CHAPTER

1 Integrated-Circuit Devices and Modelling

In this chapter, the operation and modelling of semiconductor devices are described. Although it is possible to do simple integrated-circuit design with a basic knowledge of semiconductor device modelling, for high-speed state-of-the-art design, an in-depth understanding of the second-order effects of device operation and their modelling is considered critical.

It is assumed that most readers have been introduced to transistors and their basic modelling in a previous course. Thus, fundamental semiconductor concepts are only briefly reviewed. Section 1.1 describes pn junctions (or diodes). This section is important in understanding the parasitic capacitances in many device models, such as junction capacitances. Section 1.2 covers MOS transistors and modelling. It should be noted that this section relies to some degree on the material previously presented in Section 1.1, in which depletion capacitance is covered. Section 1.4 covers bipolar-junction transistors and modelling. A summary of device models and important equations is presented in Section 1.5. This summary is particularly useful for a reader who already has a good background in transistor modelling, in which case the summary can be used to follow the notation used throughout the remainder of this book. In addition, a brief description is given of the most important process-related parameters used in SPICE modelling. Finally, this chapter concludes with an Appendix containing derivations of the more physically based device equations.

1.1 SEMICONDUCTORS AND pn JUNCTIONS

A semiconductor is a crystal lattice structure that can have free electrons (which are negative carriers) and/or free holes (which are an absence of electrons and are equivalent to positive carriers). The type of semiconductor typically used is silicon (commonly called *sand*). This material has a valence of four, implying that each atom has four free electrons to share with neighboring atoms when forming the covalent bonds of the crystal lattice. *Intrinsic* silicon (i.e., undoped silicon) is a very pure crystal structure having equal numbers of free electrons and holes. These free carriers are those electrons or holes that have gained enough energy due to thermal agitation to escape their bonds. At room temperature, there are approximately 1.5×10^{10} carriers of each type per cm^3, or equivalently 1.5×10^{16} carriers/m^3. The number of carriers approximately doubles for every 11 °C increase in temperature.

If one dopes silicon with a pentavalent impurity (i.e., atoms of an element having a valence of five, or equivalently five electrons in the outer shell, available when bonding with neighboring atoms), there will be almost one extra free electron for every impurity atom.[1] These free electrons can be used to conduct current. A pentavalent impurity is said to *donate* free electrons to the silicon crystal, and thus the impurity is known as a *donor*. Examples of donor elements are phosphorus, P, and arsenic, As. These impurities are also called n-type dopants since the free carriers resulting from their use have negative charge. When an n-type impurity is used, the total number of negative carriers or electrons is almost the same as the doping concentration, and is much greater than the number of free electrons in intrinsic silicon. In other words,

$$n_n = N_D \tag{1.1}$$

where n_n denotes the free-electron concentration in n-type material and N_D is the doping concentration (with the subscript D denoting donor). On the other hand, the number of free holes in n-doped material will be much less than the number of holes in intrinsic silicon and can be shown [Sze, 1981] to be given by

$$p_n = \frac{n_i^2}{N_D} \tag{1.2}$$

Here, n_i is the carrier concentration in intrinsic silicon.

Similarly, if one dopes silicon with atoms having a valence of three, for example, boron (B), the concentration of positive carriers or holes will be approximately equal to the *acceptor* concentration, N_A,

$$p_p = N_A \tag{1.3}$$

and the number of negative carriers in the p-type silicon, n_p, is given by

$$n_p = \frac{n_i^2}{N_A} \tag{1.4}$$

EXAMPLE 1.1

Intrinsic silicon is doped with boron at a concentration of 10^{26} atoms/m^3. At room temperature, what are the concentrations of holes and electrons in the resulting doped silicon? Assume that $n_i = 1.5 \times 10^{16}$ carriers/m^3.

Solution

The hole concentration, p_p, will approximately equal the doping concentration ($p_p = N_A = 10^{26}$ holes/m^3). The electron concentration is found from (1.4) to be

1. In fact, there will be slightly fewer mobile carriers than the number of impurity atoms since some of the free electrons from the dopants have recombined with holes. However, since the number of holes of intrinsic silicon is much less than typical doping concentrations, this inaccuracy is small.

$$n_p = \frac{(1.5 \times 10^{16})^2}{10^{26}} = 2.3 \times 10^6 \text{ electrons/m}^3 \tag{1.5}$$

Such doped silicon is referred to as p type since it has many more free holes than free electrons.

Diodes

To realize a diode, or, equivalently, a pn junction, one part of a semiconductor is doped n type, and a closely adjacent part is doped p type, as shown in Fig. 1.1. Here the diode, or junction, is formed between the p^+ region and the n region. It should be noted that the superscripts indicate the relative doping levels. For example, the p^- bulk region might have an impurity concentration of 5×10^{21} carriers/m^3, whereas the p^+ and n^+ regions would be doped more heavily to a value around 10^{25} to 10^{27} carriers/m^3. Also, note that the metal contacts to the diode (in this case, aluminum) are connected to a heavily doped region as opposed to a lightly doped region; otherwise a *Schottky diode* would occur. (Schottky diodes are discussed on page 15.) Thus, in order not to make a Schottky diode, the connection to the n region is actually made via the n^+ region.

In the p^+ side, a large number of free positive carriers are available, whereas in the n side, many free negative carriers are available. The holes in the p^+ side will tend to disperse or diffuse into the n side, whereas the free electrons in the n side will tend to diffuse to the p^+ side. This process is very similar to two gases randomly diffusing together. This diffusion lowers the concentration of free carriers in the region between the two sides. As the two types of carriers diffuse together, they recombine. Every electron that diffuses from the n side to the p side leaves behind a *bound* positive charge close to the transition region. Similarly, every hole that diffuses from the p side leaves behind a bound electron near the transition region. The end result is shown in Fig. 1.2. This diffusion of free carriers creates a *depletion region* at the junction of the two sides where no free carriers exist, and which has a net negative charge on the p^+ side and a net positive charge on the n side. The total amount of exposed or bound

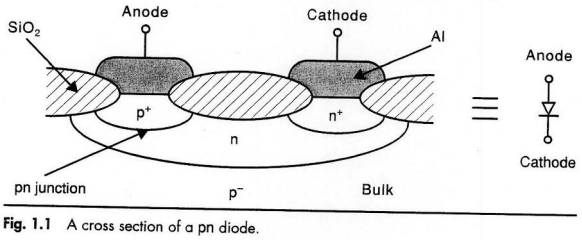


Fig. 1.1 A cross section of a pn diode.

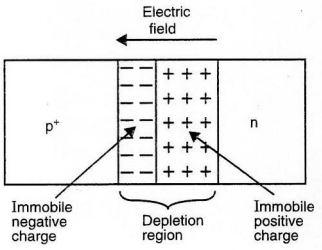


Fig. 1.2 A simplified model of a diode. Note that a depletion region exists at the junction due to diffusion and extends farther into the more lightly doped side.

charge on the two sides of the junction must be equal for charge neutrality. This requirement causes the depletion region to extend farther into the more lightly doped n side than into the p^+ side.

As these bound charges are exposed, an electric field develops going from the n side to the p side. This electric field is often called the built-in potential of the junction. It opposes the diffusion of free carriers until there is no net movement of charge under open-circuit and steady-state conditions. The built-in voltage of an open-circuit pn junction is given by Sze [1981] as

$$\Phi_0 = V_T \ln\left(\frac{N_A N_D}{n_i^2}\right) \tag{1.6}$$

where

$$V_T = \frac{kT}{q} \tag{1.7}$$

with T being the temperature in degrees Kelvin ($\cong$ 300 °K at room temperature), k being Boltzmann's constant (1.38×10^{-23} JK^{-1}), and q being the charge of an electron (1.602×10^{-19} C). At room temperature, V_T is approximately 26 mV.

EXAMPLE 1.2

A pn junction has $N_A = 10^{25}$ holes/m^3 and $N_D = 10^{22}$ electrons/m^3. What is the built-in junction potential? Assume that $n_i = 1.5 \times 10^{16}$ carriers/m^3.

Solution

Using (1.6), we obtain

$$\Phi_0 = 0.026 \times \ln\left(\frac{10^{25} \times 10^{22}}{(1.5 \times 10^{16})^2}\right) = 0.88 \text{ V} \tag{1.8}$$

This is a typical value for the built-in potential of a junction with one side heavily doped. As an approximation, we will normally use $\Phi_0 \cong 0.9$ V for the built-in potential of a junction having one side heavily doped.

Reverse-Biased Diodes

A silicon diode having an anode-to-cathode (i.e., p side to n side) voltage of 0.4 V or less will not be conducting appreciable current. In this case, it is said to be *reverse biased.* If a diode is reverse biased, current flow is primarily due to thermally generated carriers in the depletion region, and it is extremely small. Although this reverse-biased current is only weakly dependent on the applied voltage, *the reverse-biased current is directly proportional to the area of the diode junction.* However, an effect that should not be ignored, particularly at high frequencies, is the junction capacitance of a diode. In reverse-biased diodes, this junction capacitance is due to varying charge storage in the depletion regions and is modelled as a *depletion capacitance.*

To determine the depletion capacitance, we first state the relationship between the depletion widths and the applied reverse voltage, V_R [Sze, 1981].

$$x_n = \left[\frac{2K_s\varepsilon_0(\Phi_0 + V_R)}{q}\frac{N_A}{N_D(N_A + N_D)}\right]^{1/2} \tag{1.9}$$

$$x_p = \left[\frac{2K_s\varepsilon_0(\Phi_0 + V_R)}{q}\frac{N_D}{N_A(N_A + N_D)}\right]^{1/2} \tag{1.10}$$

Here, ε_0 is the permittivity of free space (equal to 8.854×10^{-12} F/m), V_R is the reverse-bias voltage of the diode, and K_s is the relative permittivity of silicon (equal to 11.8). It should be noted that these equations assume that the doping changes abruptly from the n to the p side.

From the above equations, we see that if one side of the junction is more heavily doped than the other, the depletion region will extend mostly on the lightly doped side. For example, if $N_A >> N_D$ (i.e., if the p region is more heavily doped), we can approximate (1.9) and (1.10) as

$$x_n \cong \left[\frac{2K_s\varepsilon_0(\Phi_0 + V_R)}{qN_D}\right]^{1/2} \qquad x_p \cong \left[\frac{2K_s\varepsilon_0(\Phi_0 + V_R)N_D}{qN_A^2}\right]^{1/2} \tag{1.11}$$

Indeed, for this case

$$\frac{x_n}{x_p} \cong \frac{N_A}{N_D} \tag{1.12}$$

This special case is called a *single-sided diode.*

EXAMPLE 1.3

For a pn junction having $N_A = 10^{25}$ holes/m^3 and $N_D = 10^{22}$ electrons/m^3, what are the depletion-layer depths for a 5-V reverse-bias voltage?

Solution

Since $N_A >> N_D$ and we already have found in Example 1.2 that $\Phi_0 = 0.9$ V, we can use (1.11) to find

$$x_n = \left[\frac{2 \times 11.8 \times 8.854 \times 10^{-12} \times 5.9}{1.6 \times 10^{-19} \times 10^{22}}\right]^{1/2} = 0.88\ \mu\text{m} \tag{1.13}$$

$$x_p = \frac{x_n}{(N_A/N_D)} = 0.88\ \text{nm} \tag{1.14}$$

Note that the depletion width in the lightly doped n region is 1,000 times greater than that in the more heavily doped p region.

The charge stored in the depletion region, per unit cross-sectional area, is found by multiplying the depletion-region width by the concentration of the immobile charge (which is approximately equal to q times the impurity doping density). For example, on the n side, we find the charge in the depletion region to be given by multiplying (1.9) by qN_D, resulting in

$$Q^+ = \left[2qK_s\varepsilon_0(\Phi_0 + V_R)\frac{N_AN_D}{N_A + N_D}\right]^{1/2} \tag{1.15}$$

This amount of charge must also equal Q^- on the p side since there is charge equality. In the case of a single-sided diode when $N_A >> N_D$, we have

$$Q^- = Q^+ \cong [2qK_s\varepsilon_0(\Phi_0 + V_R)N_D]^{1/2} \tag{1.16}$$

Note that this result is independent of the impurity concentration on the heavily doped side. Thus, we see from the above relation that the charge stored in the depletion region is dependent on the applied reverse-bias voltage. *It is this charge-voltage relationship that is modelled by a nonlinear depletion capacitance.*

For small changes in the reverse-biased junction voltage, about a bias voltage, we can find an equivalent *small-signal* capacitance, C_j, by differentiating (1.15) with respect to V_R. Such a differentiation results in

$$C_j = \frac{dQ^+}{dV_R} = \left[\frac{qK_s\varepsilon_0}{2(\Phi_0 + V_R)}\frac{N_AN_D}{N_A + N_D}\right]^{1/2} = \frac{C_{j0}}{\sqrt{1 + \dfrac{V_R}{\Phi_0}}} \tag{1.17}$$

where C_{j0} is the depletion capacitance per unit area at $V_R = 0$ and is given by

$$C_{j0} = \sqrt{\frac{qK_s\varepsilon_0}{2\Phi_0}\frac{N_A N_D}{N_A + N_D}} \tag{1.18}$$

In the case of a one-sided diode with $N_A >> N_D$, we have

$$C_j = \left[\frac{qK_s\varepsilon_0 N_D}{2(\Phi_0 + V_R)}\right]^{1/2} = \frac{C_{j0}}{\sqrt{1 + \frac{V_R}{\Phi_0}}} \tag{1.19}$$

where now

$$C_{j0} = \sqrt{\frac{qK_s\varepsilon_0 N_D}{2\Phi_0}} \tag{1.20}$$

It should be noted that many of the junctions encountered in integrated circuits are one-sided junctions with the lightly doped side being the substrate or sometimes what is called the *well*. The more heavily doped side is often used to form a contact to interconnecting metal. From (1.20), we see that, for these one-sided junctions, the depletion capacitance is approximately independent of the doping concentration on the heavily doped side, and is proportional to the square root of the doping concentration of the more lightly doped side. Thus, smaller depletion capacitances are obtained for more lightly doped substrates—a strong incentive to strive for lightly doped substrates.

Finally, note that by combining (1.15) and (1.18), we can express the equation for the immobile charge on either side of a reverse-biased junction as

$$Q = 2C_{j0}\Phi_0\sqrt{1 + \frac{V_R}{\Phi_0}} \tag{1.21}$$

As seen in Example 1.6, this equation is useful when one is approximating the large-signal charging (or discharging) time for a reverse-biased diode.

EXAMPLE 1.4

For a pn junction having $N_A = 10^{25}$ holes/m^3 and $N_D = 10^{22}$ electrons/m^3, what is the total zero-bias depletion capacitance for a diode of area 10 μm × 10 μm? What is its depletion capacitance for a 3-V reverse-bias voltage?

Solution

Making use of (1.20), we have

$$C_{j0} = \sqrt{\frac{1.6\times10^{-19}\times11.8\times8.854\times10^{-12}\times10^{22}}{2\times0.9}} = 304.7\ \mu\text{F/m}^2 \tag{1.22}$$

Since the diode area is 100×10^{-12} m^2, the total zero-bias depletion capacitance is

$$C_{T\text{-}j0} = 100\times10^{-12}\times304.7\times10^{-6} = 30.5\ \text{fF} \tag{1.23}$$

At a 3-V reverse-bias voltage, we have from (1.19)

$$C_{T\text{-}j} = \frac{30.5\ \text{fF}}{\sqrt{1 + \left(\frac{3}{0.9}\right)}} = 14.7\ \text{fF} \tag{1.24}$$

As expected, we see a decrease in junction capacitance as the width of the depletion region is increased.

Graded Junctions

All of the above equations assumed an abrupt junction where the doping concentration changes quickly from p to n over a small distance. Although this is a good approximation for many integrated circuits, it is not always true. For example, the collector-to-base junction of a bipolar transistor is most commonly realized as a *graded* junction. In the case of graded junctions, the exponent 1/2 in Eq. (1.15) is inaccurate, and a better value to use is an exponent closer to unity, perhaps 0.6 to 0.7. Thus, for graded junctions, (1.15) is typically written as

$$Q = \left[2qK_s\varepsilon_0(\Phi_0 + V_R)\frac{N_A N_D}{N_A + N_D}\right]^{1-m} \tag{1.25}$$

where m is a constant typically around 1/3.

Differentiating (1.25) to find the depletion capacitance, we have

$$C_j = (1-m)\left[2qK_s\varepsilon_0\frac{N_A N_D}{N_A + N_D}\right]^{1-m}\frac{1}{(\Phi_0 + V_R)^m} \tag{1.26}$$

This depletion capacitance can also be written as

$$C_j = \frac{C_{j0}}{\left(1 + \frac{V_R}{\Phi_0}\right)^m} \tag{1.27}$$

where

$$C_{j0} = (1-m)\left[2qK_s\varepsilon_0\frac{N_A N_D}{N_A + N_D}\right]^{1-m}\frac{1}{\Phi_0^m} \tag{1.28}$$

From (1.27), we see that a graded junction results in a depletion capacitance that is less dependent on V_R than the equivalent capacitance in an abrupt junction. In other words, since m is less than 0.5, the depletion capacitance for a graded junction is

more linear than that for an abrupt junction. Correspondingly, increasing the reverse-bias voltage for a graded junction is not as effective in reducing the depletion capacitance as it is for an abrupt junction.

Finally, as in the case of an abrupt junction, the depletion charge on either side of the junction can also be written as

$$Q = \frac{C_{j0}}{1-m}\Phi_0\left(1+\frac{V_R}{\Phi_0}\right)^{1-m} \tag{1.29}$$

EXAMPLE 1.5

Repeat Example 1.4 for a graded junction with $m = 0.4$.

Solution

Noting once again that $N_A >> N_D$, we approximate (1.28) as

$$C_{j0} = (1-m)[2qK_s\varepsilon_0 N_D]^{1-m}\frac{1}{\Phi_0^m} \tag{1.30}$$

resulting in

$$C_{j0} = 81.5\ \mu F/m^2 \tag{1.31}$$

which, when multiplied by the diode's area of $10\ \mu m \times 10\ \mu m$, results in

$$C_{T\text{-}j0} = 8.1\ fF \tag{1.32}$$

For a 3-V reverse-bias voltage, we have

$$C_{T\text{-}j} = \frac{8.1 fF}{(1+3/0.9)^{0.4}} = 4.5\ fF \tag{1.33}$$

Large-Signal Junction Capacitance

The equations for the junction capacitance given above are only valid for small changes in the reverse-bias voltage. This limitation is due to the fact that C_j depends on the size of the reverse-bias voltage instead of being a constant. As a result, it is extremely difficult and time consuming to accurately take this nonlinear capacitance into account when calculating the time to charge or discharge a junction over a large voltage change. A commonly used approximation when analyzing the transient response for large voltage changes is to use an *average size* for the junction capacitance by calculating the junction capacitance at the two extremes of the reverse-bias voltage. Unfortunately, a problem with this approach is that when the diode is forward biased with $V_R \cong -\Phi_0$, Eq. (1.17) "blows up" (i.e., is equal to infinity). To circumvent this

problem, one can instead calculate the charge stored in the junction for the two extreme values of applied voltage (through the use of (1.21)), and then through the use of $Q = CV$, calculate the average capacitance according to

$$C_{j\text{-}av} = \frac{Q(V_2)-Q(V_1)}{V_2-V_1} \tag{1.34}$$

where V_1 and V_2 are the two voltage extremes [Hodges, 1988].

From (1.21), for an abrupt junction with reverse-bias voltage V_i, we have

$$Q(V_i) = 2C_{j0}\Phi_0\sqrt{1+\frac{V_i}{\Phi_0}} \tag{1.35}$$

Therefore,

$$C_{j\text{-}av} = 2C_{j0}\Phi_0\frac{\left(\sqrt{1+\frac{V_2}{\Phi_0}}-\sqrt{1+\frac{V_1}{\Phi_0}}\right)}{V_2-V_1} \tag{1.36}$$

One special case often encountered is charging a junction from 0 V to 5 V. For this special case, and using $\Phi_0 = 0.9$ V, we find that

$$C_{j\text{-}av} = 0.56C_{j0} \tag{1.37}$$

Thus, as a rough approximation to quickly estimate the charging time of a junction capacitance from 0 V to 5 V (or vice versa), one can use

$$C_{j\text{-}av} = \frac{C_{j0}}{2} \tag{1.38}$$

It will be seen in the following example that (1.37) compares well with a SPICE simulation.

EXAMPLE 1.6

For the circuit shown in Fig. 1.3, where a reverse-biased diode is being charged from 0 V to 5 V, through a 10-kΩ resistor, calculate the time required to charge the diode from 0 V to 3.5 V. Assume that $C_{j0} = 0.2$ fF/(μm)^2 and that the diode has an area of $20\ \mu m \times 5\ \mu m$. Compare your answer to that obtained using SPICE. Repeat the question for the case of the diode being discharged from 5 V to 1.5 V.

Solution

The total small-signal capacitance of the junction at 0-V bias voltage is obtained by multiplying 0.2 fF/(μm)^2 by the junction area to obtain

$$C_{j0} = 0.2\times10^{-15}\times20\times5 = 0.02\ pF \tag{1.39}$$

Using (1.37), we have

$$C_{j\text{-}av} = 0.56\times0.02 = 0.011\ pF \tag{1.40}$$

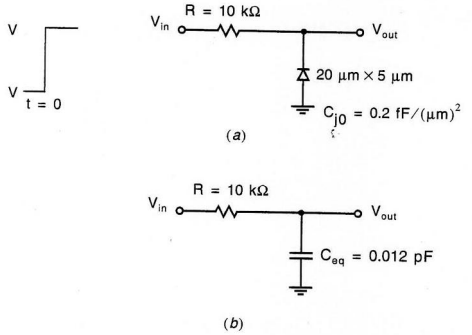


Fig. 1.3 (*a*) The circuit used in Example 1.6; (*b*) its RC approximate equivalent.

resulting in a time constant of

$$\tau = RC_{j\text{-}av} = 0.11 \text{ ns} \tag{1.41}$$

It is not difficult to show that the time it takes for a first-order circuit to rise (or fall) 70 percent of its final value is equal to 1.2τ. Thus, in this case,

$$t_{70\%} = 1.2\tau = 0.13 \text{ ns} \tag{1.42}$$

As a check, the circuit of Fig. 1.3(*a*) was analyzed using SPICE. The input data file was as follows:

```
R 1 2 10k
D 0 2 DMOD
*
VIN 1 0 dc 2.5 PULSE (0 5 0 10p 10p 0.49n 1.0n)
*
.MODEL DMOD D(CJO=0.02E-12)
*
.OPTIONS NUMDGT=5 ITL1=500
.WIDTH OUT=80
.TRAN 0.01n 1.0n
.PRINT TRAN V(2)
.END
```

The SPICE simulation gave a 0-V to 3.5-V rise time of 0.14 ns and a 5-V to 1.5-V fall time of 0.12 ns. These times compare favorably with the 0.13 ns predicted. The reason for the different values of the rise and fall times is the nonlinearity of the junction capacitance. For smaller bias voltages it is larger than that

predicted by (1.37), whereas for larger bias voltages it is smaller. If we use the more accurate approximation of (1.36) for the rise time with $V_2 = 3.5$ and $V_1 = 0$ V, we find

$$C_{j\text{-}av} = 2 \times 0.02 \times \frac{0.9}{3.5}\left(\sqrt{1 + \frac{3.5}{0.9}} - 1\right) = 0.012 \text{ pF} \tag{1.43}$$

Also, for the fall time, we find that

$$C_{j\text{-}av} = 2 \times 0.02 \times \frac{0.9}{1.5 - 5}\left(\sqrt{1 + \frac{1.5}{0.9}} - \sqrt{1 + \frac{5}{0.9}}\right) = 0.010 \text{ pF} \tag{1.44}$$

These more accurate approximations result in

$$t_{+70\%} = 0.144 \text{ ns} \tag{1.45}$$

and

$$t_{-70\%} = 0.114 \text{ ns} \tag{1.46}$$

in closer agreement with SPICE. Normally, the extra accuracy that results from using (1.36) instead of (1.37) is not worth the extra complication because one seldom knows the area of C_{j0} to better than 20 percent accuracy.

Forward-Biased Junctions

A positive voltage applied from the p side to the n side of a diode reduces the electric field opposing the diffusion of the free carriers across the depletion region. It also reduces the width of the depletion region. If this forward-bias voltage is large enough, the carriers will start to diffuse across the junction, resulting in a current flow from the anode to the cathode. For silicon, appreciable diode current starts to occur for a forward-bias voltage around 0.5 V. For germanium and gallium arsenide semiconductor materials, current conduction starts to occur around 0.3 V and 0.9 V, respectively.

When the junction potential is sufficiently lowered for conduction to occur, the carriers diffuse across the junction due to the large gradient in the mobile carrier concentrations. Note that there are more carriers diffusing from the heavily doped side to the lightly doped side than from the lightly doped side to the heavily doped side.

After the carriers cross the depletion region, they greatly increase the *minority charge* at the edge of the depletion region. These minority carriers will diffuse away from the junction toward the bulk. As they diffuse, they recombine with the majority carriers, thereby decreasing their concentration. This concentration gradient of the minority charge (which decreases the farther one gets from the junction) is responsible for the current flow near the junction.

The majority carriers that recombine with the diffusing minority carriers come from the metal contacts at the junctions because of the forward-bias voltage. These majority carriers flow across the bulk, from the contacts to the junction, due to an electric field applied across the bulk. This current flow is called *drift*. It results in

small potential drops across the bulk, especially in the lightly doped side. Typical values of this voltage drop might be 50 mV to 0.1 V, depending primarily on the doping concentration of the lightly doped side, the distance from the contacts to the junction, and the cross-sectional area of the junction.

In the forward-bias region, the current-voltage relationship is exponential and can be shown (see Appendix) to be

$$I_D = I_S e^{V_D/V_T} \tag{1.47}$$

where V_D is the voltage applied across the diode and

$$I_S \propto A_D\left(\frac{1}{N_A} + \frac{1}{N_D}\right) \tag{1.48}$$

I_S is known as the *scale current* and is seen to be proportional to the area of the diode junction, A_D, and inversely proportional to the doping concentrations.

Junction Capacitance of Forward-Biased Diode

When a junction changes from reverse biased (with little current through it) to forward biased (with significant current flow across it), the charge being stored near and across the junction changes. Part of the change in charge is due to the change in the width of the depletion region and therefore the amount of immobile charge stored in it. This change in charge is modelled by the depletion capacitance, C_j, similar to when the junction is reverse biased. An additional change in charge storage is necessary to account for the change of the minority carrier concentration close to the junction required for the diffusion current to exist. For example, if a forward-biased diode current is to double, then the slopes of the minority charge storage at the diode junction edges must double, and this, in turn, implies that the minority charge storage must double. This component is modelled by another capacitance, called the *diffusion capacitance,* and denoted C_d.

The diffusion capacitance can be shown (see Appendix) to be

$$C_d = \tau_T \frac{I_D}{V_T} \tag{1.49}$$

where τ_T is the transit time of the diode. Normally τ_T is specified for a given technology, so that one can calculate the diffusion capacitance. *Note that the diffusion capacitance of a forward-biased junction is proportional to the diode current.*

The total capacitance of the forward-biased junction is the sum of the diffusion capacitance, C_d, and the depletion capacitance, C_j. Thus, the total junction capacitance is given by

$$C_T = C_d + C_j \tag{1.50}$$

For a forward-biased junction, the depletion capacitance, C_j, can be roughly approximated by $2C_{j0}$. The accuracy of this approximation is not critical since the diffusion capacitance is typically much larger than the depletion capacitance.

Finally, it should be mentioned that as a diode is turned off for a short period of time a current will flow in the negative direction until the minority charge is removed. This behavior does not occur in Schottky diodes since they do not have minority charge storage.

Small-Signal Model of a Forward-Biased Diode

A small-signal equivalent model for a forward-biased diode is shown in Fig. 1.4. A resistor, r_d, models the change in the diode voltage, V_D, that occurs when I_D changes. Using (1.47), we have

$$\frac{1}{r_d} = \frac{dI_D}{dV_D} = I_S \frac{e^{V_D/V_T}}{V_T} = \frac{I_D}{V_T} \tag{1.51}$$

This resistance is called the incremental resistance of the diode. For very accurate modelling, it is sometimes necessary to add the series resistance due to the bulk and also the resistance associated with the contacts. Typical values for the contact resistance (caused by the work-function[2] difference between metal and silicon) might be 20 Ω to 40 Ω.

By combining (1.49) and (1.51), we see that an alternative equation for the diffusion capacitance, C_d, is

$$C_d = \frac{\tau_T}{r_d} \tag{1.52}$$

Since for moderate forward-bias currents, $C_d >> C_j$, the total small-signal capacitance is $C_T \cong C_d$, and

$$r_d C_T \cong \tau_T \tag{1.53}$$

Thus, for charging or discharging a forward-biased junction with a current source having an impedance much larger than r_d, the time constant of the charging is approximately equal to the transit time of the diode and is independent of the diode current. For smaller diode currents, where C_j becomes important, the charging or discharging time constant of the circuit becomes larger than τ_T.

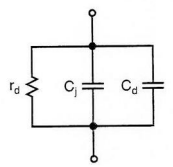


Fig. 1.4 The small-signal model for a forward-biased junction.

2. The work-function of a material is defined as the minimum energy required to remove an electron at the Fermi level to the outside vacuum region.

predicted by (1.37), whereas for larger bias voltages it is smaller. If we use the more accurate approximation of (1.36) for the rise time with $V_2 = 3.5$ and $V_1 = 0$ V, we find

$$C_{j\text{-}av} = 2 \times 0.02 \times \frac{0.9}{3.5}\left(\sqrt{1 + \frac{3.5}{0.9}} - 1\right) = 0.012 \text{ pF} \tag{1.43}$$

Also, for the fall time, we find that

$$C_{j\text{-}av} = 2 \times 0.02 \times \frac{0.9}{1.5 - 5}\left(\sqrt{1 + \frac{1.5}{0.9}} - \sqrt{1 + \frac{5}{0.9}}\right) = 0.010 \text{ pF}. \tag{1.44}$$

These more accurate approximations result in

$$t_{+70\%} = 0.144 \text{ ns} \tag{1.45}$$

and

$$t_{-70\%} = 0.114 \text{ ns} \tag{1.46}$$

in closer agreement with SPICE. Normally, the extra accuracy that results from using (1.36) instead of (1.37) is not worth the extra complication because one seldom knows the area of C_{j0} to better than 20 percent accuracy.

Forward-Biased Junctions

A positive voltage applied from the p side to the n side of a diode reduces the electric field opposing the diffusion of the free carriers across the depletion region. It also reduces the width of the depletion region. If this forward-bias voltage is large enough, the carriers will start to diffuse across the junction, resulting in a current flow from the anode to the cathode. For silicon, appreciable diode current starts to occur for a forward-bias voltage around 0.5 V. For germanium and gallium arsenide semiconductor materials, current conduction starts to occur around 0.3 V and 0.9 V, respectively.

When the junction potential is sufficiently lowered for conduction to occur, the carriers diffuse across the junction due to the large gradient in the mobile carrier concentrations. Note that there are more carriers diffusing from the heavily doped side to the lightly doped side than from the lightly doped side to the heavily doped side.

After the carriers cross the depletion region, they greatly increase the *minority charge* at the edge of the depletion region. These minority carriers will diffuse away from the junction toward the bulk. As they diffuse, they recombine with the majority carriers, thereby decreasing their concentration. This concentration gradient of the minority charge (which decreases the farther one gets from the junction) is responsible for the current flow near the junction.

The majority carriers that recombine with the diffusing minority carriers come from the metal contacts at the junctions because of the forward-bias voltage. These majority carriers flow across the bulk, from the contacts to the junction, due to an electric field applied across the bulk. This current flow is called *drift*. It results in

EXAMPLE 1.7

A given diode has a transit time of 100 ps and is biased at 1 mA. What are the values of its small-signal resistance and diffusion capacitance? Assume room temperature, so that $V_T = kT/q = 26$ mV.

Solution

We have

$$r_d = \frac{V_T}{I_D} = \frac{26 \text{ mV}}{1 \text{ mA}} = 26 \ \Omega$$

and

$$C_d = \frac{\tau_T}{r_d} = \frac{100 \text{ ps}}{26 \ \Omega} = 3.8 \text{ pF}$$

Note that this diffusion capacitance is over 100 times larger than the total depletion capacitance found in Examples 1.4 and 1.5.

Schottky Diodes

A different type of diode, one often used in microcircuit design, is realized by contacting metal to a lightly doped semiconductor region (rather than a heavily doped region) as shown in Fig. 1.5. Notice that the aluminum anode is in direct contact with a relatively lightly doped n^- region. Because the n^- region is relatively lightly doped, the work-function difference between the aluminum contact and the n^- silicon is larger than would be the case for aluminum contacting to an n^+ region, as occurs at the cathode. This causes a depletion region and, correspondingly, a diode to occur at the interface between the aluminum anode and the n^- silicon region. This diode has different characteristics than a normal pn junction diode. First, its voltage drop when forward

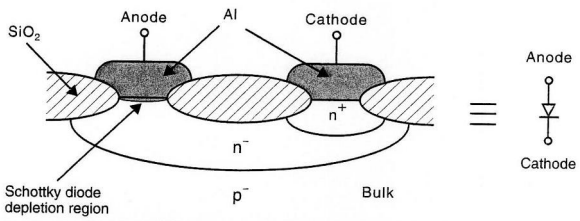


Fig. 1.5 A cross section of a Schottky diode.

biased is smaller. This voltage drop is dependent on the metal used; for aluminum it might be around 0.5 V. More importantly, when the diode is forward biased, there is no minority-charge storage in the lightly doped n^- region. Thus, the small-signal model of a forward-biased Schottky diode has $C_d = 0$ (with reference to Fig. 1.4). The absence of this diffusion capacitance makes the diode much faster. It is particularly faster when turning off, because it is not necessary to remove the minority charge first. Rather, it is only necessary to discharge the depletion capacitance through about 0.2 V.

Schottky diodes have been used extensively in bipolar logic circuits. They are also used in a number of high-speed analog circuits, particularly those realized in gallium arsenide (GaAs) technologies, rather than silicon technologies.

1.2 MOS TRANSISTORS

Presently, the most popular technology for realizing microcircuits makes use of MOS transistors. Unlike most bipolar junction transistor (BJT) technologies, which make dominant use of only one type of transistor (npn transistors in the case of BJT processes[3]), MOS circuits normally use two complementary types of transistors—n-channel and p-channel. While n-channel devices conduct with a positive gate voltage, p-channel devices conduct with a negative gate voltage. Moreover, electrons are used to conduct current in n-channel transistors, while holes are used in p-channel transistors. Microcircuits containing both n-channel and p-channel transistors are called CMOS circuits, for *complementary MOS*. The acronym MOS stands for *metal-oxide semiconductor,* which historically denoted the gate, insulator, and channel region materials, respectively. However, most present CMOS technologies utilize polysilicon gates rather than metal gates.

Before CMOS technology became widely available, most MOS processes made use of only n-channel transistors (NMOS). However, often two different types of n-channel transistors could be realized. One type, enhancement n-channel transistors, is similar to the n-channel transistors realized in CMOS technologies. Enhancement transistors require a positive gate-to-source voltage to conduct current. The other type, depletion transistors, conduct current with a gate-source voltage of 0 V. Depletion transistors were used to create high-impedance loads in NMOS logic gates.

A typical cross section of an n-channel enhancement-type MOS transistor is shown in Fig. 1.6. With no voltage applied to the gate, the n^+ source and drain regions are separated by the p^- substrate. The separation between the drain and the source is called the channel length, L. In present MOS technologies, the minimum channel length is typically between 0.3 μm and 1.0 μm. It should be noted that there is no physical difference between the drain and the source.[4] *The source terminal of an*

3. Most BJT technologies can also realize low-speed lateral pnp transistors. Normally these would only be used to realize current sources as they have low gains and poor frequency responses. Recently, bipolar technologies utilizing high-speed vertical pnp transistors, as well as high-speed npn transistors, have become available and are growing in popularity. These technologies are called complementary bipolar technologies.
4. Large MOS transistors used for power applications might not be realized with symmetric drain and source junctions.

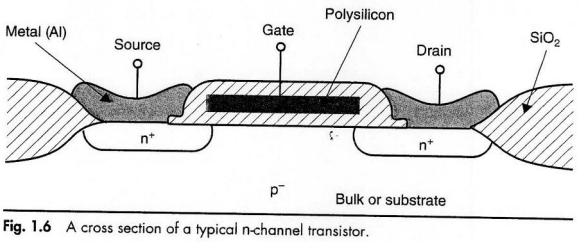


Fig. 1.6 A cross section of a typical n-channel transistor.

n-channel transistor is defined as whichever of the two terminals has a lower voltage. For a p-channel transistor, the source would be the terminal with the higher voltage. When a transistor is turned on, current flows from the drain to the source in an n-channel transistor and from the source to the drain in a p-channel transistor. In both cases, the true carriers travel from the source to drain, but the current directions are different because n-channel carriers (electrons) are negative, whereas p-channel carriers (holes) are positive.

The gate is normally realized using polysilicon, which is heavily doped noncrystalline (or amorphous) silicon. Polysilicon gates are used nowadays (instead of metal) because polysilicon allows the dimensions of the transistor to be realized much more accurately during the patterning of the transistor, which involves what is called a self-aligned process. This higher geometric accuracy results in smaller, faster transistors.

The gate is physically separated from the surface of the silicon by a thin insulator made of silicon dioxide (SiO_2). Thus, the gate is electrically isolated from the channel and affects the channel (and hence, the transistor current) only through electrostatic coupling, similar to capacitive coupling. A typical thickness for the SiO_2 insulator between the gate and the channel is presently from 0.01 μm to 0.03 μm. Since the gate is electrically isolated from the channel, it never conducts dc current. Indeed, the excellent isolation results in leakage currents being almost undetectable. However, because of the inherent capacitances in MOS transistors, transient gate currents do exist when gate voltages are quickly changing.

Normally the p^- substrate (or bulk) is connected to the most negative voltage in a microcircuit. In analog circuits, this might be the negative power supply, but in digital circuits it is normally ground or 0 V. This connection results in all transistors placed in the substrate being surrounded by reverse-biased junctions, which electrically isolate the transistors and thereby prevent conduction through the substrate between transistors (unless, of course, they are connected together through some other means).

Symbols for MOS Transistors

Many symbols have been used to represent MOS transistors. Figure 1.7 shows some of the symbols that have been used to represent n-channel MOS transistors. The symbol in Fig. 1.7(*a*) is often used; note that there is nothing in the symbol to specify

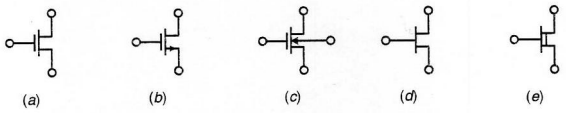


Fig. 1.7 Commonly used symbols for n-channel transistors.

whether the transistor is n-channel or p-channel. A common rule is to assume, when in doubt, that the transistor is an n-channel enhancement transistor. Figure 1.7(*b*) is the most commonly used symbol for an n-channel enhancement transistor and is used throughout this text. The arrow pointing outward on the source indicates that the transistor is n-channel, similar to the convention used for npn transistors, and indicates the direction of hole current.

MOS transistors are actually four-terminal devices, with the substrate being the fourth terminal. In n-channel devices, the p^- substrate is normally connected to the most negative voltage in the microcircuit, whereas for p-channel devices, the n^- substrate is normally connected to the most positive voltage. In these cases the substrate connection is normally not shown in the symbol. However, for CMOS technologies, at least one of the two types of transistors will be formed in a *well* substrate that need not be connected to one of the power supply nodes. For example, an n-well process would form n-channel transistors in a p^- substrate encompassing the entire microcircuit, while the p-channel transistors would be formed in many n-well substrates. In this case, most of the n-well substrates would be connected to the most positive power supply, while some might be connected to other nodes in the circuit (often the well is connected to the source of a transistor that is not connected to the power supply). In these cases, the symbol shown in Fig. 1.7(*c*) can be used to show the substrate connection explicitly. It should be noted that this case is not encountered often in digital circuits and is more common in analog circuits. Sometimes, in the interest of simplicity, the isolation of the gate is not explicitly shown, as is the case of the symbol of Fig. 1.7(*d*). This simple notation is more common for digital circuits in which a large number of transistors are present. Since this symbol is also used for JFET transistors, it will never be used to represent MOS transistors in this text. The last symbol, shown in Fig. 1.7(*e*), denotes an n-channel depletion transistor. The extra line is used to indicate that a physical channel exists for a 0-V gate-source voltage. Depletion transistors were used in older NMOS technologies but are not typically available in CMOS processes.

Figure 1.8 shows some commonly used symbols for p-channel transistors. In this text, the symbol of Fig. 1.8(*a*) will be most often used. The symbol in Fig. 1.8(*c*) is

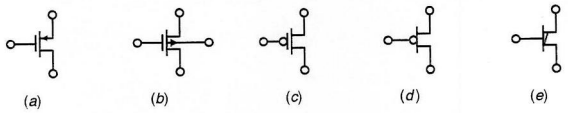


Fig. 1.8 Commonly used symbols for p-channel transistors.

sometimes used in digital circuits, where the circle indicates that a low voltage on the gate turns the transistor on, as opposed to a high voltage for an n-channel transistor (Fig. 1.7(*a*)). The symbols of Fig. 1.8(*d*) or Fig. 1.8(*e*) might be used in larger circuits where many transistors are present, to simplify the drawing somewhat. They will not be used in this text.

Basic Operation

The basic operation of MOS transistors will be described with respect to an n-channel transistor. First, consider the simplified cross sections shown in Fig. 1.9, where the source, drain, and substrate are all connected to ground. In this case, the MOS transistor operates similarly to a capacitor. The gate acts as one plate of the capacitor, and the surface of the silicon, just under the thin insulating SiO_2, acts as the other plate.

If the gate voltage is very negative, as shown in Fig. 1.9(*a*), positive charge will be attracted to the channel region. Since the substrate was originally doped p^-, this negative gate voltage has the effect of simply increasing the channel doping to p^+,

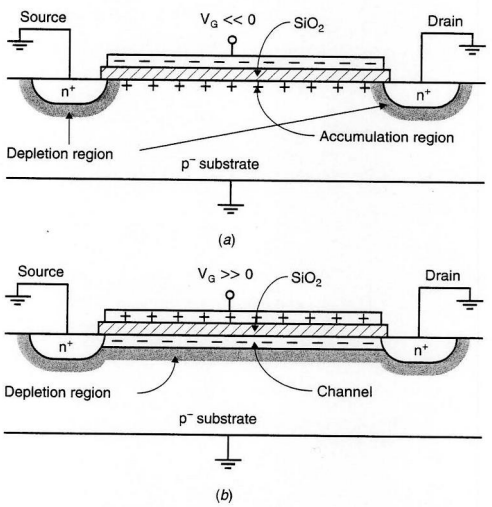


Fig. 1.9 An n-channel MOS transistor. (*a*) $V_G << 0$, resulting in an accumulated channel (no current flow); (*b*) $V_G >> 0$, and the channel is present (current flow possible from drain to source).

resulting in what is called an *accumulated channel.* The n^+ source and drain regions are separated from the p^+-channel region by depletion regions, resulting in the equivalent circuit of two back-to-back diodes. Thus, only leakage current will flow even if one of the source or drain voltages becomes large (unless the drain voltage becomes so large as to cause the transistor to break down).

In the case of a positive voltage being applied to the gate, the opposite situation occurs, as shown in Fig. 1.9(*b*). For small positive gate voltages, the positive carriers in the channel under the gate are initially repulsed and the channel changes from a p^- doping level to a depletion region. As a more positive gate voltage is applied, the gate attracts negative charge from the source and drain regions, and the channel becomes an n region with mobile electrons connecting the drain and source regions.[5] In short, a sufficiently large positive gate-source voltage changes the channel beneath the gate to an n region, and the channel is said to be *inverted.*

The gate-source voltage, for which the concentration of electrons under the gate is equal to the concentration of holes in the p^- substrate far from the gate, is commonly referred to as the *transistor threshold voltage* and denoted V_{tn} (for n-channel transistors). For gate-source voltages larger than V_{tn}, there is an n-type channel present, and conduction between the drain and the source can occur. For gate-source voltages less than V_{tn}, it is normally assumed that the transistor is off and no current flows between the drain and the source. However, it should be noted that this assumption of zero drain-source current for a transistor that is off is only an approximation. In fact, for gate voltages around V_{tn}, there is no abrupt current change, and for gate-source voltages slightly less than V_{tn}, small amounts of *subthreshold current* can flow, as discussed in Section 1.3.

When the gate-source voltage, V_{GS}, is larger than V_{tn}, the channel is present. As V_{GS} is increased, the density of electrons in the channel increases. Indeed, the carrier density, and therefore the charge density, is proportional to $V_{GS} - V_{tn}$, which is often called the *effective gate-source voltage* and denoted V_{eff}. Specifically, define

$$V_{eff} \equiv V_{GS} - V_{tn} \tag{1.54}$$

The charge density of electrons is then given by

$$Q_n = C_{ox}(V_{GS} - V_{tn}) = C_{ox}V_{eff} \tag{1.55}$$

Here, C_{ox} is the gate capacitance per unit area and is given by

$$C_{ox} = \frac{K_{ox}\varepsilon_0}{t_{ox}} \tag{1.56}$$

where K_{ox} is the relative permittivity of SiO_2 (approximately 3.9) and t_{ox} is the thickness of the thin oxide under the gate. A point to note here is that (1.55) is only accurate when both the drain and the source voltages are zero.

5. The drain and source regions are sometimes called diffusion regions or junctions for historical reasons. This use of the word *junction* is not synonymous with our previous use, in which it designated a pn interface of a diode.

To obtain the total gate capacitance, (1.56) should be multiplied by the effective gate area, WL, where W is the gate width and L is the effective gate length. These dimensions are shown in Fig. 1.10. Thus the total gate capacitance, C_{gs}, is given by

$$C_{gs} = WLC_{ox} \tag{1.57}$$

and the total charge of the channel, $Q_{T\text{-}n}$, is given by

$$Q_{T\text{-}n} = WLC_{ox}(V_{GS} - V_{tn}) = WLC_{ox}V_{eff} \tag{1.58}$$

The gate capacitance, C_{gs}, is one of the major load capacitances that circuits must be capable of driving. Gate capacitances are also important when one is calculating *charge injection,* which occurs when a MOS transistor is being turned off because the channel charge, $Q_{T\text{-}n}$, must flow from under the gate out through the terminals to other places in the circuit.

Next, if the drain voltage is increased above 0 V, a drain-source potential difference exists. This difference results in current flowing from the drain to the source.[6] The relationship between V_{DS} and the drain-source current, I_D, is the same as for a resistor, assuming V_{DS} is small. This relationship is given [Sze, 1981] by

$$I_D = \mu_n Q_n \frac{W}{L} V_{DS} \tag{1.59}$$

where $\mu_n \cong 0.06\ m^2/Vs$ is the mobility of electrons near the silicon surface, and Q_n is the charge concentration of the channel per unit area (looking from the top down). Note that as the channel length increases, the drain-source current decreases, whereas this current increases as either the charge density or the transistor width increases. Using (1.58) and (1.59) results in

$$I_D = \mu_n C_{ox}\frac{W}{L}(V_{GS} - V_{tn})V_{DS} = \mu_n C_{ox}\frac{W}{L}V_{eff}V_{DS} \tag{1.60}$$

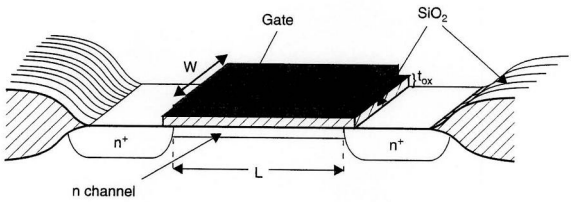


Fig. 1.10 The important dimensions of a MOS transistor.

6. The current is actually conducted by negative carriers (electrons) flowing from the source to the drain. Negative carriers flowing from source to drain results in a positive current from drain to source, I_{DS}.

where it should be emphasized that this relationship is only valid for drain-source voltages near zero (i.e., V_{DS} much smaller than V_{eff}).

As the drain-source voltage increases, the channel charge concentration decreases at the drain end. This decrease is due to the smaller gate-to-channel voltage difference across the thin gate oxide as one moves closer to the drain. In other words, since the drain voltage is assumed to be at a higher voltage than the source, there is an increasing voltage gradient from the source to the drain, resulting in a smaller gate-to-channel voltage near the drain. Since the charge density at a distance x from the source end of the channel is proportional to $V_G - V_{ch}(x) - V_{tn}$, as $V_G - V_{ch}(x)$ decreases, the charge density also decreases.[7] This effect is illustrated in Fig. 1.11.

Note that at the drain end of the channel, we have

$$V_G - V_{ch}(L) = V_{GD} \tag{1.61}$$

For small V_{DS}, we saw from (1.60) that I_D was linearly related to V_{DS}. However, as V_{DS} increases, and the charge density decreases near the drain, the relationship becomes nonlinear. In fact, the linear relationship for I_D versus V_{DS} flattens for larger V_{DS}, as shown in Fig. 1.12.

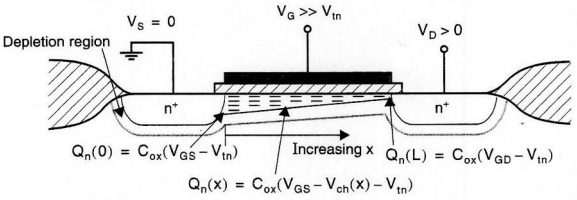


Fig. 1.11 The channel charge density for $V_{DS} > 0$.

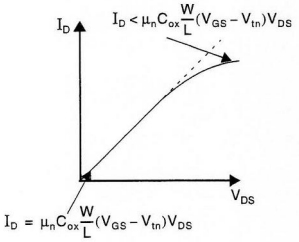


Fig. 1.12 For V_{DS} not close to zero, the I_D versus V_{DS} relationship is no longer linear.

7. $V_G - V_{CH}(x)$ is the gate-to-channel voltage drop at distance x from the source end, with V_G being the same everywhere in the gate, since the gate material is highly conductive.

As the drain voltage is increased, at some point the gate-to-channel voltage at the drain end will decrease to the threshold value V_{tn} — the minimum gate-to-channel voltage needed for n carriers in the channel to exist. Thus, at the drain end, the channel becomes *pinched off*, as shown in Fig. 1.13. This pinch-off occurs at $V_{GD} = V_{tn}$, since the channel voltage at the drain end is simply equal to V_D. Thus, pinch-off occurs for

$$V_{DG} > -V_{tn} \tag{1.62}$$

Denoting $V_{DS\text{-}sat}$ as the drain-source voltage when the channel becomes pinched off, we can substitute $V_{DG} = V_{DS} - V_{GS}$ into (1.62) and find an equivalent pinch-off expression

$$V_{DS} > V_{DS\text{-}sat} \tag{1.63}$$

where $V_{DS\text{-}sat}$ is given[8] by

$$V_{DS\text{-}sat} = V_{GS} - V_{tn} = V_{eff} \tag{1.64}$$

The electron carriers travelling through the pinched-off drain region are velocity saturated, similar to a gas under pressure travelling through a very small tube. If the drain-gate voltage rises above this critical pinch-off voltage of $-V_{tn}$, the charge concentration in the channel remains constant (to a first-order approximation) and the drain current no longer increases with increasing V_{DS}. The result is the current-voltage relationship shown in Fig. 1.14 for a given gate-source voltage. In the region of operation where $V_{DS} > V_{DS\text{-}sat}$, the drain current is independent of V_{DS} and is called the *active region*.[9] The region where I_D changes with V_{DS} is called the *triode region*. When MOS transistors are used in analog amplifiers, they almost always are biased in the active region. When they are used in digital logic gates, they often operate in both regions.

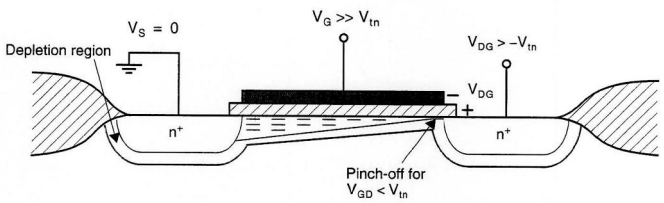


Fig. 1.13 When V_{DS} is increased so that $V_{GD} < V_{tn}$, the channel becomes pinched off at the drain end.

8. Because of the body effect, the threshold voltage at the drain end of the transistor is increased, resulting in the true value of $V_{DS\text{-}sat}$ being slightly lower than V_{eff}.
9. Historically, the active region was called the saturation region, but this led to confusion because in the case of bipolar transistors, the saturation region occurs for small V_{CE}, whereas for MOS transistors it occurs for large V_{DS}. The renaming of the saturation region to the active region is becoming widely accepted.

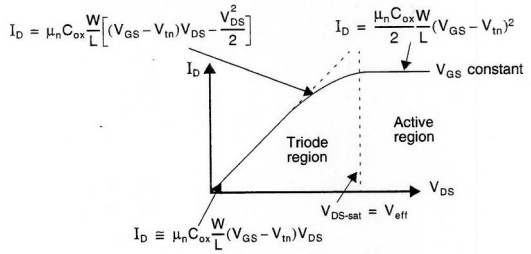


Fig. 1.14 The I_D versus V_{DS} curve for an ideal MOS transistor. For $V_{DS} > V_{DS\text{-}sat}$, I_D is approximately constant.

Before proceeding, it is worth discussing the terms *weak, moderate,* and *strong inversion.* As just discussed, a gate-source voltage greater than V_{tn} results in an inverted channel, and drain-source current can flow. However, as the gate-source voltage is increased, the channel does not become inverted (i.e., n-region) suddenly, but rather gradually. Thus, it is useful to define three regions of channel inversion with respect to the gate-source voltage. In most circuit applications, noncutoff MOSFET transistors are operated in strong inversion, with $V_{eff} > 100$ mV (many prudent circuit designers use a minimum value of 200 mV). As the name suggests, strong inversion occurs when the channel is strongly inverted. It should be noted that all the equation models in this section assume strong inversion operation. Weak inversion occurs when V_{GS} is approximately 100 mV or more below V_{tn} and is discussed as subthreshold operation in Section 1.3. Finally, moderate inversion is the region between weak and strong inversion.

Large-Signal Modelling

The *triode region equation* for a MOS transistor relates the drain current to the gate-source and drain-source voltages. It can be shown (see Appendix) that this relationship is given by

$$I_D = \mu_n C_{ox}\left(\frac{W}{L}\right)\left[(V_{GS}-V_{tn})V_{DS}-\frac{V_{DS}^2}{2}\right] \tag{1.65}$$

As V_{DS} increases, I_D increases until the drain end of the channel becomes pinched off, and then I_D no longer increases. This pinch-off occurs for $V_{DG} = -V_{tn}$, or approximately,

$$V_{DS} = V_{GS} - V_{tn} = V_{eff} \tag{1.66}$$

Right at the edge of pinch-off, the drain current resulting from (1.65) and the drain current in the active region (which, to a first-order approximation, is constant with

respect to V_{DS}) must have the same value. Therefore, the *active region equation* can be found by substituting (1.66) into (1.65), resulting in

$$I_D = \frac{\mu_n C_{ox}}{2}\left(\frac{W}{L}\right)(V_{GS}-V_{tn})^2 \tag{1.67}$$

For $V_{DS} > V_{eff}$, the current stays constant at the value given by (1.67), ignoring second-order effects such as the finite output impedance of the transistor. This equation is perhaps the most important one that describes the large-signal operation of a MOS transistor. It should be noted here that (1.67) represents a squared current-voltage relationship for a MOS transistor in the active region. In the case of a BJT transistor, an exponential current-voltage relationship exists in the active region.

As just mentioned, (1.67) implies that the drain current, I_D, is independent of the drain-source voltage. This independence is only true to a first-order approximation. The major source of error is due to the channel length shrinking as V_{DS} increases. To see this effect, consider Fig. 1.15, which shows a cross section of a transistor in the active region. A pinched-off region with very little charge exists between the drain and the channel. The voltage at the end of the channel closest to the drain is fixed at $V_{GS} - V_{tn} = V_{eff}$. The voltage difference between the drain and the near end of the channel lies across a short depletion region often called the *pinch-off region.* As V_{DS} becomes larger than V_{eff}, this depletion region surrounding the drain junction increases its width in a square-root relationship with respect to V_{DS}. This increase in the width of the depletion region surrounding the drain junction decreases the effective channel length. In turn, this decrease in effective channel length increases the drain current, resulting in what is commonly referred to as *channel-length modulation.*

To derive an equation to account for channel-length modulation, we first make use of (1.11) and denote the width of the depletion region by x_d, resulting in

$$\begin{aligned} x_d &\cong k_{ds}\sqrt{V_{D\text{-}ch} + \Phi_0} \\ &= k_{ds}\sqrt{V_{DG} + V_{tn} + \Phi_0} \end{aligned} \tag{1.68}$$

where

$$k_{ds} = \sqrt{\frac{2K_s\varepsilon_0}{qN_A}} \tag{1.69}$$

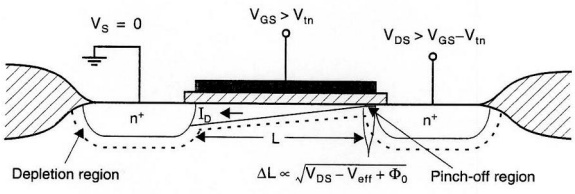


Fig. 1.15 Channel length shortening for $V_{DS} > V_{eff}$.

and has units of $m/\sqrt{V}$. Note that N_A is used here since the n-type drain region is more heavily doped than the p-type channel (i.e., $N_D >> N_A$). By writing a Taylor approximation for I_D around its operating value of $V_{DS} = V_{GS} - V_{tn} = V_{eff}$, we find I_D to be given by

$$I_D = I_{D\text{-sat}} + \left(\frac{\partial I_D}{\partial L}\right)\left(\frac{\partial L}{\partial V_{DS}}\right)\Delta V_{DS} \cong I_{D\text{-sat}}\left(1 + \frac{k_{ds}(V_{DS} - V_{eff})}{2L\sqrt{V_{DG} + V_{tn} + \Phi_0}}\right) \tag{1.70}$$

where $I_{D\text{-sat}}$ is the drain current when $V_{DS} = V_{eff}$, or equivalently, the drain current when the channel-length modulation is ignored. Note that in deriving the final equation of (1.70), we have used the relationship $\partial L/\partial V_{DS} = -\partial x_d/\partial V_{DS}$. Usually, (1.70) is written as

$$I_D = \frac{\mu_n C_{ox}}{2}\left(\frac{W}{L}\right)(V_{GS} - V_{tn})^2[1 + \lambda(V_{DS} - V_{eff})] \tag{1.71}$$

where λ is the output impedance constant (in units of V^{-1}) given by

$$\lambda = \frac{k_{ds}}{2L\sqrt{V_{DG} + V_{tn} + \Phi_0}} = \frac{k_{ds}}{2L\sqrt{V_{DS} - V_{eff} + \Phi_0}} \tag{1.72}$$

Equation (1.71) is accurate until V_{DS} is large enough to cause second-order effects, often called *short-channel effects*. For example, (1.71) assumes that current flow down the channel is not *velocity-saturated* (i.e., increasing the electric field no longer increases the carrier speed). Velocity saturation commonly occurs in new technologies that have very short channel lengths and therefore large electric fields. If V_{DS} becomes large enough so short-channel effects occur, I_D increases more than is predicted by (1.71). Of course, for quite large values of V_{DS}, the transistor will eventually break down.

A plot of I_D versus V_{DS} for different values of V_{GS} is shown in Fig. 1.16. Note that in the active region, the small (but nonzero) slope indicates the small dependence of I_D on V_{DS}.

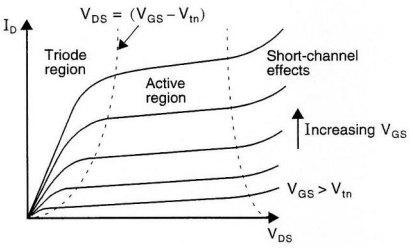


Fig. 1.16 I_D versus V_{DS} for different values of V_{GS}.

EXAMPLE 1.8

Find I_D for an n-channel transistor that has doping concentrations of $N_D = 10^{25}$, $N_A = 10^{22}$, $\mu_n C_{ox} = 92\ \mu A/V^2$, $W/L = 20\ \mu m/2\ \mu m$, $V_{GS} = 1.2$ V, $V_{tn} = 0.8$ V, and $V_{DS} = V_{eff}$. Assuming λ remains constant, estimate the new value of I_D if V_{DS} is increased by 0.5 V.

Solution

From (1.69), we have

$$k_{ds} = \sqrt{\frac{2 \times 11.8 \times 8.854 \times 10^{-12}}{1.6 \times 10^{-19} \times 10^{22}}} = 362 \times 10^{-9}\ m/\sqrt{V}$$

which is used in (1.72) to find λ as

$$\lambda = \frac{362 \times 10^{-9}}{2 \times 2 \times 10^{-6} \times \sqrt{0.9}} = 95.3 \times 10^{-3}\ V^{-1}$$

Using (1.71), we find for $V_{DS} = V_{eff} = 0.4$ V,

$$I_{D1} = \left(\frac{92 \times 10^{-6}}{2}\right)\left(\frac{20}{2}\right)(0.4)^2(1) = 73.6\ \mu A$$

In the case where $V_{DS} = V_{eff} + 0.5\ V = 0.9$ V, we have

$$I_{D2} = 73.6\ \mu A \times (1 + \lambda \times 0.5) = 77.1\ \mu A$$

Note that this example shows almost a 5 percent increase in drain current for a 0.5 V increase in drain-source voltage.

Body Effect

The large-signal equations in the preceding section were based on the assumption that the source voltage was the same as the substrate (i.e., bulk) voltage. However, often the source and substrate can be at different voltage potentials. In these situations, a second-order effect exists that is modelled as an increase in the threshold voltage, V_{tn}, as the source-to-substrate reverse-bias voltage increases. This effect, typically called the *body effect*, is more important for transistors in a well of a CMOS process where the substrate doping is higher. It should be noted that the body effect is often important in analog circuit designs and should not be ignored without consideration.

To account for the body effect, it can be shown (see Appendix at the end of this chapter) that the threshold voltage of an n-channel transistor is now given by

$$V_{tn} = V_{tn0} + \gamma(\sqrt{V_{SB} + |2\phi_F|} - \sqrt{|2\phi_F|}) \tag{1.73}$$

where V_{tn0} is the threshold voltage with zero V_{SB} (i.e., source-to-substrate voltage),

and

$$\gamma = \frac{\sqrt{2qN_A K_s \varepsilon_0}}{C_{ox}} \tag{1.74}$$

The factor γ is often called the *body-effect constant* and has units of $\sqrt{V}$. Notice that γ is proportional to $\sqrt{N_A}$,[10] so the body effect is larger for transistors in a well where typically the doping is higher than the substrate of the microcircuit.

p-Channel Transistors

All of the preceding equations have been presented for n-channel enhancement transistors. *In the case of p-channel transistors, these equations can also be used if a negative sign is placed in front of every voltage variable.* Thus, V_{GS} becomes V_{SG}, V_{DS} becomes V_{SD}, V_{tn} becomes $-V_{tp}$, and so on. The condition required for conduction is now $V_{SG} > V_{tp}$, where V_{tp} is now a negative quantity for an enhancement p-channel transistor.[11] The requirement on the source-drain voltage for a p-channel transistor to be in the active region is $V_{SD} > V_{SG} + V_{tp}$. The equations for I_D, in both regions, remain unchanged, because all voltage variables are squared, resulting in positive hole current flow from the source to the drain in p-channel transistors. For n-channel depletion transistors, the only difference is that $V_{td} < 0$ V. A typical value might be $V_{td} = -2$ V.

Small-Signal Modelling in the Active Region

The most commonly used small-signal model for a MOS transistor operating in the active region is shown in Fig. 1.17. We first consider the dc parameters in which all the capacitors are ignored (i.e., replaced by open circuits). This leads to the low-frequency, small-signal model shown in Fig. 1.18. The voltage-controlled current source, $g_m v_{gs}$, is the most important component of the model, with the transistor transconductance g_m defined as

$$g_m = \frac{\partial I_D}{\partial V_{GS}} \tag{1.75}$$

In the active region, we use (1.67), which is repeated here for convenience,

$$I_D = \frac{\mu_n C_{ox}}{2}\left(\frac{W}{L}\right)(V_{GS} - V_{tn})^2 \tag{1.76}$$

10. For an n-channel transistor. For a p-channel transistor, γ is proportional to N_D.
11. It is possible to realize depletion p-channel transistors, but these are of little value and seldom worth the extra processing involved. Depletion n-channel transistors are also seldom encountered in CMOS microcircuits, although they might be worth the extra processing involved in some applications, especially if they were in a well.

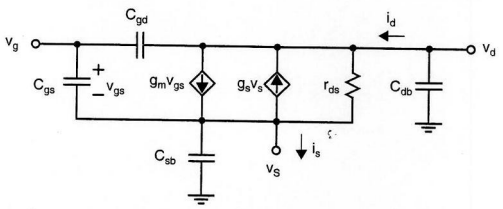


Fig. 1.17 The small-signal model for a MOS transistor in the active region.

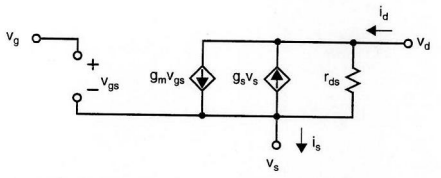


Fig. 1.18 The low-frequency, small-signal model for an active MOS transistor.

and we apply the derivative shown in (1.75) to obtain

$$g_m = \frac{\partial I_D}{\partial V_{GS}} = \mu_n C_{ox}\frac{W}{L}(V_{GS} - V_{tn}) = \mu_n C_{ox}\frac{W}{L}V_{eff} \tag{1.77}$$

or equivalently,

$$g_m = \mu_n C_{ox}\frac{W}{L}V_{eff} \tag{1.78}$$

where the effective gate-source voltage, V_{eff}, is defined as $V_{eff} \equiv V_{GS} - V_{tn}$. Thus, we see that the transconductance of a MOS transistor is directly proportional to V_{eff}.

Sometimes it is desirable to express g_m in terms of I_D rather than V_{GS}. From (1.76), we have

$$V_{GS} = V_{tn} + \sqrt{\frac{2I_D}{\mu_n C_{ox}(W/L)}} \tag{1.79}$$

The second term in (1.79) is the effective gate-source voltage, V_{eff}, where

$$V_{eff} = V_{GS} - V_{tn} = \sqrt{\frac{2I_D}{\mu_n C_{ox}(W/L)}} \tag{1.80}$$

Substituting (1.80) in (1.78) results in an alternate expression for g_m.

$$g_m = \sqrt{2\mu_n C_{ox}\frac{W}{L}I_D} \tag{1.81}$$

Thus, the transistor transconductance is proportional to $\sqrt{I_D}$ for a MOS transistor, whereas it is proportional to I_C for a BJT.

A third expression for g_m is found by rearranging (1.81) and then using (1.80) to obtain

$$g_m = \frac{2I_D}{V_{eff}} \tag{1.82}$$

Note that this expression is independent of $\mu_n C_{ox}$ and W/L, and it relates the transconductance to the ratio of drain current to effective gate-source voltage. This simple relationship can be quite useful during an initial circuit design.

The second voltage-controlled current-source in Fig. 1.18, shown as $g_s v_s$, models the body effect on the small-signal drain current, i_d. When the source is connected to small-signal ground, or when its voltage does not change appreciably, then this current source can be ignored. When the body effect cannot be ignored, we have

$$g_s = \frac{\partial I_D}{\partial V_{SB}} = \frac{\partial I_D}{\partial V_{tn}}\frac{\partial V_{tn}}{\partial V_{SB}} \tag{1.83}$$

From (1.76) we have

$$\frac{\partial I_D}{\partial V_{tn}} = -\mu_n C_{ox}\left(\frac{W}{L}\right)(V_{GS} - V_{tn}) = -g_m \tag{1.84}$$

Using (1.73), which gives V_{tn} as

$$V_{tn} = V_{tn0} + \gamma(\sqrt{V_{SB} + |2\phi_F|} - \sqrt{|2\phi_F|}) \tag{1.85}$$

we have

$$\frac{\partial V_{tn}}{\partial V_{SB}} = \frac{\gamma}{2\sqrt{V_{SB} + |2\phi_F|}} \tag{1.86}$$

The negative sign of (1.84) is eliminated by subtracting the current $g_s v_s$ from the major component of the drain current, $g_m v_{gs}$, as shown in Fig. 1.18. Thus, using (1.84) and (1.86), we have

$$g_s = \frac{\gamma g_m}{2\sqrt{V_{SB} + |2\phi_F|}} \tag{1.87}$$

Note that although g_s is nonzero for $V_{SB} = 0$, if the source is connected to the bulk, ΔV_{SB} is zero, and so the effect of g_s does not need to be taken into account. However, if the source happens to be biased at the same potential as the bulk but is not

directly connected to it, then the effect of g_s should be taken into account since ΔV_{SB} is not necessarily zero.

The resistor, r_{ds}, shown in Fig. 1.18, accounts for the finite output impedance (i.e., it models the channel-length modulation and its effect on the drain current due to changes in V_{DS}). Using (1.71), repeated here for convenience,

$$I_D = \frac{\mu_n C_{ox}}{2}\left(\frac{W}{L}\right)(V_{GS} - V_{tn})^2[1 + \lambda(V_{DS} - V_{eff})] \tag{1.88}$$

we have

$$\frac{1}{r_{ds}} = g_{ds} = \frac{\partial I_D}{\partial V_{DS}} = \lambda\left(\frac{\mu_n C_{ox}}{2}\right)\left(\frac{W}{L}\right)(V_{GS} - V_{tn})^2 = \lambda I_{D\text{-}sat} \cong \lambda I_D \tag{1.89}$$

where the approximation assumes λ is small, such that we can approximate the drain bias current as being the same as $I_{D\text{-}sat}$. Thus,

$$r_{ds} \cong \frac{1}{\lambda I_D} \tag{1.90}$$

where

$$\lambda = \frac{k_{ds}}{2L\sqrt{V_{DS} + (-V_{eff}) + \Phi_0}} \tag{1.91}$$

and

$$k_{ds} = \sqrt{\frac{2K_s\varepsilon_0}{qN_A}} \tag{1.92}$$

It should be noted here that (1.90) is often empirically adjusted to take into account second-order effects.

EXAMPLE 1.9

Derive the low-frequency model parameters for an n-channel transistor that has doping concentrations of $N_D = 10^{25}$, $N_A = 10^{22}$, $\mu_n C_{ox} = 92\ \mu A/V^2$, $W/L = 20\ \mu m/2\ \mu m$, $V_{GS} = 1.2$ V, $V_{tn} = 0.8$ V, and $V_{DS} = V_{eff}$. Assume $\gamma = 0.5\ \sqrt{V}$ and $V_{SB} = 0.5$ V. What is the new value of r_{ds} if the drain-source voltage is increased by 0.5 V?

Solution

Since these parameters are the same as in Example 1.8, we have

$$g_m = \frac{2I_D}{V_{eff}} = \frac{2 \times 73.6\ \mu A}{0.4\ V} = 0.368\ mA/V$$

and from (1.87), we have

$$g_s = \frac{0.5 \times 0.368 \times 10^{-3}}{2\sqrt{0.5 + 1.8}} = 0.061 \text{ mA/V}$$

Note that this source-bulk transconductance value is about 1/6 that of the gate-source transconductance.

For r_{ds}, we use (1.90) to find

$$r_{ds} = \frac{1}{95.3 \times 10^{-3} \times 73.6 \times 10^{-6}} = 143 \text{ k}\Omega$$

At this point, it is interesting to calculate the gain $g_m r_{ds} = 52.6$, which is the largest voltage gain this single transistor can achieve for these operating bias conditions. As we will see, this gain of 52.6 is much smaller than the corresponding single-transistor gain in a bipolar transistor.

Recalling that $V_{eff} = 0.4$ V, if V_{DS} is increased to 0.9 V, the new value for λ is

$$\lambda = \frac{362 \times 10^{-9}}{2(2 \times 10^{-6})\sqrt{1.4}} = 76.4 \times 10^{-3} \text{ V}^{-1}$$

resulting in a new value of r_{ds} given by

$$r_{ds} = \frac{1}{\lambda I_{D2}} = \frac{1}{76.4 \times 10^{-3} \times 77.1 \ \mu\text{A}} = 170 \text{ k}\Omega$$

An alternate low-frequency model, known as a T model, is shown in Fig. 1.19. This *T model* can often result in simpler equations and is most often used by experienced designers for a quick analysis. At first glance, it might appear that this model allows for nonzero gate current, but a quick check confirms that the drain current must always equal the source current, and, therefore, the gate current must always be zero. For this reason, when using the T model, one assumes from the beginning that the gate current is zero.

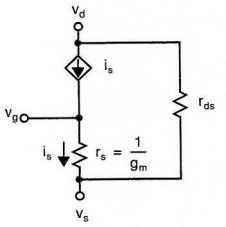


Fig. 1.19 The small-signal, low-frequency T model for an active MOS transistor (the body effect is not modelled).

EXAMPLE 1.10

Find the T model parameter, r_s, for the transistor in Example 1.9.

Solution

The value of r_s is simply the inverse of g_m, resulting in

$$r_s = \frac{1}{g_m} = \frac{1}{0.368 \times 10^{-3}} = 2.72 \text{ k}\Omega$$

The value of r_{ds} remains the same, either 143 kΩ or 170 kΩ, depending on the drain-source voltage.

Most of the capacitors in the small-signal model are related to the physical transistor. Shown in Fig. 1.20 is a cross section of a MOS transistor, where the parasitic capacitances are shown at the appropriate locations. The largest capacitor in Fig. 1.20 is C_{gs}. This capacitance is primarily due to the change in channel charge as a result of a change in V_{GS}. It can be shown [Tsividis, 1987] that C_{gs} is approximately given by

$$C_{gs} \cong \frac{2}{3} WLC_{ox} \tag{1.93}$$

When accuracy is important, an additional term should be added to (1.93) to take into account the overlap between the gate and source junction, which should include the *fringing capacitance* (fringing capacitance is due to boundary effects). This additional component is given by

$$C_{ov} = WL_{ov}C_{ox} \tag{1.94}$$

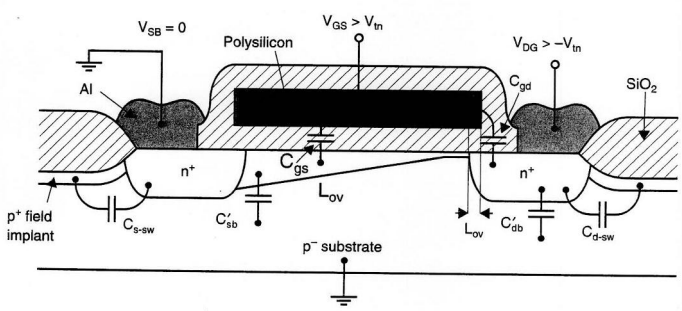


Fig. 1.20 A cross section of an n-channel MOS transistor showing the small-signal capacitances.

where L_{ov} is the overlap distance and is usually empirically derived. Thus,

$$C_{gs} = WC_{ox}\left(\frac{2}{3}L + L_{ov}\right) \tag{1.95}$$

when higher accuracy is needed.

The next largest capacitor in Fig. 1.20 is C'_{sb}, the capacitor between the source and the substrate. This capacitor is due to the depletion capacitance of the reverse-biased source junction, and it includes the channel-to-bulk capacitance (assuming the transistor is on). Its size is given by

$$C'_{sb} = (A_s + A_{ch})C_{js} \tag{1.96}$$

where A_s is the area of the source junction, A_{ch} is the area of the channel (i.e., WL) and C_{js} is the depletion capacitance of the source junction, given by

$$C_{js} = \frac{C_{j0}}{\sqrt{1 + \frac{V_{SB}}{\Phi_0}}} \tag{1.97}$$

Note that the total area of the effective source includes the original area of the junction (when no channel is present) plus the effective area of the channel.

The depletion capacitance of the drain is smaller because it does not include the channel area. Here, we have

$$C'_{db} = A_d C_{jd} \tag{1.98}$$

where

$$C_{jd} = \frac{C_{j0}}{\sqrt{1 + \frac{V_{DB}}{\Phi_0}}} \tag{1.99}$$

and A_d is the area of the drain junction.

The capacitance C_{gd}, sometimes called the *Miller-capacitor*, is important when the transistor is being used in circuits with large voltage gain. C_{gd} is primarily due to the overlap between the gate and the drain and fringing capacitance. Its value is given by

$$C_{gd} = C_{ox}WL_{ov} \tag{1.100}$$

where, once again, L_{ov} is usually empirically derived.

Two other capacitors are often important in integrated circuits. These are the source and drain *sidewall capacitances*, $C_{s\text{-}sw}$ and $C_{d\text{-}sw}$. These capacitances can be large because of some highly doped p^+ regions under the thick field oxide called *field implants*. The major reason these regions exist is to ensure there is no leakage current between transistors. Because they are highly doped and they lie beside the highly doped source and drain junctions, the sidewall capacitances can result in large additional capacitances that must be taken into account in determining C_{sb} and C_{db}. The sidewall capacitances are especially important in modern technologies as dimensions

shrink. For the source, the sidewall capacitance is given by

$$C_{s\text{-}sw} = P_s C_{j\text{-}sw} \tag{1.101}$$

where P_s is the length of the perimeter of the source junction, excluding the side adjacent to the channel, and

$$C_{j\text{-}sw} = \frac{C_{j\text{-}sw0}}{\sqrt{1 + \frac{V_{SB}}{\Phi_0}}} \tag{1.102}$$

It should be noted that $C_{j\text{-}sw0}$, the sidewall capacitance per unit length at 0-V bias voltage, can be quite large because the field implants are heavily doped.

The situation is similar for the drain sidewall capacitance, $C_{d\text{-}sw}$,

$$C_{d\text{-}sw} = P_d C_{j\text{-}sw} \tag{1.103}$$

where P_d is the drain perimeter excluding the portion adjacent to the gate.

Finally, the source-bulk capacitance, C_{sb}, is given by

$$C_{sb} = C'_{sb} + C_{s\text{-}sw} \tag{1.104}$$

with the drain-bulk capacitance, C_{db}, given by

$$C_{db} = C'_{db} + C_{d\text{-}sw} \tag{1.105}$$

EXAMPLE 1.11

An n-channel transistor is modelled as having the following capacitance parameters: $C_j = 2.4 \times 10^{-4}$ pF/(μm)^2, $C_{j\text{-}sw} = 2.0 \times 10^{-4}$ pF/μm, $C_{ox} = 1.9 \times 10^{-3}$ pF/(μm)^2, $C_{gs\text{-}ov} = C_{gd\text{-}ov} = 2.0 \times 10^{-4}$ pF/μm. Find the capacitances C_{gs}, C_{gd}, C_{db}, and C_{sb} for a transistor having W = 100 μm and L = 2 μm. Assume the source and drain junctions extend 4 μm beyond the gate, so that the source and drain areas are $A_s = A_d = 400$ (μm)^2 and the perimeter of each is $P_s = P_d = 108$ μm.

Solution

We calculate the various capacitances as follows:

$$C_{gs} = \left(\frac{2}{3}\right)WLC_{ox} + C_{gs\text{-}ov} \times W = 0.27 \text{ pF}$$

$$C_{gd} = C_{gd\text{-}ov} \times W = 0.02 \text{ pF}$$

$$C_{sb} = C_j(A_s + WL) + (C_{j\text{-}sw} \times P_s) = 0.17 \text{ pF}$$

$$C_{db} = (C_j \times A_d) + (C_{j\text{-}sw} \times P_d) = 0.12 \text{ pF}$$

Note that the source-bulk and drain-bulk capacitances are significant compared to the gate-source capacitance. Thus, for high-speed circuits, it is important to

keep the areas and perimeters of drain and source junctions as small as possible (possibly by sharing junctions between transistors, as seen in the next chapter).

Small-Signal Modelling in the Triode and Cutoff Regions

The low-frequency, small-signal model of a MOS transistor in the triode region (which is sometimes referred to as the linear region) is a resistor. Using (1.65), the large-signal equation for I_D in the triode region,

$$I_D = \mu_n C_{ox}\left(\frac{W}{L}\right)\left[(V_{GS} - V_{tn})V_{DS} - \frac{V_{DS}^2}{2}\right] \tag{1.106}$$

results in

$$\frac{1}{r_{ds}} = g_{ds} = \frac{dI_D}{dV_{DS}} = \mu_n C_{ox}\left(\frac{W}{L}\right)(V_{GS} - V_{tn} - V_{DS}) \tag{1.107}$$

where r_{ds} is the small-signal drain-source resistance (and g_{ds} is the conductance). For the common case of V_{DS} near zero, we have

$$g_{ds} = \frac{1}{r_{ds}} = \mu_n C_{ox}\left(\frac{W}{L}\right)(V_{GS} - V_{tn}) = \mu_n C_{ox}\left(\frac{W}{L}\right)V_{eff} \tag{1.108}$$

which is similar to the I_D-versus-V_{DS} relationship given earlier in (1.60).

EXAMPLE 1.12

For the transistor of Example 1.9, find the triode model parameters when V_{DS} is near zero.

Solution

From (1.108), we have

$$g_{ds} = 92 \times 10^{-6} \times \left(\frac{20}{2}\right) \times 0.4 = 0.368 \text{ mA/V}$$

Note that this conductance value is the same as the transconductance of the transistor, g_m, in the active region. The resistance, r_{ds}, is simply $1/g_{ds}$, resulting in $r_{ds} = 2.72$ kΩ.

The accurate modelling of the high-frequency operation of a transistor in the triode region is nontrivial (even with the use of a computer simulation). A moderately accurate model is shown in Fig. 1.21, where the gate-to-channel capacitance

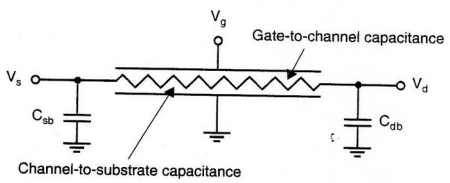


Fig. 1.21 A distributed RC model for a transistor in the active region.

and the channel-to-substrate capacitance are modelled as distributed elements. However, the I-V relationships of the distributed RC elements are highly nonlinear because the junction capacitances of the source and drain are nonlinear depletion capacitances, as is the channel-to-substrate capacitance. Also, if V_{DS} is not small, then the channel resistance per unit length should increase as one moves closer to the drain. This model is much too complicated for use in hand analysis.

A simplified model often used for small V_{DS} is shown in Fig. 1.22, where the resistance, r_{ds}, is given by (1.108). Here, the gate-to-channel capacitance has been evenly divided between the source and drain nodes,

$$C_{gs} = C_{gd} = \frac{A_{ch}C_{ox}}{2} = \frac{WLC_{ox}}{2} \tag{1.109}$$

Note that this equation ignores the gate-to-junction overlap capacitances, as given by (1.94), which should be taken into account when accuracy is very important. The channel-to-substrate capacitance has also been divided in half and shared between the source and drain junctions. Each of these capacitors should be added to the junction-to-substrate capacitance and the junction-sidewall capacitance at the appropriate

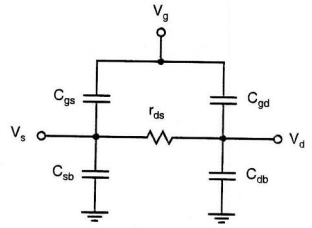


Fig. 1.22 A simplified triode-region model valid for small V_{DS}.

node. Thus, we have

$$C_{sb\text{-}0} = C_{j0}\left(A_s + \frac{A_{ch}}{2}\right) + C_{j\text{-}sw0}P_s \tag{1.110}$$

and

$$C_{db\text{-}0} = C_{j0}\left(A_d + \frac{A_{ch}}{2}\right) + C_{j\text{-}sw0}P_d \tag{1.111}$$

Also,

$$C_{sb} = \frac{C_{sb\text{-}0}}{\sqrt{1 + \frac{V_{sb}}{\Phi_0}}} \tag{1.112}$$

and

$$C_{db} = \frac{C_{db\text{-}0}}{\sqrt{1 + \frac{V_{sb}}{\Phi_0}}} \tag{1.113}$$

It might be noted that C_{sb} is often comparable in size to C_{gs} due to its larger area and the sidewall capacitance.

When the transistor turns off, the model changes considerably. A reasonable model is shown in Fig. 1.23. Perhaps the biggest difference is that r_{ds} is now infinite. Another major difference is that C_{gs} and C_{gd} are now much smaller. Since the channel has disappeared, these capacitors are now due to only overlap and fringing capacitance. Thus, we have

$$C_{gs} = C_{gd} = WL_{ov}C_{ox} \tag{1.114}$$

However, the reduction of C_{gs} and C_{gd} does not mean that the total gate capacitance is necessarily smaller. We now have a "new" capacitor, C_{gb}, which is the gate-

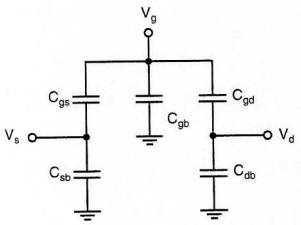


Fig. 1.23 A small-signal model for a MOSFET that is turned off.

to-substrate capacitance. This capacitor is highly nonlinear and dependent on the gate voltage. If the gate voltage has been very negative for some time and the gate is accumulated, then we have

$$C_{gb} = A_{ch}C_{ox} = WLC_{ox} \tag{1.115}$$

If the gate-to-source voltage is around 0 V, then C_{gb} is equal to C_{ox} in series with the channel-to-bulk depletion capacitance and is considerably smaller, especially when the substrate is lightly doped. Another case where C_{gb} is small is just after a transistor has been turned off, before the channel has had time to accumulate. Because of the complicated nature of correctly modelling C_{gb} when the transistor is turned off, equation (1.115) is usually used for hand analysis as a worst-case estimate.

The capacitors C_{sb} and C_{db} are also smaller when the channel is not present. We now have

$$C_{sb\text{-}0} = A_sC_{j0} \tag{1.116}$$

and

$$C_{db\text{-}0} = A_dC_{j0} \tag{1.117}$$

1.3 ADVANCED MOS MODELLING

In this section, we look at three advanced modelling concepts that a microcircuit designer is likely to encounter—short-channel effects, subthreshold operation, and leakage currents.

Short-Channel Effects

A number of short-channel effects degrade the operation of MOS transistors as device dimensions are scaled down. These effects include mobility degradation, reduced output impedance, and hot-carrier effects (such as oxide trapping and substrate currents). These short-channel effects will be briefly described here. For more detailed modelling of short-channel effects, see [Wolf, 1995].

Transistors that have short channel lengths and large electric fields experience a degradation in the effective mobility of their carriers due to several factors. One of these factors is the large lateral electric field (which has a vector in a direction perpendicular from the gate into the silicon) caused by large gate voltages and short channel lengths. This large lateral field causes the effective channel depth to change and also causes more electron collisions, thereby lowering the effective mobility. Another factor causing this degradation is that, due to large electric fields, carrier velocity begins to saturate. A first-order approximation that models this carrier-velocity saturation for electrons is given by

$$v_d \cong \frac{\mu_n E}{1 + E/E_c} \tag{1.118}$$

where E is the electric field and E_c is the critical electrical field, which might be on the order of 1.5×10^6 V/m. Using this equation in the derivation of the I_D-V_{eff}

characteristics of a MOS transistor, it can be shown [Gray, 1993] that the drain current is now given by

$$I_D = \frac{\mu_n C_{ox}}{2[1 + \theta V_{eff}]} \frac{W}{L} V_{eff}^2 \tag{1.119}$$

where $\theta = 1/(LE_c)$ and, for a 0.8-μm technology, might have a typical value of 0.6 V^{-1}. It can be shown that this mobility degradation is equivalent to a finite series source resistance given by

$$R_{SX} = \frac{1}{E_C} \frac{1}{\mu_n C_{ox}} \frac{1}{W} \tag{1.120}$$

For $\mu_n C_{ox} = 90\ \mu A/V^2$, this resistance might be on the order of 6 kΩ per μm of width (again, for a 0.8-μm-long transistor). This equivalent series source resistance is typically larger than the physical source resistance. This saturation causes the square-law characteristic of the current-voltage relationship to be inaccurate, and the true relationship will be somewhere between linear and square. In many voltage-to-current conversion circuits that rely on the square-law characteristic, this inaccuracy can be a major source of error. Taking channel lengths larger than the minimum allowed helps to minimize this degradation.

Transistors with short channel lengths also experience a reduced output impedance because depletion region variations at the drain end (which affect the effective channel length) have an increased proportional effect on the drain current. In addition, a phenomenon known as drain-induced barrier lowering (DIBL) effectively lowers V_t as V_{DS} is increased, thereby further lowering the output impedance of a short-channel device. This lower output impedance is the main reason that cascode current mirrors are becoming increasingly popular.

Another important short-channel effect is due to *hot carriers*. These high-velocity carriers can cause harmful effects, such as the generation of electron-hole pairs by impact ionization and avalanching. These extra electron-hole pairs can cause currents to flow from the drain to the substrate, as shown in Fig. 1.24. This effect can be modelled by a finite drain-to-ground impedance. As a result, this effect is one of the major limitations on achieving very high output impedances of cascode current sources. In addition, this current flow can cause voltage drops across the substrate and possibly cause latch-up, as the next section describes.

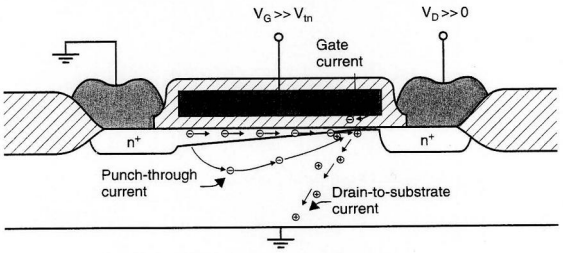


Fig. 1.24 Drain-to-substrate current caused by electron-hole pairs generated by impact ionization at drain end of channel.

Another hot-carrier effect occurs when electrons gain energies high enough so they can tunnel into and possibly through the thin gate oxide. Thus, this effect can cause dc gate currents. However, often more harmful is the fact that any charge trapped in the oxide will cause a shift in transistor threshold voltage. As a result, hot carriers are one of the major factors limiting the long-term reliability of MOS transistors.

A third hot-carrier effect occurs when electrons with enough energy *punch through* from the source to the drain. As a result, these high-energy electrons are no longer limited by the drift equations governing normal conduction along the channel. This mechanism is somewhat similar to punch-through in a bipolar transistor, where the collector depletion region extends right through the base region to the emitter. In a MOS transistor, the channel length becomes effectively zero, resulting in unlimited current flow (except for the series source and drain impedances, as well as external circuitry). This effect is an additional cause of lower output impedance and possibly transistor breakdown.

It should be noted that all of the hot-carrier effects just described are more pronounced for n-channel transistors than for their p-channel counterparts because electrons have larger velocities than holes.

Finally, it should be noted that short-channel transistors have much larger subthreshold currents than long-channel devices.

Subthreshold Operation

The device equations presented for MOS transistors in the preceding sections are all based on the assumption that V_{eff} (i.e., $V_{GS} - V_t$) is greater than about 100 mV and the device is in strong inversion. When this is not the case, the accuracy of the square-law equations is poor. If $V_{eff} < -100$ mV, the transistor is in weak inversion and is said to be operating in the *subthreshold region*. In this region, the transistor is more accurately modelled by an exponential relationship between its control voltage and current, somewhat similar to a bipolar transistor. In the subthreshold region, the drain current is approximately given by the exponential relationship [Geiger, 1990]

$$I_D \cong I_{D0}\left(\frac{W}{L}\right) e^{(qV_{GS}/nkT)} \tag{1.121}$$

where

$$n = \frac{C_{ox} + C_{depl}}{C_{ox}} \cong 1.5 \tag{1.122}$$

and it has been assumed that $V_S = 0$ and $V_{DS} > 75$ mV. The constant I_{D0} might be around 20 nA.

Although the transistors have an exponential relationship in this region, the transconductances are still small because of the small bias currents, and the transistors are slow because of small currents for charging and discharging capacitors. In addition, matching between transistors suffers because it now strongly depends on transistor-threshold-voltage matching. Normally, transistors are not operated in the subthreshold region, except in very low-frequency and low-power applications.

Leakage Currents

An important second-order device limitation in some applications is the leakage current of the junctions. For example, this leakage can be important in estimating the maximum time a sample-and-hold circuit or a dynamic memory cell can be left in hold mode. The leakage current of a reverse-biased junction (not close to breakdown) is approximately given by

$$I_{lk} \equiv \frac{qA_j n_i}{2\tau_0} x_d \tag{1.123}$$

where A_j is the junction area, n_i is the intrinsic concentration of carriers in undoped silicon, τ_0 is the effective minority carrier lifetime, and x_d is the thickness of the depletion region. τ_0 is given by

$$\tau_0 \equiv \frac{1}{2}(\tau_n + \tau_p) \tag{1.124}$$

where τ_n and τ_p are the electron and hole lifetimes. Also, x_d is given by

$$x_d = \sqrt{\frac{2K_s\varepsilon_0}{qN_A}(\Phi_0 + V_r)} \tag{1.125}$$

and n_i is given by

$$n_i \equiv \sqrt{N_C N_V}\, e^{(-E_g)/(kT)} \tag{1.126}$$

where N_C and N_V are the densities of states in the conduction and valence bands and E_g is the difference in energy between the two bands.

Since the intrinsic concentration, n_i, is a strong function of temperature (it approximately doubles for every temperature increase of 11 °C for silicon), the leakage current is also a strong function of temperature. Roughly speaking, the leakage current also doubles for every 11 °C rise in temperature. Thus, the leakage current at higher temperatures is much larger than at room temperature. This leakage current imposes a maximum time on how long a dynamically charged signal can be maintained in a high impedance state.

1.4 BIPOLAR-JUNCTION TRANSISTORS

In the early electronic years, the majority of microcircuits were realized using bipolar-junction transistors (BJTs). However, in the late 1970s, microcircuits that used MOS

transistors began to dominate the industry, with BJT microcircuits remaining popular for high-speed applications. More recently, bipolar CMOS (BiCMOS) technologies, where both bipolar and MOS transistors are realized in the same microcircuit, have grown in popularity. BiCMOS technologies are particularly attractive for mixed analog-digital applications, and thus it is important for an analog designer to become familiar with bipolar devices.

Modern bipolar transistors can have unity-gain frequencies as high as 15 to 45 GHz or more, compared to unity-gain frequencies of only 1 to 4 GHz for MOS transistors that use a technology with similar lithography resolution. Unfortunately, in bipolar transistors, the *base* control terminal has a nonzero input current when the transistor is conducting current (from the collector to the emitter for an npn transistor; from the emitter to the collector for a pnp transistor). Fortunately, at low frequencies, the base current is much smaller than the collector-to-emitter current—it may be only 1/100 of the collector current for an npn transistor. For lateral pnp transistors, the base current may be as large as 1/20 of the emitter-to-collector current.

A typical cross section of an npn bipolar-junction transistor is shown in Fig. 1.25. Although this structure looks quite complicated, it corresponds approximately to the equivalent structure shown in Fig. 1.26. In a good BJT transistor, the width of the base

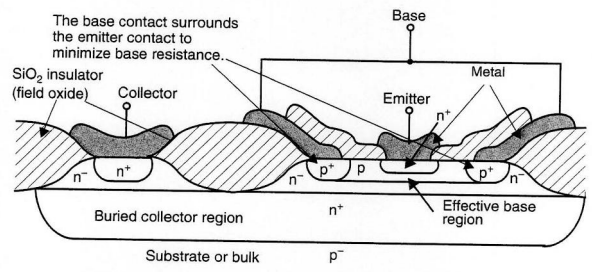


Fig. 1.25 A cross section of an npn bipolar-junction transistor.

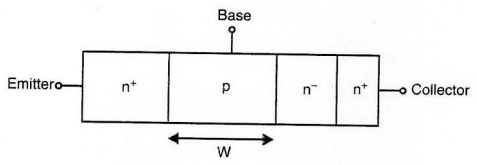


Fig. 1.26 A simplified structure of an npn transistor.

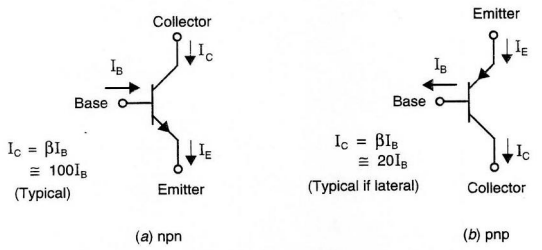


Fig. 1.27 The symbols representing (*a*) an npn bipolar-junction transistor and (*b*) a pnp bipolar-junction transistor.

region, W, is small (typically, less than 1 μm). Also, as we will see, the base must be more lightly doped than the emitter.

The circuit symbols used to represent npn and pnp transistors are shown in Fig. 1.27.

Basic Operation

To understand the operation of bipolar transistors, we consider here an npn transistor with the emitter connected to ground, as shown in Fig. 1.28. If the base voltage, V_B, is less than about 0.5 V, the transistor will be cut off, and no current will flow. We will see that when the base-emitter pn junction becomes forward biased, current will start to flow from the base to the emitter, but, partly because the base width is small, a much larger proportional current will flow from the collector to the emitter. Thus, the npn transistor can be considered a current amplifier at low frequencies. In other

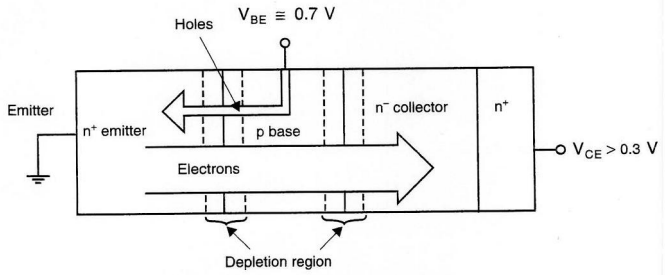


Fig. 1.28 Various components of the currents of an npn transistor.

words, if the transistor is not cut off and the collector-base junction is reverse biased, a small base current controls a much larger collector-emitter current.

A simplified overview of how an npn transistor operates follows: When the base-emitter junction becomes forward biased, it starts to conduct, similar to any forward-biased junction. The current consists of majority carriers from the base (in this case, holes) and majority carriers from the emitter (in this case, electrons) diffusing across the junction. Because the emitter is more heavily doped than the base, there are many more electrons injected from the emitter than there are holes injected from the base. Assuming the collector voltage is large enough so that the collector-base junction is reverse biased, no holes from the base will go to the collector. However, the electrons that travel from the emitter to the base, where they are now minority carriers, diffuse away from the base-emitter junction because of the minority-carrier concentration gradient in the base region. Any of these minority electrons that get close to the collector-base junction will immediately be "whisked" across the junction due to the large positive voltage on the collector, which attracts the negatively charged electrons. In a properly designed bipolar transistor, such as that shown in Fig. 1.25, the vertical base width, W, is small, and almost all of the electrons that diffuse from the emitter to the base reach the collector-base junction and are swept across the junction, thus contributing to current flow in the collector. The result is that the collector current very closely equals the electron current flowing from the emitter to the base. The much smaller base current very closely equals the current due to the holes that flow from the base to the emitter. The total emitter current is the sum of the electron collector current and the hole base current, but since the hole current is much smaller than the electron current, the emitter current is approximately equal to the collector current.

Since the collector current is approximately equal to the electron current flowing from the emitter to the base, and the amount of this electron current is determined by the base-emitter voltage, it can be shown (see Appendix at the end of this chapter) that the collector current is exponentially related to the base-emitter voltage by the relationship

$$I_C \cong I_{CS} e^{V_{BE}/V_T} \tag{1.127}$$

where I_{CS} is the *scale current*. This scale current is proportional to the area of the base-emitter junction. The base current, determined by the hole current flowing from the base to the emitter, is also exponentially related to the base-emitter voltage, resulting in the ratio of the collector current to the base current being a constant that, to a first-order approximation, is independent of voltage and current. This ratio, typically denoted as β, is defined to be

$$\beta \equiv \frac{I_C}{I_B} \tag{1.128}$$

where I_C and I_B are the collector and base currents. Typical values of β are between 50 and 200.

Note that (1.127) implies that the collector current is independent of the collector voltage. This independence ignores second-order effects such as the decrease in effective base width, W, due to the increase in the width of the collector-base depletion

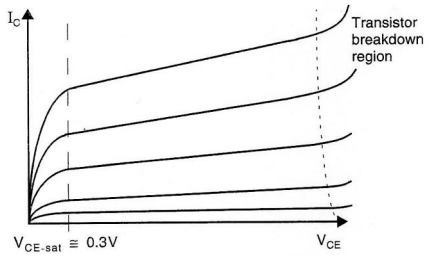


Fig. 1.29 Typical plot of I_C versus V_{CE} for a BJT.

region when the collector bias voltage is increased. To illustrate this point, a typical plot of the collector current, I_C, as a function of collector-to-emitter voltage, V_{CE}, for different values of I_B is shown in Fig. 1.29 for a practical transistor. The fact that the curves are not flat for $V_{CE} > V_{CE\text{-}sat}$ indicates the dependence of I_C on V_{CE}. Indeed, to a good approximation, the dependence is linear with a slope that intercepts the V_{CE} axis at $V_{CE} = -V_A$ for all values of I_B. The intercept voltage value, V_A, is called the *Early voltage* for bipolar transistors, with a typical value being from 50 V to 100 V. This dependency results in a finite output impedance (as in a MOS transistor) and can be modelled by modifying equation (1.127) [Sze, 1981] to be

$$I_C \cong I_{CS} e^{V_{BE}/V_T}\left(1 + \frac{V_{CE}}{V_A}\right) \tag{1.129}$$

Large-Signal Modelling

A conducting BJT that has a V_{CE} greater than $V_{CE\text{-}sat}$ (which is approximately 0.3 V) is said to be operating in the active region. Such a collector-emitter voltage is required to ensure that none of the holes from the base go to the collector. A large-signal model of a BJT operating in the active region is shown in Fig. 1.30.

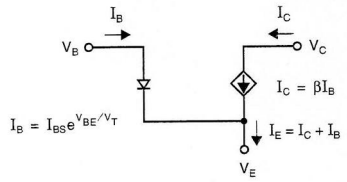


Fig. 1.30 A large-signal model for a BJT in the active region.

Since $I_B = I_C/\beta$, we have

$$I_B = \frac{I_{CS}}{\beta} e^{V_{BE}/V_T} = I_{BS} e^{V_{BE}/V_T} \tag{1.130}$$

which is similar to a diode equation, but with a multiplying constant of $I_{CS}/\beta = I_{BS}$. Since $I_E = I_B + I_C$, we have

$$I_E = I_{CS}\left(\frac{\beta + 1}{\beta}\right) e^{V_{BE}/V_T} = I_{ES} e^{V_{BE}/V_T} \tag{1.131}$$

or equivalently

$$I_C = \alpha I_E \tag{1.132}$$

where α has been defined as

$$\alpha = \frac{\beta}{\beta + 1} \tag{1.133}$$

and for large values of β, can be approximated as

$$\alpha \cong 1 - \frac{1}{\beta} \cong 1 \tag{1.134}$$

If the effect of V_{CE} on I_C is included in the model, the current-controlled source, βI_B, should be replaced by a current source given by

$$I_C = \beta I_B\left(1 + \frac{V_{CE}}{V_A}\right) \tag{1.135}$$

where V_A is the Early-voltage constant. This additional modelling of the finite output impedance is normally not done in large-signal analysis without the use of a computer due to its complexity.

As the collector-emitter voltage approaches $V_{CE\text{-}sat}$ (typically around 0.2 to 0.3 V), the base-collector junction becomes forward biased, and holes from the base will begin to diffuse to the collector. A common model for this case, when the transistor is *saturated* or in the *saturation region,* is shown in Fig. 1.31. It should be noted that the value of $V_{CE\text{-}sat}$ decreases for smaller values of collector current.

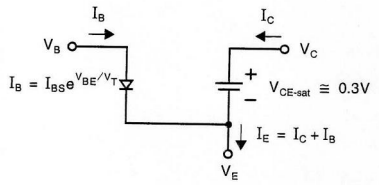


Fig. 1.31 A large-signal model for a BJT in the saturation region.

Base-Charge Storage in the Active Region

When a transistor is in the active region, many minority carriers are stored in the base region (electrons are stored in an npn transistor). Recall that this minority charge is responsible for I_C, so this charge must be removed (through the base contact) before a transistor can turn off. As in a forward-bias diode, this charge can be modelled as a diffusion capacitance, C_d, between the base and emitter given by (see Appendix at the end of this chapter)

$$C_d = \tau_b \frac{I_C}{V_T} \tag{1.136}$$

where τ_b is the base-transit-time constant. Thus, we see that the diffusion capacitance is proportional to I_C. The total base-emitter capacitance, C_{be}, will include the base-emitter depletion capacitance, C_j, in parallel with C_d. Normally, however, C_j is much less than C_d, unless the transistor current is small, and can often be ignored.

Base-Charge Storage of a Saturated Transistor

When a transistor becomes saturated, the minority-charge storage in the base and, even more so, in the lightly doped region of the collector, increases drastically. The major component of this charge storage is due to holes diffusing from the base, through the collector junction, and continuing on through the lightly doped n^- *epitaxial region* of the collector to the n^+ collector region. The n^- epitaxial region is so named because it is epitaxially grown on a p region. Most of the charge storage occurs in this region. Also, additional charge storage occurs because electrons that diffused from the collector are stored in the base, but this charge is normally smaller. The magnitude of the additional charge stored by a transistor that is saturated is given by

$$Q_s = \tau_s\left(I_B - \frac{I_C}{\beta}\right) \tag{1.137}$$

where the base overdrive current, defined to be $I_B - I_C/\beta$, is approximately equal to the hole current from the base to the collector. Normally, in saturation, $I_B >> I_C/\beta$, and (1.137) can be approximated by

$$Q_s \cong \tau_s I_B \tag{1.138}$$

The constant τ_s is approximately equal to the epitaxial-region transit time, τ_E (ignoring the storage of electrons in the base that diffused from the collector). Since the epitaxial region is much wider than the base, the constant τ_s is normally much larger than the base transit time, the constant τ_b, often by up to two orders of magnitude. The specific value of τ_s is usually found empirically for a given technology.

When a saturated transistor is being turned off, first the base current will reverse. *However, before the collector current will change, the saturation charge, Q_s, must be removed.* After Q_s is removed, the base minority charge, Q_b, will be removed. During this time, the collector current will decrease until the transistor shuts off. Typically, the time to remove Q_s greatly dominates the overall charge removal.

If the time required to remove the base saturation charge, t_s, is much shorter than the epitaxial-region transit time, τ_E, then one can derive a simple expression for the time required to remove the saturation charge. If the reverse base current (when the saturation charge is being removed), denoted by I_{BR}, remains constant while Q_s is being removed, then we have [Hodges, 1988]

$$t_s \cong \frac{Q_s}{I_{BR}} \cong \frac{\tau_s\left(I_B - \frac{I_C}{\beta}\right)}{I_{BR}} \cong \tau_s \frac{I_B}{I_{BR}} \tag{1.139}$$

where $\tau_s \cong \tau_E$.

Normally, the forward base current during saturation, I_B, will be much smaller than the reverse base current during saturation-charge removal, I_{BR}. If this were not the case, then our original assumption that $t_s << \tau_E \cong \tau_s$ would not be true. In this case, the turn-off time of the BJT would be so slow as to make the circuit unusable in most applications. Nevertheless, the turn-off time for this case, when t_s is not much less than τ_E, is given by [Hodges, 1988]

$$t_s = \tau_s \ln\left[\frac{I_{BR} + I_B}{I_{BR} + \frac{I_C}{\beta}}\right] \tag{1.140}$$

The reader should verify that for $I_{BR} >> I_B$ and $I_{BR} >> I_C/\beta$, the expression in (1.140) is approximately equivalent to the much simpler one in (1.139).

In both of the cases just described, the time required to remove the storage charge of a saturated transistor is much larger than the time required to turn off a transistor in the active region. *In high-speed microcircuit designs, one never allows bipolar transistors to saturate, to avoid the long turn-off time that would result.*

EXAMPLE 1.13

For $\tau_b = 0.2$ ns, $\tau_s = 100$ ns (a small value for τ_s), $I_B = 0.2$ mA, $I_C = 1$ mA, $\beta = 100$, and $I_{BR} = 1$ mA, calculate the time required to remove the base saturation charge using (1.139), and compare it to the time obtained using the more accurate expression of (1.140). Compare these results to the time required to remove the base minority charge for the same I_{BR}.

Solution

Using (1.139), we have

$$t_s = \frac{10^{-7}(2 \times 10^{-4})}{10^{-3}} = 20 \text{ ns} \tag{1.141}$$

Using (1.140), we have

$$t_s = 10^{-7}\ln\left[\frac{10^{-3}+2\times10^{-4}}{10^{-3}+\frac{10^{-3}}{100}}\right] = 17.2\text{ ns} \tag{1.142}$$

which is fairly close to the first result.

The time required for an active transistor to remove the base minority charge, Q_b, is given by

$$t_A = \frac{Q_b}{I_{BR}} = \frac{\tau_b I_C}{I_{BR}} = 0.2\text{ ns} \tag{1.143}$$

This is approximately 100 times shorter than the time for removing the base saturation charge!

Small-Signal Modelling

The most commonly used small-signal model is the hybrid-π model. This model is similar to the small-signal model used for MOS transistors, except it includes a finite base-emitter impedance, r_π, and it has no emitter-to-bulk capacitance. The hybrid-π model is shown in Fig. 1.32. As in the MOS case, we will first discuss the transconductance, g_m, and the small-signal resistances, and then we will discuss the parasitic capacitances.

The transistor transconductance, g_m, is perhaps the most important parameter of the small-signal model. The transconductance is the ratio of the small-signal collector current, i_c, to the small-signal base-emitter voltage, v_{be}. Thus, we have

$$g_m = \frac{i_c}{v_{be}} = \frac{\partial I_C}{\partial V_{BE}} \tag{1.144}$$

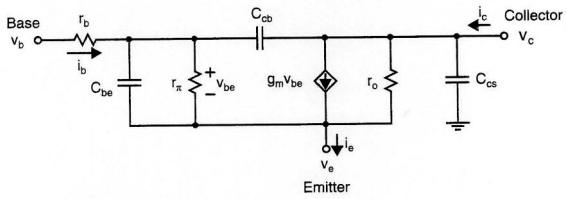


Fig. 1.32 The small-signal model of an active BJT.

Recall that in the active region

$$I_C = I_{CS}e^{V_{BE}/V_T} \tag{1.145}$$

Then

$$g_m = \frac{\partial I_C}{\partial V_{BE}} = \frac{I_{CS}}{V_T}e^{V_{BE}/V_T} \tag{1.146}$$

Using (1.145) again, we obtain

$$g_m = \frac{I_C}{V_T} \tag{1.147}$$

where V_T is given by

$$V_T = \frac{kT}{q} \tag{1.148}$$

and is approximately 26 mV at a room temperature of $T = 300$ °K. Thus, the transconductance is proportional to the bias current of a BJT. In integrated-circuit design, it is important that the transconductance (and hence speed) remain temperature independent, so the bias currents are usually made proportional to absolute temperature (since V_T is proportional to absolute temperature).

The presence of the resistor r_π reflects the fact that the base current is nonzero. We have

$$r_\pi = \frac{\partial V_{BE}}{\partial I_B} \tag{1.149}$$

Because from (1.130) we have

$$I_B = \frac{I_C}{\beta} = \frac{I_{CS}}{\beta}e^{V_{BE}/V_T} \tag{1.150}$$

we therefore have

$$\frac{1}{r_\pi} = \frac{\partial I_B}{\partial V_{BE}} = \frac{I_{CS}}{\beta V_T}e^{V_{BE}/V_T} \tag{1.151}$$

Using (1.150) again, we have

$$r_\pi = \frac{V_T}{I_B} \tag{1.152}$$

or equivalently,

$$r_\pi = \beta\frac{V_T}{I_C} = \frac{\beta}{g_m} \tag{1.153}$$

Since

$$i_e = i_c + i_b \tag{1.154}$$

we also have

$$\begin{aligned}\frac{\partial I_E}{\partial V_{BE}} &= \frac{\partial I_C}{\partial V_{BE}} + \frac{\partial I_B}{\partial V_{BE}} \\ &= g_m + \frac{g_m}{\beta} \\ &= g_m\left(\frac{1+\beta}{\beta}\right) \\ &= \frac{g_m}{\alpha}\end{aligned} \tag{1.155}$$

Some alternative models, usually called T models (see page 55), use the emitter resistance, r_e, where

$$r_e = \frac{\partial V_{BE}}{\partial I_E} = \frac{\alpha}{g_m} \tag{1.156}$$

Continuing, we have

$$\frac{1}{r_o} = \frac{\partial I_C}{\partial V_{CE}} \tag{1.157}$$

The small-signal resistance, r_o, models the dependence of the collector current on the collector-emitter voltage. Repeating (1.129) here for convenience,

$$I_C = I_{CS} e^{V_{BE}/V_T}\left(1 + \frac{V_{CE}}{V_A}\right) \tag{1.158}$$

we have

$$\frac{1}{r_o} = \frac{\partial I_C}{\partial V_{CE}} = \frac{I_{CS}}{V_A} e^{V_{BE}/V_T} \tag{1.159}$$

Thus,

$$r_o = \frac{V_A}{I_C} \tag{1.160}$$

which is inversely proportional to the collector current. As an aside, note that $g_m r_o = V_A/V_T$ is a constant value independent of the transistor operating point. This constant is usually between 2,000 and 8,000 for an npn BJT and is an upper limit on the attainable voltage gain for a single-transistor amplifier.

The resistor r_b models the resistance of the semiconductor material between the base contact and the effective base region due to the moderately lightly doped base p material (see Fig. 1.25). This resistor, although small (typically a few hundred ohms), can be important in limiting the speed of very-high-frequency low-gain BJT circuits and is a major source of noise.

EXAMPLE 1.14

For $I_C = 1$ mA, $\beta = 100$, and $V_A = 100$ V, calculate g_m, r_π, r_e, r_o, and $g_m r_o$.

Solution

We have

$$g_m = \frac{I_C}{V_T} = \frac{10 \times 10^{-3}\ \text{A}}{0.026\ \text{V}} = 38.5\ \text{mA/V} \tag{1.161}$$

$$r_\pi = \frac{\beta}{g_m} = 2.6\ \text{k}\Omega \tag{1.162}$$

$$r_e = \frac{\alpha}{g_m} = \left(\frac{100}{101}\right)26 = 25.7\ \Omega \tag{1.163}$$

$$r_o = \frac{V_A}{I_C} = \frac{100}{10^{-3}} = 100\ \text{k}\Omega \tag{1.164}$$

and $g_m r_o$, the maximum possible gain with a single-transistor amplifier, is given by

$$g_m r_o = \frac{V_A}{V_T} = 3{,}846 \tag{1.165}$$

Note that this gain is much higher than the 52.6 that was found for a single MOS transistor in Example 1.9. Also note that this BJT maximum gain is independent of the bias current. For MOS transistors, it can be shown that the maximum gain decreases with larger bias currents in a square-root relationship. This is one of the reasons why it is possible to realize a single-transistor BJT amplifier with a much larger gain than would result if a MOS transistor were used, especially at high current levels (and therefore at high frequencies).

The high-frequency operation of a BJT is limited by the capacitances of the small-signal model. We have already encountered one of these capacitances, C_{be}, in Section 1.1. Recapping, we have

$$C_{be} = C_j + C_d \tag{1.166}$$

where C_j is the depletion capacitance of the base-emitter junction. For a forward-biased junction, a rough approximation for C_j is

$$C_j \cong 2A_E C_{je0} \tag{1.167}$$

The diffusion capacitance, C_d, is given in (1.136) as

$$C_d = \tau_b \frac{I_C}{V_T} = g_m \tau_b \tag{1.168}$$

The capacitor, C_{cb}, models the depletion capacitance of the collector-base junction. Since this is a graded junction, we can approximate C_{cb} by

$$C_{cb} = \frac{A_C C_{jc0}}{\left(1 + \frac{V_{CB}}{\Phi_{c0}}\right)^{1/3}} \tag{1.169}$$

where A_C is the effective area of the collector-base interface.

Due to the lower doping levels in the base and especially in the collector (perhaps 5×10^{22} acceptors/m^3 and 10^{21} donors/m^3, respectively), Φ_{c0}, the built-in potential for the collector-base junction, will be less than that of the base-emitter junction (perhaps 0.75 V as opposed to 0.9 V). It should be noted that the cross-sectional area of the collector-base junction, A_C, is typically much larger than the effective area of the base-emitter junction, A_E, which is shown in Fig. 1.25. This size differential results in $A_C C_{jc0}$ being larger than $A_E C_{je0}$, the base-emitter junction capacitance at 0 V bias, despite the lower doping levels.

Finally, another large capacitor is C_{cs}, the capacitance of the collector-to-substrate junction. Since this area is quite large, C_{cs}, which is the depletion capacitance that results from this area, will be much larger than either C_{cb} or the depletion capacitance component of C_{be}, that is, C_j. The value of C_{cs} can be calculated using

$$C_{cs} = \frac{A_T C_{js0}}{\left(1 + \frac{V_{CS}}{\Phi_{s0}}\right)^{1/2}} \tag{1.170}$$

where A_T is the effective transistor area and C_{js0} is the collector-to-substrate capacitance per unit area at 0-V bias voltage.

A common indicator for the *speed* of a BJT is the frequency at which the transistor's current gain drops to unity, when its collector is connected to a small-signal ground. This frequency is denoted f_t and is called the transistor *unity-gain frequency*. We can see how this frequency is related to the transistor model parameters by analyzing the small-signal circuit of Fig. 1.33. In the simplified model in part *b*, the resistor r_b is ignored because it has no effect on i_b since the circuit is being driven by a perfect current source. We have

$$V_{be} = i_b \left(r_\pi \left\| \frac{1}{sC_{be}} \right\| \frac{1}{sC_{cb}} \right) \tag{1.171}$$

and

$$i_c = g_m V_{be} \tag{1.172}$$

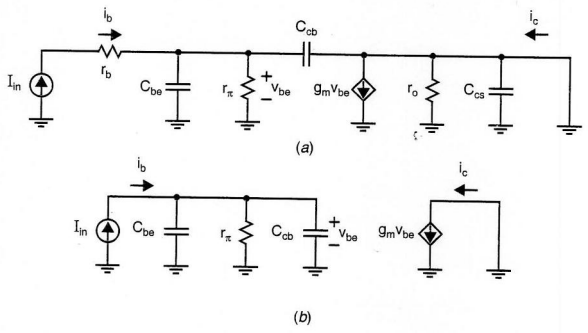


Fig. 1.33 (*a*) A small-signal model used to find f_t; (*b*) an equivalent simplified model.

Solving for i_c / i_b gives

$$\frac{i_c}{i_b} = \frac{g_m r_\pi}{1 + s(C_{be} + C_{cb}) r_\pi} \tag{1.173}$$

At low frequencies, the current gain is $g_m r_\pi$, which equals the expected value of β (using (1.153)). At high frequencies, i_c / i_b is approximately given by

$$\left| \frac{i_c}{i_b}(\omega) \right| \cong \frac{g_m r_\pi}{\omega(C_{be} + C_{cb}) r_\pi} = \frac{g_m}{\omega(C_{be} + C_{cb})} \tag{1.174}$$

To find the unity-gain frequency, we set $|(i_c/i_b)(\omega_t)| = 1$ and solve for ω_t, which results in

$$\omega_t = \frac{g_m}{C_{be} + C_{cb}} \tag{1.175}$$

or

$$f_t = \frac{g_m}{2\pi(C_{be} + C_{cb})} \tag{1.176}$$

Often, either f_t, ω_t, or $\tau_t = 1/\omega_t$ will be specified for a transistor at a particular bias current. These values indicate an upper limit on the maximum frequency at which the transistor can be effectively used.

The hybrid-π model is only one of a number of small-signal models that can be used. One common alternative is the low-frequency T model shown in Fig. 1.34. Use of this T model often results in a much simplified analysis, compared to use of the hybrid-π model, and thus it is useful for hand analysis.

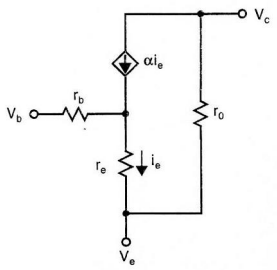


Fig. 1.34 A low-frequency, small-signal T model for an active BJT.

1.5 DEVICE MODEL SUMMARY

As a useful aid, all of the equations for the large-signal and small-signal modelling of diodes, MOS transistors, and bipolar transistors, along with values for the various constants, are listed in the next few pages.

Constants

$q = 1.602 \times 10^{-19}$ C	$k = 1.38 \times 10^{-23}$ JK^{-1}
$n_i = 1.1 \times 10^{16}$ carriers/m^3 at T = 300 °K	$\varepsilon_0 = 8.854 \times 10^{-12}$ F/m
$K_{ox} \cong 3.9$	$K_s \cong 11.8$
$\mu_n = 0.05$ m^2/V · s	$\mu_p = 0.02$ m^2/V · s

Diode Equations

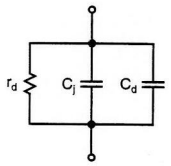
Reverse-Biased Diode (Abrupt Junction)

$C_j = \dfrac{C_{j0}}{\sqrt{1 + \dfrac{V_R}{\Phi_0}}}$	$Q = 2C_{j0}\Phi_0\sqrt{1 + \dfrac{V_R}{\Phi_0}}$
$C_{j0} = \sqrt{\dfrac{qK_s\varepsilon_0}{2\Phi_0}\dfrac{N_D N_A}{N_A + N_D}}$	$C_{j0} = \sqrt{\dfrac{qK_s\varepsilon_0 N_D}{2\Phi_0}}$ if $N_A >> N_D$
$\Phi_0 = \dfrac{kT}{q}\ln\left(\dfrac{N_A N_D}{n_i^2}\right)$	

Forward-Biased Diode

$I_D = I_S e^{V_D/V_T}$	$I_S = A_D q n_i^2\left(\dfrac{D_n}{L_n N_A} + \dfrac{D_p}{L_p N_D}\right)$
$V_T = \dfrac{kT}{q} \cong 26$ mV at 300 °K	

Small-Signal Model of Forward-Biased Diode



$r_d = \dfrac{V_T}{I_D}$	$C_T = C_d + C_j$
$C_d = \tau_T\dfrac{I_D}{V_T}$	$C_j \cong 2C_{j0}$
$\tau_T = \dfrac{L_n^2}{D_n}$	

MOS Transistor Equations

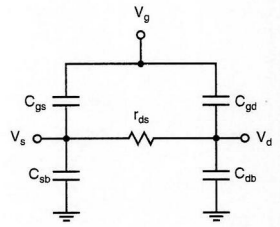
The following equations are for n-channel devices—for p-channel devices, put negative signs in front of all voltages. These equations do not account for short-channel effects (i.e., $L < 2L_{min}$).

Triode Region ($V_{GS} > V_{tn}$, $V_{DS} \leq V_{eff}$)

$I_D = \mu_n C_{ox}\left(\dfrac{W}{L}\right)\left[(V_{GS} - V_{tn})V_{DS} - \dfrac{V_{DS}^2}{2}\right]$	
$V_{eff} = V_{GS} - V_{tn}$	$V_{tn} = V_{tn\text{-}0} + \gamma(\sqrt{V_{SB} + 2\phi_F} - \sqrt{2\phi_F})$

$\phi_F = \frac{kT}{q}\ln\left(\frac{N_A}{n_i}\right)$	$\gamma = \frac{\sqrt{2qK_{si}\varepsilon_0 N_A}}{C_{ox}}$
$C_{ox} = \frac{K_{ox}\varepsilon_0}{t_{ox}}$	

Small-Signal Model in Triode Region (for $V_{DS} << V_{eff}$)

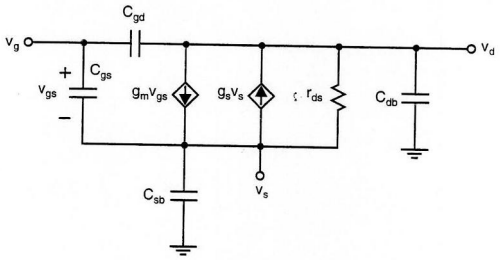


$r_{ds} = \frac{1}{\mu_n C_{ox}\left(\frac{W}{L}\right)V_{eff}}$	
$C_{gd} = C_{gs} \cong \frac{1}{2}WLC_{ox} + WL_{ov}C_{ox}$	$C_{sb} = C_{db} = \frac{C_{j0}(A_s + WL/2)}{\sqrt{1 + \frac{V_{sb}}{\Phi_0}}}$

Active (or Pinch-Off) Region ($V_{GS} > V_{tn}$, $V_{DS} \geq V_{eff}$)

$I_D = \frac{\mu_n C_{ox} W}{2 L}(V_{GS} - V_{tn})^2[1 + \lambda(V_{DS} - V_{eff})]$	
$\lambda \propto \frac{1}{L\sqrt{V_{DS} - V_{eff} + \Phi_0}}$	$V_{tn} = V_{tn\text{-}0} + \gamma(\sqrt{V_{SB} + 2\phi_F} - \sqrt{2\phi_F})$
$V_{eff} = V_{GS} - V_{tn} = \sqrt{\frac{2I_D}{\mu_n C_{ox} W/L}}$	

Small-Signal Model (Active Region)



$g_m = \mu_n C_{ox}\left(\frac{W}{L}\right)V_{eff}$	$g_m = \sqrt{2\mu_n C_{ox}(W/L)I_D}$
$g_m = \frac{2I_D}{V_{eff}}$	$g_s = \frac{\gamma g_m}{2\sqrt{V_{SB} + \lvert 2\phi_F\rvert}}$
$r_{ds} = \frac{1}{\lambda I_D}$	$g_s \cong 0.2 g_m$
$\lambda = \frac{k_{rds}}{2L\sqrt{V_{DS} - V_{eff} + \Phi_0}}$	$k_{rds} = \sqrt{\frac{2K_s\varepsilon_0}{qN_A}}$
$C_{gs} = \frac{2}{3}WLC_{ox} + WL_{ov}C_{ox}$	$C_{gd} = WL_{ov}C_{ox}$
$C_{sb} = (A_s + WL)C_{js} + P_s C_{j\text{-}sw}$	$C_{js} = \frac{C_{j0}}{\sqrt{1 + V_{SB}/\Phi_0}}$
$C_{db} = A_d C_{jd} + P_d C_{j\text{-}sw}$	$C_{jd} = \frac{C_{j0}}{\sqrt{1 + V_{DB}/\Phi_0}}$

Typical Values for a 0.8-μm Process

$V_{tn} = 0.8$ V	$V_{tp} = -0.9$ V
$\mu_n C_{ox} = 90\ \mu A/V^2$	$\mu_p C_{ox} = 30\ \mu A/V^2$

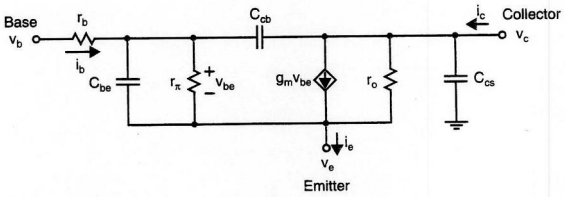
$C_{ox} = 1.9 \times 10^{-3}$ pF/(μm)^2	$C_j = 2.4 \times 10^{-4}$ pF/(μm)^2
$C_{j\text{-}sw} = 2.0 \times 10^{-4}$ pF/μm	$C_{gs(overlap)} = 2.0 \times 10^{-4}$ pF/μm
$\phi_F = 0.34$ V	$\Phi_0 = 0.9$ V
$\gamma = 0.5$ V$^{1/2}$	$t_{ox} = 0.02$ μm
$N_B = 6 \times 10^{21}$ impurities/m^3	

Bipolar-Junction Transistors

Active Transistor

$I_C = I_{CS} e^{V_{BE}/V_T}$	$V_T = \frac{kT}{q} \cong 26$ mV at 300 °K
For more accuracy, $I_C = I_{CS} e^{V_{BE}/V_T}\left(1 + \frac{V_{CE}}{V_A}\right)$	
$I_{CS} = \frac{A_E q D_n n_i^2}{W N_A}$	$I_B = \frac{I_C}{\beta}$
$I_E = \left(1 + \frac{1}{\beta}\right) I_C = \frac{1}{\alpha} I_C = (\beta + 1) I_B$	$\beta = \frac{I_C}{I_B} = \frac{D_n N_D L_p}{D_p N_A W} \cong 2.5 \frac{N_D L_p}{N_A W}$
$\alpha = \frac{\beta}{1 + \beta}$	

Small-Signal Model of an Active BJT



$g_m = \frac{I_C}{V_T}$	$r_\pi = \frac{V_T}{I_B} = \frac{\beta}{g_m}$
$r_e = \frac{\alpha}{g_m}$	$r_o = \frac{V_A}{I_C}$
$g_m r_o = \frac{V_A}{V_T}$	$C_d = \tau_b \frac{I_C}{V_T} = g_m \tau_b$
$C_{be} = C_j + C_d$	$C_{cs} = \frac{A_T C_{js0}}{\left(1 + \frac{V_{CS}}{\Phi_{s0}}\right)^{1/2}}$
$C_{cb} = \frac{A_C C_{jc0}}{\left(1 + \frac{V_{CB}}{\Phi_{c0}}\right)^{1/3}}$	

1.6 SPICE-MODELLING PARAMETERS

This section briefly describes some of the important model parameters for diodes, bipolar transistors, and MOS transistors used during a SPICE simulation. It should be noted here that not all SPICE model parameters are described. However, enough are described to enable the reader to understand the relationship between the relative parameters and the corresponding constants used when doing hand analysis.

Diode Model

There are a number of important dc parameters. The constant I_S is specified using either the parameter IS or JS in SPICE. These two parameters are synonyms, and only one should be specified. A typical value specified for I_S might be between 10^{-18} A and 10^{-15} A for small diodes in a microcircuit. Another important parameter is called the *emission coefficient,* n. This constant multiplies V_T in the exponential diode I-V relationship given by

$$I_D = I_S e^{V_{BE}/(nV_T)} \tag{1.177}$$

The SPICE parameter for n is N and is defaulted to 1 when not specified (1 is a reasonable value for junctions in a microcircuit). A third important dc characteristic is the *series resistance,* which is specified in SPICE using RS. It should be noted here that some SPICE programs allow the user to specify the area of the diode, whereas

others expect absolute parameters that already take into account the effective area. The manual for the program being used should be consulted.

The diode transit time is specified using the SPICE parameter TT. The most important capacitance parameter specified is CJ. CJO and CJ are synonyms—one should never specify both. This parameter specifies the capacitance at 0-V bias. Once again, it may be specified as absolute or as relative to the area (i.e., F/m^2), depending on the version of SPICE used. Also, the area junction grading coefficient, MJ, might be specified to determine the exponent used in the capacitance equation. Typical values are 0.5 for abrupt junctions and 0.33 for graded junctions. In some SPICE versions, it might also be possible to specify the sidewall capacitance at 0-V bias as well as its grading junction coefficient. Finally, the built-in potential of the junction, which is also used in calculating the capacitance, can be specified using PB. PHI, VJ, and PHA are all synonyms of PB.

Reasonably accurate diode simulations can usually be obtained by specifying only IS, CJ, MJ, and PB. However, most modern versions of SPICE have many more parameters that can be specified if one wants accurate temperature and noise simulations. Users should consult their manuals for more information.

Table 1.1 summarizes some of the more important diode parameters. This set of parameters constitutes a minimal set for reasonable simulation accuracy under ordinary conditions.

MOS Transistors

Modern MOS models are quite complicated, so only some of the more important MOS parameters used in SPICE simulations are described here. These parameters are used in what are called the Level 2 or Level 3 models. The model level can be chosen by setting the SPICE parameter LEVEL to either 2 or 3. The oxide thickness, t_{ox}, is specified using the SPICE parameter TOX. If it is specified, then it is not necessary to specify the thin gate-oxide capacitance (C_{ox}, specified by parameter COX). The mobility, μ_n, can be specified using UO. If UO is specified, the intrinsic transistor conductance ($\mu_n C_{ox}$) will be calculated automatically, unless this automatic calculation is overridden by specifying either KP (or its synonym, BETA). The transistor threshold voltage at $V_S = 0$ V, V_{tn}, is specified by VTO. The body-effect parameter, γ,

Table 1.1 Important SPICE parameters for modelling diodes

SPICE Parameter	Model Constant	Brief Description	Typical Value
IS	I_S	Transport saturation current	10^{-17} A
RS	R_d	Series resistance	30 Ω
TT	τ_T	Diode transit time	12 ps
CJ	C_{j0}	Capacitance at 0-V bias	0.01 pF
MJ	m_j	Diode grading coefficient exponent	0.5
PB	Φ_0	Built-in diode contact potential	0.9 V

can be specified using GAMMA, or it will be automatically calculated if the substrate doping, N_A, is specified using NSUB. Normally, one would not want SPICE to calculate γ because the effective substrate doping under the channel can differ significantly from the substrate doping in the bulk due to threshold-voltage adjust implants. The output impedance constant, λ, can be specified using LAMBDA. Normally, LAMBDA should not be specified since it takes precedence over internal calculations and does not change the output impedance as a function of different transistor lengths or bias voltages (which should be the case). Indeed, modelling the transistor output impedance is one of weakest points in SPICE. If LAMBDA is not specified, it is calculated automatically. The surface inversion potential, $|2\phi_F|$, can be specified using PHI, or it will be calculated automatically. Another parameter usually specified is the lateral diffusion of the junctions under the gate, L_D, which is specified by LD. For accurate simulations, one might also specify the resistances in series with the source and drain by specifying RS and RD (typically only the source resistance is important). Many other parameters exist to model such things as short-channel effects, subthreshold effects, and channel-width effects, but these parameters are outside the scope of this book.

The modelling of parasitic capacitances in SPICE is quite involved. Originally, this modelling was not very accurate since it did not include charge conservation for the gate charge. However, this modelling has greatly improved in recent commercial versions of SPICE. The capacitances under the junctions per unit area at 0-V bias, (i.e., C_{j0}) can be specified using CJ or can be calculated automatically from the specified substrate doping. The sidewall capacitances at 0 V, $C_{j\text{-}sw0}$, should normally be specified using CJSW because this parameter is used to calculate significant parasitic capacitances. The bulk grading coefficient specified by MJ can usually be defaulted to 0.5. Similarly, the sidewall grading coefficient specified by MJSW can usually be defaulted to 0.33 (SPICE assumes a graded junction). The built-in bulk-to-junction contact potential, Φ_0, can be specified using PB or defaulted to 0.8 V (note that 0.9 V would typically be more accurate, but the resulting simulation differences are small). Sometimes the gate-to-source or drain-overlap capacitances can be specified using CGSO or CGDO, but normally these would be left to be calculated automatically using COX and LD.

Some of the more important parameters that should result in reasonable simulations (except for modelling short-channel effects) are summarized in Table 1.2 for both n- and p-channel transistors. Table 1.2 lists reasonable parameters for a typical 0.8-μm technology.

Bipolar Junction Transistors

For historical reasons, most parameters for modelling bipolar transistors are specified absolutely. Also, rather than specifying the emitter area of a BJT in $(\mu m)^2$ on the line where the individual transistor connections are specified, most SPICE versions have multiplication factors. These multiplication factors can be used to automatically multiply parameters when a transistor is composed of several transistors connected in parallel. This multiplying parameter is normally called M.

Table 1.2 A reasonable set of MOS parameters for a typical 0.8-μm technology

SPICE Parameter	Model Constant	Brief Description	Typical Value
VTO	V_{tn}:V_{tp}	Transistor threshold voltage (in V)	0.7:–0.9
UO	μ_n:μ_p	Carrier mobility in bulk (in $cm^2/V{\cdot}s$)	500:175
TOX	t_{ox}	Thickness of gate oxide (in m)	1.8×10^{-8}
LD	L_D	Lateral diffusion of junction under gate (in m)	6×10^{-8}
GAMMA	γ	Body-effect parameter	0.5: 0.8
NSUB	N_A:N_D	The substrate doping (in cm^{-3})	3×10^{16}:7.5×10^{16}
PHI	$\lvert 2\phi_F \rvert$	Surface inversion potential (in V)	0.7
PB	Φ_0	Built-in contact potential of junction to bulk (in V)	0.9
CJ	C_{j0}	Junction-depletion capacitance at 0-V bias (in F/m^2)	2.5×10^{-4}:4.0×10^{-4}
CJSW	$C_{j\text{-}sw0}$	Sidewall capacitance at 0-V bias (in F/m)	2.0×10^{-10}:2.8×10^{-10}
MJ	m_j	Bulk-to-junction exponent (grading coefficient)	0.5
MJSW	$m_{j\text{-}sw}$	Sidewall-to-junction exponent (grading coefficient)	0.3

The most important dc parameters are the transistor current gain, β, specified by the SPICE parameter BF; the transistor-transport saturation current, I_{CS}, specified using the parameter IS; and the Early-voltage constant, specified by the parameter VAF. Typical values for these might be 100, 10^{-17} A, and 50 V, respectively. If one wants to model the transistor in reverse mode (where the emitter voltage is higher than the collector voltage for an npn), then one might specify BR, ISS, and VAR, as well; these are the parameters that correspond to BIF, IS, and VAF in the reverse direction. Typically, this reverse-mode modelling is not important for most circuits. Some other important dc parameters for accurate simulations are the base, emitter, and collector resistances, which are specified by RB, RE, and RC, respectively. It is especially important to specify RB (which might be 200 Ω to 500 Ω).

The important capacitance parameters and their corresponding SPICE parameters include the depletion capacitances at 0-V bias voltage, CJE, CJC, CJS; their grading coefficients, MJE, MJC, MJS; and their built-in voltages, VJE, VJC, VJS, for base-emitter, base-collector, and collector-substrate junctions. Again, the 0-V depletion capacitances should be specified in absolute values for a unit-sized transistor. Normally the base-emitter and base-collector junctions are graded (i.e., MJE, MJC = 0.33), whereas the collector-substrate junction may be either abrupt (MJS = 0.5) or graded (MJS = 0.5), depending on processing details. Typical built-in voltages might be 0.75 V to 0.8 V. In addition, for accurate simulations, one should specify the forward-base transit time, τ_F, specified by TF, and, if the transistor is to be operated in reverse mode or under saturated conditions, the reverse-base transit time, τ_R, specified by TR.

The most important of the model parameters just described are summarized in Table 1.3.

Once again, many other parameters can be specified if accurate simulation is desired. Other parameters might include those to model β degradation under high or

Table 1.3 The most important SPICE parameters for modelling BJTs

SPICE Parameter	Model Constant	Brief Description	Typical Value
BF	β	Transistor current gain in forward direction	100
ISS	I_{CS}	Transport saturation current in forward direction	2×10^{-18} A
VAF	V_A	Early voltage in forward direction	50 V
RB	r_b	Series base resistance	500 Ω
RE	R_E	Series emitter resistance	30 Ω
CJE	C_{je0}	Base-emitter depletion capacitance at 0 V	0.015 pF
CJC	C_{je0}	Base-collector depletion capacitance at 0 V	0.018 pF
CJS	C_{je0}	Collector-substrate depletion capacitance at 0 V	0.040 pF
MJE	m_e	Base-emitter junction exponent (grading factor)	0.30
MJC	m_c	Base-collector junction exponent (grading factor)	0.35
MJS	m_s	Collector-substrate junction exponent (grading factor)	0.29
TF	τ_F	Forward-base transit time	12 ps
TR	τ_R	Reverse-base transit time	4 ns

low current applications and parameters for accurate noise and temperature analysis. Readers should refer to their SPICE manuals for descriptions of these parameters.

1.7 APPENDIX

The purpose of this appendix is to present derivations for device equations that rely heavily on device physics knowledge. Specifically, equations are derived for the exponential relationship and diffusion capacitance of diodes, for the threshold voltage and triode relationship for MOS transistors, and for the exponential relationship and base charge storage for bipolar transistors.

Diode Exponential Relationship

The concentration of minority carriers in the bulk, far from the junction, is given by Eqs. (1.2) and (1.4). Close to the junction, the minority-carrier concentrations are much larger. Indeed, the concentration next to the junction increases exponentially with the external voltage, V_D, that is applied in the forward direction. The concentration of holes in the n side next to the junction, p_n, is given by [Sze, 1981]

$$p_n = p_{n0} e^{V_D/V_T} = \frac{n_i^2}{N_D} e^{V_D/V_T} \tag{1.178}$$

Similarly, the concentration of electrons in the p side next to the junction is given by

$$n_p = n_{p0}e^{V_D/V_T} = \frac{n_i^2}{N_A}e^{V_D/V_T} \tag{1.179}$$

As the carriers diffuse away from the junction, their concentration exponentially decreases. The relationship for holes in the n side is

$$p_n(x) = p_n(0)e^{-x/L_p} \tag{1.180}$$

where x is the distance from the junction and L_p is a constant known as the diffusion length for holes in the n side. Similarly, for electrons in the p side we have

$$n_p(x) = n_p(0)e^{-x/L_n} \tag{1.181}$$

where L_n is a constant known as the diffusion length of electrons in the p side. Note that $p_n(0)$ and $n_p(0)$ are given by (1.178) and (1.179), respectively. Note also that the constants L_n and L_p are dependent on the doping concentrations N_A and N_D, respectively.

The current density of diffusing carriers moving away from the junction is given by the well-known diffusion equations [Sze, 1981]. For example, the current density of diffusing electrons is given by

$$J_{D\text{-}n} = -qD_n\frac{dn_p(x)}{dx} \tag{1.182}$$

where D_n is the diffusion constant of electrons in the p side of the junction. The negative sign is present because electrons have negative charge. Note that $D_n = (kT/q)\mu_n$, where μ_n is the mobility of electrons. Using (1.181), we have

$$\frac{dn_p(x)}{dx} = \frac{n_p(0)}{L_n}e^{-x/L_n} = -\frac{n_p(x)}{L_n} \tag{1.183}$$

Therefore

$$J_{D\text{-}n} = \frac{qD_n}{L_n}n_p(x) \tag{1.184}$$

Thus, the current density due to diffusion is proportional to the minority-carrier concentration. Next to the junction, all the current flow results from the diffusion of minority carriers. Further away from the junction, some of the current flow is due to diffusion and some is due to majority carriers drifting by to replace carriers that recombined with minority carriers or diffused across the junction.

Continuing, we use (1.179) and (1.184) to determine the current density next to the junction of electrons in the p side:

$$\begin{aligned} J_{D\text{-}n} &= \frac{qD_n}{L_n}n_p(0) \\ &= \frac{qD_n n_i^2}{L_n N_A}e^{V_D/V_T} \end{aligned} \tag{1.185}$$

For the total current of electrons in the p side, we multiply (1.185) by the effective junction area, A_D. The total current remains constant as we move away from the junction since, in the steady state, the minority carrier concentration at any particular location remains constant with time. In other words, if the current changed as we moved away from the junction, the charge concentrations would change with time.

Using a similar derivation, we obtain the total current of holes in the n side, $I_{D\text{-}p}$, as

$$I_{D\text{-}p} = \frac{A_D qD_n n_i^2}{L_p N_D}e^{V_D/V_T} \tag{1.186}$$

where D_n is the diffusion constant of electrons in the p side of the junction, L_p is the diffusion length of holes in the n side, and N_D is the impurity concentration of donors in the n side. This current, consisting of positive carriers, flows in the direction opposite to that of the flow of minority electrons in the p side. However, since electron carriers are negatively charged, the direction of the current flow is the same. Note also that if the p side is more heavily doped than the n side, most of the carriers will be holes, whereas if the n side is more heavily doped than the p side, most of the carriers will be electrons.

The total current is the sum of the minority currents at the junction edges:

$$I_D = A_D qn_i^2\left(\frac{D_n}{L_n N_A} + \frac{D_p}{L_p N_D}\right)e^{V_D/V_T} \tag{1.187}$$

Equation (1.187) is often expressed as

$$I_D = I_S e^{V_D/V_T} \tag{1.188}$$

where

$$I_S = A_D qn_i^2\left(\frac{D_n}{L_n N_A} + \frac{D_p}{L_p N_D}\right) \tag{1.189}$$

Equation (1.188) is the well-known exponential current-voltage relationship of forward-biased diodes.

The concentrations of minority carriers near the junction and the direction of current flow are shown in Fig. 1.35.

Diode-Diffusion Capacitance

To find the diffusion capacitance, C_d, we first find the minority charge close to the junction, Q_d, and then differentiate it with respect to V_D. The minority charge close to the junction, Q_d, can be found by integrating either (1.180) or (1.181) over a few diffusion lengths. For example, if we assume n_{p0}, the minority electron concentration in the p side far from the junction is much less than $n_p(0)$, the minority electron con-

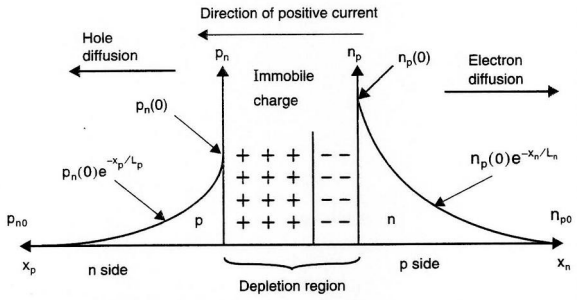


Fig. 1.35 The concentration of minority carriers and the direction of diffusing carriers near a forward-biased junction.

centration at the junction edge, we can use (1.181) to obtain

$$\begin{aligned} Q_n &= qA_D\int_0^\infty n_p(x)\,dx \\ &= qA_D\int_0^\infty n_p(0)e^{-X/L_n}\,dx \\ &= qA_DL_nn_p(0) \end{aligned} \tag{1.190}$$

Using (1.4) for $n_p(0)$ results in

$$Q_n = \frac{qA_DL_nn_i^2}{N_A}e^{V_D/V_T} \tag{1.191}$$

In a similar manner, we also have

$$Q_p = \frac{qA_DL_nn_i^2}{N_D}e^{V_D/V_T} \tag{1.192}$$

For a typical junction, one side will be much more heavily doped than the other side, and therefore the minority charge storage in the heavily doped side can be ignored since it will be much less than that in the lightly doped side. Assuming the n side is heavily doped, we find the total charge, Q_d, to be approximately given by Q_n, the minority charge in the p side. Thus, the small-signal diffusion capacitance, C_d, is given by

$$C_d = \frac{dQ_d}{dV_D} \equiv \frac{dQ_n}{dV_D} = \frac{qA_DL_nn_i^2}{N_AV_T}e^{V_D/V_T} \tag{1.193}$$

Using (1.187) and again noting that $N_D >> N_A$, we have

$$C_d = \frac{L_n^2I_D}{D_nV_T} \tag{1.194}$$

Equation (1.194) is often expressed as

$$C_d = \tau_T\frac{I_D}{V_T} \tag{1.195}$$

where τ_T is the transit time of the diode given by

$$\tau_T = \frac{L_n^2}{D_n} \tag{1.196}$$

for a single-sided diode in which the n side is more heavily doped.

MOS Threshold Voltage and the Body Effect

Many factors affect the gate-source voltage at which the channel becomes conductive. These factors are as follows:

1. The work-function difference between the gate material and the substrate material
2. The voltage drop between the channel and the substrate required for the channel to exist
3. The voltage drop across the thin oxide required for the depletion region, with its immobile charge, to exist
4. The voltage drop across the thin oxide due to unavoidable charge trapped in the thin oxide
5. The voltage drop across the thin oxide due to implanted charge at the surface of the silicon. The amount of implanted charge is adjusted in order to realize the desired threshold voltage.

The first factor affecting the transistor threshold voltage, V_{th}, is the built-in *Fermi potential* due to the different materials and doping concentrations used for the gate material and the substrate material. If one refers these potentials to that of intrinsic silicon [Tsividis, 1987], we have

$$\phi_{F\text{-}Gate} = \frac{kT}{q}\ln\left(\frac{N_D}{n_i}\right) \tag{1.197}$$

for a polysilicon gate with doping concentration N_D, and

$$\phi_{F\text{-}Sub} = \frac{kT}{q}\ln\left(\frac{n_i}{N_A}\right) \tag{1.198}$$

for a p substrate with doping concentration N_A. The work-function difference is then given by

$$\phi_{MS} = \phi_{F\text{-}Sub} - \phi_{F\text{-}Gate} = \frac{kT}{q}\ln\left(\frac{N_D N_A}{n_i^2}\right) \tag{1.199}$$

The next factor that determines the transistor threshold voltage is the voltage drop from the channel to the substrate, which is required for the channel to exist. The question of exactly when the channel exists does not have a precise answer. Rather, the channel is said to exist when the concentration of electron carriers in the channel is equal to the concentration of holes in the substrate. At this gate voltage, the channel is said to be *inverted.* As the gate voltage changes from a low value to the value at which the channel becomes inverted, the voltage drop in the silicon also changes, as does the voltage drop in the depletion region between the channel and the bulk. After the channel becomes inverted, any additional increase in gate voltage is closely equal to the increase in voltage drop across the thin oxide. In other words, after channel inversion, gate voltage variations have little effect on the voltage drop in the silicon or the depletion region between the channel and the substrate.

The electron concentration in the channel is equal to the hole concentration in the substrate when the voltage drop from the channel to the substrate is equal to two times the difference between the Fermi potential of the substrate and intrinsic silicon, ϕ_F, where

$$\phi_F = -\frac{kT}{q}\ln\left(\frac{N_A}{n_i}\right) \tag{1.200}$$

Equation (1.200) is a factor in several equations used in modelling MOS transistors. For typical processes, ϕ_F can usually be approximated as 0.35 V for typical doping levels at room temperature.

The third factor that affects the threshold voltage is due to the immobile negative charge in the depletion region left behind after the p mobile carriers are repelled. This effect gives rise to a voltage drop across the thin oxide of $-Q_B/C_{ox}$, where

$$Q_B = -qN_A x_d \tag{1.201}$$

and x_d is the width of the depletion region. Since

$$x_d = \sqrt{\frac{2K_s\varepsilon_0|2\phi_F|}{qN_A}} \tag{1.202}$$

we have

$$Q_B = -\sqrt{2qN_AK_s\varepsilon_0|2\phi_F|} \tag{1.203}$$

The fourth factor that determines V_{tn} is due to the unavoidable charge trapped in the thin oxide. Typical values for the effective ion density of this charge, N_{ox}, might be 2×10^{14} to 10^{15} ions/m^3. These ions are almost always positive. This effect gives rise to a voltage drop across the thin oxide, V_{ox}, given by

$$V_{ox} = \frac{-Q_{ox}}{C_{ox}} = \frac{-qN_{ox}}{C_{ox}} \tag{1.204}$$

The *native transistor threshold voltage* is the threshold voltage that would occur naturally if one did not include a special ion implant used to adjust the threshold voltage. This value is given by

$$V_{t\text{-native}} = \phi_{MS} - 2\phi_F - \frac{Q_B}{C_{ox}} - \frac{Q_{ox}}{C_{ox}} \tag{1.205}$$

A typical native threshold value might be around –0.1 V. It should be noted that transistors that have native transistor threshold voltages are becoming more important in circuit design where they might be used in transmission gates or in source-follower buffers.

The fifth factor that affects threshold voltage is a charge implanted in the silicon under the gate to change the threshold voltage from that given by (1.205) to the desired value, which might be 0.7 V for an n-channel transistor.

For the case in which the source-to-substrate voltage is increased, the effective threshold voltage is increased. This is known as the *body effect.* The body effect occurs because, as the source-bulk voltage, V_{SB}, becomes larger, the depletion region between the channel and the substrate becomes wider, and therefore more immobile negative charge becomes uncovered. This increase in charge changes the third factor in determining the transistor threshold voltage. Specifically, instead of using (1.203) to determine Q_B, one should now use

$$Q_B = -\sqrt{2qN_AK_s\varepsilon_0(V_{SB} + |2\phi_F|)} \tag{1.206}$$

If the threshold voltage when $V_{SB} = 0$ is denoted V_{tn0}, then, using (1.205) and (1.206), one can show that

$$\begin{aligned} V_{tn} &= V_{tn0} + \Delta V_{tn} \\ &= V_{tn0} + \frac{\sqrt{2qN_AK_s\varepsilon_0}}{C_{ox}}\left[\sqrt{V_{SB} + |2\phi_F|} - \sqrt{|2\phi_F|}\right] \\ &= V_{tn0} + \gamma\left(\sqrt{V_{SB} + |2\phi_F|} - \sqrt{|2\phi_F|}\right) \end{aligned} \tag{1.207}$$

where

$$\gamma = \frac{\sqrt{2qN_AK_s\varepsilon_0}}{C_{ox}} \tag{1.208}$$

The factor γ is often called the *body-effect constant.*

MOS Triode Relationship

The current flow in a MOS transistor is due to *drift* current rather than diffusion current. This type of current flow is the same mechanism that determines the current in a resistor. The current density, J, is proportional to the electrical field, E, where the constant of proportionality, σ, is called the *electrical permittivity.* Thus,

$$J = \sigma E \tag{1.209}$$

This constant for an n-type material is given by

$$\sigma = qn\mu_n \tag{1.210}$$

where n is the concentration per unit volume of negative carriers and μ_n is the mobility of electrons. Thus, the current density is given by

$$J = qn\mu_n E \tag{1.211}$$

Next, consider the current flow through the volume shown in Fig. 1.36, where the volume has height H and width W. The current is flowing perpendicular to the plane H × W down the length of the volume, L. The current, I, everywhere along the length of the volume is given by

$$I = JWH \tag{1.212}$$

The voltage drop along the length of the volume in the direction of L for a distance dx is denoted dV and is given by

$$dV = E(x)\,dx \tag{1.213}$$

Combining (1.211), (1.212), and (1.213), we obtain

$$q\mu_n WHn(x)\,dV = I\,dx \tag{1.214}$$

where the carrier density n(x) is now assumed to change along the length L and is therefore a function of x.

As an aside, we examine the case of a resistor where n(x) is usually constant. A resistor of length L would therefore have a current given by

$$I = \frac{q\mu_n WH}{L}\Delta V \tag{1.215}$$

Thus, the resistance is given by

$$R = \frac{L}{q\mu_n WH} \tag{1.216}$$

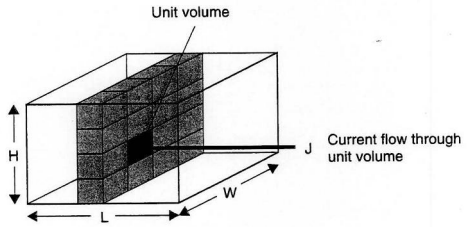


Fig. 1.36 Current flowing through a unit volume.

Often this resistance is presented in a relative manner, in which the length and width are removed (since they can be design parameters) but the height remains included. In this case, the resulting expression is commonly referred to as the *resistance per square* and designated as $R_\square$ where

$$R_\square = q\mu_n H \tag{1.217}$$

The total resistance is then given by

$$R_{total} = R_\square \frac{L}{W} \tag{1.218}$$

This equation is important when calculating the resistance of interconnects used in integrated circuits.

In the case of a MOS transistor, the charge density is not constant down the channel. If, instead of the carrier density per unit volume, one expresses n(x) as a function of charge density per square area from the top looking down, we have

$$Q_n(x) = qHn(x) \tag{1.219}$$

Substituting (1.219) into (1.214) results in

$$\mu_n WQ_n(x)\,dV = I\,dx \tag{1.220}$$

Equation (1.220) applies to drift current through any structure that has varying charge density in the direction of the current flow. It can also be applied to a MOS transistor in the triode region to derive its I-V relationship. It should be noted here that in this derivation, it is assumed the source voltage is the same as the substrate voltage.

Since the transistor is in the triode region, we have $V_{DG} < -V_{tn}$. This requirement is equivalent to $V_{DS} < V_{GS} - V_{tn} = V_{eff}$. It is assumed that the effective channel length is L. Assuming the voltage in the channel at distance x from the source is given by $V_{ch}(x)$, from Fig. 1.37, we have

$$Q_n(x) = C_{ox}[V_{GS} - V_{ch}(x) - V_{tn}] \tag{1.221}$$

Substituting (1.221) into (1.220) results in

$$\mu_n WC_{ox}[V_{GS} - V_{ch}(x) - V_{tn}]dV_{ch} = I_D\,dx \tag{1.222}$$

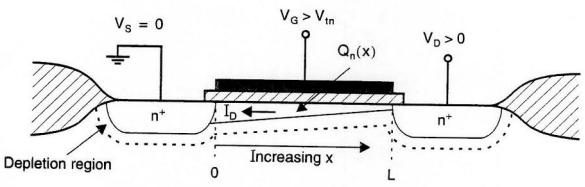


Fig. 1.37 The transistor definitions used in developing the transistor's I-V relationship.

Integrating both sides of (1.222), and noting that the total voltage along the channel of length L is V_{DS}, we obtain

$$\int_0^{V_{DS}} \mu_n W C_{ox}[V_{GS} - V_{ch}(x) - V_{tn}]\, dV_{ch} = \int_0^L I_D\, dx \tag{1.223}$$

which results in

$$\mu_n W C_{ox}\left[(V_{GS} - V_{tn})V_{DS} - \frac{V_{DS}^2}{2}\right] = I_D L \tag{1.224}$$

Thus, solving for I_D results in the well-known triode relationship for a MOS transistor:

$$I_D = \mu_n C_{ox}\left(\frac{W}{L}\right)\left[(V_{GS} - V_{tn})V_{DS} - \frac{V_{DS}^2}{2}\right] \tag{1.225}$$

It should be noted that taking into account the body effect along the channel, the triode model of (1.225) is modified to

$$I_D = \mu C_{ox}\left(\frac{W}{L}\right)\left[(V_{GS} - V_{tn})V_{DS} - \alpha\frac{V_{DS}^2}{2}\right] \tag{1.226}$$

where $\alpha \cong 1.7$ [Tsividis, 1987].

Bipolar Transistor Exponential Relationship

The various components of the base, collector, and emitter currents were shown in Fig. 1.28, on page 44. Figure 1.38 shows plots of the minority-carrier concentrations

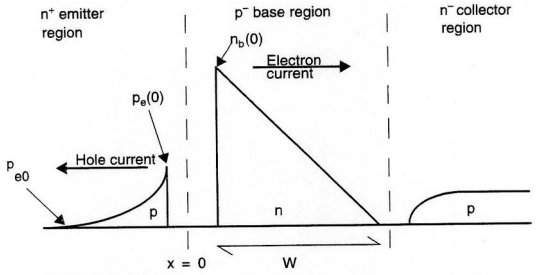


Fig. 1.38 The concentrations of minority carriers in the emitter, base, and collector.

in the emitter, base, and collector regions. The current flow of these minority carriers is due to diffusion. By calculating the gradient of the minority-carrier concentrations near the base-emitter junction in a manner similar to that used for diodes, it is possible to derive a relationship between the electron current and the hole current of Fig. 1.28.

The concentration of holes in the emitter at the edge of the base-emitter depletion region is denoted $p_e(0)$. This concentration decreases exponentially the farther one gets from the junction, in a manner similar to that described for diodes. The concentration far from the junction, p_e0, is given by

$$p_{e0} = \frac{n_i^2}{N_D} \tag{1.227}$$

where N_D is the doping density of the n^+ emitter. At a distance x from the edge of the emitter-base depletion region, we have

$$p_e(x) \cong p_e(0)e^{-x/L_p} \tag{1.228}$$

where

$$\begin{aligned} p_e(0) &= p_{e0}e^{V_{BE}/V_T} \\ &= \frac{n_i^2}{N_D}e^{V_{BE}/V_T} \end{aligned} \tag{1.229}$$

and where V_{BE} is the forward-bias voltage of the base-emitter junction.

At the edge of the base-emitter depletion region, the gradient of the hole concentration in the emitter is found, using (1.228), to be

$$\left.\frac{dp_e(x)}{dx}\right|_{x=0} = \frac{p_e(0)}{L_p} \tag{1.230}$$

Using (1.229), we can rewrite this as

$$\left.\frac{dp_e(0)}{dx}\right|_{x=0} = \frac{n_i^2}{L_p N_D}e^{V_{BE}/V_T} \tag{1.231}$$

The hole current is now found using the diffusion equation

$$I_{pe} = A_E q D_p \left.\frac{dp_e(x)}{dx}\right|_{x=0} \tag{1.232}$$

where A_E is the effective area of the emitter. Recall that the minority-hole current in the emitter, I_{pe}, is closely equal to the base current, I_B. After combining (1.231) and (1.232), we obtain

$$I_B = \frac{A_E q D_p n_i^2}{L_p N_D}e^{V_{BE}/V_T} \tag{1.233}$$

The situation on the base side of the base-emitter junction is somewhat different. The concentration of the minority carriers, in this case electrons that diffused from the emitter, is given by a similar equation,

$$n_b(0) = \frac{n_i^2}{N_A} e^{V_{BE}/V_T} \tag{1.234}$$

However, the gradient of this concentration at the edge of the base-emitter depletion region is calculated differently. This difference in gradient concentration is due to the close proximity of the collector-base junction, where the minority carrier (electron) concentration, $n_b(W)$, must be zero. This zero concentration at the collector-base junction occurs because any electrons diffusing to the edge of the collector-base depletion region immediately drift across the junction to the collector, as stated previously. If the base "width" is much shorter than the diffusion length of electrons in the base, L_n, then almost no electrons will recombine with base majority carriers (holes) before they diffuse to the collector-base junction. Given this fact, the decrease in electron or minority concentration from the base-emitter junction to the collector-base junction is a linear relationship decreasing from $n_b(0)$ at the emitter junction to zero at the collector junction in distance W. This assumption ignores any recombination of electrons in the base as they travel to the collector, which is a reasonable assumption for modern transistors that have very narrow bases. Thus, throughout the base region, the gradient of the minority-carrier concentration is closely given by

$$\begin{aligned} \frac{dn_b(x)}{dx} &= -\frac{n_b(0)}{W} \\ &= -\frac{n_i^2}{WN_A} e^{V_{BE}/V_T} \end{aligned} \tag{1.235}$$

Combining (1.235) with the diffusion equation, we obtain

$$\begin{aligned} I_{nb} &= -A_E q D_n \frac{dn_b(0)}{dx} \\ &= \frac{A_E q D_n n_i^2}{WN_A} e^{V_{BE}/V_T} \end{aligned} \tag{1.236}$$

Remembering that I_{nb} is closely equal to the collector current I_C, we have

$$I_C \cong I_{CS} e^{V_{BE}/V_T} \tag{1.237}$$

where

$$I_{CS} = \frac{A_E q D_n n_i^2}{WN_A} \tag{1.238}$$

The ratio of the collector current to the base current, commonly called the *transistor common-emitter current gain* and denoted β, is found using (1.237), (1.238), and

(1.233). We have

$$\beta = \frac{I_C}{I_B} = \frac{D_n N_D L_p}{D_p N_A W} \cong 2.5 \frac{N_D L_p}{N_A W} \tag{1.239}$$

which is a constant independent of voltage and current. Noting that $N_D >> N_A$, $L_p > W$, and $D_n \cong 2.5\, D_p$, we have $\beta >> 1$. A typical value might be between 50 and 200. The derivation of β just presented ignores many second-order effects that make β somewhat current and voltage dependent and are beyond the scope of this book. Interested readers should see Roulston, 1990, for more details. Regardless of second-order effects, equation (1.239) does reflect the approximate relationships among β, doping levels, and base width. For example, (1.239) explains why heavily doped emitters are important to achieve large current gain.

Base Charge Storage of an Active BJT

Figure 1.38 shows a minority-carrier storage in the base region, Q_b, given by

$$Q_b = A_E q \frac{n_b(0) W}{2} \tag{1.240}$$

Using (1.234) for $n_b(0)$, we have

$$Q_b = \frac{A_E q n_i^2 W}{2N_A} e^{V_{BE}/V_T} \tag{1.241}$$

This equation can be rewritten using (1.237) and (1.238) to obtain

$$Q_b = \frac{W^2}{2D_n} I_C = \tau_b I_C \tag{1.242}$$

where τ_b, called the *base-transit time constant,* is given approximately by

$$\tau_b = \frac{W^2}{2D_n} \tag{1.243}$$

ignoring second-order effects. Normally, the base-transit time constant is specified for a given technology and takes into account other charge-storage effects not considered here, and is therefore often denoted τ_T. However, since the base storage of electrons dominates the other effects, we have $\tau_T \cong \tau_b$.

If the current in a BJT changes, the base charge storage must also change. This change can be modelled by a diffusion capacitance, C_d, between the base and the emitter terminals. Using (1.242), we have

$$C_d \cong \frac{dQ_b}{dV_{BE}} = \frac{d(\tau_b I_C)}{dV_{BE}} \tag{1.244}$$

Using (1.244) and $I_C = I_{CS}e^{V_{BE}/V_T}$ results in

$$C_d = \tau_b \frac{I_C}{V_T} \tag{1.245}$$

This equation is similar to that for a diode.

1.8 REFERENCES

R. Geiger, P. Allen, and N. Strader. *VLSI: Design Techniques for Analog and Digital Circuits*. McGraw-Hill, New York, 1990.

P. Gray and R. G. Meyer. *Analog Integrated Circuits*, 3rd. ed. John Wiley & Sons, New York, 1993.

D. Hodges and H. Jackson. *Analysis and Design of Digital Integrated Circuits*, 2nd ed. McGraw-Hill, New York, 1988.

D. Roulston. *Semiconductor Devices*. McGraw-Hill, New York, 1990.

S. M. Sze. *Physics of Semiconductor Devices*. Wiley Interscience, New York, 1981.

Y. Tsividis. *Operation and Modeling of the MOS Transistor*. McGraw-Hill, New York, 1987.

S. Wolf. *Silicon Processing for the VLSI Era—Volume 3: The Submicron MOSFET*. Lattice Press, Sunset Beach, California, 1995.

1.9 PROBLEMS

Unless otherwise stated, assume the following hold throughout the problems section:

- Room temperature = 300 °K
- npn bipolar transistors:
  $\beta = 100$
  $V_A = 80$ V
  $\tau_b = 13$ ps
  $\tau_s = 4$ ns
  $r_b = 330\ \Omega$
- n-channel MOS transistors:
  $\mu_n C_{ox} = 92\ \mu A/V^2$
  $V_{tn} = 0.8$ V
  $\gamma = 0.5\ V^{1/2}$
  $r_{ds}\ (\Omega) = 8000L\ (\mu m)/I_D\ (mA)$ in active region
  $C_j = 2.4 \times 10^{-4}\ pF/(\mu m)^2$
  $C_{j\text{-}sw} = 2.0 \times 10^{-4}\ pF/\mu m$
  $C_{ox} = 1.9 \times 10^{-3}\ pF/(\mu m)^2$
  $C_{gs(overlap)} = C_{gd(overlap)} = 2.0 \times 10^{-4}\ pF/\mu m$
- p-channel MOS transistors:
  $\mu_p C_{ox} = 30\ \mu A/V^2$
  $V_{tp} = -0.9$ V
  $\gamma = 0.8\ V^{1/2}$
  $r_{ds}\ (\Omega) = 12{,}000\ L\ (\mu m)/I_D\ (mA)$ in active region

$C_j = 4.5 \times 10^{-4}\ pF/(\mu m)^2$
$C_{j\text{-}sw} = 2.5 \times 10^{-4}\ pF/\mu m$
$C_{ox} = 1.9 \times 10^{-3}\ pF/(\mu m)^2$
$C_{gs(overlap)} = C_{gd(overlap)} = 2.0 \times 10^{-4}\ pF/\mu m$

1.1 Estimate the hole and electron concentrations in silicon doped with arsenic at a concentration of 10^{25} atoms/m^3 at a temperature 22 °C above room temperature. Is the resulting material n type or p type?

1.2 For the pn junction of Example 1.2, does the built-in potential, Φ_0, increase or decrease when the temperature is increased 11 °C above room temperature?

1.3 Calculate the amount of charge per $(\mu m)^2$ in each of the n and p regions of the pn junction of Example 1.2 for a 5-V reverse-bias voltage. How much charge on each side would be present in a 10 $\mu m \times 10\ \mu m$ diode?

1.4 A silicon diode has $\tau_t = 12$ ps and $C_{j0} = 15$ fF. It is biased by a 43-kΩ resistor connected between the cathode of the diode and the input signal, as shown in Fig. P1.4. Initially, the input is 5 V, and then at time 0 it changes to 0 V. Estimate the time it takes for the output voltage to change from 5 V to 1.5 V (i.e., the $\Delta t_{-70\%}$ time). Repeat for an input voltage change from 0 V to 5 V and an output voltage change from 0 V to 3.5 V.

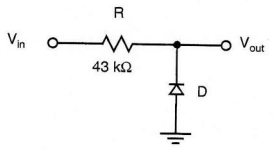


Fig. P1.4

1.5 Compare your answers for Problem 1.4 to those obtained using a SPICE simulation.

1.6 Verify that when $V_{DS} = V_{eff}$ is used in the triode equation for a MOS transistor, the current equals that of the active region equation given in (1.67).

1.7 Find I_D for an n-channel transistor having doping concentrations of $N_D = 10^{25}$ and $N_A = 10^{22}$ with $W = 50\ \mu m$, $L = 1.5\ \mu m$, $V_{GS} = 1.1$ V, and $V_{DS} = V_{eff}$. Assuming λ remains constant, estimate the new value of I_D if V_{DS} is increased by 0.3 V.

1.8 A MOS transistor in the active region is measured to have a drain current of 20 μA when $V_{DS} = V_{eff}$. When V_{DS} is increased by 0.5 V, I_D increases to 23 μA. Estimate the output impedance, r_{ds}, and the output impedance constant, λ.

1.9 Derive the low-frequency model parameters for an n-channel transistor having doping concentrations of $N_D = 10^{25}$ and $N_A = 10^{22}$ with $W = 10\ \mu m$, $L = 1.2\ \mu m$, $V_{GS} = 1.1$ V, and $V_{DS} = V_{eff}$. Assume that $V_{SB} = 1.0$ V.

1.10 Find the capacitances C_{gs}, C_{gd}, C_{db}, and C_{sb} for an active transistor having $W = 50\ \mu m$ and $L = 1.2\ \mu m$. Assume that the source and drain junctions extend $4\ \mu m$ beyond the gate, resulting in source and drain areas being $A_s = A_d = 200\ (\mu m)^2$ and the perimeter of each being $P_s = P_d = 58\ \mu m$.

1.11 Consider the circuit shown in Fig. P1.11, where V_{in} is a dc signal of 1 V. Taking into account only the channel charge storage, determine the final value of V_{out} when the transistor is turned off, assuming half the channel charge goes to C_L.

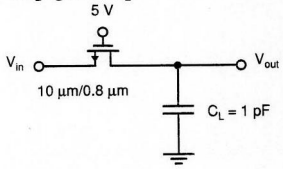


Fig. P1.11

1.12 For the same circuit as in Problem 1.11, the input voltage has a step voltage change at time 0 from 1 V to 1.2 V (the gate voltage remains at 5 V). Find its 99 percent settling time (the time it takes to settle to within 1 percent of the total voltage change). You may ignore the body effect and all capacitances except C_L. Also assume that $V_{tn} = V_{tn0}$. Repeat the question for V_{in} changing from 3 V to 3.1 V.

1.13 Repeat Problem 1.12, but now take into account the body effect on V_{tn}.

1.14 For an npn transistor having $I_C = 0.1$ mA, calculate g_m, r_π, r_e, r_0, and $g_m r_0$.

1.15 A bipolar junction transistor has the following SPICE parameters (the SPICE name for the parameter is included in parentheses):

I_S (IS) $= 2.0 \times 10^{-18}$
B_F (BF) $= 100$
B_R (BR) $= 1$
V_A (VA) $= 50$ V
τ_F (TF) $= 12 \times 10^{-12}$ s
τ_R (TR) $= 4 \times 10^{-9}$ s
C_{je0} (CJE) $= 15 \times 10^{-15}$ F
Φ_e (VJE) $= 0.9$ V
m_e (MJE) $= 0.27$
C_{jc0} (CJC) $= 18 \times 10^{-15}$ F
Φ_c (VJC) $= 0.7$ V
m_c (MJC) $= 0.37$
C_{js0} (CJS) $= 40 \times 10^{-15}$ F
Φ_s (VJS) $= 0.64$ V
m_s (MJS) $= 0.29$
R_e (RE) $= 30$

R_b (RB) $= 500$
R_c (RC) $= 90$

Initially, the circuit shown in Fig. P1.15 has a 0-V input. At time 0 its input changes to 5 V. Estimate the time it takes its output voltage to saturate, using the concepts of average capacitance and first-order transient solutions for each node. The time constants of the individual nodes can be added to arrive at an overall time constant for the approximate first-order transient response of the circuit. Next, assume that the input changes from 5 V to 0 V at time 0. How long does it take the output voltage to change to 3.5 V?

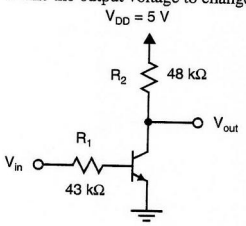


Fig. P1.15

1.16 Compare your answers to Problem 1.15 to those obtained using SPICE.

1.17 Verify that for $I_{BR} >> I_B$ and $I_{BR} >> I_C / \beta$, Eq. (1.140) simplifies to Eq. (1.139).