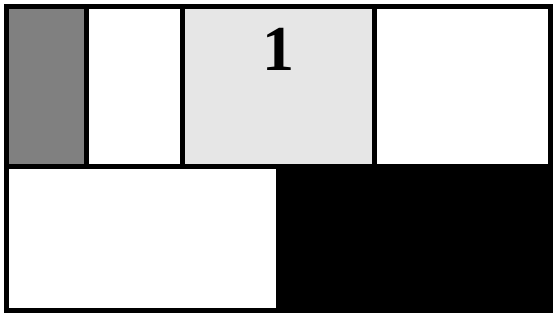
CHAPTER



THE DIODE

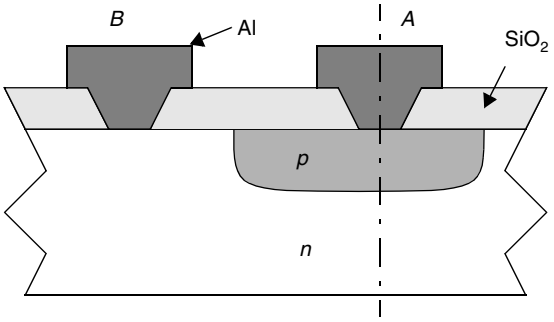
A Short Treatise on Diodes

1.1 The Diode

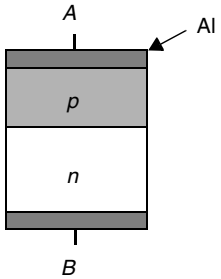
Although diodes rarely occur directly in the schematic diagrams of present-day digital gates, they are still omnipresent. For instance, each MOS transistor implicitly contains a number of reverse-biased diodes. Diodes are used to protect the input devices of an IC against static charges. Also, a number of bipolar gates use diodes as a means to adjust voltage levels. Therefore, a brief review of the basic properties and device equations of the diode is appropriate.

1.1.1 A First Glance at the Device

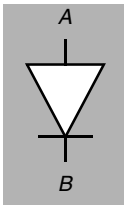
The *pn*-junction diode is the simplest of the semiconductor devices. Figure 1.1a shows a cross-section of a typical *pn*-junction. It consists of two homogeneous regions of *p*- and *n*-type material, separated by a region of transition from one type of doping to another, which is assumed thin. Such a device is called a *step* or *abrupt junction*. The *p*-type material is doped with *acceptor* impurities (such as boron), which results in the presence of holes as the dominant or majority carriers. Similarly, the doping of silicon with *donor* impurities (such as phosphorus or arsenic) creates an *n*-type material, where electrons are the majority carriers. Aluminum contacts provide access to the *p*- and *n*-terminals of the device. The circuit symbol of the diode, as used in schematic diagrams, is introduced in Figure 1.1c.



(a) Cross-section of *pn*-junction in an IC process



(b) One-dimensional representation



(c) Diode symbol

Figure 1.1 Abrupt *pn*-junction diode and its schematic symbol.

To understand the behavior of the *pn*-junction diode, we often resort to a one-dimensional simplification of the device (Figure 1.1b). Bringing the *p*- and *n*-type materials together causes a large concentration gradient at the boundary. The electron concentration changes from a high value in the *n*-type material to a very small value in the *p*-type material The reverse is true for the hole concentration. This gradient causes electrons to *diffuse* from *n* to *p* and holes to diffuse from *p* to *n*. When the holes leave the *p*-type material, they leave behind immobile acceptor ions, which are negatively charged. Consequently, the *p*-type material is negatively charged in the vicinity of the *pn*-boundary. Similarly, a positive charge builds up on the *n*-side of the boundary as the diffusing electrons leave behind the positively charged donor ions. The region at the junction, where the majority carriers have been removed, leaving the fixed acceptor and donor ions, is called the *depletion* or *space-charge region*. The charges create an electric field across the boundary, directed from the *n* to the *p*-region. This field counteracts the diffusion of holes and electrons, as it causes electrons to *drift* from *p* to *n* and holes to drift from *n* to *p*. Under equilibrium, the depletion charge sets up an electric field such that the drift currents are equal and opposite to the diffusion currents, resulting in a zero net flow.

The above analysis is summarized in Figure 1.2 that plots the current directions, the charge density, the electrical field, and the electrostatic field of the abrupt *pn*-junction under zero-bias conditions. In the device shown, the *p* material is more heavily doped than the *n*, or $N_A > N_D$, with N_A and N_D the acceptor and donor concentrations, respectively. Hence, the charge concentration in the depletion region is higher on the *p*-side of the junction. Figure 1.2 also shows that under zero bias, there exists a voltage ϕ_0 across the junction, called the *built-in potential*. This potential has the value

$$\phi_0 = \phi_T \ln\left[\frac{N_A N_D}{n_i^2}\right] \tag{1.1}$$

where ϕ_T is the *thermal voltage*

$$\phi_T = \frac{kT}{q} = 26\,\text{mV at 300 K} \tag{1.2}$$

The quantity n_i is the intrinsic carrier concentration in a pure sample of the semiconductor and equals approximately 1.5×10^{10} cm^{-3} at 300 K for silicon.

Example 1.1 Built-in Voltage of *pn*-junction

An abrupt junction has doping densities of $N_A = 10^{15}$ atoms/cm^3, and $N_D = 10^{16}$ atoms/cm^3. Calculate the built-in potential at 300 K.

From Eq. (1.1),

$$\phi_0 = 26 \ln\left[\frac{10^{15} \times 10^{16}}{2.25 \times 10^{20}}\right] \text{mV} = 638 \text{ mV}$$

Assume now that a forward voltage V_D is applied to the junction or, in other words, that the potential of the *p*-region is raised with respect to the *n*-zone. The applied potential lowers the potential barrier. Consequently, the flow of mobile carriers across the junction increases as the diffusion current dominates the drift component. These carriers traverse

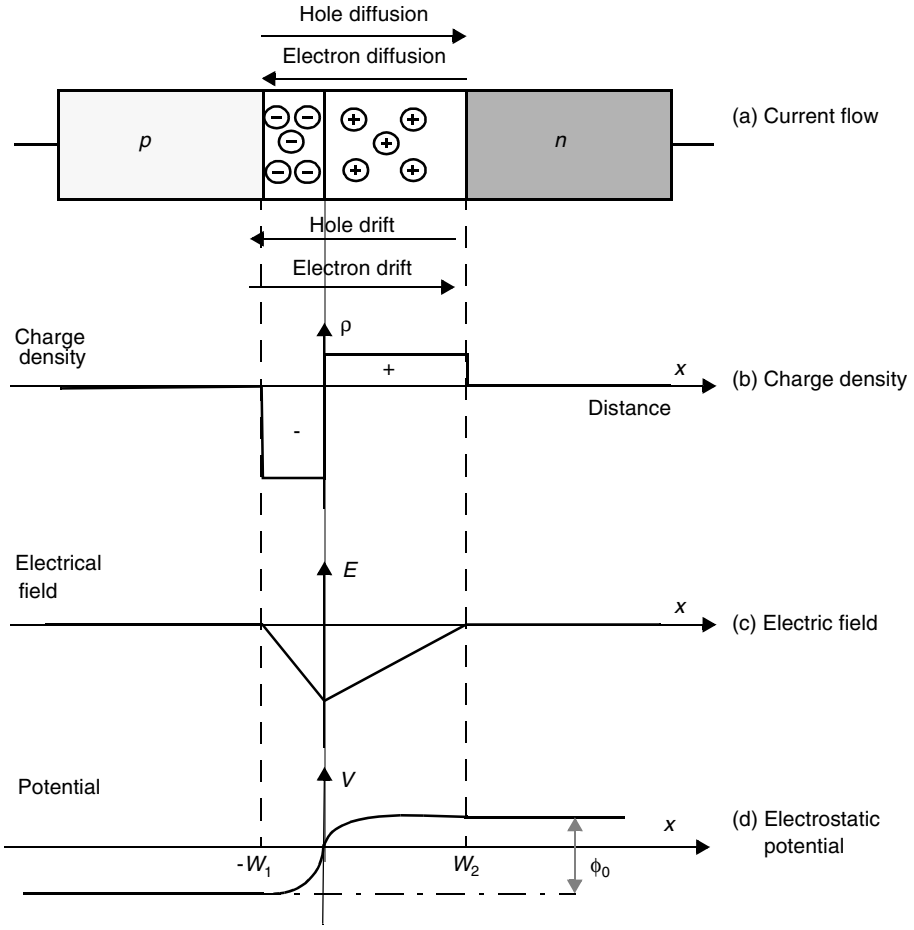
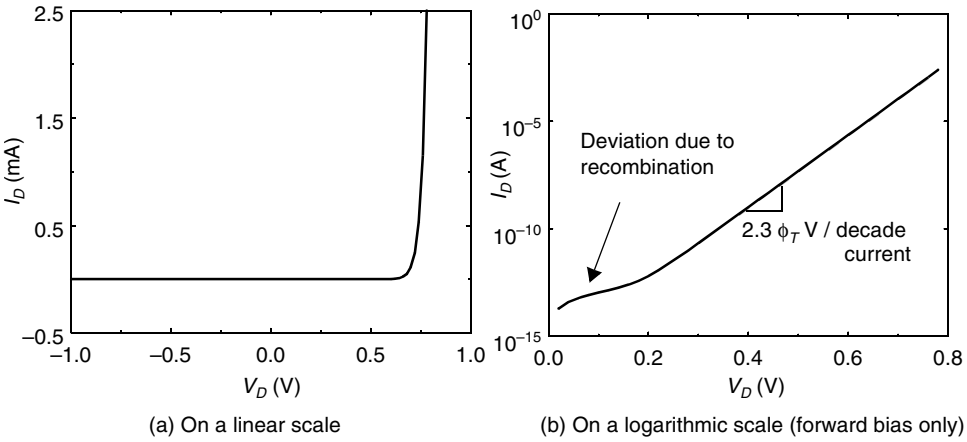


Figure 1.2 The abrupt *pn*-junction under equilibrium bias.

the depletion region and are injected into the neutral *n*- and *p*-regions, where they become minority carriers. Under the assumption that no voltage gradient exists over the neutral regions, which is approximately the case for most modern devices, these minority carriers will diffuse through the region as a result of the concentration gradient until they get recombined with a majority carrier. The net result is a current flowing through the diode from the *p*-region to the *n*-region. The most important property of this current is its *exponential dependence* upon the applied bias voltage.

On the other hand, when a reverse voltage V_D is applied to the junction or when the potential of the *p*-region is lowered with respect to the *n*-region, the potential barrier is raised. This results in a reduction in the diffusion current, and the drift current becomes dominant. A current flows from the *n*-region to the *p*-region. Since the number of minority carriers in the neutral regions (electrons in the *p*-zone, holes in the *n*-region) is very small, this drift current component is virtually ignorable. It is fair to state that in the reverse-bias mode the diode operates as a nonconducting, or blocking, device. The diode thus acts as a one-way conductor. This is illustrated in Figure 1.3, which plots the diode current I_D as a



(a) On a linear scale (b) On a logarithmic scale (forward bias only)

Figure 1.3 Diode current as a function of the bias voltage V_D.

function of the bias voltage V_D. The exponential behavior for positive-bias voltages is shown in Figure 1.3b, where the current is plotted on a logarithmic scale. The current increases by a factor of 10 for every extra 60 mV (= 2.3 ϕ_T) of forward bias. At small voltage levels (V_D < 0.15 V), a deviation from the exponential dependence can be observed, which is due to the recombination of holes and electrons in the depletion region as discussed in more detail later in the chapter.

After this intuitive introduction, we present analytical expressions for the behavior of the *pn*-junction. A distinction is made between the *static* (or steady-state) and the *dynamic* (or transient) behavior of the device.

1.1.2 Static Behavior

From earlier encounters with semiconductor devices [e.g., Sedra87], the reader is most probably familiar with the *ideal diode equation*, which relates the current through the diode I_D to the diode bias voltage V_D

$$I_D = I_S(e^{V_D/\phi_T} - 1) \tag{1.3}$$

I_S represents a constant value, called the *saturation current* of the diode. Under reverse-bias conditions, where $V_D << 0$, $I_D \approx -I_S$ and equals the reverse-bias leakage current. ϕ_T is the thermal voltage of Eq. (1.2) and is equal to 26 mV at room temperature. The remainder of this section presents a physical background for this equation.

Forward Bias

When a positive voltage is applied to the junction, mobile carriers drift through the depletion region and are injected into the neutral regions, where they become *excess minority carriers* and diffuse in the direction of the terminal connections. It is the distribution of these excess minority carriers that dictates the static behavior of the *pn*-diode. Figure 1.4 shows the minority carrier concentrations in the neutral region near a *pn*-junction for the

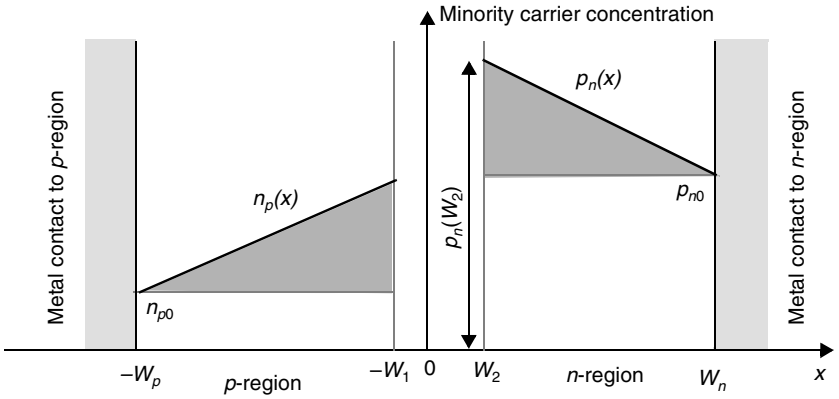


Figure 1.4 Minority carrier concentrations in the neutral region near an abrupt *p*n-junction under forward-bias conditions.

forward-bias condition. Observe that the majority carrier concentrations have to proceed along the same line, because charge neutrality dictates that any local increase in the electron (hole) concentration is matched by a similar increase in the hole (electron) density. While the fractional increase of the minority carrier concentration is substantial, it is largely ignorable for the majority carriers.

Figure 1.4 shows a linear decay in the minority carrier concentrations when moving away from the junction. At the metal contacts (which can be assumed to be infinite sources or sinks of holes or electrons), the minority carrier concentrations are at their equilibrium values (n_{p0} and p_{n0}), independent of the applied bias. The linear decay model is valid under the assumption that the width of the *n*- or *p*-regions is sufficiently small so that injected minority carriers diffuse to the metal contact before recombining with majority carriers. This operation condition is called the *short-base diode model* and is valid for most contemporary semiconductor diodes.

The gradient in the minority concentrations causes a diffusion current in the neutral (also called *bulk*) regions that is proportional to that gradient. The constant of proportionality is called the *diffusion coefficient* (D_p and D_n for holes and electrons, respectively). Based on these observations, an expression for the diode current can be derived. In this derivation, we initially consider the *n*-region only. Similar expressions can be derived for the *p*-region.

$$I_{D,p} \sim \frac{\mathrm{d}p_n}{\mathrm{d}x} = -qA_D D_p \frac{\mathrm{d}p_n}{\mathrm{d}x} \tag{1.4}$$

with

$$p_n(x) = -\left[\frac{p_n(W_2) - p_{n0}}{W_n - W_2}\right]x + \left[\frac{p_n(W_2)W_n - p_{n0}W_2}{W_n - W_2}\right] \tag{1.5}$$

where $p_n(x)$ represents the hole concentration in the *n*-region as a function of the position x, A_D is the junction area, and q *is* the electron (hole) charge. Combining the two equations results in an expression of the diode current as a function of the minority carrier concen-

tration at the boundary of the depletion region. The latter is determined by the *law of the junction*, which states that the concentration at the edge of the depletion region is an exponential function of the applied bias voltage

$$p_n(W_2) = p_{n0}e^{V_D/\phi_T} \tag{1.6}$$

with p_{n0} the hole concentration in the *n*-region under equilibrium conditions. For $N_A >> n_i$, the equilibrium minority hole concentration (in the *n*-region) is obtained from the following expression

$$p_{n0} \approx n_i^2 / N_D \tag{1.7}$$

and, similarly, for the *p*-region

$$n_{p0} \approx n_i^2 / N_A \tag{1.8}$$

Combining Equations (1.4), (1.5) and (1.6) yields the diode current,

$$I_{Dp} = qA_D D_p \frac{p_{n0}}{W_n - W_2}(e^{V_D/\phi_T} - 1) \tag{1.9}$$

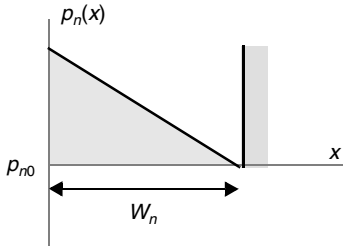
Repeating the same analysis for the *p*-region and summing the *p* and *n* current-contributions produces the ideal diode current expression of Eq. (1.3). It also yields an expression for the saturation current I_S

$$I_S = qA_D\left(\frac{D_p p_{n0}}{W_n - W_2} + \frac{D_n n_{p0}}{W_p - W_1}\right) \tag{1.10}$$

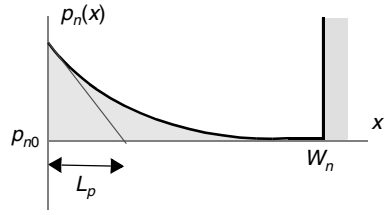
Keep in mind that the above equation is based on a number of assumptions, which might not be valid for all actual devices. First of all, it is assumed that the length of the neutral regions is short enough that recombination does not occur (*short-base diode* model). For this to be valid, the widths of the *p*- and *n*-regions must be substantially smaller than a material constant called the *diffusion length* (denoted as L_p and L_n for holes and electrons, respectively). If this is not the case, the diode becomes of the *long-base* type. Minority carriers diffusing through the neutral region gradually recombine with majority carriers. This affects the minority-carrier concentration as illustrated in Figure 1.5. Instead of a linear decay, the concentration drops in an exponential fashion. In one diffusion length, the excess minority concentration drops to $1/e$ (= 0.37) of its original value. After a few diffusion lengths, virtually all injected carriers have recombined, and the minority carrier concentration reaches its thermal equilibrium value. The current equation remains essentially unchanged. The only modification is that the W_n and W_p parameters in the saturation current (Eq. (1.10)) are replaced by the diffusion lengths L_p and L_n.

Other assumptions are that the resistance of the neutral regions is negligible, and that the minority carrier injection levels are substantially lower than the majority concentration levels (*low-injection condition*). Later in the chapter we discuss how violating these conditions affects the device operation.

Eq. (1.10) clearly shows that the diode current is the composite result of a hole and an electron current. In most practical diodes, one of the sides has a substantially lighter doping level and hence produces a larger number of minority carriers. The corresponding



(a) Short-base diode: $W_n << L_p$



(b) Long-base diode: $W_n >> L_p$

Figure 1.5 Minority carrier concentrations in the *n*-region near a *pn*-junction.

current component dominates the overall value. For instance, in the example of Figure 1.2, the *p*-region has a heavier doping than the *n*-region. Consequently, $p_{n0} >> n_{p0}$, and the hole current dominates.

Problem 1.1 Diode Current

For a diode with the following properties, compute the saturation current I_S. Also, solve V_D for $I_D = 0.1$ mA, assuming that $\phi_T = 26$ mV.

$A_D = 9\ \mu m^2$,
$N_D = 5 \times 10^{15}\ cm^{-3}$,
$N_A = 2.5 \times 10^{16}\ cm^{-3}$,
$\phi_0 = 0.795$ V,
$D_n = 25\ cm^2/sec$,
$D_p = 10\ cm^2/sec$,
$W_n = 5\ \mu m$ and $W_p = 0.7\ \mu m$,
$W_2 = 0.15\ \mu m$ and $W_1 = 0.03\ \mu m$.
$n_i = 1.5 \times 10^{10}\ cm^{-3}$ and $q = 1.6 \times 10^{-19}$ C.
Also, $L_n = 5\ \mu m$ and $L_p = 31\ \mu m$.

Reverse Bias

When applying a reverse-bias voltage to the junction, the ideal diode equation predicts that the diode current I_D approaches $-I_S$ for $|V_D| >> \phi_T$. This is readily understood when analyzing the minority carrier concentration distribution under reverse-bias conditions, as shown in Figure 1.6.[1] From the law of the junction (which is equally valid under reverse-bias con-

[1] It is worth observing that all equations derived for the forward-bias apply just as well under reverse-bias conditions.

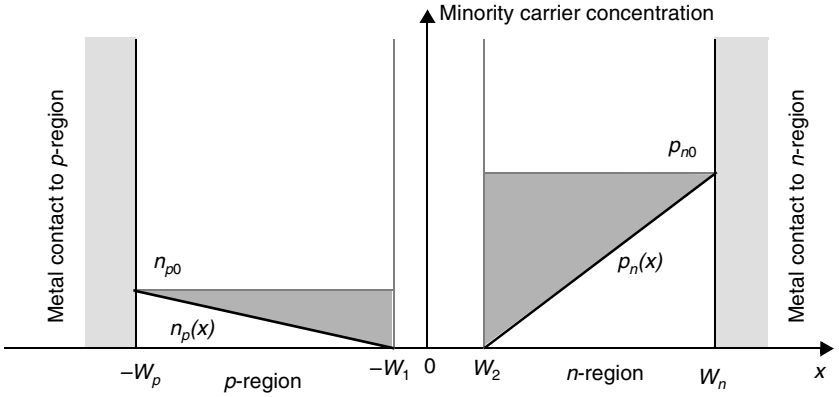


Figure 1.6 Minority carrier concentration in the neutral regions near the *pn*-junction under reverse-bias conditions.

ditions), it can be derived that the concentration of minority carriers at the depletion-region boundaries is small and actually approaches 0 when sufficient reverse bias is applied. At the metallic contacts, the concentration is restored to the thermal equilibrium value.

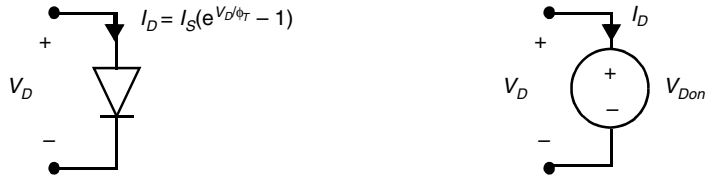
The resulting gradient causes a diffusion of minority carriers towards the junction. Once they reach the depletion region they are swept across the junction by the electric field of the depletion region (which is actually increased by the reversed bias) and transported to their majority zone (holes to the *p*-region, electrons to the *n*-region). This reverse current is restricted by two factors: the limited availability of minority carriers (p_{n0}, n_{p0}) and the fact that the concentration gradient does not change much once the reverse-bias voltage is sufficiently large (which typically means > 4 ϕ_T), as is obvious when taking the derivative of Eq. (1.6) as well as from Figure 1.6.

It is worth mentioning that in actual devices, the reverse currents are substantially larger than the saturation current I_S. This is due to the thermal generation of hole and electron pairs in the depletion region. The electric field present sweeps these carriers out of the region, causing an additional current component. For typical silicon junctions, the saturation current is nominally in the range of 10^{-17} A/μm^2, while the actual reverse currents are approximately three orders of magnitude higher. Actual device measurements are, therefore, necessary to determine realistic values for the reverse diode leakage currents.

Models for Manual Analysis

The derived current-voltage equations can be summarized in a set of simple models that are useful in the manual analysis of diode circuits. A first model, shown in Figure 1.7a, is based on the ideal diode equation Eq. (1.3). While this model yields accurate results, it has the disadvantage of being strongly nonlinear. This prohibits a fast, first-order analysis of the dc-operation conditions of a network. An often-used, simplified model is derived by inspecting the diode current plot of Figure 1.3. For a "fully conducting" diode, the voltage drop over the diode V_D lies in a narrow range, approximately between 0.6 and 0.8 V. To a first degree, it is reasonable to assume that a conducting diode has a fixed voltage drop V_{Don} over it. Although the value of V_{Don} depends upon I_S, a value of 0.7 V is typically

assumed. This gives rise to the model of Figure 1.7b, where a conducting diode is replaced by a fixed voltage source.



(a) Ideal diode model (b) First-order diode model

Figure 1.7 Diode models.

Example 1.2 Analysis of Diode Network

Consider the simple network of Figure 1.8 and assume that $V_S = 3$ V, $R_S = 10$ kΩ and $I_S = 0.5 \times 10^{-16}$ A. The diode current and voltage are related by the following network equation

$$V_S - R_S I_D = V_D$$

Inserting the ideal diode equation and (painfully) solving the nonlinear equation using either numerical or iterative techniques yields the following solution: $I_D = 0.224$ mA, and $V_D = 0.757$ V. The simplified model with $V_{Don} = 0.7$ V produces similar results ($V_D = 0.7$ V, $I_D = 0.23$ A) with far less effort. It hence makes considerable sense to use this model when determining a first-order solution of a diode network.

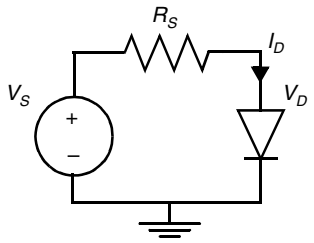


Figure 1.8 A simple diode circuit.

1.1.3 Dynamic, or Transient, Behavior

So far, we have mostly been concerned with the static, or steady-state, characteristics of the diode. Just as important in the design of digital circuits is the response of the device to changes in its bias conditions. The transient, or dynamic, response determines the maximum speed at which the device can be operated. Because the operation mode of the diode is a function of the amount of charge present in both the neutral and the space-charge regions, its dynamic behavior is strongly determined by how fast charge can be moved around. An accurate model of the charge distribution in a diode is, therefore, essential and will be presented first.

Depletion-Region Capacitance

In the ideal model, the depletion region is void of mobile carriers, and its charge is determined by the immobile donor and acceptor ions. The corresponding charge distribution under zero-bias conditions was plotted in Figure 1.2. This picture can be easily extended to incorporate the effects of forward or reverse biasing. At an intuitive level the following observations can be easily verified—under forward-bias conditions, the potential barrier is reduced, which means that less space charge is needed to produce the potential difference. This corresponds to a reduced depletion-region width. On the other hand, under reverse conditions, the potential barrier is increased corresponding to an increased space charge and a wider depletion region. These observations are confirmed by the well- known depletion-region expressions given below (a derivation of these expressions, which are valid for abrupt junctions, is either simple or can be found in any textbook on devices such as [Sedra87]). One observation is crucial—due to the global charge neutrality requirement of the diode, the total acceptor and donor charges must be numerically equal.

1. Depletion-region charge (V_D is positive for forward bias).

$$Q_j = A_D\sqrt{\left(2\varepsilon_{si}q\frac{N_AN_D}{N_A+N_D}\right)(\phi_0 - V_D)} \tag{1.11}$$

2. Depletion-region width.

$$W_j = W_2 - W_1 = \sqrt{\left(\frac{2\varepsilon_{si}}{q}\frac{N_A+N_D}{N_AN_D}\right)(\phi_0 - V_D)} \tag{1.12}$$

3. Maximum electric field.

$$E_j = \sqrt{\left(\frac{2q}{\varepsilon_{si}}\frac{N_AN_D}{N_A+N_D}\right)(\phi_0 - V_D)} \tag{1.13}$$

In the preceding equations ε_{si} stands for the electrical permittivity of silicon and equals 11.7 times the permittivity of a vacuum, or 1.053×10^{-12} F/cm. The ratio of the *n*- versus *p*-side of the depletion-region width is determined by the doping-level ratios: $\mathbf{W}_2/(-W_1) = N_A/N_D$.

From an abstract point of view, it is possible to visualize the depletion region as a capacitance, albeit one with very special characteristics. Because the space-charge region contains few mobile carriers, it can be conceived as an insulator with a dielectric constant ε_{si} of the semiconductor material. The *n*- and *p*-regions act as the capacitor plates. A small change in the voltage applied to the junction dV_D causes a change in the space charge dQ_j. Hence, a depletion-layer capacitance can be defined

$$\begin{aligned} C_j = \frac{dQ_j}{dV_D} &= A_D\sqrt{\left(\frac{\varepsilon_{si}q}{2}\frac{N_AN_D}{N_A+N_D}\right)(\phi_0 - V_D)^{-1}} \\ &= \frac{C_{j0}}{\sqrt{1 - V_D/\phi_0}} \end{aligned} \tag{1.14}$$

where C_{j0} is the capacitance under zero-bias conditions and is only a function of the physical parameters of the device.

$$C_{j0} = A_D \sqrt{\left(\frac{\varepsilon_{si} q}{2} \frac{N_A N_D}{N_A + N_D}\right) \phi_0^{-1}} \tag{1.15}$$

Notice that the same capacitance value is obtained when using the standard parallel-plate capacitor equation $C_j = \varepsilon_{si} A_D / W_j$ (with W_j given in Eq. (1.12)) Typically, the A_D factor is omitted, and C_j and C_{j0} are expressed as a capacitance/unit area.

The resulting junction capacitance is plotted in the function of the bias voltage in Figure 1.9 for a typical silicon diode found in MOS circuits. A strong *nonlinear dependence* can be observed. Note also that the capacitance decreases with an increasing reverse bias: a reverse bias of 5 V reduces the capacitance by more than a factor of two.

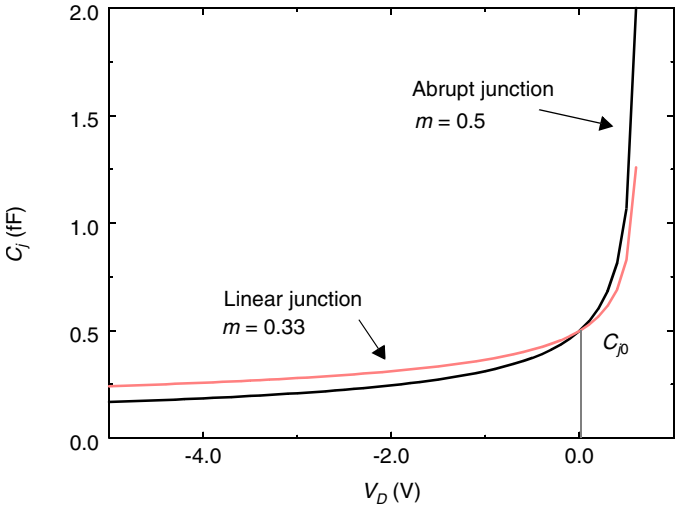


Figure 1.9 Junction capacitance (in fF/μm^2) as a function of the applied bias voltage.

Example 1.3 Junction Capacitance

Consider the following silicon junction diode: C_{j0} = 0.5 fF/μm^2, A_D = 12 μm^2 and ϕ_0 = 0.64 V. A reverse bias of –5 V results in a junction capacitance of 0.17 fF/μm^2, or, for the total diode, a capacitance of 2.02 fF.

Equation (1.14) is only valid under the condition that the *pn*-junction is an *abrupt junction*, where the transition from *n* to *p* material is instantaneous. This is often not the case in actual integrated-circuit *pn*-junctions, where the transition from *n* to *p* material can be gradual. In those cases, a linear distribution of the impurities across the junction is a better approximation than the step function of the abrupt junction. An analysis of the *linearly-graded junction* shows that the junction capacitance equation of Eq. (1.14) still holds, but with a variation in order of the denominator. A more generic expression for the junction capacitance can hence be provided,

$$C_j = \frac{C_{j0}}{(1 - V_D/\phi_0)^m} \tag{1.16}$$

where m is called the *grading coefficient* and equals 1/2 for the abrupt junction and 1/3 for the linear or graded junction. Both cases are illustrated in Figure 1.9.

The reader should be aware that the junction capacitance is actually a small-signal parameter whose value varies over bias points. In digital circuits, operating voltages tend to move rapidly over wide ranges. Under those circumstances, it is more attractive to replace the voltage-dependent, nonlinear capacitance C_j by an equivalent, linear capacitance C_{eq}. C_{eq} is defined such that, for a given voltage swing from voltages V_{high} to V_{low}, the same amount of charge is transferred as would be predicted by the nonlinear model

$$C_{eq} = \frac{\Delta Q_j}{\Delta V_D} = \frac{Q_j(V_{high}) - Q_j(V_{low})}{V_{high} - V_{low}} = K_{eq} C_{j0} \tag{1.17}$$

Combining Eq. (1.11) (extended to accommodate the grading coefficient m) and Eq. (1.17) yields the value of K_{eq}.

$$K_{eq} = \frac{-\phi_0^m}{(V_{high} - V_{low})(1-m)} [(\phi_0 - V_{high})^{1-m} - (\phi_0 - V_{low})^{1-m}] \tag{1.18}$$

Example 1.4 Average Junction Capacitance

The diode of Example 1.3 is switched between 0 and –5 V. Compute the average junction capacitance (m = 0.5).

For the defined voltage range and for ϕ_0 = 0.64 V, K_{eq} evaluates to 0.502. The average capacitance hence equals 0.25 fF/μm^2.

Diffusion Capacitance

Under forward bias, the *pn*-junction exhibits a capacitive effect much larger than just the junction capacitance. This extra capacitive effect is due to the excess minority carrier charge stored at the boundaries of the depletion region.

It now turns out that this *excess charge is directly related to the current flowing through the diode*. The total excess minority charge stored in a region can be derived by integrating Eq. (1.5) over the complete region and the total diode area and taking into account that each carrier carries a charge q (= 1.6×10^{19} C). For the *n*-region, for instance, this results in the following relation:

$$\begin{aligned} Q_p &= qA_D \int_{W_2}^{W_n} (p_n(x) - p_{n0}) dx \\ &= qA_D \frac{(W_n - W_2) p_{n0} (e^{V_D/\phi_T} - 1)}{2} \\ &= \frac{(W_n - W_2)^2}{2D_p} I_{Dp} \approx \frac{W_n^2}{2D_p} I_{Dp} \end{aligned} \tag{1.19}$$

with I_{Dp} the hole component of the diode current. The ratio of the squared width of the neutral region and the diffusion coefficient is another important device parameter called the *mean transit time*.

$$\tau_{Tp} = \frac{W_n^2}{2D_p} \text{ sec}$$

and (1.20)

$$\tau_{Tn} = \frac{W_p^2}{2D_n} \text{ sec}$$

The total diode current can now be expressed as a function of the excess minority carrier charge

$$I_D = \frac{Q_p}{\tau_{Tp}} + \frac{Q_n}{\tau_{Tn}} = \frac{Q_D}{\tau_T} \tag{1.21}$$

This equation simply states that, in the steady state, the current I_D is inversely proportional to the time it takes a carrier to transport from the junction to the metallic contact.

Example 1.5 Mean Transit Times

For the diode of Problem 1.1, the mean transit times evaluate to the following values:

$\tau_{Tp} = (5\ \mu\text{m} - 0.15\ \mu\text{m})^2/\ (2 \times 10\ \text{cm}^2/\text{sec}) = 11.7$ nsec

$\tau_{Tn} = (0.7\ \mu\text{m} - 0.03\ \mu\text{m})^2/\ (2 \times 25\ \text{cm}^2/\text{sec}) = 0.09$ nsec

---

Similar expressions can be derived for the long-base diode. In that case, the transit time is replaced by the *excess minority carrier lifetime* parameter, which indicates the mean time it takes for an injected minority carrier to recombine with a majority carrier.

$$\tau_p = L_p^2/D_p \text{ sec}$$

and (1.22)

$$\tau_n = L_n^2/D_n \text{ sec}$$

In silicon, typical values of L_p and L_n range from 1 to 100 μm, and the corresponding values of the lifetime are in the range of 1 to 10,000 nsec.

Under transient conditions, a change in current translates into a change in the excess minority carrier charge. In correspondence with the approach used for the depletion region, we model the effect of this charge by an equivalent capacitance called the *diffusion capacitance* C_d. The value of C_d is easily derived

$$C_d = \frac{dQ_D}{dV_D} = \tau_T \frac{dI_D}{dV_D} \approx \frac{\tau_T I_D}{\phi_T} \tag{1.23}$$

which shows a linear dependence upon I_D (as could be expected). For reverse bias, it is fair to assume that C_d is ignorable. Observe, once again, that ***C_d* is a small-signal capacitance** and is only valid around a given bias voltage.

Similar to the junction capacitance, an average diffusion capacitance can be defined for a voltage range of interest.

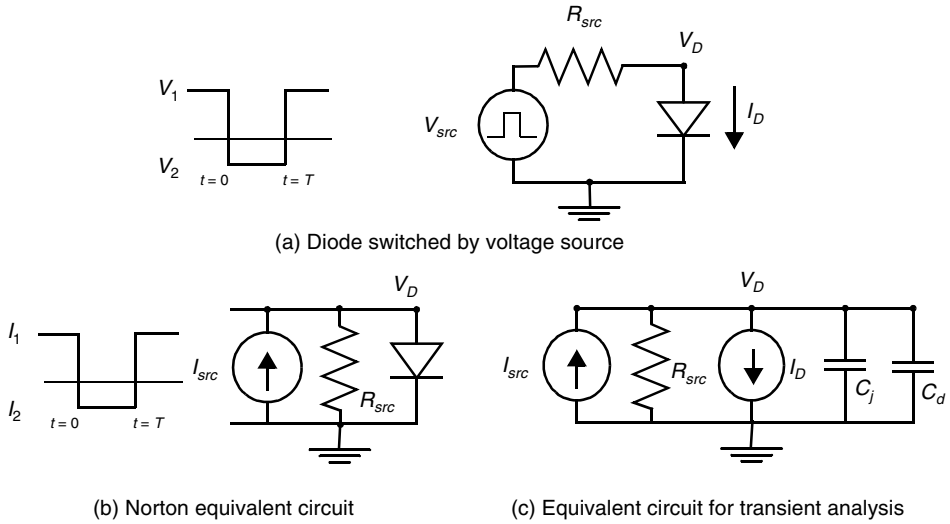
$$C_D = \frac{\Delta Q_D}{\Delta V_D} = \tau_T \frac{(I_D(V_{high}) - I_D(V_{low}))}{V_{high} - V_{low}} = \phi_T \frac{(C_{d(high)} - C_{d(low)})}{V_{high} - V_{low}} \quad (1.24)$$

Example 1.6 Diffusion Capacitance

For $I_S = 0.5 \times 10^{-16}$A, $\tau_T = 1$ nsec, and $\phi_T = 26$ mV, C_d evaluates to a capacitance of 6.5 pF for a forward bias of 0.75V.

Diode Switching Time: A Case Study

The presented models can now be employed to determine the time it takes to switch a diode between two different states. Consider the circuit of Figure 1.10a. Before time 0, the voltage source V_{src} provides a strong forward bias for the diode ($V_1 >> V_{D,on}$). At time $t = 0$, the voltage source switches to a negative voltage, such that the diode goes into reverse bias. At time $t = T$, the voltage source is reversed again, turning on the diode. We simplify the analysis by replacing the voltage source and its resistance by the Norton equivalent circuit (Figure 1.10b). It is assumed that the source resistance R_{src} is large enough so that virtually all current flows through the diode in forward-bias conditions.



(a) Diode switched by voltage source

(b) Norton equivalent circuit

(c) Equivalent circuit for transient analysis

Figure 1.10 Diode switching time.

To determine how fast the circuit will move to a new steady state, we analyze the equivalent circuit of Figure 1.10c, where the diode is replaced by a nonlinear current source (representing the ideal diode equation) and two capacitances, representing the space charge (C_j) and the excess minority carrier charge (C_d), respectively. Once again, observe that C_j and C_d are small signal-capacitances. To be applicable to large-signal analysis, averaged capacitance values must be used. The model for the reverse-bias operation mode is obtained by simply eliminating the current source I_D and the diffusion capacitance C_D.

Deriving the transient response of this network seems simple, as it requires the "mere" solution of the following differential equation:

$$I_{src} = I_D(t) + (C_d + C_j)\frac{dV_D}{dt} = I_S(e^{V_D(t)/\phi_T} - 1) + (C_d + C_j)\frac{dV_D}{dt} \text{ o} \qquad (1.25)$$

Unfortunately, this equation is heavily nonlinear again, and finding an analytical solution is a daunting task, which is easily solved by a computer but hard to perform manually. Observe that C_d and C_j are nonlinear functions of V_D as well. Some simplifications are hence at hand. A glimpse of how to tackle this is offered by an inspection of a simulated response, shown in Figure 1.11.

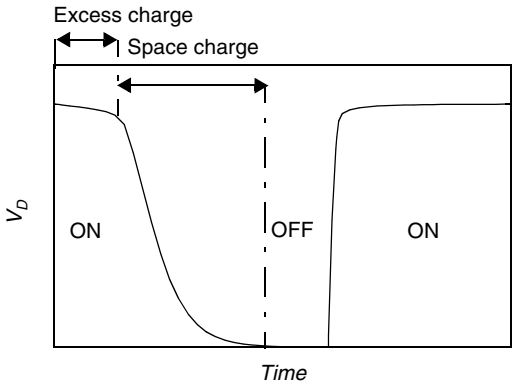


Figure 1.11 Simulated transient response of diode.

The turn-off transient, for instance, clearly displays two operation intervals:

1. Initially, the reverse current I_2 is used to remove the excess minority carrier charge from the neutral regions. During that time, the diode remains on, and the voltage over the diode is approximately constant. This is easily understood: a linear drop in voltage requires an exponential drop in current (and equivalently in excess charge). The constant voltage means that the space charge remains approximately constant as well. The effect of C_j can be ignored during this interval.
2. Once the diode has been turned off ($I_D \approx 0$), the circuit evolves towards steady state. While building a reverse bias over the diode, the space charge changes. In this region, the junction capacitance C_j dominates the performance.

The reader should be aware that the partitioning of the transient into two intervals is somewhat artificial and that both intervals overlap. For instance, in the later phases of the diode turn-off, not all excess charge is removed, yet the voltage over the diode starts to drop, changing the space charge. It is fair to assume that the impact of the space charge is dominant at that time, as the excess charge has been reduced to minuscule amounts (once again, this is a consequence of the exponential relationship between excess charge and diode voltage). The assumption is therefore very reasonable.

Based on these observations, we can derive the duration of both intervals. We first address the turn-off transient.

Removal of the excess charge. Instead of trying to solve Eq. (1.25), a more tangible way of monitoring the excess minority carrier charge is to operate in the charge domain. Using the charge-control expression of the diode current (Eq. (1.21)) we can derive the following expression

$$I_{src} = \frac{Q_D(t)}{\tau_T} + \frac{dQ_D}{dt} \tag{1.26}$$

This equation states that in transient operation, the current supplied to the diode splits in two fractions, as illustrated in Figure 1.12. A first component sustains the normal diffusion current, which is proportional to the excess minority carrier charge present. The second component adds (turn-on) or removes (turn-off) excess carrier charge. This component obviously drops to zero when the steady-state condition is reached.

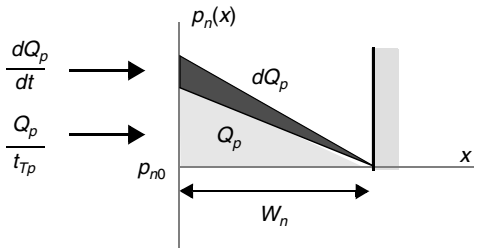


Figure 1.12 Incremental change in minority carrier charge during diode turn-on (showing *n*-region only) and the resulting current components.

Solving the differential equation, while taking into account that the initial value of Q_D equals $I_1 \times \tau_T$ and $I_{src} = I_2$, yields $Q_D(t)$

$$Q_D(t) = \tau_T[I_2 + (I_1 - I_2)e^{-t/\tau_T}] \tag{1.27}$$

The turn-off time is derived by solving for the time $t = t_1$, for which Q_D evaluates to 0.

$$t_1 = \tau_T \ln\left(\frac{I_1 - I_2}{-I_2}\right) \tag{1.28}$$

Changing the Space Charge. Once the diode is turned off, the circuit will evolve to a steady-state solution, where all the source current is flowing through the resistor R_{src}, or $V_D(t = \infty) = I_{src}R_{src} = I_2R_{src}$. A reverse voltage is built over the diode, which means that extra space charge has to be provided. The circuit operation during this period is modeled by the equivalent circuit of Figure 1.10b, where the diode current I_d is set to zero (when the diode is off, the reverse current can generally be ignored). Since the change in excess minority carrier charge in reverse mode is negligible, C_d is ignored as well. This remaining circuit, which is a simple *RC* circuit) is described by the following differential equation:

$$I_{src} = \frac{V_D(t)}{R_{src}} + C_j\frac{dV_D}{dt} \tag{1.29}$$

where C_j is the average junction capacitance over the voltage region of interest. Assuming that the value of V_D at time $t = t_1$ equals 0, the solution of this equation is the well-known exponential expression[2]

$$V_D(t) = I_2 R_{src}\left(1 - e^{-\frac{t-t_1}{R_{src}C_j}}\right) \tag{1.30}$$

The diode voltage reaches its final value in an asymptotic fashion. For such a waveform, the 90% point is often used to determine the end of the transition (as the 100% point would take infinitely long). It is easily determined that this 90% point is reached after 2.3 time-constants $R_{src}C_j$.

Turn-on Transient. Similar considerations hold for the turn-on transient, as is illustrated in Figure 1.11. Before the diode can be turned on, it is necessary to change the space charge first. The transient waveform for the diode voltage is easily derived (using a similar approach as above, and assuming that the transient starts at time $t = 0$):

$$V_D(t) = R_{src}\left(I_1 - (I_2 - I_1)e^{-\frac{t}{R_{src}C_j}}\right) \tag{1.31}$$

Solving this equation for $V_D = 0$ (the time the diode starts to turn on) yields

$$t_3 = R_{src}C_j \ln\left(\frac{I_1 - I_2}{I_1}\right) \tag{1.32}$$

The build-up of the space-charge is still described by the differential equation (1.26). With the proper initial condition ($Q_D(t = t_3) = 0$), solving this equation shows that the excess charge increases in an exponential fashion and asymptotically approaches its final value

$$Q_D(t) = I_1\tau_T\left[1 - e^{-\frac{t-t_3}{\tau_T}}\right] \tag{1.33}$$

It takes approximately 2.2 time constants τ_T for Q_D to reach 90% of its final value.

The procedures used above are representative of the analysis and derivation techniques to be used in the chapters to come—using a number of approximations and linearizations, a simple, tractable model is constructed of a complex, nonlinear circuit. Although inaccurate, this model fosters insight into the circuit operation and identifies the dominant parameters. The first-order solution obtained from this analysis can then be further optimized, or fine-tuned, using computer-aided tools.

Example 1.7 Diode Transient Response

Consider the diode circuit of Figure 1.10 for the following parameters: R_{src} = 50 kΩ, I_1 = 1 mA, I_2 = –0.1 mA. The following parameters are used for the diode: $I_S = 2 \times 10^{-16}$A, C_{j0} = 0.2 pF, τ_T = 5 nsec, and ϕ_0 = 0.65.

---

[2] One can argue that the junction capacitance starts to dominate at the point where $V_D = \phi_0$ and that this should be the starting point of the second interval. Adopting this assumption does not affect the final result in a major way.

The steady-state voltages over the diode are easily derived,

$$V_D\,(\text{diode on}) = \phi_T \ln (I_D/I_S) = 0.75\ \text{V}$$

$$V_D\,(\text{diode off}) = I_2 R_{src} = -0.1\ \text{mA} \times 50\ \text{k}\Omega = -5\ \text{V}.$$

Using the expressions derived above, we can further estimate the lengths of the various intervals in the turn-off and turn-on transients:

$$t_1 \;=\; 5\ \text{nsec}\ \times \ln\frac{1+0.1}{0.1} \;=\; 12\ \text{nsec}$$

Finding an approximation of $(t_2 - t_1)$ requires first of all a value for the average junction capacitance C_j, which can be obtained with the aid of Eq. (1.18),

$$C_j \;=\; K_{eq} C_{j0} \;=\; 0.51 \times 0.2 \;=\; 0.102\ \text{pF}$$

assuming a voltage swing from 0 to –5V. Using this information we can easily derive the 90% transition point

$$t_2 - t_1 \;=\; 2.2 \times 50k\Omega \times 0.102pF \;=\; 11.2\ \text{nsec}$$

This yields a total turn-off time of 23.2 nsec. For the turn-on, we obtain the following delays

$$t_3 \;=\; 50\ \text{k}\Omega \times 0.10\dot{2}\ \text{pF} \times \ln\frac{1+0.1}{1} \;=\; 0.49\ \text{nsec}$$

and

$$t_4 - t_3 \;=\; 2.2 \times 5\ \text{nsec} = \;11\ \text{nsec}$$

which translates into a turn-on time of approximately 11.5 nsec. The faster response is due to larger current (1 mA) available for turning on the device (versus the 0.1 mA for turn-off).

A SPICE simulation of the same circuit produces a transient response similar to the one shown in Figure 1.11 and the following numerical results: t_1 (0 voltage crossing) = 13.2 nsec, t_2 (–4.5 V or 90% of the final value) = 23.6 nsec and t_3 (0 voltage crossing) = 0.57 nsec. Determining the correct value of t_4 is hard as there no direct way to derive the value of Q_D (and its 90% point) from the simulation results.

The obtained numbers are in perfect agreement with the simulated ones. The closeness of the match is, however, partially due to good luck. While building the estimation model, we made a number of approximations and simplifications, which obviously cause deviations from the actual result. The art of approximation is to keep the errors within reasonable bounds.

1.1.4 The Actual Diode—Secondary Effects

In practice, the diode current is less than what is predicted by the ideal diode equation. Not all applied bias voltage appears directly across the junction, as there is always some voltage drop over the neutral regions. Fortunately, the resistivity of the neutral zones is generally small (between 1 and 100 Ω, depending upon the doping levels) and the voltage drop

only becomes significant for large currents (>1 mA). This effect can be modeled by adding a resistor in series with the *n*- and *p*-region diode contacts.

In the discussion above, it was assumed that under sufficient reverse bias, the reverse current reaches a constant value, which is essentially zero. When the reverse bias exceeds a certain level, called the *breakdown voltage*, the reverse current shows a dramatic increase as shown in Figure 1.13. In the diodes found in typical MOS and bipolar

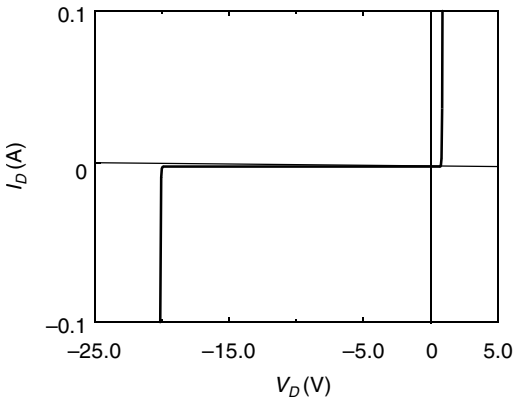


Figure 1.13 *I-V* characteristic of junction diode, showing breakdown under reverse-bias conditions (Breakdown voltage = 20 V).

processes, this increase is caused by the *avalanche breakdown*. The increasing reverse bias causes the magnitude of the electrical field across the junction to increase. Consequently, carriers crossing the depletion region are accelerated to high velocity. At a critical field E_{crit}, the carriers reach a high enough energy level that electron-hole pairs are created on collision with immobile silicon atoms. These carriers create, in turn, more carriers before leaving the depletion region. The value of E_{crit} is approximately 2×10^5 V/cm for impurity concentrations of the order of 10^{16} cm^{-3}. While avalanche breakdown in itself is not destructive and its effects disappear after the reverse bias is removed, maintaining a diode for a long time in avalanche conditions is not recommended as the high current levels (and the associated power dissipation) might cause permanent damage to the structure. Observe that avalanche breakdown is not the only breakdown mechanism encountered in diodes. For highly doped diodes, another mechanism, called Zener breakdown, can occur. Discussion of this phenomenon is beyond the scope of this text.

Finally, it is worth mentioning that the diode current is affected by the operating *temperature* in a dual way:

1. The thermal voltage ϕ_T, which appears in the exponent of the current equation, is linearly dependent upon the temperature. An increase in ϕ_T causes the current to drop.
2. The saturation current I_S is also temperature-dependent, as the thermal equilibrium carrier concentrations increase with increasing temperature. Theoretically, the saturation current approximately doubles every 5 °C. Experimentally, the reverse current has been measured to double every 8 °C.

This dual dependence has a significant impact on the operation of a digital circuit. First of all, current levels (and hence power consumption) can increase substantially. For instance, for a forward bias of 0.7 V at 300 K, the current increases approximately 6%/°C, and dou-

bles every 12 °C. Secondly, integrated circuits rely heavily on reverse-biased diodes as isolators. Increasing the temperature causes the leakage current to increase and decreases the isolation quality.

1.1.5 The SPICE Diode Model

In the preceding sections, we have presented a model for manual analysis of a diode circuit. For more complex circuits, or when a more accurate modeling of the diode that takes into account second-order effects is required, manual circuit evaluation becomes intractable, and computer-aided simulation is necessary. While different circuit simulators have been developed over the last decades, the SPICE program, developed at the University of California at Berkeley, is definitely the most successful [Nagel75]. Simulating an integrated circuit containing active devices requires a mathematical model for those devices (which is called the *SPICE model* in the rest of the text). The accuracy of the simulation depends directly upon the quality of this model. For instance, one cannot expect to see the result of a second-order effect in the simulation if this effect is not present in the device model. Creating accurate and computation-efficient SPICE models has been a long process and is by no means finished. Every major semiconductor company has developed their own proprietary models, which it claims have either better accuracy or computational efficiency and robustness.

The standard SPICE model for a diode is simple, as shown in Figure 1.14. The steady-state characteristic of the diode is modeled by the nonlinear current source I_D, which is a modified version of the ideal diode equation

$$I_D = I_S(e^{V_D/n\phi_T} - 1) \tag{1.34}$$

The extra parameter n is called the *emission coefficient*. It equals 1 for most common diodes but can be somewhat higher than 1 for others. The resistor R_s models the series resistance contributed by the neutral regions on both sides of the junction. For higher current levels, this resistance causes the internal diode V_D to differ from the externally applied voltage, hence causing the current to be lower than what would be expected from the ideal diode equation.

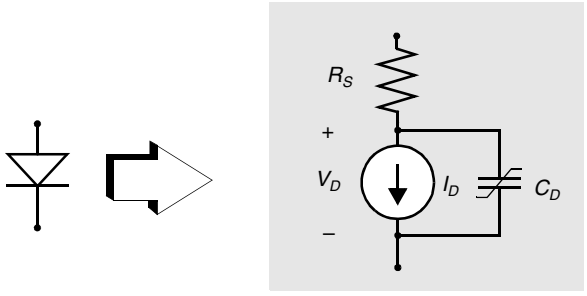


Figure 1.14 SPICE diode model.

The dynamic behavior of the diode is modeled by the nonlinear capacitance C_D, which combines the two different charge-storage effects in the diode: the excess minority carrier charge and the space charge.

$$C_D = \frac{\tau_T I_S}{\phi_T} e^{V_D/n\phi_T} + \frac{C_{j0}}{(1 - V_D/\phi_0)^m} \tag{1.35}$$

We can verify that this equation is, aside from the introduction of the emission coefficient, nothing else than a combination of Eq. (1.23) and Eq. (1.16). The parameter τ_T is called the *transit time* and represents, depending upon the diode type, the excess minority carrier lifetime (τ_n, τ_p) for long-base diodes, or the mean transit time τ_T for short-base diodes.

A listing of the parameters used in the diode model is given in Table 1.1. Besides the parameter name, symbol, and SPICE name, the table contains also the default value used by SPICE in case the parameter is left undefined. Observe that this table is by no means complete. Other parameters are available to govern second-order effects such as breakdown, high-level injection, and noise. To be concise, we chose to limit the listing to the parameters of direct interest to this text. For a complete description of the device models (as well as the usage of SPICE), we refer to the numerous textbooks devoted to SPICE (e.g., [Banhzaf92], [Thorpe92]).

Table 1.1 First-order SPICE diode model parameters.

Parameter Name	Symbol	SPICE Name	Units	Default Value
Saturation current	I_S	IS	A	1.0 E–14
Emission coefficient	n	N	–	1
Series resistance	R_S	RS	Ω	0
Transit time	τ_T	TT	sec	0
Zero-bias junction capacitance	C_{j0}	CJ0	F	0
Grading coefficient	m	M	–	0.5
Junction potential	ϕ_0	VJ	V	1