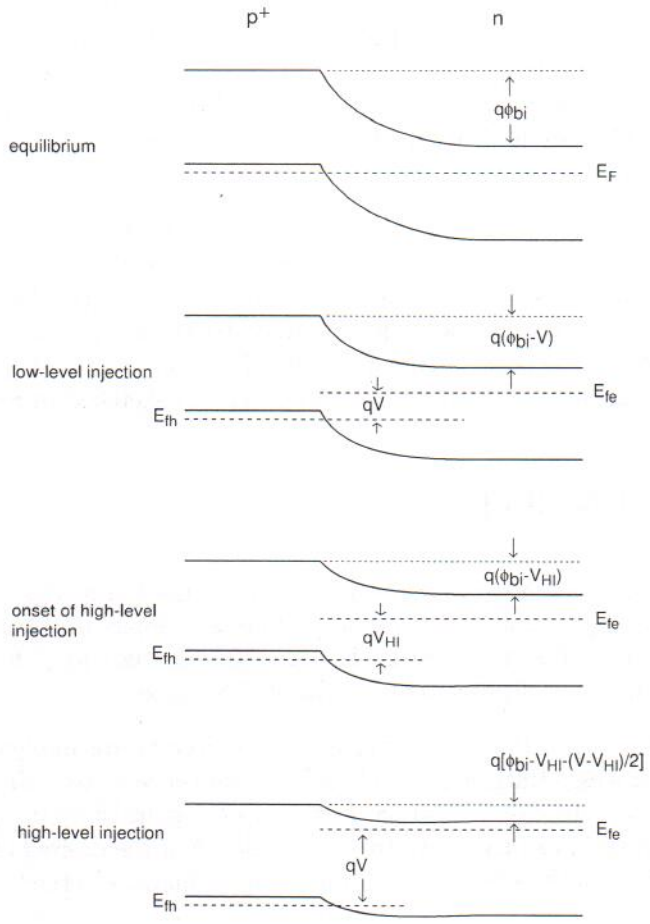


Figure 6.42: Energy band diagram of a PN junction as a function of injection level. In the low-level injection regime, the energy barrier across the junction is reduced in a proportional way to the applied voltage. In the high-level injection regime, the energy barrier is reduced only at half the rate of increase of the forward voltage.

in the high injection regime. This result can be easily derived from the expression of the np product inside the SCR given by Eq. 6.43. This equation remains valid in the high-level injection regime because it simply states that electrons and holes are in equilibrium among each other due to the fact that the SCR is very thin. At the edge of the SCR that goes into high-level injection first, $n \simeq p$, and:

$$n \simeq p \simeq n_i \exp \frac{qV}{2kT} \tag{6.122}$$

The nature of the diode current is still carrier recombination in the quasi-neutral regions. The boundary condition of Eq. 6.122 implies that the voltage dependence of the current will be of the form $e^{qV/2kT}$.

Physically, one can understand the origin of this result by examining the evolution of the

energy band diagrams around the SCR as the voltage increases. This is sketched in Fig. 6.42. The key is to examine the dependence of the energy barrier that carriers face in crossing the junction. In the low-level injection regime, as the voltage increases, the energy barrier is reduced by exactly qV. This is because the majority carrier quasi-Fermi levels are "rigidly" tied up to the respective band edges. This results from the fact that in low-level injection, the majority carrier concentrations remain unchanged.

As one of the sides of the diode goes into high-level injection, this is not the case any longer. The large minority carrier injection forces an increase in the majority carrier concentration. As a consequence, in high-level injection, the majority carrier quasi-Fermi level on the high-level injected side gets closer and closer to the majority carrier band edge (conduction band for a p^+-n diode). After this happens, the energy barrier presented by the junction is only reduced at half the rate of the quasi-Fermi level splitting (always qV). This is sketched in Fig. 6.42.

6.7 Integrated PN diode

PN diodes are present in many integrated circuits. They are used in rectifying circuits, as detectors in communications applications, and as bias shifters and input protection devices (against electrostatic discharge or ESD). This section is dedicated to studying a few important issues involved in the design and analysis of actual integrated PN diodes.

In spite of their usefulness, PN diodes in integrated circuits are implemented without the luxury of a dedicated process. Rather, some of the existing process steps that are used to make the n- and p-MOSFETs in the case of CMOS, or the bipolar transistors in the case of a bipolar process, must be combined in creative ways to achieve diodes of the desired characteristics. This brings in some interesting issues. We discuss the most important ones in the following subsections.

6.7.1 Isolation

A first issue to think about is device isolation. In integrated electronics, multiple devices sit side by side on a common substrate. For circuits to operate properly, all these devices must be properly isolated from each other. This is particularly hard with the devices packed tightly as is desired in very dense integrated circuits. There are several techniques to achieve device isolation in ICs. One of them is using trenches dug into the substrate and filled with a dielectric. A second way is using a PN junction. This is called *junction isolation*. A combination of trenches and junction isolation can be seen in the diode of Fig. 6.43 which is built on a bipolar process (more about this in Ch. 11).

In this diode, the p-type base of the transistor is used as anode and the n-type collector is used as cathode. Isolation with other devices is provided through a dielectric trench that rings around the sides and a substrate PN junction at the bottom. Using a PN junction for isolation makes sense since we know that in equilibrium or under small reverse bias, a PN junction has negligible current flowing through it. PN junction isolation is very widely used in IC design. We can see it in the integrated resistor design of Fig. 5.11 and in the CMOS transistor pair of Fig.

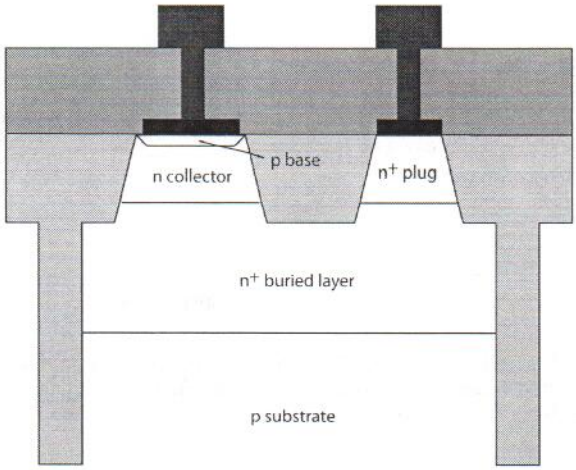


Figure 6.43: Sketch of integrated Si PN diode fabricated in a BJT process.

6.2.

For a substrate PN junction in which the substrate is p-type, reverse-bias can be assured if the substrate is connected to the most negative voltage available in the circuit (if the substrate was n-type, the most positive voltage would be used). In any case, the substrate junction contributes a parasitic depletion capacitance that must be accounted for when modeling this device.

The main challenge in implementing integrated PN diodes is the presence of a parasitic "bipolar" transistor that can compromise isolation. The problem is illustrated in Fig. 6.44 for a PN diode implemented on a CMOS process on a p-type substrate. In this case, the most negative voltage available in the circuit is applied to the substrate. As a consequence, the cathode/substrate PN junction is reverse biased or at zero volts. Under these conditions, a parasitic "bipolar" action can take place between the two PN junctions. The details of the problem need to be postponed until we study the bipolar transistor in Ch. 11. Basically, with the anode-cathode junction under forward bias, holes are injected downwards into the cathode. If the hole diffusion length in the cathode is shorter than the cathode depth, all the holes recombine in the cathode. This is the ideal case that we have studied so far. However, if the hole diffusion length is comparable or longer to the depth of the cathode, some of these holes will reach the back of the n-region. With the cathode/substrate junction in reverse bias, the high electric field prevalent in its SCR "extracts" or "collects" holes away into the substrate where they are majority carriers. The result is that a fraction of the diode current is being diverted to the substrate. This is an undesirable current path that if unaccounted for, can disrupt or impede proper operation of the diode in its circuit.

A solution to this problem in a CMOS process is actually very difficult. In a typical CMOS process, the wells have depths that are much shorter than the minority carrier diffusion lengths and that means that a great deal of the minority carriers injected into the well will be collected by the substrate instead of recombining inside the well. Integrated diodes just simply cannot be implemented in CMOS. The only exception are diodes that have a terminal at the same voltage as the substrate. Obviously, in this case, current diversion to the substrate is not an issue.

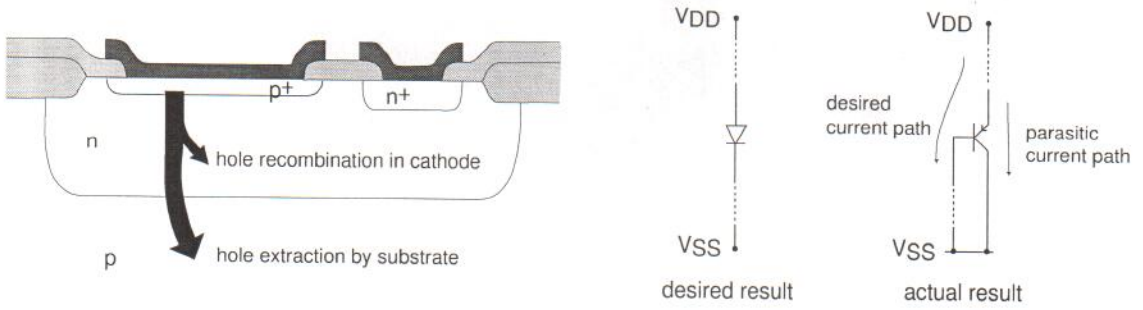


Figure 6.44: Sketch of isolation problem in integrated PN junction. If the hole diffusion length in the anode is longer than its depth, holes injected from the cathode can be extracted by the substrate in a *bipolar-type effect*. This is undesirable in most circuits.

In a bipolar transistor process the problem much easier to handle. In this case, as shown in Fig. 6.43, the presence of the n^+-buried collector effectively prevents hole transport down to the substrate. There are two reasons for it. First is the potential barrier associated with the $n - n^+$ high-low junction that confines holes to the n-side. Second is the small diffusion length prevalent in the buried layer as a consequence of its high doping level. Any hole that is injected into the buried layer will recombine within it before it can diffuse down towards the substrate. If properly designed, diodes such as that of Fig. 6.43 can prevent injected holes from reaching the n^+-p substrate junction.

6.7.2 Series resistance

A second issue to consider in a practical integrated diode is its series resistance. In an actual diode both contacts are made from the top surface. This allows the diode to be connected to other devices on the substrate. Unlike in an ideal PN diode, in a real integrated diode current flow is two dimensional. For example, in the diode of Fig. 6.43 electron majority carrier current in the cathode enters through the top contact, goes down the n^+ plug, then turns sideways to flow through the buried layer and then turns around again to eventually flows upwards towards the intrinsic device. In a situation like this, how does one estimate the series resistance?

A sketch of the parasitic resistance components in the device of Fig. 6.43 is shown in Fig. 6.45. There are essentially eight resistance components. Starting from the contact to the region, we have:

- contact resistance to p-region, R_{cp},
- body resistance of p-region, R_p,
- body resistance of n-region, R_n,
- spreading resistance of n^+-buried layer, R_{b1},
- lateral resistance of n^+-buried layer, R_{b2},

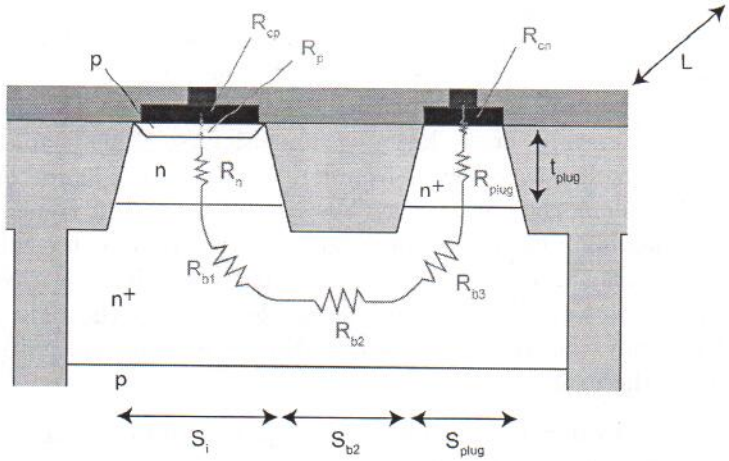


Figure 6.45: Parasitic resistances associated with integrated Si PN diode fabricated in a BJT process.

- spreading resistance of n^+-buried layer, R_{b3},
- n^+-plug region resistance, R_{plug}, and
- contact resistance to n-region, R_{cn}.

Of these, R_{cp}, R_{cn}, R_p and R_n are computed as indicated earlier in this chapter for the ideal diode. New resistances are those associated with the n^+-plug, R_{plug}, and the n^+-buried layer, $R_{b1} + R_{b2} + R_{b3}$.

R_{plug} can be easily computed simply by knowing the geometry of the plug and its resistivity, ρ_{plug} (we assume the doping level to be uniform). If we denote the depth of the plug as t_{plug}, its width as S_{plug}, and its length (which is also the length of the entire diode) as L then we have:

$$R_{plug} \simeq \rho_{plug} \frac{t_{plug}}{S_{plug} L} \tag{6.123}$$

The resistance associated with the buried layer is more complex because the current flow is eminently two dimensional. We can simplify the problem if we consider that the lateral dimensions of a typical diode are much larger than the vertical dimensions. In consequence, we can neglect the component of resistance associated with vertical current flow in the buried layer and just focus on the lateral resistance. In a problem like this, it suffices to characterize the buried layer through its sheet resistance R_{shb}.

Of the three components of the buried layer resistance, R_{b2} is easy to compute. In this region, electron flow is essentially one dimensional. We can then write:

$$R_{b2} \simeq R_{shb} \frac{S_{b2}}{L} \tag{6.124}$$

We could think of using a similar approach to compute the resistance associated with lateral carrier flow through the two other regions of the buried layer, but doing this would overestimate it. This is because the total diode current does not flow through the entire lateral extent of these regions S_i and S_{plug}. For example, in the intrinsic device region, we can assume that the diode current flows downwards uniformly across the entire dimension S_i. When it hits the buried layer, it turns around and it starts flowing sideways from left to right in the figure. As a result, as we advance from left to right along the buried layer, the current through it is increasing in a linear way until it reaches its maximum value (the total diode current) right at the edge of the intrinsic region on the right hand side. Similarly, the lateral current through the buried layer under the plug region drops linearly as we advance from its left edge towards the right as the current is deviated upwards. Obviously, the resistance associated with these current paths is *less* than the geometrical resistance of the regions.

Solving this problem is not difficult but it requires some careful work. We postpone a solution until we treat the base resistance in a bipolar transistor which is a similar problem. This is done in Sec. 11.5.6. The result is actually quite simple. For a problem like this, the lateral resistance is one third of the geometrical resistance. Then:

$$R_{b1} \simeq \frac{1}{3} R_{shb} \frac{S_i}{L} \tag{6.125}$$

$$R_{b3} \simeq \frac{1}{3} R_{shb} \frac{S_{plug}}{L} \tag{6.126}$$

With this, we now have a complete first order model for the series resistance of an integrated diode.

6.7.3 High-low junction

An additional important issue that emerges in an integrated PN diode implemented in a bipolar process derives from the presence of the n^+ buried layer below the n-type region and the boundary condition that this represents for holes injected into this region. In a diode of this kind, there is a sudden increase in doping level inside the n-type region. This is what is known as a *high-low junction*. Its presence can have a substantial impact on the minority carrier distribution in the n side of the diode under forward bias and therefore on the diode current.

The relevant case to study is that of a "short" diode in which the diffusion length of holes in the n-type region is much longer than the extent of this region. It is in this case that injected holes diffuse downwards and reach the high-low junction.

Qualitatively, it is easy to appreciate the impact of a high-low junction on minority carrier flow by drawing an energy band diagram of the structure in equilibrium. This is shown in Fig. 6.46. The sudden increase of n-type doping level across the high-low junction brings the conduction band closer to the Fermi level. This creates an energy barrier for holes in the valence band that acts to prevent hole injection into the n^+-region. Therefore, with the diode in forward bias,

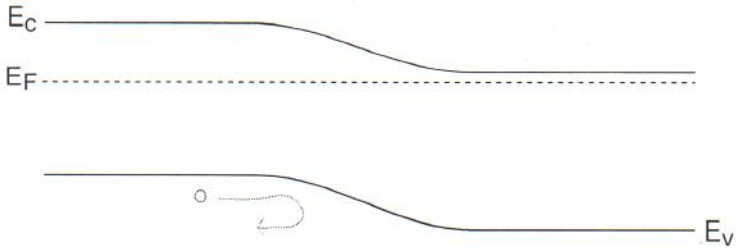


Figure 6.46: Energy band diagram of an n-n$^+$ high-low junction in equilibrium. For holes in the n-region, the high-low junction appears as an energy barrier that blocks their flow into the n$^+$ region.

when holes are injected into the n-type region, they diffuse away from the junction and reach the high-low junction where they pile up.

Intuitively we can then understand that a high-low junction exerts a blocking action for the minority carriers. It confines them to the lightly doped side of the junction where carrier lifetimes tend to be long. In consequence, minority carrier current tends to be suppressed.

Let us now develop a model for the situation. Let us assume the simple situation depicted in Fig. 6.47. This figure shows the n-side of a PN diode with a doping level distribution that abruptly changes from N_D to N_D^+ (with $N_D^+ > N_D$) at $x = x_{hl}$. The lightly doped region is short in the scale of the relevant minority carrier diffusion length. The highly doped side, on the other hand, is long. Let us label the diffusion length there as L_h^+.

Under forward bias, holes are injected into the n-side of the diode and they diffuse in. The presence of the high-low junction restricts the number of holes that can overcome the energy barrier there to a small number. These diffuse and recombine in the highly doped region and the excess hole concentration drops with a characteristic exponential shape with L_h^+ as the decay length. In the lightly-doped side of the junction, the profile is a straight line since this region is short in the scale of the diffusion length.

In order to solve this problem, we need to first think about the boundary conditions at the high-low junction. The first one is about the excess hole concentration. It is clear that there is a sudden drop as we cross the junction. We can think of this in two ways. At the high-low junction, there is a potential barrier of magnitude:

$$\Delta E_{hl} = kT \ln \frac{N_D^+}{N_D} \tag{6.127}$$

In consequence, the ratio of the hole concentration across the junction is:

$$p(x_{hl}^+) = p(x_{hl}^-) \exp(-\frac{\Delta E_{hl}}{kT}) = p(x_{hl}^-) \frac{N_D}{N_D^+} \tag{6.128}$$

A second way to think about this is to assume quasi-equilibrium across the high-low junction. What do we mean by this? This means that a single hole quasi-Fermi level completely describes the hole distribution in the vicinity of the high-low junction. In this case, we can assume that

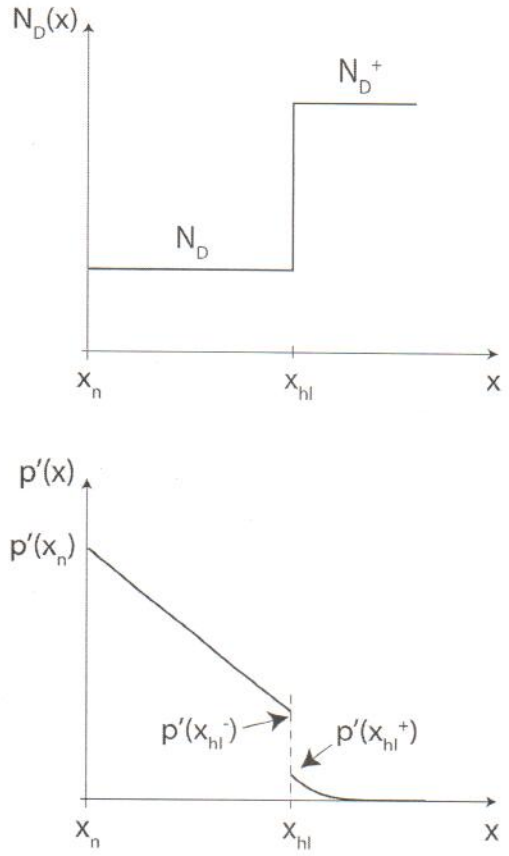


Figure 6.47: Doping distribution and excess hole concentration across high-low junction.

the pn product does not change across the junction. Since the n-type doping changes from N_D to N_D^+, the hole concentration must drop by the same ratio. That is exactly what Eq. 6.128 says.

The second boundary condition at $x = x_{hl}$ is that of continuity of hole current. Over the very short distance from $x = x_{hl}^-$ to $x = x_{hl}^+$, hole recombination is negligible. Then, $J_h(x_{hl}^-) = J_h(x_{hl}^+)$. This implies that:

$$D_h \frac{dp'}{dx}|_{x_{hl}^-} = D_h^+ \frac{dp'}{dx}|_{x_{hl}^+} \tag{6.129}$$

where D_h and D_h^+ represent the diffusion coefficients for holes in the n- and n^+-regions respectively. In general, these are not equal since the doping level is not the same.

This is important, for example, in a case in which there is an ohmic contact at the other side of the high-low junction. In the absence of a high-low junction, excess minority carrier recombination at the ohmic contact is very efficient. Inserting a high-low junction in the path of the minority carriers blocks their flow towards the ohmic contact and reduces the overall recombination current.

This is very important in minimizing recombination as in solar cells where high-low junctions are widely used.

With these two boundary conditions at x_{hl}, the problem can be solved quickly. The hole current density injected into the n-region is continuous all the way to $x = x_{hl}^+$. Then, at this point, we have:

$$J_h = q\frac{D_h^+}{L_h^+}p'(x_{hl}^+) = q\frac{D_h^+}{L_h^+}\frac{N_D}{N_D^+}p'(x_{hl}^-) \tag{6.130}$$

where we have used Eq. 6.128. In the n-type region, we can write:

$$J_h = qD_h\frac{p'(x_n) - p'(x_{hl}^-)}{x_{hl} - x_n} \tag{6.131}$$

Equating these two expressions of J_h allows us to solve for $p'(x_{hl}^-)$. Inserting this result into either one of them, finally yields the following expression for the hole current:

$$J_h = q\frac{n_i^2}{N_D}\frac{D_h}{x_{hl} - x_n}\frac{1}{1 + \frac{D_h}{D_h^+}\frac{L_h^+}{x_{hl}-x_n}\frac{N_D^+}{N_D}}(\exp\frac{qV}{kT} - 1) \tag{6.132}$$

If we compare this result with that of the short diode with an ohmic contact placed at x_{hl}, obtained in Sec. 6.6.1, we can see that the hole injection current has been suppressed by the factor in the denominator of the third fraction. If $N_D^+ \gg N_D$, as often in practice, this factor is large and the injection current is suppressed in a very significant way. This is important in solar cells and other applications where high-low junctions are widely used.

6.8 Summary

- In a PN junction, a space-charged region forms around the metallurgical interface with a volume charge density that is set by the ionized dopant concentration. The space charge region is surrounded by two quasi-neutral regions.
- In thermal equilibrium, there is an energy barrier that appears for holes and electrons across the metallurgical junction. The height of this barrier is modulated when voltage is applied. This results in excess minority carrier concentrations at the edges of the SCR that go as $e^{qV/kT} - 1$. This is the origin of the rectifiying behavior of a PN diode.
- In forward bias, the current in an ideal diode is dominated by minority carrier injection and recombination in the quasi-neutral regions. In reverse bias, the current is dominated by minority carrier extraction and generation in quasi-neutral regions.
- The width of the space charge region is also modulated by the applied voltage. This yields a charge storage effect that is described by the depletion capacitance.

- In a forward bias PN junction, carrier storage takes place in the quasi-neutral regions. This also introduces a capacitive effect described by the diffusion capacitance.
- For strong forward bias, diffusion capacitance dominates. For moderate forward bias or reverse bias, depletion capacitance dominates.
- Space-charge region generation and recombination currents are frequently observed, respectively, in reverse bias and small forward bias.
- In most PN diodes, breakdown at high reverse bias occurs due to an avalanche process. The breakdown voltage is set by the doping level.
- A parisitic bipolar effect compromises the isolation of integrated PN diodes and makes it difficult to synthesize PN diodes in a CMOS process.

6.9 Further reading

The treatment of the PN diode is fairly standard in most textbooks. The approach followed here mirrors this well established path. It is difficult to find any books that offer something substantially different or that goes beyond the treatment presented here.

Semiconductor Device Fundamentals by R. F. Pierret, Addison-Wesley, Dover, 1996. (ISBN 0-201-54393-1, TK7871.85.P484). The description of the PN diode in this book is similar to that of the present text. The presentation is clear and emphasizes physical intuition. There are also a number of practical problems that do a good job in placing the PN junction and PN diodes in the real world. Pierret goes further than the present text in a number of areas such as the admittance of the PN diode as a function of frequency as well as a more detailed study of the turn on and turn off transitories.

AT6.1 Quasi-neutral region resistance in ideal diode

In Sec. 6.6.3 we wrote expressions for the resistance associated with the quasi-neutral regions of a long diode. Deriving them is the goal of this section. The derivation represents a great review of key materials from Ch. 5.

We do the analysis for the n-type QNR which extends from $x = x_n$, the edge of the SCR, to $x = w_n$, the ohmic contact. We follow an approach similar to that of Sec. 5.6.1. As we do this, we will think of carrier injection with $V > 0$, but our development also includes carrier extraction for $V < 0$.

In essence, what we need to accomplish is to derive an expression for the excess electric field, $\mathcal{E}'$, that appears across the n-QNR and then integrate this to obtain the voltage drop. The origin of $\mathcal{E}'$ in Sec. 5.6.1 was to balance out currents in a situation in which $D_e \neq D_h$, as is often the case. Here, we will also find that $\mathcal{E}'$ is also required to support the diode current through majority carrier drift.

The starting point is the excess minority carrier profile (holes) across the n-QNR. This was derived in Eq. 6.33. The hole current density associated with them is:

$$J_h \simeq J_{h,diff} = q\frac{D_h}{L_h}p'(x_n)\exp\frac{-x+x_n}{L_h} \tag{6.133}$$

The excess majority carrier profile (electrons) is to the first order identical to the excess minority carrier profile. This leads to an electron diffusion current given by:

$$J_{e,diff} = -q\frac{D_e}{L_h}p'(x_n)\exp\frac{-x+x_n}{L_h} \tag{6.134}$$

The total current density at any point is I/A. Then the electron drift current is given by:

$$J_{e,drift} \simeq q\mu_e N_D\mathcal{E}' = \frac{I}{A} - J_{e,diff} - J_h = \frac{I}{A} + q\frac{D_e - D_h}{L_h}p'(x_n)\exp\frac{-x+x_n}{L_h} \tag{6.135}$$

where we have assumed that $n \simeq N_D$ consistent with our low-level injection assumption.

We can now solve for $\mathcal{E}'$ and integrate across the n-QNR to obtain the ohmic drop there:

$$V_n = -\int_{x_n}^{w_n}\mathcal{E}'dx = -\frac{I(w_n - x_n)}{Aq\mu_e N_D} - \frac{D_e - D_h}{\mu_e N_D}p'(x_n) \tag{6.136}$$

From Eqs. 6.37 and 6.41 we can write:

$$p'(x_n) = \frac{I_n L_h}{qAD_h} \tag{6.137}$$

Substituting this into Eq. 6.136 we obtain:

$$V_n = -\frac{I(w_n - x_n)}{Aq\mu_e N_D} - \frac{I_n L_h}{Aq\mu_e N_D}\frac{D_e - D_h}{D_h} \tag{6.138}$$

To the extent that I_n is a fraction of I and L_h is much smaller than $w_n - x_n$, the second term in Eq. 6.138 is much smaller than the first term and the ohmic drop that is obtained for the n-QNR is consistent with the expression for its resistance that we wrote in Eq. 6.88. Once again, if D_e was precisely equal to D_h, then this expression would be rigorously correct. To the extent that $D_e \neq D_h$, a small electric field needs to be set up to keep the semiconductor quasi-neutral. By the way, the negative sign in 6.138 simply is a reflection of the fact that in our notation, the p-region, located on the left, is positive with respect to the n-region located on the right. The axis is pointing in the direction of decreasing voltages.

AT6.2 Equivalent circuit model for circuit design

Circuit computer-aided design (CAD) tools, such as SPICE, capture the behavior of a PN diode in an equivalent circuit model that is described by a system of equations. These models tend to be proprietary and differ somehow from company to company. In a given CAD system, a device equivalent circuit is "coded in," that is, its topology cannot be changed. Typically, a CAD system offers different topologies suitable for different technological implementations. The value of the elements of the equivalent circuit model can be adjusted by means of *model parameters* (in **bold** below) selected to best reproduce the I-V and C-V characteristics and any other measurements that are taken on the diodes. A successful CAD model for a device reproduces the experimental characteristics accurately, is computationally efficient and the parameters that describe the model can be easily extracted from experiments. Effective CAD models tend to be strongly based on physical models.

CAD models can get quite complex as engineers attempt to describe subtle but important features of the device, strive for their models to predict the device characteristics over a broad range of temperatures, and allow them to appropriately scale for different device layouts. In this section, we describe a typical, though relatively simple, CAD model for a PN diode.

Typically, the I-V characteristics are captured by:

$$I = I_S(\exp\frac{qV_j}{\mathbf{N}kT} - 1) \tag{6.139}$$

where V_j is the internal voltage applied across the ideal diode. V_j is computed by the CAD program in the course of the simulations after accounting for the series resistance modeled using model parameter **RS**. V_j is computed from the external voltage V by means of:

$$V_j = V - I\,\mathbf{RS} \tag{6.140}$$

Equation 6.139 contains several parameters. **N** is a model parameter called the *ideality factor* or *emission coefficient* and it should be unity for an ideal PN diode. Its value can be adjusted however to accommodate non-idealities. The value of the physical constants q and k are programmed in the simulator. The temperature T is specified by the user.

I_S in Eq. 6.139 plays the role of the diode saturation current. It is specified at any temperature by three model parameters through the following expression:

$$I_S = \mathbf{IS}\left(\frac{T}{T_M}\right)^{\mathbf{XTI}} \exp[\frac{-q\ \mathbf{EG}}{k}(\frac{1}{T} - \frac{1}{T_M})] \tag{6.141}$$

IS is a model parameter that represents the saturation current measured at a specified temperature T_M. The parameters **EG** and **XTI** account for the temperature dependence of the prefactor of I_S in the PN diode current. This arises mainly from the temperature dependence of n_i^2 which we know goes as:

$$I_S \propto n_i^2 \propto T^3 \exp -\frac{E_g}{kT} \tag{6.142}$$

Hence, the energy bandgap parameter **EG** should to first order have a value equal to the bandgap of the specific semiconductor (hence the name of this parameter) and the exponent **XTI** should be 3. To the second order, however, **EG** is a bit higher than the semiconductor bandgap due to the temperature dependence of E_g (see Sec. AT2.1) and **XTI** is not exactly 3 due to the temperature dependence of the rest of the parameters in I_S, such as mobility and carrier lifetime.

The junction capacitance in a CAD model of the PN diode is typically modeled as:

$$C_j = \frac{\mathbf{CJO}}{(1 - \frac{V_j}{\mathbf{VJ}})^{\mathbf{M}}} \tag{6.143}$$

CJO is the capacitance of the junction at zero bias. **VJ** is the junction *built-in potential* [4]. **M** is the so-called *grading coefficient*. For the uniformly doped case studied in this book, **M**= 0.5. This value can be adjusted if the doping profile is not flat. A separate perimeter capacitance term is frequently available with different voltage dependence. This comes handy to accurately model PN diodes with different area to perimeter ratios fabricated on the same process. Note also that Eq. 6.143 blows up if V_j exceeds **VJ**. This often results in convergence problems in the course of circuit simulation. Different simulators have different ways of dealing with this problem.

In addition to the junction capacitance, there is the diffusion capacitance that accounts for minority carrier storage in the PN diode. This is modeled in the following way:

$$C_d = \frac{q}{kT}\mathbf{TT}\, I_S \exp\frac{qV_j}{kT} \qquad \text{for } V_j \geq 0 \tag{6.144}$$

[4] **VJ**, the circuit model parameter for the junction built-in potential, should not be confused with V_j, the internal voltage across the edges of the PN diode space-charge region.

name	parameter description	units	ideal value
IS	saturation current	A	-
N	ideality factor	-	1
EG	energy bandgap	V	$E_g(300\ K)$
RS	series resistance	Ω	-
CJO	zero bias depletion capacitance	F	-
VJ	built-in potential	V	-
M	grading coefficient	-	0.5
XTI	saturation current temperature exponent	-	-
TT	transit time	s	-
BV	breakdown voltage	V	-

Table 6.1: Parameters for PN-diode model for circuit CAD.

where V_j is the internal voltage across the PN junction. The parameter **TT** is called the *transit time*. By comparing Eq. 6.144 with Eqs. 6.54 and 6.66, we can determine the relationship between the CAD parameter **TT** and the physics of the device:

$$\mathbf{TT} = \frac{\tau_e I_{ps} + \tau_h I_{ns}}{I_s} \tag{6.145}$$

where I_{ps} and I_{ns} represent the contributions to the saturation current originating from the p- and the n-regions, respectively and τ_e and τ_h represent the carrier lifetimes in these regions.

If these regions are short in comparison with the diffusion length, then the dominant time constant is the transit time. The form of Eq. 6.144 does not need to change to accommodate this or other more complex situations.

The total capacitance of the PN diode is given by the sum of the diffusion and the depletion capacitances.

Table 6.1 summarizes the most important PN-diode parameters for circuit CAD. The table also gives the ideal value of these parameters according with the models presented in this Chapter.

AT6.3 Switching characteristics of PN diode

Diodes are attractive devices for switching applications. These are situations, such as in electrical power management, in which the diode is supposed to change from a highly-conducting ON state to a highly blocking OFF state. The transition often needs to be very fast. In these kinds of applications, a fundamental weakness of the PN diode emerges. Typical PN diodes are quite slow, that is, it takes a significant length of time for the transitories associated with the switching events to be extinguised. In contrast, as we will see in the next chapter, Schottky diodes offer a very significant advantage and are the device of choice for many of these applications.

The main reason for the slow switching of the PN diode is that in forward bias, there is a

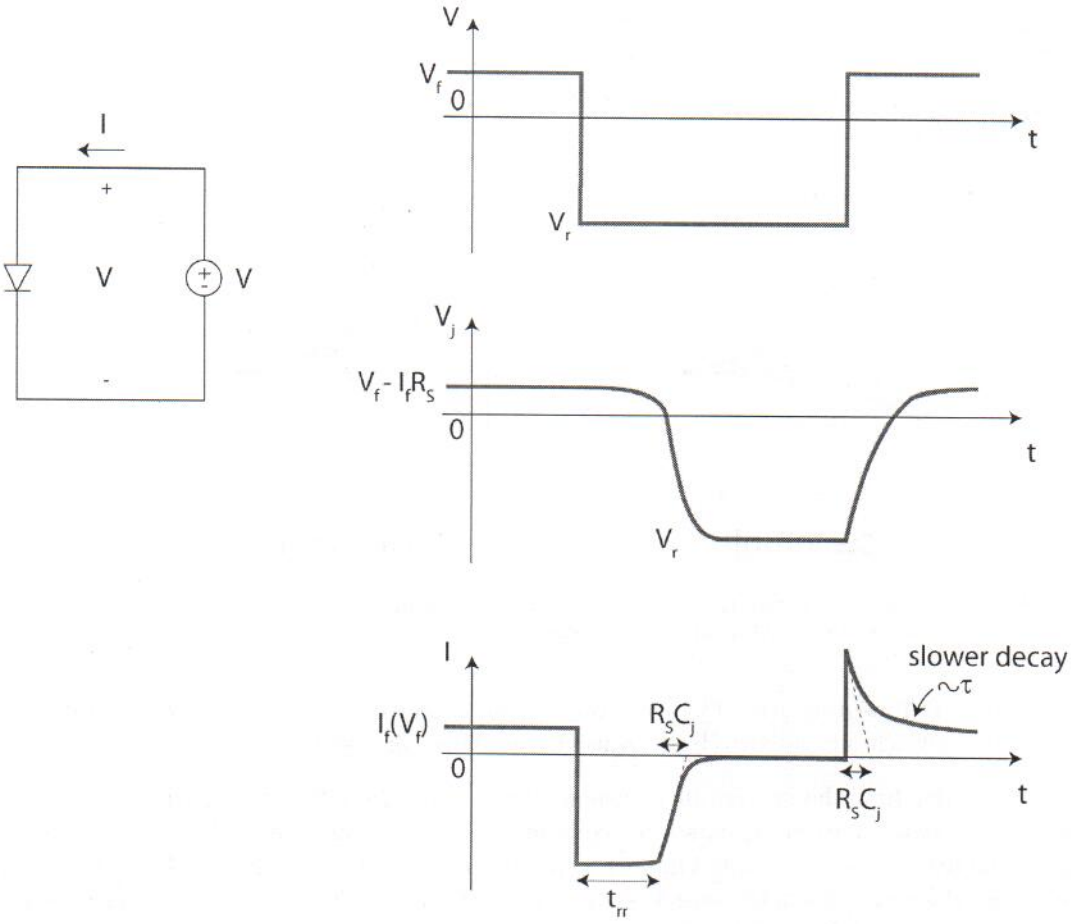


Figure 6.48: Switching transients in a PN diode. In addition to the R_sC_j time constants, the recovery of the PN diode is delayed by the need to extract (in the switch-off transient) or supply (in the switch-on transient) minority carrier charge to the quasi-neutral regions.

substantial amount of minority carrier charge stored in the device. This is negligible in reverse bias. In order for the PN diode to switch from a forward state to a reverse state, the minority carrier charge must be disposed off. Similarly, in switching from a reverse state to a forward state, minority carrier charge needs to be provided. Depending on the details of the diode and the outside circuit, it can take significant time to achieve steady state after a switching event.

In order to appreciate the significance of this, consider the case illustrated in Fig. 6.48 in which a PN diode is switched back and forth between a forward voltage V_f and a reverse voltage V_r (note that, defined this way, V_r is negative). Under DC conditions, the forward state is characterized by a large forward current flow, $I_f(V_f)$, while the reverse state is characterized by small reverse current, $I_r(V_r)$.

Let us further assume that this is an abrupt diode where one side (the lowly-doped one) dominates its minority carrier behavior and that this side is "long" in the scale of the relevant

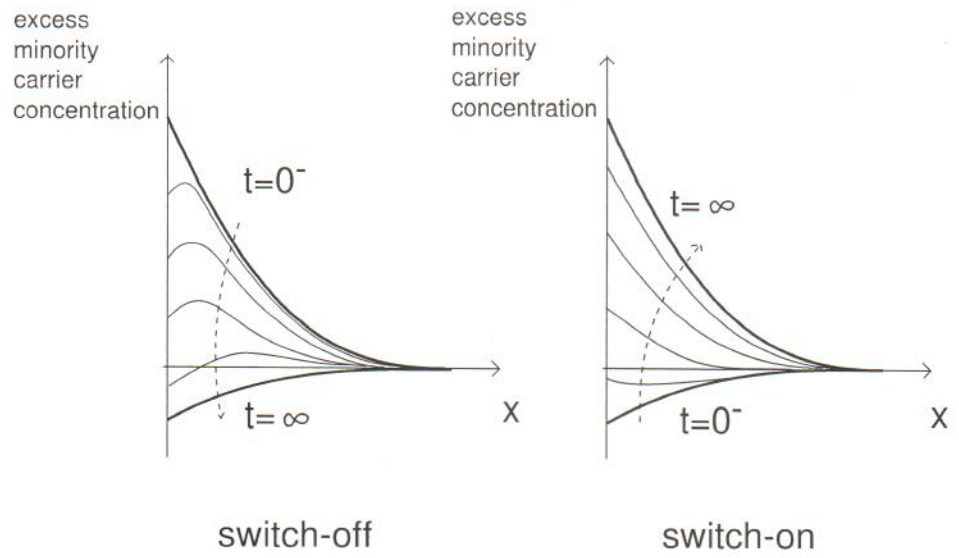


Figure 6.49: Evolution of the excess minority carrier concentrations in the quasi-neutral region of a PN diode during a switch-off transient (left) and a switch on transient (right).

minority carrier diffusion length. The evolution of the excess minority carrier concentration in time during the switch-on and switch-off transients is shown in Fig. 6.49.

Let us consider first the switch off transient depicted on the left of Fig. 6.49. At $t = 0^-$, a steady-state forward current I_f flows. In consequence, if we denote as τ the minority carrier lifetime in this region, the total stored minority carrier charge is $Q = \tau I_f$. At $t = 0^+$, the external voltage across the diode abruptly switches from V_f to $V_r < 0$. The internal voltage, however, cannot change immediately since for that to happen, the excess minority carrier concentration at the edge of the quasi-neutral region must change too. This can only be accomplished by one of two ways. One is to wait until they recombine. A second one is to make them flow out the diode. The fastest of these mechanisms controls the minority carrier discharge of the diode.

Minority carrier recombination is characterized by the carrier lifetime τ. Minority carrier discharge through the outside circuit is limited by the series resistance R_s of the diode. As the circuit schematic in Fig. 6.50 suggests, right after switching, at $t = 0^+$, the current through the diode suddenly reverses sign and becomes:

$$I(t = 0^+) = \frac{V_r - (V_f - I_f R_S)}{R_s} \simeq \frac{V_r}{R_s} \tag{6.146}$$

Here we have assumed that $V_f \ll |V_r|$. Note that $I(t = 0^+)$ is negative.

For $t > 0^+$, the minority carrier charge discharges through the resistor. Because of the exponential dependence of minority carrier charge and junction voltage, as the discharge proceeds, V_j initially changes very slowly. As a result, as Fig. 6.48 indicates, $I(t > 0^+)$ essentially remains unchanged from the value given in Eq. 6.146. Only when the minority carriers have bled off, the

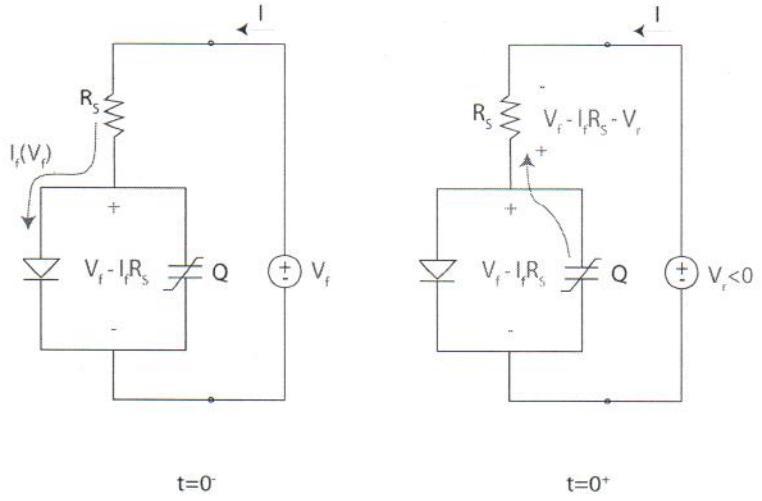


Figure 6.50: Internal and external currents in a PN diode right before and after a switch-off event.

junction voltage will reach zero voltage and eventually start evolving towards V_r.

This understanding allows us to estimate, to the first order, the time that it takes for the minority carrier charge to bleed off. This is given by the ratio of the stored charge to the discharge current:

$$t_{rr} \simeq \frac{Q}{|I(t=0^+)|} \simeq \tau \frac{I_f(V_f)R_s}{|V_r|} \tag{6.147}$$

This is called the *reverse recovery time*. Note that in a well designed diode, the ohmic drop across the series resistance in forward bias, $I_f R_S$, is much smaller than the forward voltage V_f. Since we are assuming that $V_f \ll |V_r|$, it then follows that $I_f R_S \ll |V_r|$. As a consequence, $t_{rr} \ll \tau$. Getting the minority carriers out through the external circuit is faster than waiting for them to recombine.

When the internal voltage of the diode goes through zero, the excess minority carrier discharge is finished. From there on, the discharge of the junction capacitance is what limits the transient characteristics of the diode. This is a simple exponential process characterized by a time constant $R_s C_j$.

All together, the switch off transient evolves as sketched in Fig. 6.48. For a period of time equal to t_{rr}, a fairly constant reverse current flows, after which an exponential decay to the final steady state value follows.

The dependencies observed in Eq. 6.147 make sense. The higher τ or $I_f(V_f)$, the higher the amount of minority carrier charge that is stored and the longer it takes to bleed it off. The higher R_s or the lower $|V_r|$, the lower the magnitude of the discharge current during the recovery period $|I(t=0^+)|$ and the longer it takes for the discharge to be completed.

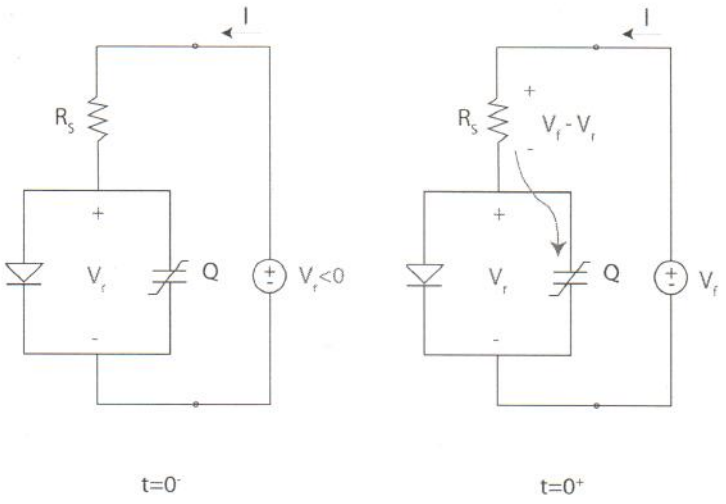


Figure 6.51: Internal and external currents in a PN diode right before and after a switch-on event.

Let us now shift our attention to the turn-on transient (Fig. 6.51). When the external voltage abruptly switches from V_r to V_f, the junction capacitance starts charging through R_s. This produces a peak of forward current of a magnitude:

$$I(t=0^+) = \frac{V_f - V_r}{R_S} \simeq \frac{|V_r|}{R_S} \tag{6.148}$$

As the junction capacitor charges up and the internal voltage rises towards V_f, the forward current drops in an exponential way. The characteristic time constant of this process is of the order of $R_s C_j$, where C_j is again the junction capacitance. When the diode becomes forward biased, two things happen that delay the attainment of steady state. First, the diode starts to conduct. This robs current that could be used to charge up the capacitor. Second, the battery must supply extra current to store the required minority carrier charge to the quasi-neutral regions. This process has a characteristic time that is again on the order of the dominant time constant for minority carrier-type behavior, τ. It then takes a few τ's to attain the final steady state. The complete transient is sketched in Fig. 6.48. The temporal evolution of the excess minority carrier concentration is sketched in Fig. 6.49.

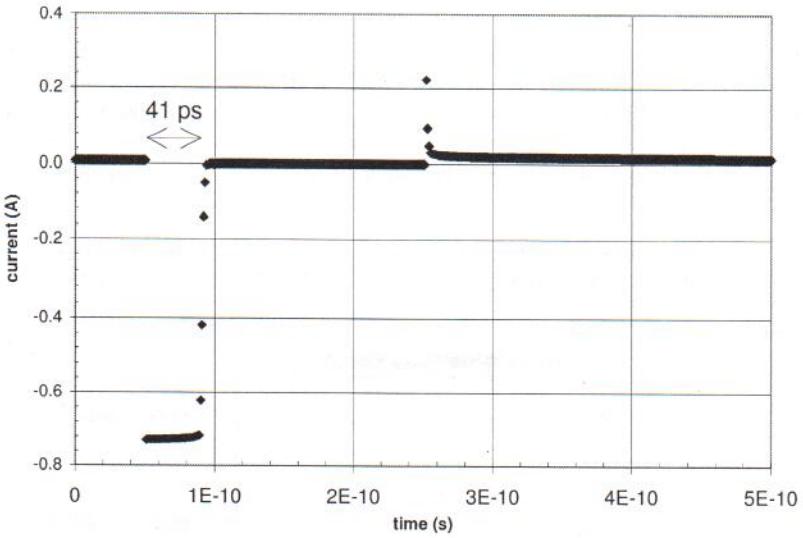


Figure 6.52: Switching transients in a PN diode as simulated by HSPICE (see exercise). For $0 \leq t \leq 50\ ps$, $I = I_f = 7.2\ mA$.

Exercise 6.6: *Simulate a switching transient of a PN diode in SPICE, as sketched in Fig. 6.48. The forward voltage is* $+0.89\ V$. *The reverse voltage is* $-5\ V$. *The SPICE parameters of the diode are:* **IS**$= 6e-17$, **N**$= 1$, **EG**$= 1.12$, **RS**$= 8$, **CJO**$= 1.2e-13$, **VJ**$= 0.6$, **M**$= 0.41$, **XTI**$= 3.814$, *and* **TT**$= 4e-9$. *Extract the characteristic time constants of the transients and compare them with simple estimates.*

Fig. 6.52 shows the output obtained from HSPICE for a case in which there is a delay time of 50 ps, the reverse voltage is applied during 200 ps and the forward voltage is applied during 250 ps.

The turn-off characteristics shows pronounced minority carrier storage. The reverse recovery time that is extracted from the simulations is about 41 ps. This compares very well with the value that can be calculated using Eq. 6.147, which is 46 ps (the DC forward current that corresponds to $V_f = 0.89\ V$ is 7.2 mA; in order to compute this number, you need to account for the series resistance of the diode). Following the reverse recovery time, the current through the diode is seen to drop very fast to the value that corresponds to the reverse bias of $-5\ V$. The expected value of the R_sC time constant is about 1 ps.

The turn-on transient displays an initial fast decay associated with the charging of the junction capacitance followed by a much slower tail towards the final $I_f(V_f)$ value. If you blow up this region and examine the time constant of this tail, you will find that it is about 2.6 ns. This compares rather well with the dominant time constant of the diode which is 4 ns.

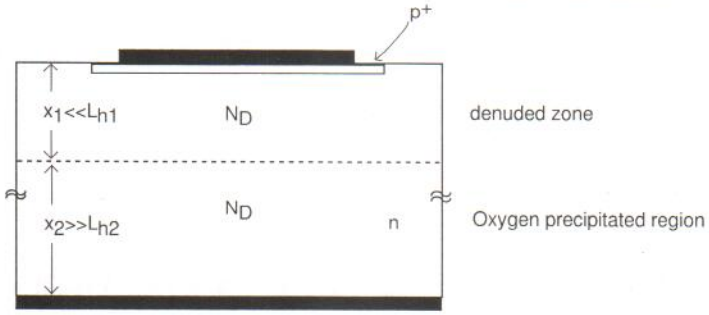
The reverse recovery current obtained in SPICE is about 730 mA. This compares well with the expected value of 736 mA. The peak current following the turn-on transient should also have this magnitude. However, the initial decay is so fast, that with the selected time discretization in SPICE this peak is not completely resolved.

This example dramatically illustrates how drastically the switching of a PN diode is slowed down by minority carrier storage.

Problems

6.1 Modern IC processing utilizes a technique called "Oxygen precipitation" to getter metallic contaminants out of a thin layer at the top surface of the wafer. It is in this so-called "denuded zone" where the devices are fabricated. Typically, the denuded zone is thin and is characterized by a fairly long lifetime, while the rest of the wafer where Oxygen precipitation has taken place exhibits much lower lifetimes. The denuded zone results in device reverse-bias leakage currents that are much reduced.

The change in lifetime in the wafer presents an interesting problem for the computation of the current injected downwards in a bulk p^+-n diode, as sketched in the figure below.

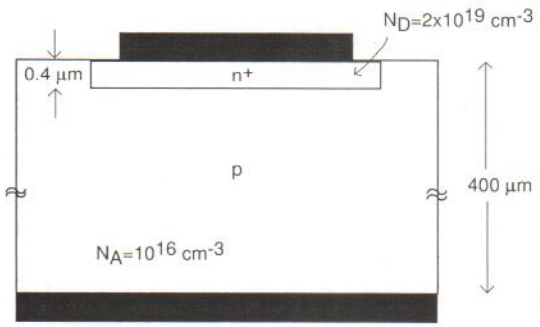


Make the following assumptions. The doping level is the same in both regions of the n substrate. The junction depth of the p^+ region and its associated depletion region are much thinner than x_1, the denuded zone thickness. x_1 is much shorter than the local diffusion length L_{h1}. The thickness of the Oxygen precipitated region x_2 is much longer than the local diffusion length L_{h2}.

a) Sketch the shape of the excess hole concentration in the n region in forward bias.

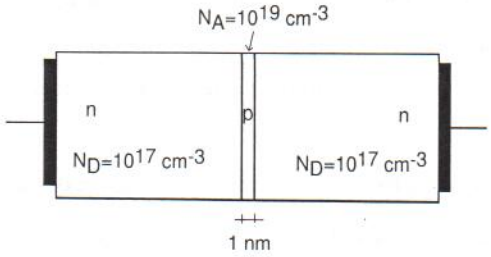
b) Derive an expression for the hole current density injected into the n-type substrate in forward bias in terms of the hole diffusion length in the Oxygen precipitated region, L_{h2}, the denuded zone thickness, x_1, and the usual parameters.

6.2 In the list of SPICE parameters that has been given to you for the abrupt PN diode sketched below, you read **TT** = 24 ns. This seems to you so way off that you think it is a typo.



Estimate **TT** for this diode.

6.3 Consider a Si device with a structure as sketched in the following diagram.



This device basically consists of a thin p-type region inserted in the middle of an n-type body. Two ohmic contacts are provided to the ends of the n regions. The device sits at room temperature.

a) Quantitatively sketch the volume charge density througout the structure in thermal equilibrium. Justify any assumptions you make.

b) Quantitatively sketch the electric field througout the structure in thermal equilibrium.

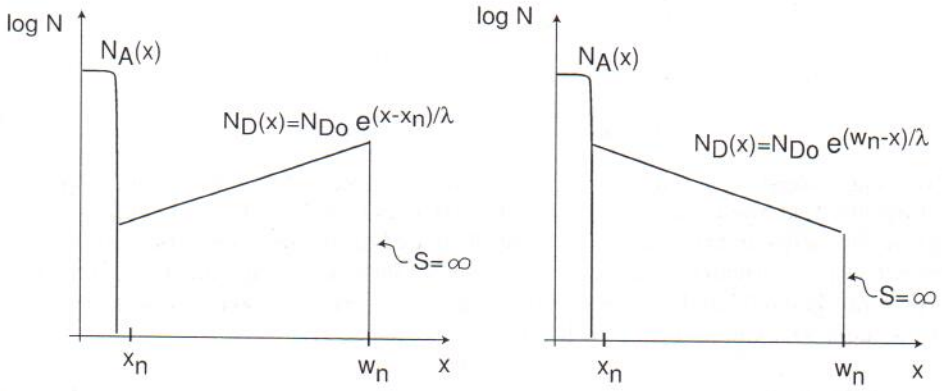
c) Quantitatively sketch the electrostatic potential throughout the structure in thermal equilibrium.

d) Quantitatively sketch the energy band diagram throughout the structure in thermal equilibrium.

e) Now apply a voltage of 0.1 V to the right contact with respect to the left contact. Explains what happens. Qualitatively sketch the electrostatics of the device (volume charge density, electric field, and electrostatic potential).

f) Quantitatively sketch the energy band diagram.

6.4 Consider two p^+-n junctions with non-uniform donor distributions as sketched below. Notice that the doping profiles are basically identical but flipped over in x. $x = x_n$ is the edge of the depletion region on the n-side and $x = w_n$ is the ohmic contact which is characterized by $S = \infty$. The extent of the quasi-neutral region on the n-side, w_n, is much shorter than the diffusion length for holes, *i.e.*, this is "short" region for the minority carriers. The doping levels are such that $x_n \ll w_n - x_n$.



Assuming that D_h is independent of position,

a) calculate expressions for the Gummel number and the hole saturation current density, $J_{hs}(x_n)$, injected into both n-regions;

b) for both profiles, sketch $J_{hs}(x_n)$ vs. $\lambda/(w_n - x_n)$, discuss limits when $\lambda \to \infty$;

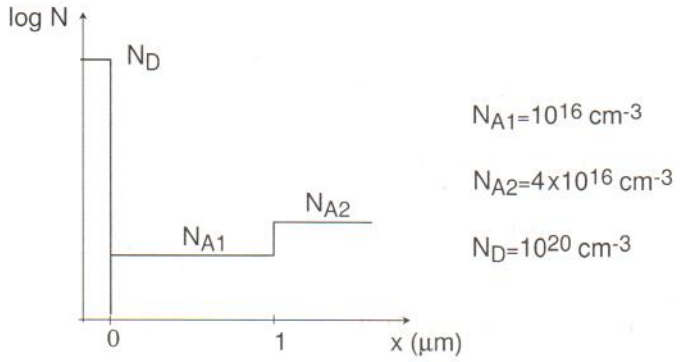
c) calculate expressions for the hole transit time, τ_{th}, across both n-regions;

d) for both profiles, sketch τ_{th} vs. $\lambda/(w_n - x_n)$, discuss limits when $\lambda \to \infty$;

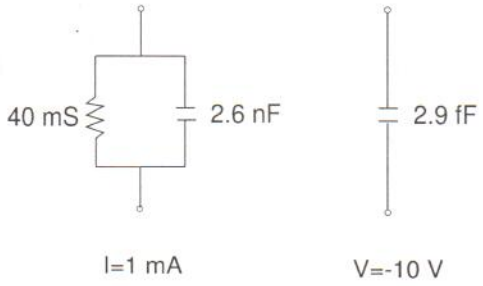
e) for both profiles, sketch the excess minority carrier profiles and comment on the differences that you found between them.

It will be helpful to use the Taylor series expansions of Appendix E.

6.5 A PN junction diode has a heavily doped emitter and a lightly doped base. The base doping profile changes in space as sketched below. Quantitatively graph the $1/C^2$ vs. V characteristics of this PN diode in reverse bias between $V = 0$ and $V = -10\ V$.



6.6 The small-signal equivalent circuit of a p^+-n diode that you are trying to reverse engineer is measured at room temperature for two different bias points, one forward ($I = 1\ mA$) and another one reverse ($V = -10\ V$). The results are sketched below:



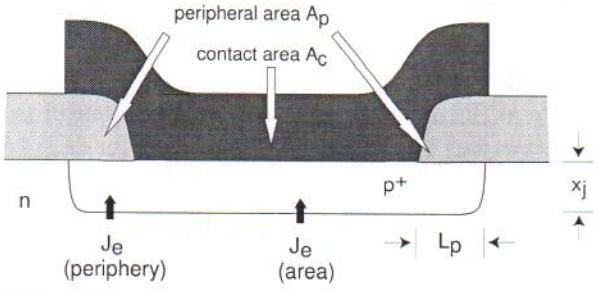
Based on your understanding of the technology that has been used, you are pretty sure that you can assume that the diode is very abrupt (that is $N_A \gg N_D$), that the doping level is uniform on the n-side, and that, from the point of view of the minority carriers, the n-side can be considered short with an ohmic contact at the end. Furthermore, assume that all minority carrier behavior is dominated by the lowly-doped side. On the microscope, you measure the area of the diode and you find it to be 10 μm^2.

a) Estimate the doping level on the n side, N_D.

b) Estimate the thickness of the n region, w_n.

c) Verify whatever assumptions you needed to make.

6.7 Consider the emitter of a PN junction diode as sketched below. The central portion of the emitter is covered by a metal contact where $S = \infty$. This region has an area A_c. The peripheral portion of the diode is covered by oxide and is characterized by $S = 0$. The area of this region is A_p. The length of the oxide covered area in the periphery, L_p, is much longer than the junction depth, x_j.



Let us do some relevant calculations for a case in which $N_A = 10^{19}\ cm^{-3}$, $x_j = 0.1\ \mu m$, $A_c = 50\ \mu m^2$, and $A_p = 5\ \mu m^2$.

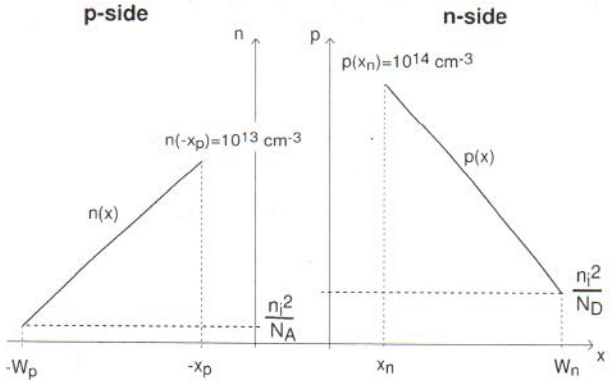
a) Calculate the ratio of the forward-bias current injected to the peripheral portion of the emitter over that injected to the contacted portion.

b) Calculate the ratio of the minority carrier charge stored in the peripheral portion of the emitter over that stored under the contact area.

c) Identify and evaluate the dominant minority carrier time constant for the region underneath the contact.

d) Identify and evaluate the dominant minority carrier time constant for the peripheral region of the emitter.

6.8 Below is a sketch *not to scale* of the minority carrier distribution across the quasi-neutral regions of a forward-biased PN diode. For this diode, $W_p - x_p = 4\ \mu m$, $W_n - x_n = 3\ \mu m$, $D_e = 25\ cm^2/s$, and $D_h = 10\ cm^2/s$. The area of the junction is $10\ \mu m^2$.



a) Calculate the hole current injected into the n-side of the diode.

b) Calculate the electron current injected into the p-side of the diode.

c) Calculate the diffusion capacitance associated with carrier storage on the n-side of the diode.

d) Calculate the diffusion capacitance associated with carrier storage on the p-side of the diode.

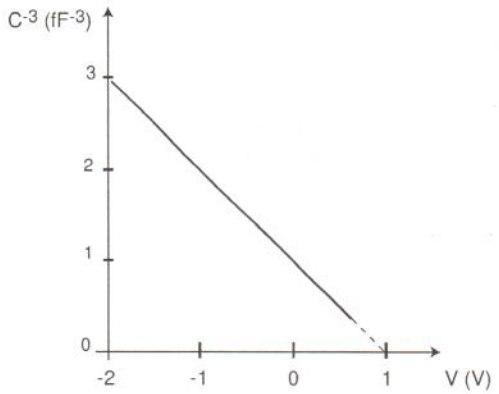
e) How much should the voltage across the junction increase if we wish to double the total current through the diode?

f) If we increase the voltage in the manner suggested in the previous question, what happens to the total diffusion capacitance of the diode?

g) What is the ratio of the doping levels across the junction: N_A/N_D?

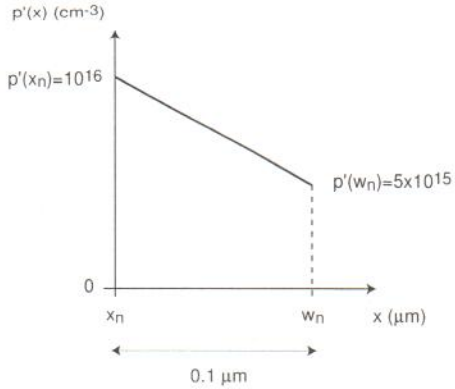
h) In what direction should N_A/N_D change if we wish to redesign the diode so as to get less diffusion capacitance at the same current level? (Assume that in redesigning the diode D_e, D_h, $W_n - x_n$, and $W_p - x_p$ do not change). Choose one: N_A/N_D *must increase.* N_A/N_D *must decrease.* Explain.

6.9 A PN diode with a junction area $A = 1\ \mu m^2$ has the following C-V characteristics:



Estimate the extent of the depletion region at $V = 0\ V$.

6.10 Consider an abrupt asymmetric p^+-n junction diode. All the action in this device is dominated by the lowly-doped n-type region. At a certain bias point, the excess minority concentration in the quasi-neutral n-type region has a distribution as sketched below:



x_n is the edge of the SCR. w_n is the surface. In this region, $D_h = 5\ cm^2/s$. Assume low-level injection.

Answer the following questions for the bias point that corresponds to the above diagram.

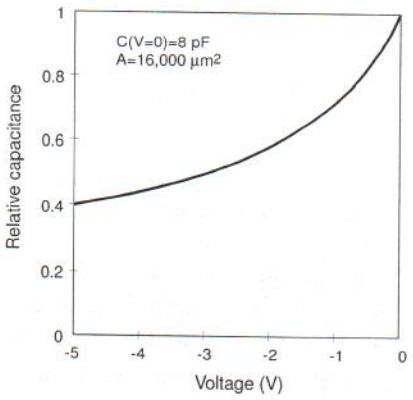
a) Calculate the diode current density.

b) Calculate the total amount of excess minority carrier charge.

c) Calculate the diffusion capacitance.

d) Calculate the dominant time constant.

6.11 In a paper on Si PN junction varactors, you see the following graph with the capacitance-voltage characteristics of the diode at room temperature:



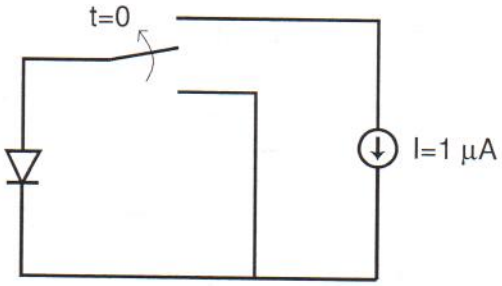
Assuming that the diode is highly asymmetrically doped, reverse engineer the diode.

a) Estimate the built-in potential of the junction.

b) Estimate the doping level of the lowly-doped side, N_L.

c) Estimate the doping level of the highly-doped side, N_H.

6.12 Consider an ideal p^+-n Si diode with a doping level in the lowly-doped n-type region of $10^{15}\ cm^{-3}$. The saturation current is 1 pA, the diode area is $10^{-4}\ cm^2$. Initially, the diode is in thermal equilibrium with its terminal shorted. At $t = 0$, a current source bias of 1 μA is applied in the reverse conduction mode as indicated in the figure below.



a) After a while, what happens? Explain.

b) how much time does it take for *it* to happen? State any assumptions you might need to make.

6.13 This problem is about testing the quality of the quasi-equilibrium approximation across the space-charge region of an ideal PN junction under bias. This is to be done by comparing the

magnitude of the minority carrier current, which is equal to the imbalance between drift and diffusion currents inside the SCR, with either one of these two current components.

Consider an n^+-p junction with "long" quasi-neutral regions. Focus on electron current. Assume negligible SCR generation and recombination.

a) Derive a first-order expression for the electron drift current at the metallurgical junction, $J_e^{drift}(x = 0)$. Assume linearity between electron drift velocity and electric field (small electric field). Make other suitable approximations.

b) Derive an expression for the net electron current across the SCR, J_e.

c) Express the ratio $J_e/J_e^{drift}(x = 0)$ in terms of $\phi_{bi} - V$, the electron diffusion length, and the Debye length in the p-QNR.

d) For typical doping levels, discuss the conditions that must be met for the quasi-equilibrium assumption to hold.

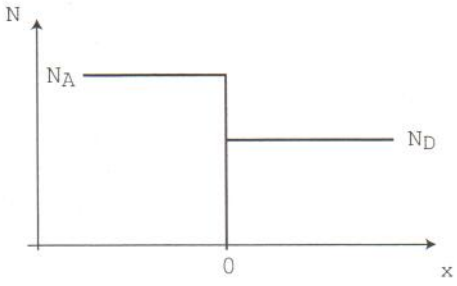
6.14 Consider an asymmetric p^+-n diode at room temperature. This diode is designed in a way that its lightly-doped n-type side dominates all minority carrier behavior. From a minority carrier point of view, this n-type side can be considered *long*.

This diode has been designed for a switching application from a certain forward voltage V_f to a certain reverse voltage V_r. Upon testing, it has been found that the *reverse recovery time*, t_{rr}, is too long. This problem is about considering options to reduce t_{rr}.

Indicate below the impact in t_{rr} of the specified change. Circle one: t_{rr} increases $= t_{rr} \uparrow$, t_{rr} decreases $= t_{rr} \downarrow$, or no effect in t_{rr}. For each item, give the reason for your choice.

increase doping level, $N_D \uparrow$:	$t_{rr} \uparrow$	$t_{rr} \downarrow$	no effect in t_{rr}
increase operational temperature, $T \uparrow$:	$t_{rr} \uparrow$	$t_{rr} \downarrow$	no effect in t_{rr}
increase n-type quasi-neutral region width, $W_n \uparrow$:	$t_{rr} \uparrow$	$t_{rr} \downarrow$	no effect in t_{rr}
irradiate diode so that carrier lifetime is reduced, $\tau \downarrow$:	$t_{rr} \uparrow$	$t_{rr} \downarrow$	no effect in t_{rr}
increase area of diode, $A \uparrow$:	$t_{rr} \uparrow$	$t_{rr} \downarrow$	no effect in t_{rr}

6.15 Consider an abrupt PN junction with $N_A = 10^{17}\ cm^{-3}$ and $N_D = 10^{16}\ cm^{-3}$, as sketched below.



a) In thermal equilibrium, compute the value of the electrostatic potential difference between $x = 0$ and deep inside the p-type quasi-neutral region.

b) Compute n_o and p_o at $x = 0$ in thermal equilibrium.

c) Compute the value of x for which $n_o = p_o = n_i$ in thermal equilibrium.

d) Compute the total amount of charge per unit area on the p side of the junction when a reverse bias voltage of 5 V is applied to the diode.

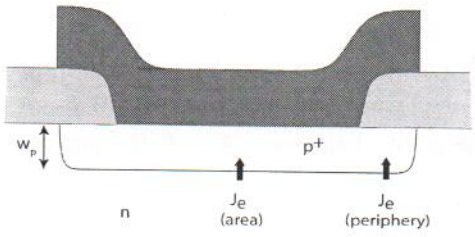
6.16 A PN junction diode in forward bias can be used as a thermometer. This is best done by biasing the diode at *constant forward current* and then measuring the voltage across as the temperature changes. This problem is about deriving a relationship for the forward bias voltage vs. temperature for an ideal PN diode.

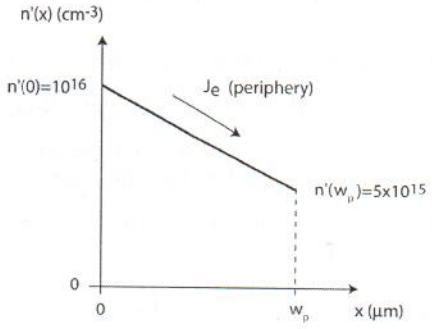
Make the following assumptions: no series resistance, sufficient forward voltage ($V \gg \frac{kT}{q}$), neglect power law-type temperature dependences (i.e., $\sim T^{\alpha}$) next to Boltzmann-type temperature dependences (i.e., $\sim \exp\frac{-E}{nkT}$).

a) Consider an ideal PN diode at a temperature T_o, biased at a forward current I_f with a voltage across $V(I_f, T_o)$. Derive an expression for V vs. T in terms of $V(I_f, T_o)$, T, T_o and fundamental constants. Sketch.

b) Calculate the temperature sensitivity, in units of mV/K, of an ideal PN diode with $I_s = 10\ pA$ biased at a current of $I_f = 1\ mA$ at $T_o = 300\ K$.

6.17 Consider a planar PN junction as sketched below on the left. This problem is about quantifying recombination at the peripheral surface of the p^+-type region, right next to the ohmic contact. In this region, at a certain forward bias point, an excess electron concentration profile is set as indicated in the figure below on the right. The depth of the p^+ region is $w_p = 0.1\ \mu m$. You can assume that the extent of the depletion region into the p^+ region is a negligible portion of this. You can also assume that the region is short to the electrons ($L_e >> w_p$) and that the electron mobility in this region is 250 $cm^2/V.s$.





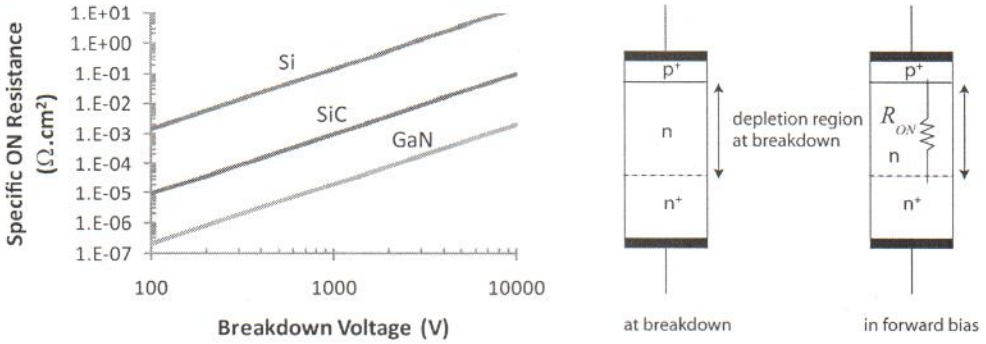
a) Estimate the electron current density in the peripheral portion of the p^+ region.

b) Estimate the net recombination rate at the surface of the peripheral portion of the p^+ region.

c) Estimate the surface recombination velocity at the surface of the peripheral portion of the p^+ region.

6.18 When comparing the suitability of different semiconductors for power electronics applications, a common graph that is used plots the specific ON resistance (in $\Omega.cm^{-2}$) against the breakdown voltage of an ideal PN diode. The specific ON resistance is the minimum resistance in the forward regime you could possibly get out of a PN diode with unity cross-sectional area that has been designed to deliver a certain reverse bias breakdown voltage. This graph is shown below. Lines for Si, SiC and GaN are sketched. For the same breakdown voltage,

GaN offers far less ON resistance than Si which is why there is so much excitement these days about GaN power electronics. Your task in this problem is to understand the meaning of this graph by calculating these parameters for an ideal Si PN diode.



Consider an ideal strongly asymmetric p^+-n-n^+ diode of unity cross-sectional area at room temperature. The n^+ region is introduced to minimize the resistance of the quasi-neutral n region. The breakdown voltage and the resistance characteristics of this diode are dominated by the lowly doped n-type region. The two key design parameters of this diode from the point of view of power electronics are the doping level of the n region and its thickness. A higher reverse breakdown voltage requires a thicker layer with a lower doping level. Hence, it offers higher resistance in the forward regime. The optimum thickness of the n region is that required to accommodate the depletion region at breakdown.

Let us do some numbers. Consider an ideal p^+-n-n^+ diode with a doping level in the lightly doped region of $N_D = 10^{15}\ cm^{-3}$.

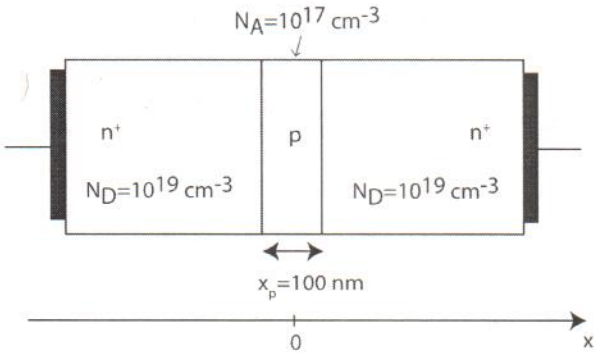
a) Estimate the breakdown voltage of this device.

b) Estimate the extent of the depletion region at breakdown.

c) Estimate the specific ON resistance contributed by the lowly-doped n-type region of a thickness equal to the one required to support the breakdown voltage of part (a).

d) Plot the point corresponding to the design in the graph at the beginning of this problem and comment.

6.19 Consider the Si structure sketched below at room temperature.



a) Evaluate the built-in potential of this structure, from n^+ region on the left to n^+ region on the right.

b) Estimate the hole concentration at the center of the p-type region.

c) Carefully sketch the energy band diagram across the structure, from n^+ region to n^+ region.

d) Calculate the capacitance of this structure at a bias voltage of 0 V.

6.20 In a certain PN junction diode at room temperature at a particular forward bias voltage, the current supported by hole injection in the n-side of the diode is 100 μA. The width of the quasi-neutral n-type region is 1 μm. This is much shorter than the corresponding hole diffusion length. At the surface of the n-type region there is an ohmic contact. The hole diffusion coefficient is 10 cm^2/s. The PN junction area is 10 μm^2.

In answering the questions that follow, make and state suitable approximations.

a) Estimate the excess hole concentration at the space-charge region edge of the n quasi-neutral region.

b) Estimate the velocity at which holes are injected at the edge of the n quasi-neutral region.

c) Estimate the hole flux arriving at the ohmic contact of the n-type quasi-neutral region.

d) Estimate the diffusion capacitance associated with hole storage in the n quasi-neutral region.