as discussed in Chapter 1. Amorphous Si:H has an E_g of about 1.8 eV. The alloying of a-Si:H with Ge to produce a-SiGe:H decreases E_g. Further, E_g of a-SiGe:H can be graded by controlling the Ge content.

6.11 BIPOLAR TRANSISTOR (BJT)

6.11.1 COMMON BASE (CB) DC CHARACTERISTICS

As an example, we will consider the *pnp* bipolar junction transistor (BJT) whose basic structure is shown in Figure 6.48a. The *pnp* transistor has three differently doped semiconductor regions. These regions of different doping occur within the same single crystal by the variation of acceptor and donor concentrations resulting from the fabrication process. The most heavily doped *p*-region (p^+) is called the **emitter.** In contact with this region is the lightly doped *n*-region, which is called the **base.** The next region is the *p*-type doped **collector.** The base region has the most narrow width for reasons discussed below. Although the three regions in Figure 6.48a have identical cross-sectional areas, in practice, due to the fabrication process, the cross-sectional area increases from the emitter to the collector and the collector region has an extended width. For simplicity, we will assume that the cross-sectional area is uniform, as in Figure 6.48a.

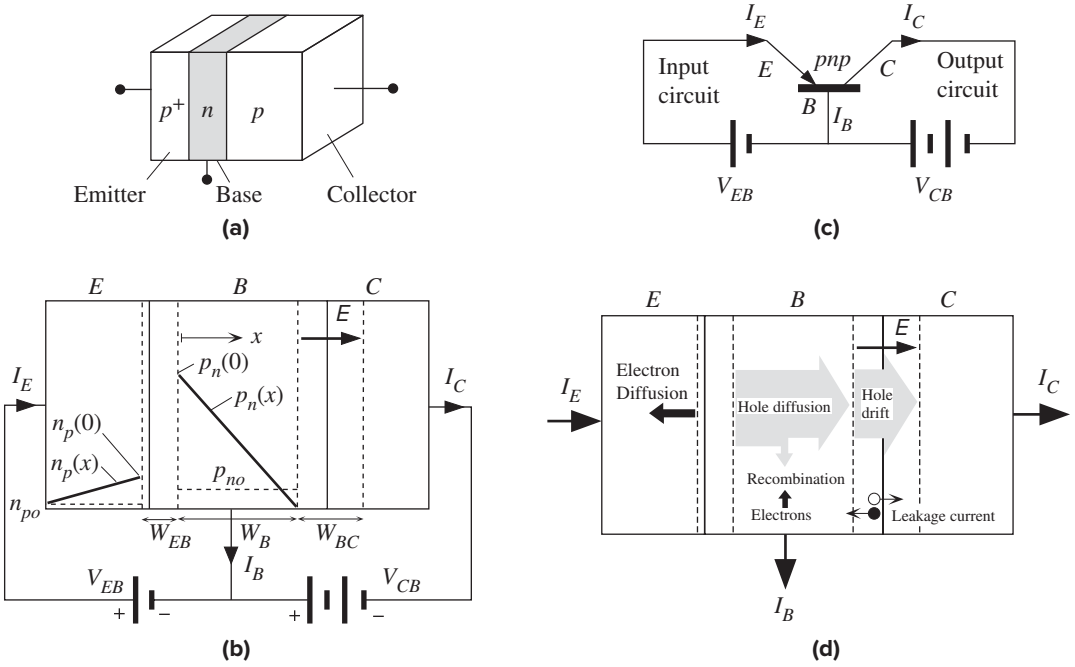
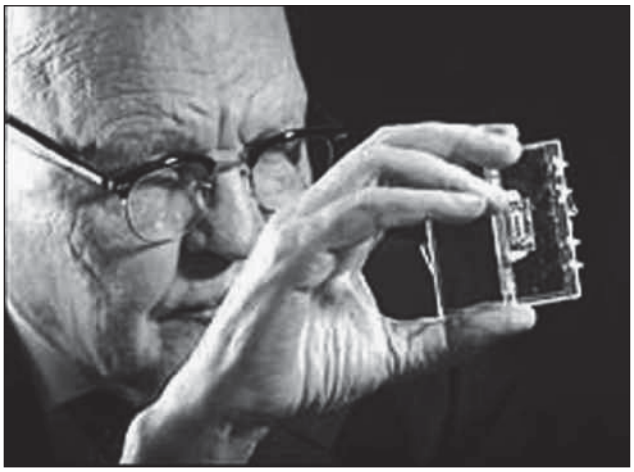
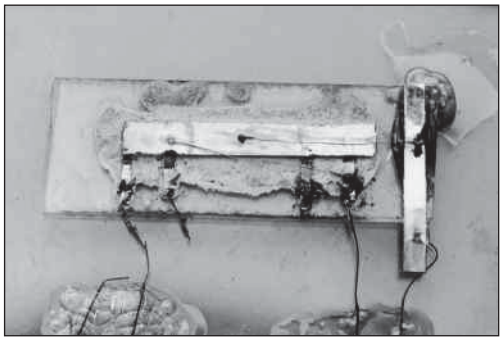


Figure 6.48 (a) A schematic illustration of the *pnp* bipolar transistor with three differently doped regions. (b) The *pnp* bipolar operated under normal and active conditions. (c) The CB configuration with input and output circuits identified. (d) The illustration of various current components under normal and active conditions.



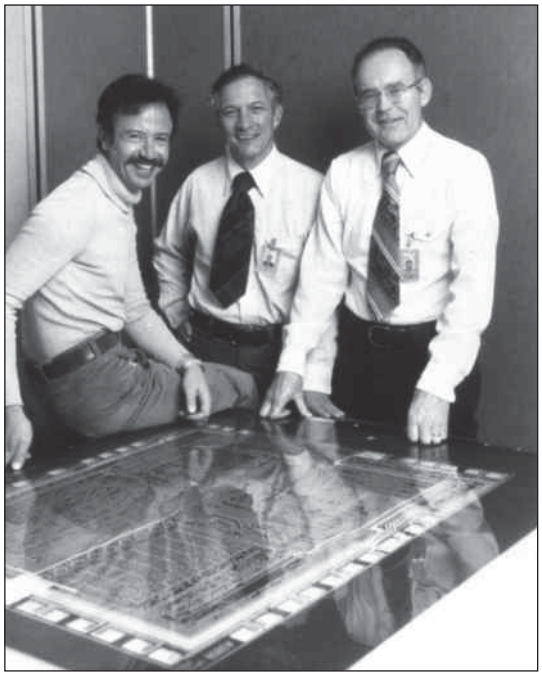


The first monolithic integrated circuit, about the size of a fingertip, was documented and developed at Texas Instruments by Jack Kilby in 1958; he won the 2000 Nobel prize in physics for his contribution to the development of the first integrated circuit. The IC was a chip of a single Ge crystal containing one transistor, one capacitor, and one resistor. Left: Jack Kilby holding his IC (photo, 1998). Right. The photo of the chip.

| Left: © AP Photo. Right: © Fotosearch/Getty Images.

This first commercial pocket transistor radio (Regency TR- 1) was released in 1954. It had 4 *npn* Ge transistors from Texas Instruments and was sold at $49.99, roughly $450 in today's dollars.

| © Bettmann/Getty Images.



Left to right: Andrew Grove (1936–2016), Robert Noyce (1927–1990), and Gordon Moore (born 1929), who founded Intel in 1968. Andrew Grove's book *Physics and Technology of Semiconductor Devices* (Wiley, 1967) was one of the classic texts on devices in the sixties and seventies. "Moore's law" that started as a rough rule in 1965 states that the number of transistors in a chip will double every 18 months; Moore updated it in 1995 to every couple of years.

| Courtesy of Intel Corp.

The *pnp* BJT connected as shown in Figure 6.48b is said to be operating under normal and active conditions, which means that the base–emitter (BE) junction is forward biased and the base–collector (BC) junction is reverse biased. The circuit in Figure 6.48b, in which the base is common to both the collector and emitter bias voltages, is known as the common base (CB) configuration.[15] Figure 6.48c shows the CB transistor circuit with the BJT represented by its circuit symbol. The arrow identifies the emitter junction and points in the direction of current flow when the EB junction is forward biased. Figure 6.48c also identifies the emitter circuit, where V_{EB} is connected, as the input circuit. The collector circuit, where V_{CB} is connected, is the output circuit.

The base–emitter junction is simply called the **emitter junction** and the base-collector junction is called the **collector junction.** As the emitter is heavily doped, the base–emitter depletion region W_{EB} extends almost entirely into the base. Generally, the base and collector regions have comparable doping, so the base–collector depletion region W_{BC} extends to both sides. The width of the neutral base region outside the depletion regions is labeled as W_B. All these parameters are shown and defined in Figure 6.48b.

We should note that all the applied voltages drop across the depletion widths. The applied collector–base voltage V_{CB} reverse biases the BC junction and hence increases the field in the depletion region at the collector junction.

Since the EB junction is forward biased, minority carriers are then injected into the emitter and base exactly as they are in the forward-biased diode. Holes are injected into the base and electrons into the emitter, as depicted in Figure 6.48d. Hole injection into the base, however, far exceeds the electron injection into the emitter because the emitter is heavily doped. We can then assume that the emitter current is almost entirely due to holes injected from the emitter into the base. Thus, when forward biased, the emitter "emits," that is, injects holes into the base.

Injected holes into the base must diffuse toward the collector junction because there is a hole concentration gradient in the base. Hole concentration $p_n(W_B)$ just outside the depletion region at the collector junction is negligibly small because the increased field sweeps nearly all the holes here across the junction into the collector (the collector junction is reverse biased).

The hole concentration $p_n(0)$ in the base just outside the emitter junction depletion region is given by the law of the junction. Measuring x from this point (Figure 6.48b),

$$p_n(0) = p_{no} \exp\left(\frac{eV_{EB}}{kT}\right) \qquad [6.55]$$

whereas at the collector end, $x = W_B$, $p_n(W_B) \approx 0$.

If no holes are lost by recombination in the base, then all the injected holes diffuse to the collector junction. There is no field in the base to drift the holes. Their motion is by diffusion. When they reach the collector junction, they are quickly swept across into the collector by the internal field E in W_{BC}. It is apparent that all the injected holes from the emitter become collected by the collector. The collector

[15] CB should not be confused with the conduction band abbreviation.

current is then the same as the emitter current. The only difference is that the emitter current flows across a smaller voltage difference V_{EB}, whereas the collector current flows through a larger voltage difference V_{CB}. This means a *net gain in power* from the emitter (input) circuit to the collector (output) circuit.

Since the current in the base is by diffusion, to evaluate the emitter and collector currents we must know the hole concentration gradient at $x = 0$ and $x = W_B$ and therefore we must know the hole concentration profile $p_n(x)$ across the base.[16] In the first instance, we can approximate the $p_n(x)$ profile in the base as a straight line from $p_n(0)$ to $p_n(W_B) = 0$, as shown in Figure 6.48b. This is only true in the absence of any recombination in the base as in the short diode case. The emitter current is then

$$I_E = -eAD_h\left(\frac{dp_n}{dx}\right)_{x=0} = eAD_h\frac{p_n(0)}{W_B}$$

We can substitute for $p_n(0)$ from Equation 6.55 to obtain

Emitter current

$$I_E = \frac{eAD_h p_{no}}{W_B}\exp\left(\frac{eV_{EB}}{kT}\right) \quad [6.56]$$

It is apparent that I_E is determined by V_{EB}, the forward bias applied across the EB junction, and the base width W_B. In the absence of recombination, the collector current is the same as the emitter current, $I_C = I_E$. The control of the collector current I_C in the output (collector) circuit by V_{EB} in the input (emitter) circuit is what constitutes the **transistor action.** The common base circuit has a **power gain** because I_C in the output in Figure 6.48c flows around a larger voltage difference V_{CB} compared with I_E in the input, which flows across V_{EB} (about 0.6 V).

The ratio of the collector current I_C to the emitter current I_E is defined as the **CB current gain** or **current transfer ratio** α of the transistor,

Definition of CB current gain

$$\alpha = \frac{I_C}{I_E} \quad [6.57]$$

Typically, α is less than unity, in the range 0.990–0.999, due to two reasons. First is the limitation due to the emitter injection efficiency. When the *BE* junction is forward biased, holes are injected from the emitter into the base, giving an emitter current $I_{E(\text{hole})}$, and electrons are injected from the base into the emitter, giving an emitter current $I_{E(\text{electron})}$. The total emitter current is, therefore,

Total emitter current

$$I_E = I_{E(\text{hole})} + I_{E(\text{electron})}$$

Only the holes injected into the base are useful in giving a collector current because only they can reach the collector. The emitter injection efficiency is defined as

Emitter injection efficiency

$$\gamma = \frac{I_{E(\text{hole})}}{I_{E(\text{hole})} + I_{E(\text{electron})}} = \frac{1}{1 + \dfrac{I_{E(\text{electron})}}{I_{E(\text{hole})}}} \quad [6.58]$$

[16] The actual concentration profile can be calculated by solving the steady-state continuity equation, which can be found in more advanced texts.

Consequently, the collector current, which depends on $I_{E(\text{hole})}$ only, is less than the emitter current. We would like γ to be as close to unity as possible; $I_{E(\text{hole})} \gg I_{E(\text{electron})}$. γ can be readily calculated for the forward-biased *pn* junction current equations as shown in Example 6.19.

Secondly, a small number of the diffusing holes in the narrow base inevitably become lost by recombination with the large number of electrons present in this region as depicted in Figure 6.48d. Thus, a fraction of $I_{E(\text{hole})}$ is lost in the base due to recombination, which further reduces the collector current. We define the **base transport factor** α_T as

Base transport factor

$$\alpha_T = \frac{I_C}{I_{E(\text{hole})}} = \frac{I_C}{\gamma I_E} \tag{6.59}$$

If the emitter were a perfect injector, $I_E = I_{E(\text{hole})}$, then the current gain α would be α_T. If τ_h is the hole (minority carrier) lifetime in the base, then $1/\tau_h$ is the probability per unit time that a hole will recombine and disappear. We also know that in time t, a particle diffuses a distance x, given by $x = \sqrt{2Dt}$ where D is the diffusion coefficient. The time τ_t it takes for a hole to diffuse across W_B is then given by

Base minority carrier transit time

$$\tau_t = \frac{W_B^2}{2D_h} \tag{6.60}$$

This diffusion time is called the **transit time** of the minority carriers across the base.

The probability of recombination in time τ_t is then τ_t/τ_h. The probability of not recombining and therefore diffusing across is $(1 - \tau_t/\tau_h)$. Since $I_{E(\text{hole})}$ represents the holes entering the base per unit time, $I_{E(\text{hole})}(1 - \tau_t/\tau_h)$ represents the number of holes leaving the base per unit time (without recombining) which is the collector current I_C. Substituting for I_C and $I_{E(\text{hole})}$ in Equation 6.59 gives the base transport factor α_T,

Base transport factor

$$\alpha_T = \frac{I_C}{I_{E(\text{hole})}} = 1 - \frac{\tau_t}{\tau_h} \tag{6.61}$$

Using Equations 6.57, 6.59, and 6.61 we can find the total **CB current gain** α:

CB current gain

$$\alpha = \alpha_T \gamma = \left(1 - \frac{\tau_t}{\tau_h}\right)\gamma \tag{6.62}$$

The recombination of holes with electrons in the base means that the base must be replenished with electrons, which are supplied by the external battery in the form of a small base current I_B, as shown in Figure 6.48d. In addition, the base current also has to supply the electrons injected from the base into the emitter, that is, $I_{E(\text{electron})}$, and shown as electron diffusion in the emitter in Figure 6.48d. The number of holes entering the base per unit time is represented by $I_{E(\text{hole})}$, and the number recombining per unit time is then $I_{E(\text{hole})}(\tau_t/\tau_h)$. Thus, I_B is

Base current

$$I_B = \left(\frac{\tau_t}{\tau_h}\right) I_{E(\text{hole})} + I_{E(\text{electron})} = \gamma \frac{\tau_t}{\tau_h} I_E + (1 - \gamma) I_E \tag{6.63}$$

which further simplifies to $I_E - I_C$; the difference between the emitter current and the collector current is the base current. (This is exactly what we expect from Kirchoff's current law.)

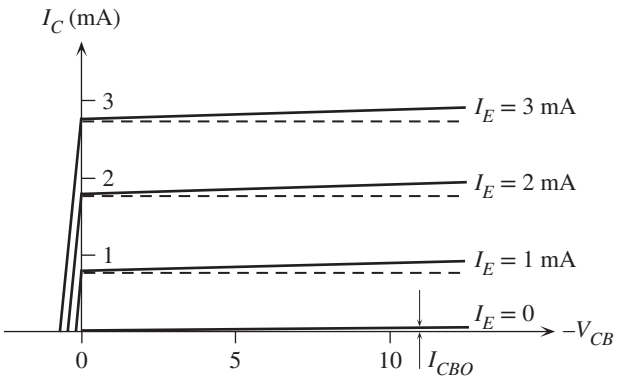


Figure 6.49 DC *I–V* characteristics of the *pnp* bipolar transistor (exaggerated to highlight various effects).

The ratio of the collector current to the base current is defined as the **current gain** β of the transistor.[17] By using Equations 6.57, 6.62, and 6.63, we can relate β to α:

Base-to-collector current gain

$$\beta = \frac{I_C}{I_B} = \frac{\alpha}{1-\alpha} \approx \frac{\gamma\tau_h}{\tau_t} \qquad [6.64]$$

The base–collector junction in Figure 6.48b is reverse biased, which leads to a leakage current into the collector terminal even in the absence of an emitter current. This leakage current is due to thermally generated EHPs in the depletion region W_{BC} being drifted by the internal field, as schematically illustrated in Figure 6.48d. Suppose that we open circuit the emitter ($I_E = 0$). Then the collector current is simply the leakage current, denoted by I_{CBO}. The base current is then $-I_{CBO}$ (flowing out from the base terminal). In the presence of an emitter current I_E, we have

Active region collector current

$$I_C = \alpha I_E + I_{CBO} \qquad [6.65]$$

Active region base current

$$I_B = (1-\alpha)I_E - I_{CBO} \qquad [6.66]$$

Equations 6.65 and 6.66 give the collector and base currents in terms of the input current I_E, which in turn depends on V_{EB}. They only hold when the collector junction is reverse biased and the emitter junction is forward biased, which is defined as the **active region** of the BJT. It should be emphasized that what constitutes the **transistor action** is the control of I_E, and hence I_C, by V_{EB}.

The dc characteristics of the CB-connected BJT as in Figure 6.48b are normally represented by plotting the collector current I_C as a function of V_{CB} for various fixed values of the emitter current. A typical example of such dc characteristics for a *pnp* transistor is illustrated in Figure 6.49. The following characteristics are apparent. The collector current when $I_E = 0$ is the CB junction leakage current I_{CBO}, typically a fraction of a microampere. As long as the collector is negatively biased with respect to the base, the CB junction is reverse biased and the collector current is given by

[17] β is a useful parameter when the transistor is used in what is called the common emitter (CE) configuration, in which the input current is made to flow into the base of the transistor, and the collector current is made to flow in the output circuit.

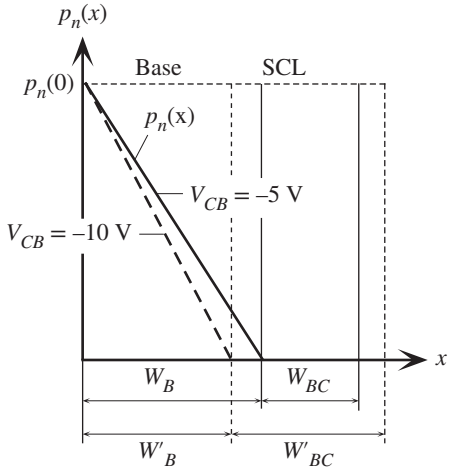


Figure 6.50 The Early effect.
When the BC reverse bias increases, the depletion width W_{BC} increases to W'_{BC}, which reduces the base width W_B to W'_B. As $p_n(0)$ is constant (constant V_{EB}), the minority carrier concentration gradient becomes steeper and the collector current, I_C increases.

$I_C = \alpha I_E + I_{CBO}$, which is close to the emitter current when $I_E \gg I_{CBO}$. When the polarity of V_{CB} is changed, the CB junction becomes forward biased. The collector junction is then like a forward-biased diode and the collector current is the difference between the forward-biased CB junction current and the forward-biased EB junction current. As they are in opposite directions, they subtract.

We note that I_C increases slightly with the magnitude of V_{CB} even when I_E is constant. In our treatment above I_C did not directly depend on V_{CB}, which simply reverse biased the collector junction to collect the diffusing holes. In our discussions we assumed that the base width W_B does not depend on V_{CB}. This is only approximately true. Suppose that we increase the reverse bias V_{CB} (for example, from -5 to -10 V). Then the base–collector depletion width W_{BC} also increases, as schematically depicted in Figure 6.50. Consequently the base width W_B gets slightly narrower, which leads to a slightly shorter base transit time τ_t. The base transport factor α_T in Equation 6.61 and hence α are then slightly larger, which leads to a small increase in I_C. The modulation of the base width W_B by V_{CB} is not very strong, which means that the slopes of the I_C versus V_{CB} lines at a fixed I_E are very small in Figure 6.49. The base width modulation by V_{CB} is called the **Early effect.**

EXAMPLE 6.18

A *pnp* TRANSISTOR Consider a *pnp* Si BJT that has the following properties. The emitter region mean acceptor doping is 2×10^{18} cm^{-3}, the base region mean donor doping is 1×10^{16} cm^{-3}, and the collector region mean acceptor doping is 1×10^{16} cm^{-3}. The hole drift mobility in the base is 400 cm^2 V^{-1} s^{-1}, and the electron drift mobility in the emitter is 200 cm^2 V^{-1} s^{-1}. The transistor emitter and base neutral region widths are about 2 μm each when the transistor is under normal operating conditions, that is, when the EB junction is forward biased and the BC junction is reverse biased. The effective cross-sectional area of the device is 0.02 mm^2. The hole lifetime in the base is approximately 400 ns. Assume that the emitter has 100 percent injection efficiency, $\gamma = 1$. Calculate the CB current transfer ratio α and the current gain β. What is the emitter–base voltage if the emitter current is 1 mA?

SOLUTION

The hole drift mobility $\mu_h = 400\ \text{cm}^2\ \text{V}^{-1}\ \text{s}^{-1}$ (minority carriers in the base). From the Einstein relationship we can easily find the diffusion coefficient of holes,

$$D_h = \left(\frac{kT}{e}\right)\mu_h = (0.02585\ \text{V})(400\ \text{cm}^2\ \text{V}^{-1}\ \text{s}^{-1}) = 10.34\ \text{cm}^2\ \text{s}^{-1}$$

The minority carrier transit time τ_t across the base is

$$\tau_t = \frac{W_B^2}{2D_h} = \frac{(2 \times 10^{-4}\ \text{cm})^2}{2(10.34\ \text{cm}^2\ \text{s}^{-1})} = 1.93 \times 10^{-9}\ \text{s} \qquad \text{or} \qquad 1.93\ \text{ns}$$

The base transport factor and hence the CB current gain is

$$\alpha = \gamma\alpha_T = 1 - \frac{\tau_t}{\tau_h} = 1 - \frac{1.93 \times 10^{-9}\ \text{s}}{400 \times 10^{-9}\ \text{s}} = 0.99517$$

The current gain β of the transistor is

$$\beta = \frac{\alpha}{1-\alpha} = \frac{0.99517}{1 - 0.99517} = 206.2$$

The emitter current is due to holes diffusing in the base ($\gamma = 1$),

$$I_E = I_{EO}\exp\left(\frac{eV_{EB}}{kT}\right)$$

where

$$\begin{aligned} I_{EO} &= \frac{eAD_hP_{no}}{W_B} = \frac{eAD_hn_i^2}{N_dW_B} \\ &= \frac{(1.6 \times 10^{-19}\ \text{C})(0.02 \times 10^{-2}\ \text{cm}^2)(10.34\ \text{cm s}^{-1})(1.0 \times 10^{10}\ \text{cm}^{-3})^2}{(1 \times 10^{16}\ \text{cm}^{-3})(2 \times 10^{-4}\ \text{cm})} \\ &= 1.66 \times 10^{-14}\ \text{A} \end{aligned}$$

Thus,

$$V_{EB} = \frac{kT}{e}\ln\left(\frac{I_E}{I_{EO}}\right) = (0.02585\ \text{V})\ \ln\left(\frac{1 \times 10^{-3}\ \text{A}}{1.66 \times 10^{-14}\ \text{A}}\right) = 0.64\ \text{V}$$

The major assumption is $\gamma = 1$, which is generally not true, as shown in Example 6.19. The actual α and hence β will be smaller due to less than 100 percent emitter injection. Note also that W_B is the *neutral region width*, that is, the region of base outside the depletion regions. It is not difficult to calculate the depletion layer widths within the base, which are about 0.2 μm on the emitter side and roughly about 0.7 μm on the collector side, so that the total base width junction to junction is 2 + 0.2 + 0.7 = 2.9 μm.

The transit time of minority carriers across the base is τ_t. If the input signal changes before the minority carriers have diffused across the base, then the collector current cannot respond to the changes in the input. Thus, if the frequency of the input signal is greater than $1/\tau_t$, the minority carriers will not have time to transit the base and the collector current will remain unmodulated by the input signal. One can set the upper frequency limit at $\sim 1/\tau_t$ which is 518 MHz.

EXAMPLE 6.19 EMITTER INJECTION EFFICIENCY γ

a. Consider a *pnp* transistor with the parameters as defined in Figure 6.48. Show that the **injection efficiency of the emitter,** defined as

$$\gamma = \frac{\text{Emitter current due to minority carriers injected into the base}}{\text{Total emitter current}}$$

is given by

$$\gamma = \frac{1}{1 + \dfrac{N_d W_B \mu_{e(\text{emitter})}}{N_a W_E \mu_{h(\text{base})}}}$$

b. How would you modify the CB current gain α to include the emitter injection efficiency?

c. Calculate the emitter injection efficiency for the *pnp* transistor in Example 6.18, which has an acceptor doping of 2×10^{18} cm^{-3} in the emitter, donor doping of 1×10^{16} cm^{-3} in the base, emitter and base neutral region widths of 2 μm, and a minority carrier lifetime of 400 ns in the base. What are its α and β taking into account the emitter injection efficiency?

SOLUTION

When the BE junction is forward biased, holes are injected into the base, giving an emitter current $I_{E(\text{hole})}$, and electrons are injected into the emitter, giving an emitter current $I_{E(\text{electron})}$. The total emitter current is therefore

$$I_E = I_{E(\text{hole})} + I_{E(\text{electron})}$$

Only the holes injected into the base are useful in giving a collector current because only they can reach the collector. Injection efficiency is defined as

Emitter injection efficiency definition

$$\gamma = \frac{I_{E(\text{hole})}}{I_{E(\text{hole})} + I_{E(\text{electron})}} = \frac{1}{1 + \dfrac{I_{E(\text{electron})}}{I_{E(\text{hole})}}}$$

But, provided that W_E and W_B are shorter than minority carrier diffusion lengths,

$$I_{E(\text{hole})} = \frac{eAD_{h(\text{base})} n_i^2}{N_d W_B} \exp\left(\frac{eV_{EB}}{kT}\right) \quad \text{and} \quad I_{E(\text{electron})} = \frac{eAD_{e(\text{emitter})} n_i^2}{N_a W_E} \exp\left(\frac{eV_{EB}}{kT}\right)$$

When we substitute into the definition of γ and use $D = \mu kT/e$, we obtain

Emitter injection efficiency

$$\gamma = \frac{1}{1 + \dfrac{N_d W_B \mu_{e(\text{emitter})}}{N_a W_E \mu_{h(\text{base})}}}$$

Emitter-to-collector current transfer ratio

The hole component of the emitter current is given as γI_E. Of this, a fraction $\alpha_T = (1 - \tau_t/\tau_h)$ will give a collector current. Thus, the emitter-to-collector current transfer ratio α, taking into account the emitter injection efficiency, is

$$\alpha = \gamma\left(1 - \frac{\tau_t}{\tau_h}\right)$$

In the emitter, $N_{a(\text{emitter})} = 2 \times 10^{18}$ cm^{-3} and $\mu_{e(\text{emitter})} = 200$ cm^2 V^{-1} s^{-1}, and in the base, $N_{d(\text{base})} = 1 \times 10^{16}$ cm^{-3} and $\mu_{h(\text{base})} = 400$ cm^2 V^{-1} s^{-1}. The emitter injection efficiency is

$$\gamma = \frac{1}{1 + \dfrac{(1 \times 10^{16})(2)(200)}{(2 \times 10^{18})(2)(400)}} = 0.99751$$

The transit time $\tau_t = W_B^2/2D_h = 1.93 \times 10^{-9}$ s (as before), so the overall α is

$$\alpha = 0.99751\left(1 - \frac{1.93 \times 10^{-9}}{400 \times 10^{-9}}\right) = 0.99269$$

and the overall β is

$$\beta = \frac{\alpha}{(1 - \alpha)} = 135.8$$

The same transistor with 100 percent emitter injection in Example 6.18 had a β of 206. It is clear that the emitter injection efficiency γ and the base transport factor α_T have comparable impacts in controlling the overall gain in this example. We neglected the recombination of electrons and holes in the EB depletion region. In fact, if we were to also consider this recombination component of the emitter current, $I_{E(\text{hole})}$ would have to be even smaller compared with the total I_E, which would make γ and hence β even lower.

6.11.2 COMMON BASE AMPLIFIER

According to Equation 6.56 the emitter current depends exponentially on V_{EB},

$$I_E = I_{EO} \exp\left(\frac{eV_{EB}}{kT}\right) \qquad [6.67]$$

It is therefore apparent that small changes in V_{EB} lead to large changes in I_E. Since $I_C \approx I_E$, we see that small variations in V_{EB} cause large changes in I_C in the collector circuit. This can be fruitfully used to obtain voltage amplification as shown in Figure 6.51. The battery V_{CC}, through R_C, provides a reverse bias for the base–collector junction. The dc voltage V_{EE} forward biases the EB junction, which means that it provides a dc current I_E. The input signal is the ac voltage v_{eb} applied in series with the dc bias voltage V_{EE} to the EB junction. The applied signal v_{eb} modulates the total voltage V_{EB} across the EB junction and hence, by virtue of Equation 6.55, modulates the injected hole concentration $p_n(0)$ up and down about the dc value determined by V_{EE} as depicted in Figure 6.51. This variation in $p_n(0)$ alters the concentration gradient and therefore gives rise to a change in I_E, and hence a nearly identical change in I_C. The change in the collector current can be converted to a voltage change by using a resistor R_C in the collector circuit as shown in Figure 6.51. However, the output is commonly taken between the collector, and the base and this voltage V_{CB} is

$$V_{CB} = -V_{CC} + R_C I_C$$

Increasing the emitter–base voltage V_{EB} (by increasing v_{eb}) increases I_C, which increases V_{CB}. Since we are interested in ac signals, that voltage variation across CB is tapped out through a dc blocking capacitor in Figure 6.51.

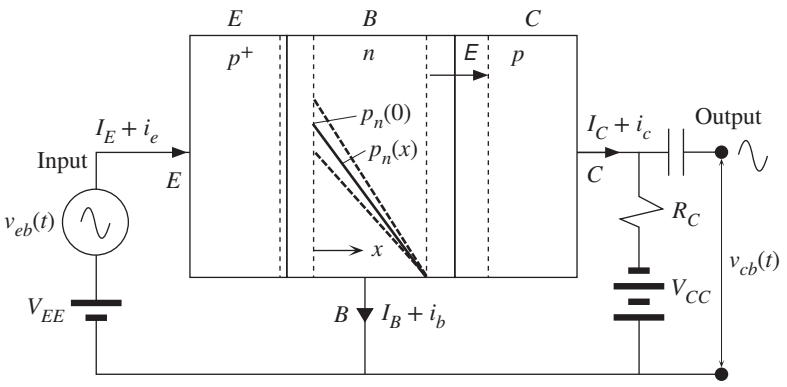


Figure 6.51 A *pnp* transistor operated in the active region in the common base amplifier configuration.

The applied (input) signal v_{eb} modulates the dc voltage across the EB junction and hence modulates the injected hole concentration up and down about the dc value $p_n(0)$. The solid line shows $p_n(x)$ when only the dc bias V_{EE} is present. The dashed lines show how $p_n(x)$ is modulated up and down by the signal v_{eb} superimposed on V_{EE}.

For simplicity we will assume that changes δV_{EB} and δI_E in the dc values of V_{EB} and I_E are small, which means that δV_{EB} and δI_E can be related by differentiating Equation 6.67. We are hence tacitly assuming an operation under small signals. Further, we will take the changes to represent the ac signal magnitudes, $v_{eb} = \delta V_{EB}$, $i_e = \delta I_E$, $i_c = \delta I_C \approx \delta I_E \approx i_e$, $v_{cb} = \delta V_{CB}$.

The output signal voltage v_{cb} corresponds to the change in V_{CB},

$$v_{cb} = \delta V_{CB} = R_C\, \delta I_C = R_C\, \delta I_E$$

The variation in the emitter current δI_E depends on the variation δV_{EB} in V_{EB}, which can be determined by differentiating Equation 6.67,

$$\frac{\delta I_E}{\delta V_{EB}} = \frac{e}{kT} I_E$$

By definition, δV_{EB} is the input signal v_{eb}. The change δI_E in I_E is the input signal current (i_e) flowing into the emitter as a result of δV_{EB}. Therefore, the quantity $\delta V_{EB}/\delta I_E$ represents an ac input resistance r_e seen by the source v_{eb}.

Small signal input resistance

$$r_e = \frac{\delta V_{EB}}{\delta I_E} = \frac{kT}{eI_E} = \frac{25}{I_E(\text{mA})} \qquad [6.68]$$

The output signal is then

$$v_{cb} = R_C\, \delta I_E = R_C \frac{v_{eb}}{r_e}$$

so the voltage amplification is

CB voltage gain

$$A_V = \frac{v_{cb}}{v_{eb}} = \frac{R_C}{r_e} \qquad [6.69]$$

To obtain a voltage gain we obviously need $R_C > r_e$, which is invariably the case by the appropriate choice of I_E, hence r_e, and R_C. For example, when the BJT is biased so that I_E is 10 mA and r_e is 2.5 Ω, and if R_C is chosen to be 50 Ω, then the gain is 20.

EXAMPLE 6.20

A COMMON BASE AMPLIFIER Consider a *pnp* Si BJT that has been connected as in Figure 6.51. The BJT has a $\beta = 135$ and has been biased to operate with a 10 mA collector current. What is the small-signal input resistance? What is the required R_C that will provide a voltage gain of 100? What is the base current? What should be the V_{CC} in Figure 6.51? Suppose $V_{CC} = -6$ V, what is the largest swing in the output voltage V_{CB} in Figure 6.51 as the input signal is increased and decreased about the bias point V_{EE}, taken as 0.65 V?

SOLUTION

The emitter and collector currents are approximately the same. From Equation 6.68,

$$r_e = \frac{25}{I_E\,(\text{mA})} = \frac{25}{10} = 2.5\ \Omega$$

The voltage gain A_V from Equation 6.69 is

$$A_V = \frac{R_C}{r_e} \qquad \text{or} \qquad 100 = \frac{R_C}{2.5\ \Omega}$$

so a gain of 100 requires $R_C = 250\ \Omega$.

$$\text{Base current } I_B = \frac{I_C}{\beta} = \frac{10\text{ mA}}{135} = 0.074\text{ mA} \qquad \text{or} \qquad 74\ \mu\text{A}$$

There is a dc voltage across R_C given by $I_C R_C = (0.010\text{ A})(250\ \Omega) = 2.5$ V. V_{CC} has to provide the latter voltage across R_C and also a sufficient voltage to keep the BC junction reverse biased at all times under normal operation. Let us set $V_{CC} = -6$ V. Thus, in the absence of any input signal v_{eb}, V_{CB} is set to -6 V + 2.5 V = -3.5 V. As we increase the signal v_{eb}, V_{EB} and hence I_C increase until the collector point C becomes nearly zero,[18] that is, $V_{CB} = 0$, which occurs when I_C is maximum at $I_{C\max} = |V_{CC}|/R_C$ or 24 mA. As v_{eb} decreases, so does V_{EB} and hence I_C. Eventually I_C will simply become zero, and point C will be at -6 V, so $V_{CB} = V_{CC}$. Thus, V_{CB} can only swing from -3.5 V to 0 V (for increasing input until $I_C = I_{C\max}$), or from -3.5 to -6 V (for decreasing input until $I_C = 0$).

6.11.3 COMMON EMITTER (CE) DC CHARACTERISTICS

An *npn* bipolar transistor when connected in the common emitter (CE) configuration has the emitter common to both the input and output circuits, as shown in Figure 6.52a. The dc voltage V_{BE} forward biases the BE junction and thereby injects electrons as minority carriers into the base. These electrons diffuse to the collector junction where the field E sweeps them into the collector to constitute the collector current I_C. V_{BE} controls the current I_E and hence I_B and I_C. The advantage of the CE configuration is that the **input current** is the current flowing between the ac source and the base, which is the base current I_B. This current is much smaller than the

[18] Various saturation effects are ignored in this approximate discussion.

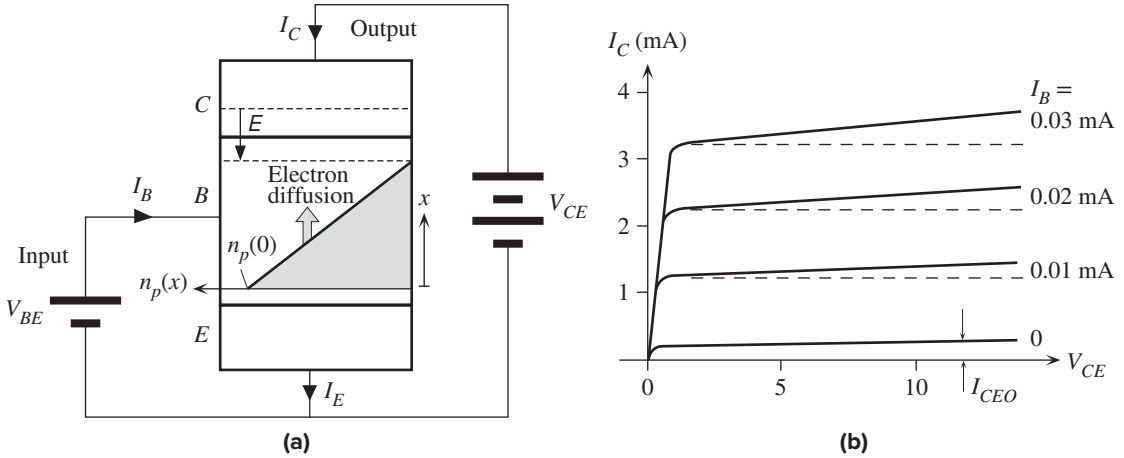


Figure 6.52 (a) An *npn* transistor operated in the active region in the common emitter configuration. The input current is the current that flows between V_{BE} and the base which is I_B. (b) DC I–V characteristics of the *npn* bipolar transistor in the CE configuration. (Exaggerated to highlight various effects.)

emitter current by about a factor of β. The output current is the current flowing between V_{CE} and the collector, which is I_C. In the CE configuration, the dc voltage V_{CE} must be greater than V_{BE} to reverse bias the collector junction and collect the diffusing electrons in the base.

The dc characteristics of the BJT in the CE configuration are normally given as I_C versus V_{CE} for various values of fixed base currents I_B, as shown in Figure 6.52b. The characteristics can be readily understood by Equations 6.65 and 6.66. We should note that, in practice, we are essentially adjusting V_{BE} to obtain the desired I_B because, by Equation 6.66,

$$I_B = (1 - \alpha)I_E - I_{CBO}$$

and I_E depends on V_{BE} via Equation 6.67.

Increasing I_B requires increasing V_{BE}, which increases I_C. Using Equations 6.65 and 6.66, we can obtain I_C in terms of I_B alone,

$$I_C = \beta I_B + \frac{1}{(1 - \alpha)} I_{CBO}$$

Active region collector current

or

$$I_C = \beta I_B + I_{CEO} \qquad [6.70]$$

where

$$I_{CEO} = \frac{I_{CBO}}{(1 - \alpha)} \approx \beta I_{CBO}$$

is the leakage current into the collector when the base is open circuited. This is much larger in the CE circuit than in the CB configuration.

Even when I_B is kept constant, I_C still exhibits a small increase with V_{CE}, which, according to Equation 6.70 indicates an increase in the current gain β with V_{CE}. This

is due to the Early effect or modulation of the base width by V_{CB}, shown in Figure 6.50. Increasing V_{CE} increases V_{CB}, which increases W_{BC}, reduces W_B, and hence shortens τ_t. The resulting effect is a larger β ($\approx \tau_h/\tau_t$).

When V_{CE} is less than V_{BE}, the collector junction becomes forward biased and Equation 6.70 is not valid. The collector current is then the difference between forward currents of emitter and collector junctions. The transistor operating in this region is said to be **saturated.**

6.11.4 LOW-FREQUENCY SMALL-SIGNAL MODEL

The *npn* bipolar transistor in the CE (common emitter) amplifier configuration is shown in Figure 6.53. The input circuit has a dc bias V_{BB} to forward bias the base–emitter (BE) junction and the output circuit has a dc voltage V_{CC} (larger than V_{BB}) to reverse bias the base–collector (BC) junction through a collector resistor R_C. The actual reverse bias voltage across the BC junction is $V_{CE} - V_{BE}$, where V_{CE} is

$$V_{CE} = V_{CC} - I_C R_C$$

An input signal in the form of a small ac signal v_{be} is applied in series with the bias voltage V_{BB} and modulates the voltage V_{BE} across the BE junction about its dc value V_{BB}. The varying voltage across the BE modulates $n_p(0)$ up and down about its dc value, which leads to a varying emitter current and hence to an almost identically varying collector current in the output circuit. The variation in the collector current is converted to an output voltage signal by the collector resistance R_C. Note that increasing V_{BE} increases I_C, which leads to a decrease in V_{CE}. Thus, the output voltage is 180° out of phase with the input voltage.

Since the BE junction is forward biased, the relationship between I_E and V_{BE} is exponential,

Emitter current and V_{BE}

$$I_E = I_{EO} \exp\left(\frac{eV_{BE}}{kT}\right) \qquad [6.71]$$

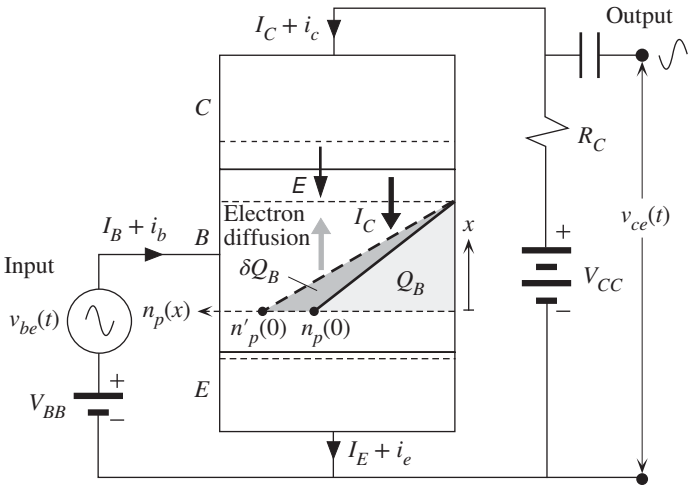


Figure 6.53 An *npn* transistor operated in the active region in the common emitter amplifier configuration.

The applied signal v_{be} modulates the dc voltage across the BE junction and hence modulates the injected electron concentration up and down about the dc value $n_p(0)$. The solid line shows $n_p(x)$ when only the dc bias V_{BB} is present. The dashed line shows how $n_p(x)$ is modulated up by a positive small signal v_{be} superimposed on V_{BB}.

where I_{EO} is a constant. We can differentiate this expression to relate small variations in I_E and V_{BE} as in the presence of small signals superimposed on dc values. For small signals, we have $v_{be} = \delta V_{BE}$, $i_b = \delta I_B$, $i_e = \delta I_E$, $i_c = \delta I_C$. Then from Equation 6.70 we see that $\delta I_C = \beta\ \delta I_B$, so $i_c = \beta i_b$. Since $\alpha \approx 1$, $i_e \approx i_c$.

What is the advantage of the CE circuit over the common base (CB) configuration? First, the input current is the base current, which is about a factor of β smaller than the emitter current. The ac input resistance of the CE circuit is therefore a factor of β higher than that of the CB circuit. This means that the amplifier does not load the ac source; the input resistance of the amplifier is much greater than the internal (or output) resistance of the ac source at the input. The small-signal input resistance r_{be} is

CE input resistance

$$r_{be} = \frac{v_{be}}{i_b} = \frac{\delta V_{BE}}{\delta I_B} \approx \beta \frac{\delta V_{BE}}{\delta I_E} = \frac{\beta kT}{eI_E} \approx \frac{\beta 25}{I_C(\text{mA})} \qquad [6.72]$$

where we differentiated Equation 6.71.

The output ac signal v_{ce} develops across the CE and is tapped out through a capacitor. Since $V_{CE} = V_{CC} - I_C R_C$, as I_C increases, V_{CE} decreases. Thus,

$$v_{ce} = \delta V_{CE} = -R_C\ \delta I_C = -R_C i_c$$

The voltage amplification is

CE voltage gain

$$A_V = \frac{v_{ce}}{v_{be}} = \frac{-R_C i_c}{r_{be} i_b} = \frac{-R_C \beta}{r_{be}} \approx -\frac{R_C I_C(\text{mA})}{25} \qquad [6.73]$$

which is the same as that in the CB configuration. However, in the CE configuration the output to input current ratio $i_c/i_b = \beta$, whereas this is almost unity in the CB configuration. Consequently, the CE configuration provides a greater power amplification, which is the second advantage of the CE circuit.

The input signal v_{be} gives rise to an output current i_c. This input voltage to output current conversion is defined in a parameter called the **mutual conductance,** or **transconductance,** g_m.

Transconductance

$$g_m = \frac{i_c}{v_{be}} \approx \frac{\delta I_E}{\delta V_{BE}} = \frac{I_E(\text{mA})}{25} = \frac{1}{r_e} \qquad [6.74]$$

The voltage amplification of the CE amplifier is then

Voltage gain

$$A_V = -g_m R_C \qquad [6.75]$$

We generally find it convenient to use a small-signal equivalent circuit for the low-frequency behavior of a BJT in the CE configuration. Between the base and emitter, the applied ac source voltage v_s sees only an input resistance of r_{be}, as shown in Figure 6.54. To underline the importance of the transistor input resistance, the output (or the internal) resistance R_s of the ac source is also shown. In the output circuit there is a voltage-controlled current source i_c which generates a current of $g_m v_{be}$. The current i_c passes through the load (or collector) resistance R_C across which the voltage signal develops. As we are only interested in ac signals, the batteries are taken as a short-circuit path for the ac current, which means that the internal

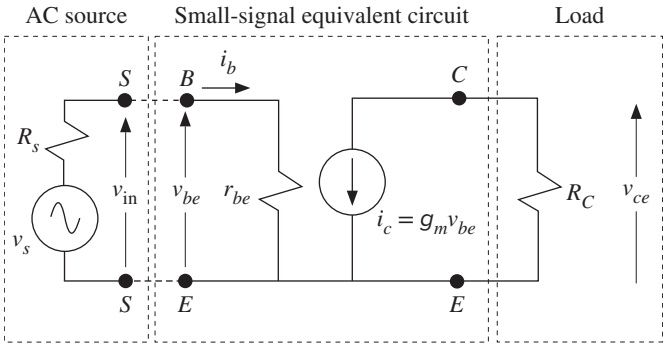


Figure 6.54 Low-frequency small-signal simplified equivalent circuit of the bipolar transistor in the CE configuration with a load resistor R_C in the collector circuit.

resistances of the batteries are taken as zero. This model, of course, is valid only under normal and active operating conditions and small signals about dc values, and at low frequencies.

The bipolar transistor general dc current equation $I_C = \beta I_B$, where $\beta \approx \tau_h/\tau_t$ is a material-dependent constant, implies that the ac small-signal collector current is

$$\delta I_C = \beta \delta I_B \qquad \text{or} \qquad i_c = \beta i_b$$

Thus the CE dc and ac small-signal current gains are the same. This is a reasonable approximation in the low-frequency range, typically at frequencies below $1/\tau_h$. It is useful to have a relationship between β, g_m, and r_{be}. Using Equations 6.72 and 6.74, we have

β at low frequencies

$$\beta = g_m r_{be} \tag{6.76}$$

In transistor data books, the dc current gain I_C/I_B is denoted as h_{FE} whereas the small-signal ac current gain i_c/i_b is denoted as h_{fe}. Except at high frequencies, $h_{fe} \approx h_{FE}$.

EXAMPLE 6.21

CE LOW-FREQUENCY SMALL-SIGNAL EQUIVALENT CIRCUIT Consider a BJT with a β of 100, used in a CE amplifier in which the collector current is 2.5 mA and R_C is 1 kΩ. If the ac source has an rms voltage of 1 mV and an output resistance R_s of 50 Ω, what is the rms output voltage? What is the input and output power and the overall power amplification?

SOLUTION

As the collector current is 2.5 mA, the input resistance and the transconductance are

$$r_{be} = \frac{\beta 25}{I_C(\text{mA})} = \frac{(100)(25)}{2.5} = 1000\ \Omega$$

and

$$g_m = \frac{I_C(\text{mA})}{25} = \frac{2.5}{25} = 0.1\ \text{A/V}$$

The *magnitude* of the voltage gain of the BJT small-signal equivalent circuit is

$$A_V = \frac{v_{ce}}{v_{be}} = g_m R_C = (0.1)(1000) = 100$$

When the ac source is connected to the B and E terminals (Figure 6.54), the input resistance r_{be} of the BJT loads the ac source, so v_{be} across BE is

$$v_{be} = v_s \frac{r_{be}}{(r_{be} + R_s)} = (1\text{ mV})\frac{1000\ \Omega}{(1000\ \Omega + 50\ \Omega)} = 0.952\text{ mV}$$

The output voltage (rms) is, therefore,

$$v_{ce} = A_V v_{be} = 100(0.952\text{ mV}) = 95.2\text{ mV}$$

The loading effect makes the output less than 100 mV. To reduce the loading of the ac source, we need to increase r_{be}, *i.e.*, reduce the collector current, but that also reduces the gain. So to keep the gain the same, we need to reduce I_C and increase R_C. However, R_C cannot be increased indefinitely because R_C itself is loaded by the input of the next stage and, in addition, there is an incremental resistance between the collector and emitter terminals (typically ~100 kΩ) that shunts R_C (not shown in Figure 6.54).

The power amplification of the CE BJT itself is

$$A_P = \frac{i_c v_{ce}}{i_b v_{be}} = \beta A_V = (100)(100) = 10{,}000$$

The input power into the BE terminals is

$$P_{in} = v_{be} i_b = \frac{v_{be}^2}{r_{be}} = \frac{(0.952 \times 10^{-3}\text{ V})^2}{1000\ \Omega} = 9.06 \times 10^{-10}\text{ W} \qquad \text{or} \qquad 0.906\text{ nW}$$

The output power is

$$P_{out} = P_{in} A_P = (9.06 \times 10^{-10})(10{,}000) = 9.06 \times 10^{-6}\text{ W} \qquad \text{or} \qquad 9.06\ \mu\text{W}$$

6.12 JUNCTION FIELD EFFECT TRANSISTOR (JFET)

6.12.1 GENERAL PRINCIPLES

The basic structure of the junction field effect transistor (JFET) with an n-type channel (n-channel) is depicted in Figure 6.55a. An n-type semiconductor slab is provided with contacts at its ends to pass current through it. These terminals are called **source** (S) and **drain** (D). Two of the opposite faces of the n-type semiconductor are heavily p-type doped to some small depth so that an n-type channel is formed between the source and drain terminals, as shown in Figure 6.55a. The two p^+ regions are normally electrically connected and are called the **gate** (G). As the gate is heavily doped, the depletion layers extend almost entirely into the n-channel, as shown in Figure 6.55a. For simplicity we will assume that the two gate regions are identical (both p^+ type) and that the doping in the n-type semiconductor is uniform. We will define the n-channel to be the region of conducting n-type material contained between the two depletion layers.

The basic and idealized symmetric structure in Figure 6.55a is useful in explaining the principle of operation as discussed later but does not truly represent the

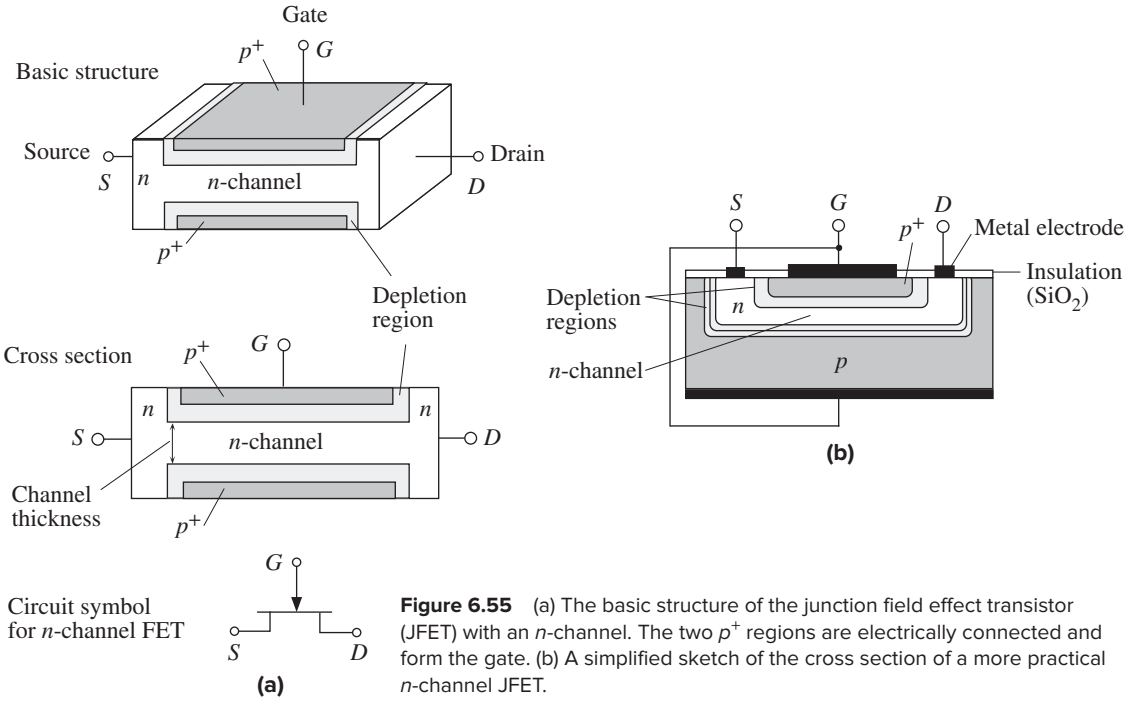


Figure 6.55 (a) The basic structure of the junction field effect transistor (JFET) with an *n*-channel. The two p^+ regions are electrically connected and form the gate. (b) A simplified sketch of the cross section of a more practical *n*-channel JFET.

structure of a typical practical device. A simplified schematic sketch of the cross section of a more practical device (as, for example, fabricated by the planar technology) is shown in Figure 6.55b where it is apparent that the two gate regions do not have identical doping and that, except for one of the gates, all contacts are on one surface.

We first consider the behavior of the JFET with the gate and source shorted ($V_{GS} = 0$), as shown in Figure 6.56a. The resistance between S and D is essentially the resistance of the conducting *n*-channel between A and B, R_{AB}. When a positive voltage is applied to D with respect to S ($V_{DS} > 0$), then a current flows from D to S, which is called the **drain current** I_D. There is a voltage drop along the channel, between A and B, as indicated in Figure 6.56a. The voltage in the *n*-channel is zero at A and V_{DS} at B. As the voltage along the *n*-channel is positive, the p^+n junctions between the gates and the *n*-channel become progressively more reverse-biased from A to B. Consequently the depletion layers extend more into the channel and thereby decrease the thickness of the conducting channel from A to B.

Increasing V_{DS} increases the widths of the depletion layers, which penetrate more into the channel and hence result in more channel narrowing toward the drain. The resistance of the *n*-channel R_{AB} therefore increases with V_{DS}. The drain current therefore does not increase linearly with V_{DS} but falls below it because

$$I_D = \frac{V_{DS}}{R_{AB}}$$

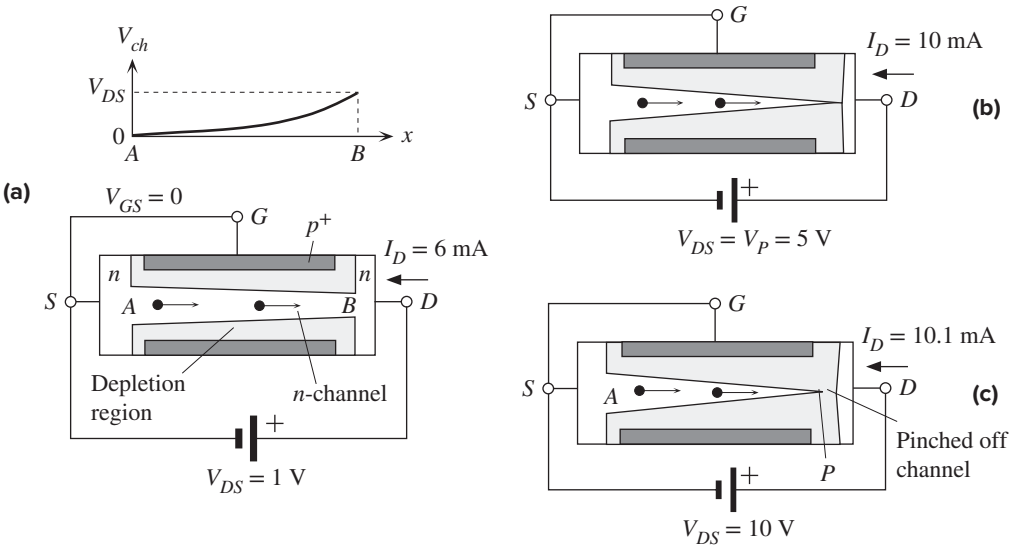


Figure 6.56 (a) The gate and source are shorted ($V_{GS} = 0$) and V_{DS} is small. (b) V_{DS} has increased to a value that allows the two depletion layers to just touch, when $V_{DS} = V_P (= 5\text{ V})$ and the p^+n junction voltage at the drain end, $V_{GD} = -V_{DS} = -V_P = -5$ V. (c) V_{DS} is large ($V_{DS} > V_P$), so a short length of the channel is pinched off.

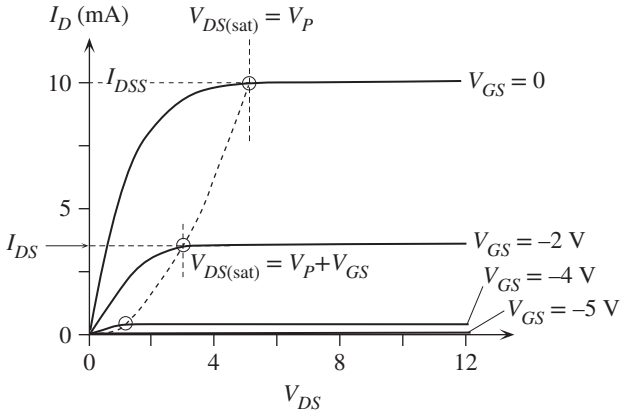


Figure 6.57 Typical I_D versus V_{DS} characteristics of a JFET for various fixed gate voltages V_{GS}.

and R_{AB} increases with V_{DS}. Thus I_D versus V_{DS} exhibits a sublinear behavior, as shown in the $V_{DS} < 5$ V region in Figure 6.57.

As V_{DS} increases further, the depletion layers extend more into the channel and eventually, when $V_{DS} = V_P$ (= 5 V), the two depletion layers around B meet at point P at the drain end of the channel, as depicted in Figure 6.56b. The channel is then said to be "pinched off" by the two depletion layers. The voltage V_P is called the **pinch-off voltage.** It is equal to the magnitude of reverse bias needed across the p^+n junctions to make them just touch at the drain end. Since the actual bias

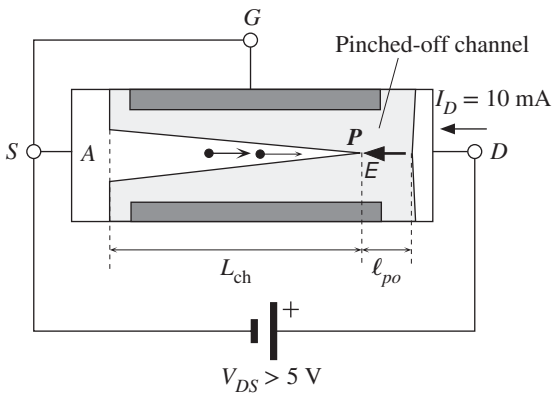


Figure 6.58 The pinched-off channel and conduction for $V_{DS} > V_P$ (= 5 V).

voltage across the p^+n junctions at the drain end (B) is V_{GD}, the pinch-off occurs whenever

Pinch-off condition

$$V_{GD} = -V_P \qquad [6.77]$$

In the present case, gate to source is shorted, $V_{GS} = 0$, so $V_{GD} = -V_{DS}$ and pinch-off occurs when $V_{DS} = V_P$ (5 V). The drain current from pinch-off onwards, as shown in Figure 6.57, does not increase significantly with V_{DS} for reasons given below. Beyond $V_{DS} = V_P$, there is a short pinched-off channel of length ℓ_{po}.

The pinched-off channel is a reverse-biased depletion region that separates the drain from the n-channel, as depicted in Figure 6.58. There is a very strong electric field E in this pinched-off region in the D to S direction. This field is the vector sum of the fields from positive donors to negative acceptors in the depletion regions of the channel and the gate on the drain side. Electrons in the n-channel drift toward P, and when they arrive at P, they are swept across the pinched-off channel by E. This process is similar to minority carriers in the base of a BJT reaching the collector junction depletion region, where the internal field there sweeps them across the depletion layer into the collector. Consequently the drain current is actually determined by the resistance of the conducting n-channel over L_{ch} from A to P in Figure 6.58 and not by the pinched-off channel.

As V_{DS} increases, most of the additional voltage simply drops across ℓ_{po} as this region is depleted of carriers and hence highly resistive. Point P, where the depletion layers first meet, moves slightly toward A, thereby slightly reducing the channel length L_{ch}. Point P must still be at a potential V_P because it is this potential that just makes the depletion layers touch. Thus the voltage drop across L_{ch} remains as V_P. Beyond pinch-off then

$$I_D = \frac{V_P}{R_{AP}} \qquad (V_{DS} > V_P)$$

Since R_{AP} is determined by L_{ch}, which decreases slightly with V_{DS}, I_D increases slightly with V_{DS}. In many cases, I_D is conveniently taken to be saturated at a value I_{DSS} for $V_{DS} > V_P$. Typical I_D versus V_{DS} behavior is shown in Figure 6.57.

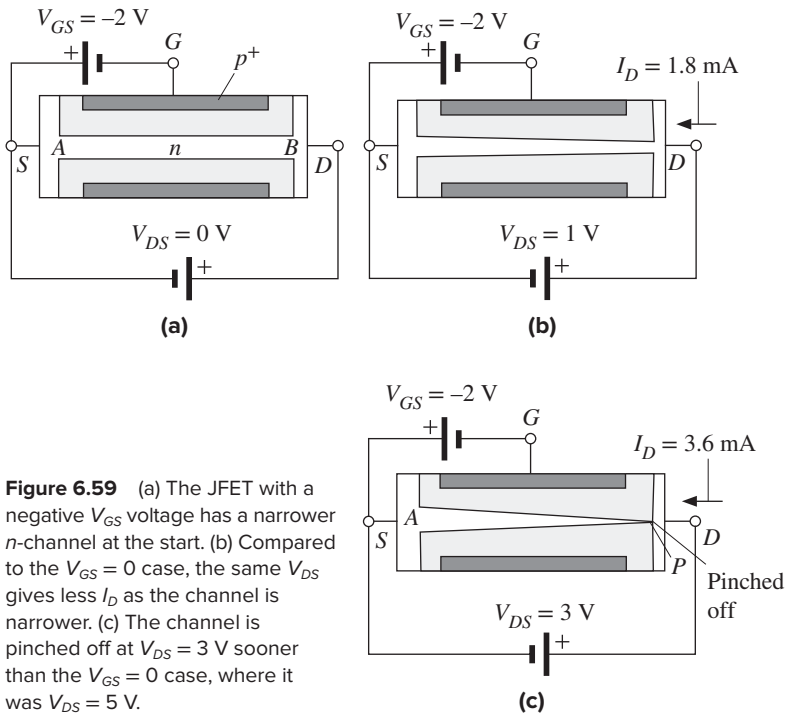


Figure 6.59 (a) The JFET with a negative V_{GS} voltage has a narrower *n*-channel at the start. (b) Compared to the $V_{GS} = 0$ case, the same V_{DS} gives less I_D as the channel is narrower. (c) The channel is pinched off at $V_{DS} = 3$ V sooner than the $V_{GS} = 0$ case, where it was $V_{DS} = 5$ V.

We now consider what happens when a negative voltage, say $V_{GS} = -2$ V, is applied to the gate with respect to the source, as shown in Figure 6.59a with $V_{DS} = 0$. The p^+n junctions are now reverse biased from the start, the channel is narrower, and the channel resistance is now larger than in the $V_{GS} = 0$ case. The drain current that flows when a small V_{DS} is applied, as in Figure 6.59b, is now smaller than in the $V_{GS} = 0$ case as apparent in Figure 6.57. The p^+n junctions are now progressively more reverse biased from V_{GS} at the source end to $V_{GD} = V_{GS} - V_{DS}$ at the drain end. We therefore need a smaller V_{DS} (= 3 V) to pinch off the channel, as shown in Figure 6.59c. When $V_{DS} = 3$ V, the G to D voltage V_{GD} across the p^+n junctions at the drain end is -5 V, which is $-V_P$, so the channel becomes pinched off. Beyond pinch-off, I_D is nearly saturated just as in the $V_{GS} = 0$ case, but its magnitude is obviously smaller as the thickness of the channel at A is smaller; compare Figures 6.56 and 6.59. In the presence of V_{GS}, as apparent from Figure 6.57, the pinch-off occurs at $V_{DS} = V_{DS(\text{sat})}$, and from Equation 6.77.

Pinch-off condition

$$V_{DS(\text{sat})} = V_P + V_{GS} \qquad [6.78]$$

where V_{GS} is a negative voltage (reducing V_P). Beyond pinch-off when $V_{DS} > V_{DS(\text{sat})}$, the point P where the channel is *just pinched* still remains at potential $V_{DS(\text{sat})}$, given by Equation 6.78.

For $V_{DS} > V_{DS(\text{sat})}$, I_D becomes nearly saturated at a value denoted as I_{DS}, which is indicated in Figure 6.57. When G and S are shorted ($V_{GS} = 0$), I_{DS} is called I_{DSS} (which stands for I_{DS} with shorted gate to source). Beyond pinch-off, with negative

V_{GS}, the drain current I_D is

$$I_D \approx I_{DS} \approx \frac{V_{DS(\text{sat})}}{R_{AP}(V_{GS})} = \frac{V_P + V_{GS}}{R_{AP}\,(V_{GS})} \qquad V_{DS} > V_{DS(\text{sat})} \qquad [6.79]$$

where $R_{AP}(V_{GS})$ is the effective resistance of the conducting n-channel from A to P (Figure 6.59b), which depends on the channel thickness and hence on V_{GS}. The resistance increases with more negative gate voltage as this increases the reverse bias across the p^+n junctions, which leads to the narrowing of the channel. For example, when $V_{GS} = -4$ V, the channel thickness at A becomes narrower than in the case with $V_{GS} = -2$ V, thereby increasing the resistance, R_{AP}, of the conducting channel and therefore decreasing I_{DS}. Further, there is also a reduction in the drain current by virtue of $V_{DS(\text{sat})}$ decreasing with negative V_{GS}, as apparent in Equation 6.79. Figure 6.57 shows the effect of the gate voltage on the I_D versus V_{DS} behavior. The two effects, that from $V_{DS(\text{sat})}$ and that from $R_{AP}(V_{GS})$ in Equation 6.79, lead to I_{DS} almost decreasing parabolically with $-V_{GS}$.

When the gate voltage is such that $V_{GS} = -V_P$ ($= -5$ V) with the source and drain shorted ($V_{DS} = 0$), then the two depletion layers touch over the entire channel length and the whole channel is closed, as illustrated in Figure 6.60. The channel is said to be off. The only drain current that flows when a V_{DS} is applied is due to the thermally generated carriers in the depletion layers. This current is very small.

Figure 6.57 summarizes the full I_D versus V_{DS} characteristics of the n-channel JFET at various gate voltages V_{GS}. It is apparent that I_{DS} is relatively independent of V_{DS} and that it is controlled by the gate voltage V_{GS}, as expected by Equation 6.79. This is analogous to the BJT in which the collector current I_C is controlled by the base–emitter bias voltage V_{BE}. Figure 6.61a shows the dependence of I_{DS} on the gate voltage V_{GS}. The transistor action is the control of the drain current I_{DS}, in the drain–source (output) circuit by the voltage V_{GS} in the gate–source (input circuit), as shown in Figure 6.61b. This control is only possible if $V_{DS} > V_{DS(\text{sat})}$. When $V_{GS} = -V_p$, the drain current is nearly zero because the channel has been totally pinched off. This gate–source voltage is denoted by $V_{GS(\text{off})}$ as the drain current has been switched off. Furthermore, we should note that as V_{GS} reverse biases the p^+n junction, the current into the gate I_G is the reverse leakage current of these junctions. It is usually very small. In some JFETs, I_G is as low as a fraction of a nanoampere. We should also note that the circuit symbol for the JFET, as shown in Figure 6.55a, has an arrow to identify the gate and the pn junction direction.

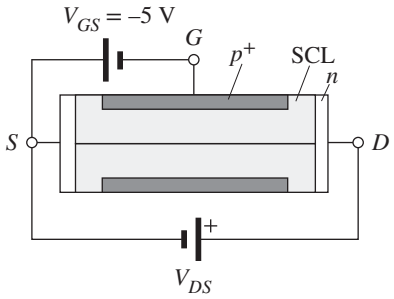


Figure 6.60 When $V_{GS} = -5$ V, the depletion layers close the whole channel from the start, at $V_{DS} = 0$. As V_{DS} is increased, there is a very small drain current, which is the small reverse leakage current due to thermal generation of carriers in the depletion layers.

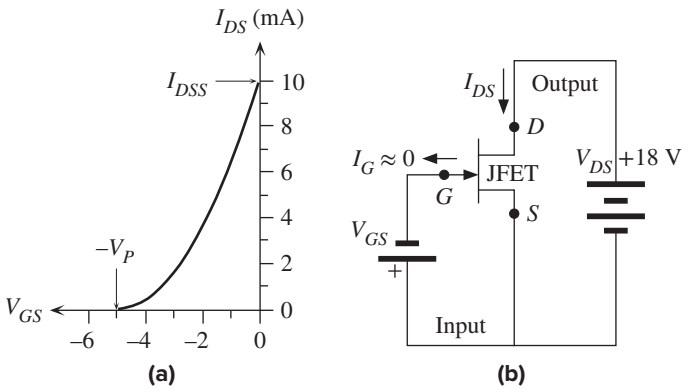


Figure 6.61 (a) Typical I_{DS} versus V_{GS} characteristics of a JFET. (b) The dc circuit where V_{GS} in the gate–source circuit (input) controls the drain current I_{DS} in the drain–source (output) circuit in which V_{DS} is kept constant and large ($V_{DS} > V_P$).

JFET equation beyond pinch-off

Is there a convenient relationship between I_{DS} and V_{GS}? If we calculate the effective resistance R_{AP} of the n-channel between A and P, we can obtain its dependence on the channel thickness, and thus on the widths of the depletion layers and hence on V_{GS}. We can then find I_{DS} from Equation 6.79. It turns out that a simple parabolic dependence seems to represent the data reasonably well,

$$I_{DS} = I_{DSS}\left[1 - \left(\frac{V_{GS}}{V_{GS(\text{off})}}\right)\right]^2 \qquad [6.80]$$

where I_{DSS} is the drain current when $V_{GS} = 0$ (Figure 6.61) and $V_{GS(\text{off})}$ is defined as $-V_P$, that is, that gate–source voltage that just pinches off the channel. The pinch-off voltage V_P here is a positive quantity because it was introduced through $V_{DS(\text{sat})}$. $V_{GS(\text{off})}$ however is negative, $-V_P$. We should note two important facts about the JFET. Its name originates from the effect that modulating the electric field in the reverse-biased depletion layers (by changing V_{GS}) varies the depletion layer penetration into the channel and hence the resistance of the channel. The transistor action hence can be thought of as being based on a **field effect.** Since there is a p^+n junction between the gate and the channel, the name has become JFET. This junction in reverse bias provides the isolation between the gate and channel.

Secondly, the region beyond pinch-off, where Equations 6.79 and 6.80 hold, is commonly called the **current saturation region,** as well as **constant current region** and **pentode region.** The term **saturation** should not be confused with similar terms used for saturation effects in bipolar transistors. A saturated BJT cannot be used as an amplifier, but JFETs are invariably used as amplifiers in the saturated current region.

6.12.2 JFET AMPLIFIER

The transistor action in the JFET is the control of I_{DS} by V_{GS}, as shown in Figure 6.61. The input circuit is therefore the gate–source circuit containing V_{GS} and the output circuit is the drain–source circuit in which the drain current I_{DS} flows. The JFET is almost never used with the *pn* junction between the gate and channel forward biased ($V_{GS} > 0$) as this would lead to a very large gate current and near shorting of the

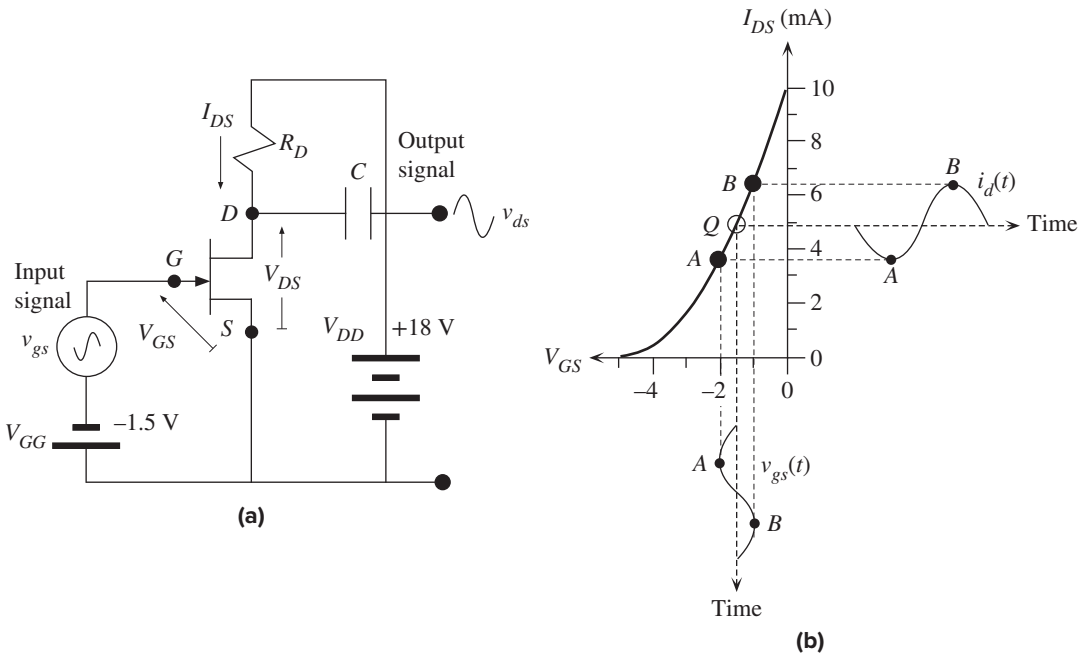


Figure 6.62 (a) Common source (CS) ac amplifier using a JFET. (b) Explanation of how I_D is modulated by the signal v_{gs} in series with the dc bias voltage V_{GG}.

gate to source voltage. With V_{GS} limited to negative voltages, the maximum current in the output circuit can only be I_{DSS}, as shown in Figure 6.61a. The maximum input voltage V_{GS} should therefore give an I_{DS} less than I_{DSS}.

Figure 6.62a shows a simplified illustration of a typical JFET voltage amplifier. As the source is common to both the input and output circuits, this is called a **common source** (CS) **amplifier.** The input signal is the ac source v_{gs} connected in series with a negative dc bias voltage V_{GG} of −1.5 V in the *GS* circuit. First we will find out what happens when there is no ac signal in the circuit ($v_{gs} = 0$). The dc supply (−1.5 V) in the input provides a negative dc voltage to the gate and therefore gives a dc current I_{DS} in the output circuit (less than I_{DSS}). Figure 6.62b shows that when $V_{GS} = -1.5$ V, point Q on the I_{DS} versus V_{GS} characteristics gives $I_{DS} = 4.9$ mA. Point Q, which determines the dc operation, is called the **quiescent point.**

The ac source v_{gs} is connected in series with the negative dc bias voltage V_{GS}. It therefore modulates V_{GS} up and down about −1.5 V with time, as shown in Figure 6.62b. Suppose that v_{gs} varies sinusoidally between −0.5 V and +0.5 V. Then, as shown in Figure 6.62b when v_{gs} is −0.5 V (point *A*), $V_{GS} = -2.0$ V and the drain current is given by point *A* on the I_{DS}–V_{GS} curve and is about 3.6 mA. When v_{gs} is +0.5 V (point *B*), then $V_{GS} = -1.0$ V and the drain current is given by point *B* on the I_{DS}–V_{GS} curve and is about 6.4 mA. The input variation from −0.5 V to +0.5 V has thus been converted to a drain current variation from 3.6 mA to 6.4 mA as indicated in Figure 6.62b. We could have just as easily calculated the drain current from Equation 6.80.

Table 6.6 Voltage and current in the common source amplifier of Figure 6.62a

v_{gs} (V)	V_{GS} (V)	I_{DS} (mA)	i_d (mA)	$V_{DS} = V_{DD} - I_{DS}R_D$	v_{ds} (V)	Voltage Gain	Comment
0	−1.5	4.9	0	8.2	0		dc conditions, point *Q*
−0.5	−2.0	3.6	−1.3	10.8	+2.6	−5.2	Point *A*
+0.5	−1.0	6.4	+1.5	5.2	−3.0	−6	Point *B*

NOTE: V_{DD} = 18 V and R_D = 2000 Ω.

Table 6.6 summarizes what happens to the drain current as the ac input voltage is varied about zero.

The change in the drain current with respect to its dc value is the output signal current denoted as i_d. Thus at *A*,

$$i_d = 3.6 - 4.9 = -1.3 \text{ mA}$$

and at *B*,

$$i_d = 6.4 - 4.9 = 1.5 \text{ mA}$$

The variation in the output current is not quite symmetric as that in the input signal v_{gs} because the I_{DS}–V_{GS} relationship, Equation 6.80, is not linear.

The drain current variations in the *DS* circuit are converted to voltage variations by the resistance R_D. The voltage across *DS* is

$$V_{DS} = V_{DD} - I_{DS} R_D \qquad [6.81]$$

where V_{DD} is the bias battery voltage in the *DS* circuit. Thus, variations in I_{DS} result in variations in V_{DS} that are in the opposite direction or 180° out of phase. The ac output voltage between *D* and *S* is tapped out through a capacitor *C*, as shown in Figure 6.62a. The capacitor *C* simply blocks the dc. Suppose that $R_D = 2000\ \Omega$ and $V_{DD} = 18$ V, then using Equation 6.81 we can calculate the dc value of V_{DS} and also the minimum and maximum values of V_{DS}, as shown in Table 6.6.

It is apparent that as v_{gs} varies from −0.5 V, at *A*, to +0.5 V, at *B*, V_{DS} varies from 10.8 V to 5.2 V, respectively. The change in V_{DS} with respect to dc is what constitutes the output signal v_{ds}, as only the ac is tapped out. From Equation 6.81, the change in V_{DS} is related to the change in I_{DS} by

$$v_{ds} = -R_D i_d \qquad [6.82]$$

Thus, the output, v_{ds}, changes from −3.0 V to 2.6 V. The peak-to-peak voltage amplification is

$$A_{V(\text{pk-pk})} = \frac{\Delta V_{DS}}{\Delta V_{GS}} = \frac{v_{ds(\text{pk-pk})}}{v_{gs(\text{pk-pk})}} = \frac{-3 \text{ V} - (2.6 \text{ V})}{0.5 \text{ V} - (-0.5 \text{ V})} = -5.6$$

The negative sign represents the fact that the output and input voltages are out of phase by 180°. This can also be seen from Table 6.6 where a negative v_{gs} results in a positive v_{ds}. Even though the ac input signal v_{gs} is symmetric about zero, ±0.5 V,

the ac output signal v_{ds} is not symmetric, which is due to the I_{DS} versus V_{GS} curve being nonlinear, and thus varies between −3.0 V and 2.6 V. If we were to calculate the voltage amplification for the most negative input signal, we would find −5.2, whereas for the most positive input signal, it would be −6. The peak-to-peak voltage amplification, which was −5.6, represents a mean gain taking both negative and positive input signals into account.

The amplification can of course be increased by increasing R_D, but we must maintain V_{DS} at all times above $V_{DS(\text{sat})}$ (beyond pinch-off) to ensure that the drain current I_{DS} in the output circuit is only controlled by V_{GS} in the input circuit.

When the signals are small about dc values, we can use differentials to represent small signals. For example, $v_{gs} = \delta V_{GS}$, $i_d = \delta I_{DS}$, $v_{ds} = \delta V_{DS}$, and so on. The variation δI_{DS} due to δV_{GS} about the dc value may be used to define a **mutual transconductance** g_m (sometimes denoted as g_{fs}) for the JFET,

Definition of JFET transconductance

$$g_m = \frac{dI_{DS}}{dV_{GS}} \approx \frac{\delta I_{DS}}{\delta V_{GS}} = \frac{i_d}{v_{gs}}$$

This transconductance can be found by differentiating Equation 6.80,

JFET transconductance

$$g_m = \frac{dI_{DS}}{dV_{GS}} = -\frac{2I_{DSS}}{V_{GS(\text{off})}}\left[1 - \left(\frac{V_{GS}}{V_{GS(\text{off})}}\right)\right] = -\frac{2[I_{DSS}I_{DS}]^{1/2}}{V_{GS(\text{off})}} \quad \textbf{[6.83]}$$

The output signal current is

$$i_d = g_m v_{gs}$$

so using Equation 6.82, the small-signal voltage amplification is

Small-signal voltage gain

$$A_V = \frac{v_{ds}}{v_{gs}} = \frac{-R_D(g_m v_{gs})}{v_{gs}} = -g_m R_D \quad \textbf{[6.84]}$$

Equation 6.84 is only valid under small-signal conditions in which the variations about the dc values are small compared with the dc values themselves. The negative sign indicates that v_{ds} and v_{gs} are 180° out of phase.

EXAMPLE 6.22

THE JFET AMPLIFIER Consider the *n*-channel JFET common source amplifier shown in Figure 6.62a. The JFET has an I_{DSS} of 10 mA and a pinch-off voltage V_P of 5 V as in Figure 6.62b. Suppose that the gate dc bias voltage supply $V_{GG} = -1.5$ V, the drain circuit supply $V_{DD} = 18$ V, and $R_D = 2000\ \Omega$. What is the voltage amplification for small signals? How does this compare with the peak-to-peak amplification of −5.6 found for an input signal that had a peak-to-peak value of 1 V?

SOLUTION

We first calculate the operating conditions at the bias point with no ac signals. This corresponds to point Q in Figure 6.62b. The dc bias voltage V_{GS} across the gate to source is −1.5 V. The resulting dc drain current I_{DS} can be calculated from Equation 6.80 with $V_{GS(\text{off})} = -V_P = -5$ V:

$$I_{DS} = I_{DSS}\left[1 - \left(\frac{V_{GS}}{V_{GS(\text{off})}}\right)\right]^2 = (10\text{ mA})\left[1 - \left(\frac{-1.5}{-5}\right)\right]^2 = 4.9\text{ mA}$$

The transconductance at this dc current (at Q) is given by Equation 6.83,

$$g_m = -\frac{2(I_{DSS}I_{DS})^{1/2}}{V_{GS(\text{off})}} = -\frac{2[(10\times10^{-3})(4.9\times10^{-3})]^{1/2}}{-5} = 2.8\times10^{-3}\ \text{A/V}$$

The voltage amplification of small signals about point Q is

$$A_V = -g_m R_D = -(2.8\times10^{-3})(2000) = -5.6$$

This turns out to be the same as the peak-to-peak voltage amplification we calculated in Table 6.6. When the input ac signal v_{gs} varies between −0.5 and +0.5 V, as in Table 6.6, the output signal is not symmetric. It varies between −3 V and 2.8 V, so the voltage gain depends on the input signal. The amplifier is then said to exhibit **nonlinearity.**

6.13 METAL-OXIDE-SEMICONDUCTOR FIELD EFFECT TRANSISTOR (MOSFET)

6.13.1 FIELD EFFECT AND INVERSION

The metal-oxide-semiconductor field effect transistor is based on the effect of a field penetrating into a semiconductor. Its operation can be understood by first considering a parallel plate capacitor with metal electrodes and a vacuum as insulation in between, as shown in Figure 6.63a. When a voltage V is applied between the plates, charges $+Q$ and $-Q$ (where $Q = CV$) appear on the plates and there is an electric field given by $E = V/L$. The origins of these charges are the conduction electrons for $-Q$ and exposed positively charged metal ions for $+Q$. Metallic bonding is based on all the valence electrons forming a sea of conduction electrons and permeating the space between metal ions that are fixed at crystal lattice sites. Since the electrons are mobile, they are readily displaced by the field. Thus, in the lower plate E displaces some of the conduction electrons to the surface to form $-Q$. In the top plate E displaces some of the electrons from the surface into the bulk to expose positively charged metal ions to form $+Q$.

Suppose that the plate area is 1 cm^2 and spacing is 0.1 μm and that we apply 2 V across it. The capacitance C is 8.85 nF and the magnitude of charge Q on each plate is 1.77×10^{-8} C, which corresponds to 1.1×10^{11} electrons. A typical metal such as copper has something like 2×10^{15} atoms per cm^2 on the surface. Thus, there will be that number of positive metal ions and electrons on the surface (assuming one conduction electron per atom). The charges $+Q$ and $-Q$ can therefore be generated by the electrons and metal ions at the surface alone. For example, if one in every 1.7×10^4 electrons on the surface moves one atomic spacing (∼0.3 nm) into the bulk, then the surface will have a charge of $+Q$ due to exposed positive metal ions. It is clear that, for all practical purposes, the electric field does not penetrate into the metal and terminates at the metal surface.

The same is not true when one of the electrodes is a semiconductor, as shown in Figure 6.63b where the "capacitor" now is of a **metal-insulator-semiconductor** (MOS) device. Suppose that we replace the lower metal in Figure 6.63a with a p-type semiconductor with an acceptor concentration of 10^{15} cm^{-3}. The number of acceptor atoms on the surface[19] is 1×10^{10} cm^{-2}. We may assume that at room temperature

[19] Surface concentration of atoms (atoms per unit area) can be found from $n_{\text{surf}} \approx (n_{\text{bulk}})^{2/3}$.

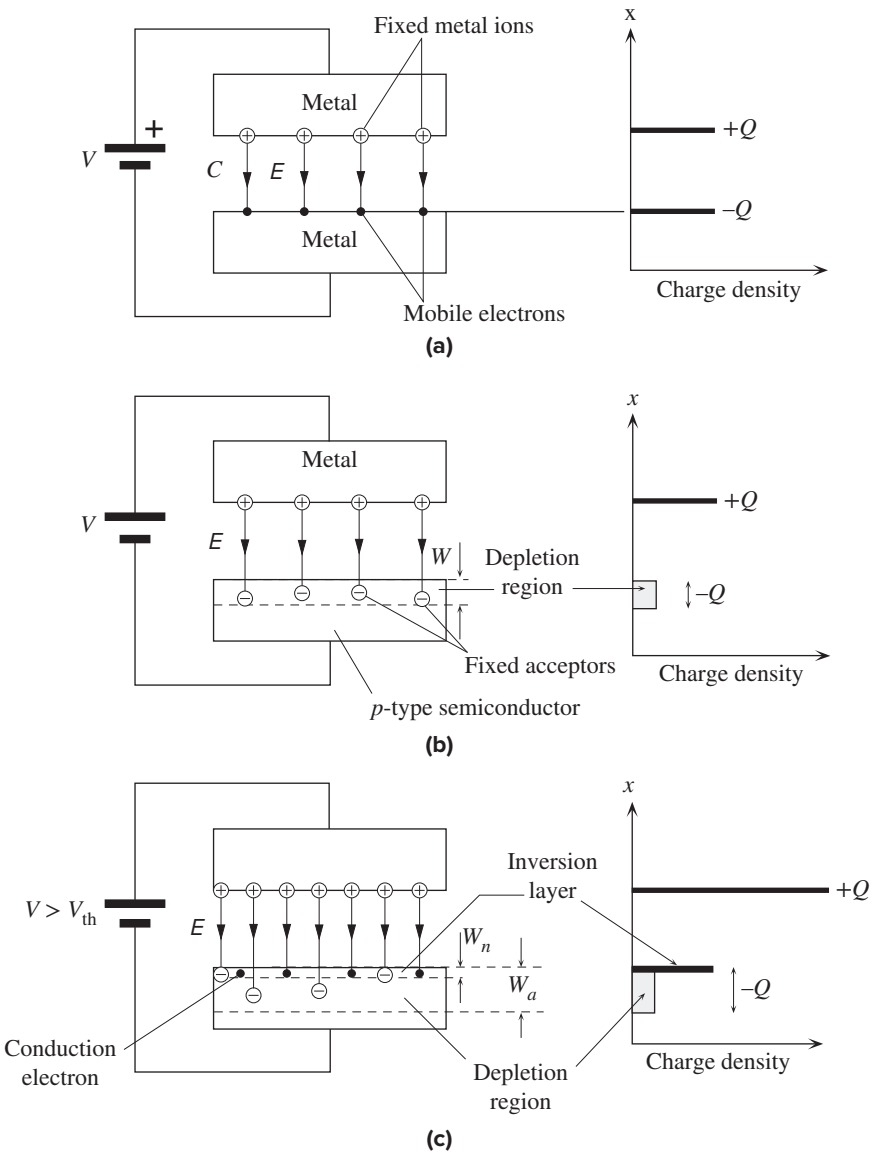


Figure 6.63 The field effect. (a) In a metal-air-metal capacitor, all the charges reside on the surface. (b) Illustration of field penetration into a *p*-type semiconductor. (c) As the field increases, eventually when $V > V_{th}$, an inversion layer is created near the surface in which there are conduction electrons.

all the acceptors are ionized and thus negatively charged. It is immediately apparent that we do not have a sufficient number of negative acceptors at the surface to generate the charge $-Q$. We must therefore also expose negative acceptors in the bulk, which means that the field must penetrate into the semiconductor. Holes in the surface region of the semiconductor become repelled toward the bulk and thereby

expose more negative acceptors. We can estimate the width W into which the field penetrates since the total negative charge exposed $eAWN_a$ must be Q. We find that W is of the order of 1 μm, which is something like 4000 atomic layers. Our conclusion is that the field penetrates into a semiconductor by an amount that depends on the doping concentration.

The penetrating field into the semiconductor drifts away most of the holes in this region and thereby exposes negatively charged acceptors to make up the charge $-Q$. The region into which the field penetrates has lost holes and is therefore depleted of its equilibrium concentration of holes. We refer to this region as a **depletion layer.** As long as $p > n$ even though $p \ll N_a$, this region still has p-type characteristics as holes are in the majority.

If the voltage increases further, $-Q$ also increases in magnitude, as the field becomes stronger and penetrates more into the semiconductor but eventually it becomes more difficult to make up the charge $-Q$ by simply extending the depletion layer width W into the bulk. It becomes possible (and more favorable) to attract conduction electrons into the depletion layer and form a thin electron layer of width W_n near the surface. The charge $-Q$ is now made up of the fixed negative charge of acceptors in W_a and of conduction electrons in W_n, as shown in Figure 6.63c. Further increases in the voltage do not change the width W_a of the depletion layer but simply increase the electron concentration in W_n. Where do these electrons come from as the semiconductor is doped p-type? Some are attracted into the depletion layer from the bulk, where they were minority carriers. But most are thermally generated by the breaking of Si–Si bonds (*i.e.*, across the bandgap) in the depleted layer. Thermal generation in the depletion layer generates EHPs that become separated by the field. The holes are then drifted by the field into the bulk and the electrons toward the surface. Recombination of the thermally generated electrons and holes with other carriers is greatly reduced because the depletion layer has so few carriers. Since the electron concentration in the electron layer exceeds the hole concentration and this layer is within a normally p-type semiconductor, we call this an **inversion layer.**

It is now apparent that increasing the field in the metal-insulator-semiconductor device first creates a depletion layer and then an inversion layer at the surface when the voltage exceeds some threshold value V_{th}. This is the basic principle of the field effect device. As long as $V > V_{\text{th}}$, any increase in the field and hence $|-Q|$ leads to more electrons in the inversion layer, whereas the width of the depletion layer W_a and hence the quantity of fixed negative charge remain constant. The insulator between the metal and the semiconductor, that is, a vacuum in Figure 6.63, is typically SiO_2 in many devices.

6.13.2 ENHANCEMENT MOSFET

Figure 6.64 shows the basic structure of an enhancement n-channel MOSFET device (NMOSFET). A metal-insulator-semiconductor structure is formed between a p-type Si substrate and a metal electrode, which is called the gate (G). The insulator is the SiO_2 oxide grown during fabrication. There are two n^+ doped regions at the ends of the MOS device that form the source (S) and drain (D). A metal contact is also made

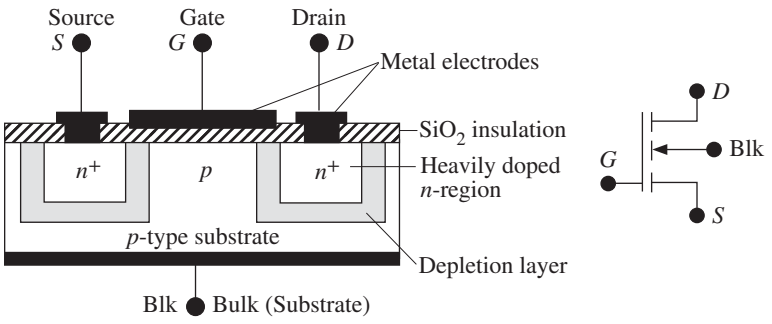


Figure 6.64 The basic structure of the enhancement MOSFET and its circuit symbol.

to the p-type Si substrate (or the bulk), which in many devices is connected to the source terminal as shown in Figure 6.64. Further, many MOSFETs have a degenerately doped polycrystalline Si material as the gate that serves the same function as the metal electrode.

With no voltage applied to the gate, S to D is an n^+pn^+ structure that is always reverse biased whatever the polarity of the source to drain voltage. However, if the substrate (bulk) is connected to the source, a negative V_{DS} will forward bias the n^+p junction between the drain and the substrate. As the n-channel MOSFET device is not normally used with a negative V_{DS}, we will not consider this polarity.

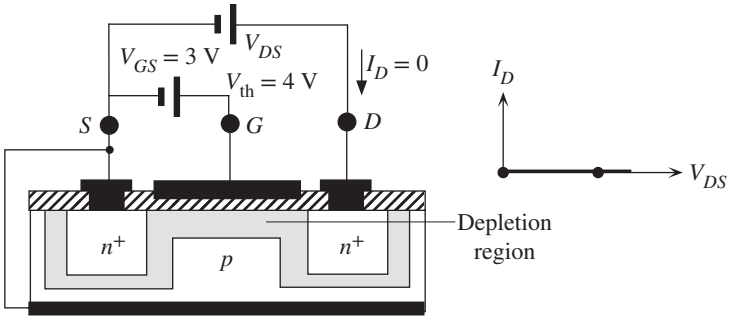
When a positive voltage less than V_{th} is applied to the gate, $V_{GS} < V_{\text{th}}$, as shown in Figure 6.65a, the p-type semiconductor under the gate develops a depletion layer as a result of the expulsion of holes into the bulk, just as in Figure 6.63b. Since S and D are isolated by a low-conductivity p-doped region that has a depletion layer from S to D, no current can flow for any positive V_{DS}.

With $V_{DS} = 0$, as soon as V_{GS} is increased beyond the threshold voltage V_{th}, an n-channel inversion layer is formed within the depletion layer under the gate and immediately below the surface, as shown in Figure 6.65b. This n-channel links the two n^+ regions of source and drain. We then have a continuous n-type material with electrons as mobile carriers between the source and drain. When a small V_{DS} is applied, a drain current I_D flows that is limited by the resistance of the n-channel $R_{n\text{-ch}}$:

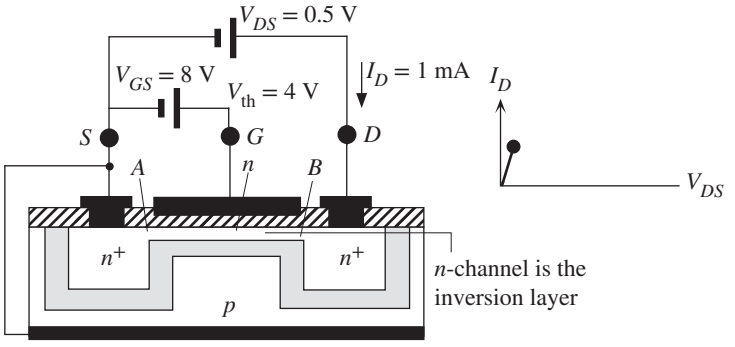
$$I_D = \frac{V_{DS}}{R_{n\text{-ch}}} \qquad [6.85]$$

Thus, I_D initially increases with V_{DS} almost linearly, as shown in Figure 6.65b.

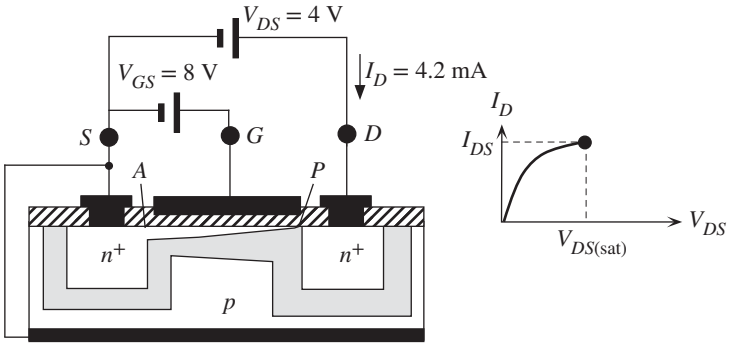
The voltage variation along the channel is from zero at A (source end) to V_{DS} at B (drain end). The gate to the n-channel voltage is then V_{GS} at A and $V_{GD} = V_{GS} - V_{DS}$ at B. Thus point A depends only on V_{GS} and remains undisturbed by V_{DS}. As V_{DS} increases, the voltage at B (V_{GD}) decreases and thereby causes less inversion. This means that the channel gets narrower from A to B and its resistance $R_{n\text{-ch}}$, increases with V_{DS}. I_D versus V_{DS} then falls increasingly below the $I_D \propto V_{DS}$ line. Eventually when the gate to n-channel voltage at B decreases to just below V_{th}, the inversion layer at B disappears and a depletion layer is exposed, as illustrated in



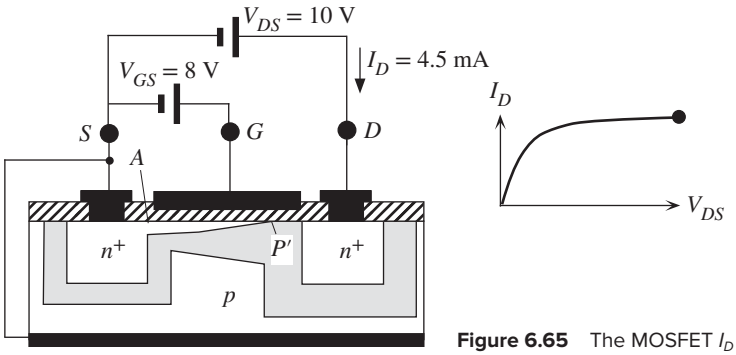
(a) Below threshold $V_{GS} < V_{th}$ and $V_{DS} > 0$



(b) Above threshold $V_{GS} > V_{th}$ and $V_{DS} < V_{DS(sat)}$



(c) Above threshold $V_{GS} > V_{th}$ and saturation, $V_{DS} = V_{DS(sat)}$



(d) Above threshold $V_{GS} > V_{th}$ and saturation region, $V_{DS} > V_{DS(sat)}$

Figure 6.65 The MOSFET I_D versus V_{DS} characteristics.

Figure 6.65c. The *n*-channel becomes pinched off at this point *P*. This occurs when $V_{DS} = V_{DS(\text{sat})}$, satisfying

$$V_{GD} = V_{GS} - V_{DS(\text{sat})} = V_{\text{th}} \qquad [6.86]$$

It is apparent that the whole process of the narrowing of the *n*-channel and its eventual pinch-off is similar to the operation of the *n*-channel JFET. When the drifting electrons in the *n*-channel reach *P*, the large electric field within the very narrow depletion layer at *P* sweeps the electrons across into the n^+ drain. The current is limited by the supply of electrons from the *n*-channel to the depletion layer at *P*, which means that it is limited by the effective resistance of the *n*-channel between *A* and *P*.

When V_{DS} exceeds $V_{DS(\text{sat})}$, the additional V_{DS} drops mainly across the highly resistive depletion layer at *P*, which extends slightly to *P′* toward *A*, as shown in Figure 6.65d. At *P′*, the gate to channel voltage must still be just V_{th} as this is the voltage required to just pinch off the channel and just eliminate inversion. The widening of the depletion layer (from *B* to *P′*) at the drain end with V_{DS}, however, is small compared with the channel length *AB*. The resistance of the channel from *A* to *P′* does not change significantly with increasing V_{DS}, which means that the drain current I_D is then nearly saturated at I_{DS},

$$I_D \approx I_{DS} \approx \frac{V_{DS(\text{sat})}}{R_{AP'n\text{-ch}}} \qquad V_{DS} > V_{DS(\text{sat})} \qquad [6.87]$$

As $V_{DS(\text{sat})}$ depends on V_{GS}, so does I_{DS}. The overall I_{DS} versus V_{DS} characteristics for various fixed gate voltages V_{GS} of a typical enhancement MOSFET is shown in Figure 6.66a. It can be seen that there is only a slight increase in I_{DS} with V_{DS} beyond $V_{DS(\text{sat})}$. The I_{DS} versus V_{GS} when $V_{DS} > V_{DS(\text{sat})}$ characteristics are shown in Figure 6.66b. It is apparent that as long as $V_{DS} > V_{DS(\text{sat})}$, the saturated drain current I_{DS} in the source–drain (or output) circuit is almost totally controlled by the gate voltage V_{GS} in the source–gate (or input) circuit. This is what constitutes the MOSFET

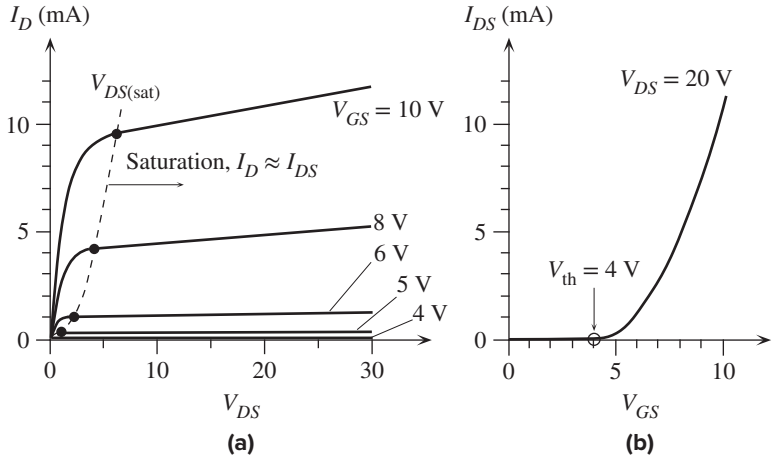


Figure 6.66 (a) Typical I_D versus V_{DS} characteristics of an enhancement MOSFET ($V_{\text{th}} = 4$ V) for various fixed gate voltages V_{GS}. (b) Dependence of I_{DS} on V_{GS} at a given V_{DS} ($>V_{DS(\text{sat})}$).

action. Variations in V_{GS} then lead to variations in the drain current I_{DS} (just as in the JFET), which forms the basis of the MOSFET amplifier. The term **enhancement** refers to the fact that a gate voltage exceeding V_{th} is required to enhance a conducting channel between the source and drain. This contrasts with the JFET where the gate voltage depletes the channel and decreases the drain current.

The experimental relationship between I_{DS} and V_{GS} (when $V_{DS} > V_{DS(\text{sat})}$) has been found to be best described by a parabolic equation similar to that for the JFET, except that now V_{GS} enhances the channel when $V_{GS} > V_{\text{th}}$ so I_{DS} exists only when $V_{GS} > V_{\text{th}}$,

Enhancement NMOSFET

$$I_{DS} = K(V_{GS} - V_{\text{th}})^2 \quad [6.88]$$

where K is a constant. For an ideal MOSFET, it can be expressed as

Enhancement NMOSFET constant

$$K = \frac{Z\mu_e\varepsilon}{2Lt_{\text{ox}}} \quad [6.89]$$

where μ_e is the electron drift mobility in the channel, L and Z are the length and width of the gate controlling the channel, and ε and t_{ox} are the permittivity ($\varepsilon_r\varepsilon_o$) and thickness of the oxide insulation under the gate. According to Equation 6.88, I_{DS} is independent of V_{DS}. The shallow slopes of the I_D versus V_{DS} lines beyond $V_{DS(\text{sat})}$ in Figure 6.66a can be accounted for by writing Equation 6.88 as

Enhancement NMOSFET

$$I_{DS} = K(V_{GS} - V_{\text{th}})^2(1 + \lambda V_{DS}) \quad [6.90]$$

where λ is a constant that is typically 0.01 V^{-1}. If we extend the I_{DS} versus V_{DS} lines, they intersect the $-V_{DS}$ axis at $1/\lambda$, which is called the **Early voltage.** It should be apparent that I_{DSS}, which is I_{DS} with the gate and source shorted ($V_{GS} = 0$), is zero and is not a useful quantity in describing the behavior of the enhancement MOSFET.

The drift mobility μ_e in Equation 6.89 represents the drift of electrons in the channel near the surface of the semiconductor. This region also has the field from the gate penetrating into it as well as a longitudinal field along the channel. μ_e is not the same as the drift mobility in the bulk of p-Si but depends on the field penetrating into the channel, and defects and dopants in this region, especially near the semiconductor–oxide interface. μ_e is therefore a field effect mobility and should be viewed as an *effective mobility in the channel.*

EXAMPLE 6.23 **THE ENHANCEMENT NMOSFET** A particular discrete enhancement NMOS transistor has a gate with a width (Z) of 50 μm, length (L) of 10 μm, and SiO_2 thickness of 450 Å. The relative permittivity of SiO_2 is 3.9. Its threshold voltage is 4 V. Estimate the drain current when $V_{GS} = 8$ V and $V_{DS} = 20$ V, given $\lambda = 0.01$. The effective electron drift mobility μ_e is roughly 700 $\text{cm}^2\ \text{V}^{-1}\ \text{s}^{-1}$.

SOLUTION

Since $V_{DS} > V_{\text{th}}$, we can assume that the drain current is saturated and we can use the I_{DS} versus V_{GS} relationship in Equation 6.90,

$$I_{DS} = K(V_{GS} - V_{\text{th}})^2(1 + \lambda V_{DS})$$

where the constant K is given by Equation 6.89

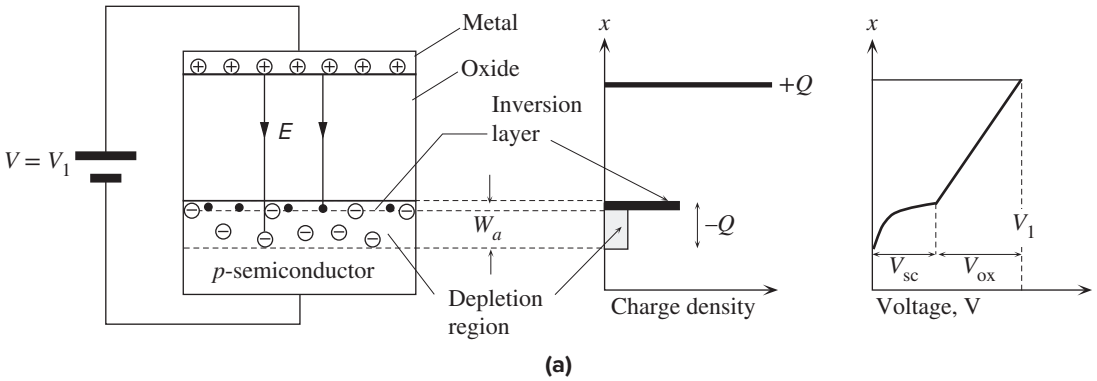
$$K = \frac{Z\mu_e\varepsilon_r\varepsilon_o}{2Lt_{ox}} = \frac{(50 \times 10^{-6})(700 \times 10^{-4})(3.9 \times 8.85 \times 10^{-12})}{2(10 \times 10^{-6})(450 \times 10^{-10})} = 0.000134 \text{ AV}^{-1}$$

When $V_{GS} = 8$ V and $V_{DS} = 20$ V, with $\lambda = 0.01$, from Equation 6.90,

$$I_{DS} = 0.000134(8 - 4)^2[1 + (0.01)(20)]$$
$$= 0.0026 \text{ A} \quad \text{or} \quad 2.6 \text{ mA}$$

6.13.3 THRESHOLD VOLTAGE

The threshold voltage is an important parameter in MOSFET devices. Its control in device fabrication is therefore essential. Figure 6.67a shows an idealized MOS structure where all the electric field lines from the metal pass through the oxide and penetrate the p-type semiconductor. The charge $-Q$ is made up of fixed negative acceptors in a surface region of W_a and of conduction electrons in the inversion layer at the surface, as shown in Figure 6.67a. The voltage drop across the MOS structure, however, is not uniform. As the field penetrates the semiconductor, there is a voltage drop V_{sc} across the field penetration region of the semiconductor by virtue of $E = -dV/dx$, as shown in Figure 6.67a. The field terminates on both electrons in the



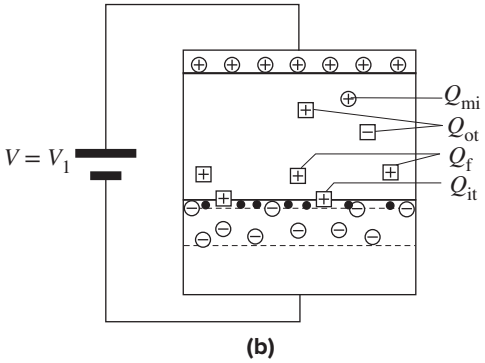


Figure 6.67 (a) The threshold voltage and the ideal MOS structure. (b) In practice, there are several charges in the oxide and at the oxide–semiconductor interface that affect the threshold voltage: Q_{mi} = mobile ionic charge (*e.g.*, Na^+), Q_{ot} = trapped oxide charge, Q_f = fixed oxide charge, and Q_{it} = charge trapped at the interface.

inversion layer and acceptors in W_a, so within the semiconductor E is not uniform and therefore the voltage drop is not constant. But the field in the oxide is uniform, as we assumed there were no charges inside the oxide. The voltage drop across the oxide is constant and is V_{ox}, as shown in Figure 6.67a. If the applied voltage is V_1, we must have $V_{sc} + V_{ox} = V_1$. The actual voltage drop V_{sc} across the semiconductor determines the condition for inversion. We can show this as follows. If the acceptor doping concentration is 10^{16} cm^{-3}, then the Fermi level E_F in the bulk of the p-type semiconductor must be 0.347 eV below E_{Fi} in intrinsic Si. To make the surface n-type we need to shift E_F at the surface to go just above E_{Fi}. Thus we need to shift E_F from bulk to surface by at least 0.347 eV. We have to bend the energy band by 0.347 eV at the surface. Since the voltage drop across the semiconductor is V_{sc} and the corresponding electrostatic PE change is eV_{sc}, this must be 0.347 eV or $V_{sc} =$ 0.347 V. The gate voltage for the start of inversion will then be $V_{ox} + 0.347$ V. By inversion, however, we generally infer that the electron concentration at the surface is comparable to the hole concentration in the bulk. This means that we actually have to shift E_F above E_{Fi} by another 0.347 eV, so the gate threshold voltage V_{th} must be $V_{ox} + 0.694$ V.

In practice there are a number of other important effects that must be considered in evaluating the threshold voltage. Invariably there are charges both within the oxide and at the oxide–semiconductor interface that alter the field penetration into the semiconductor and hence the threshold voltage needed at the gate to cause inversion. Some of these are depicted in Figure 6.67b and can be qualitatively summarized as follows.

There may be some mobile ions within the SiO_2, such as alkaline ions (Na^+, K^+), which are denoted as Q_{mi} in Figure 6.67b. These may be introduced unintentionally, for example, during cleaning and etching processes in the fabrication. In addition there may be various trapped (immobile) charges within the oxide Q_{ot} due to structural defects, for example, an interstitial Si^+. Frequently these oxide trapped charges are created as a result of radiation damage (irradiation by X-rays or other high-energy beams). They can be reduced by annealing the device.

A significant number of fixed positive charges (Q_f) exist in the oxide region close to the interface. They are believed to originate from the nonstoichiometry of the oxide near the oxide–semiconductor interface. They are generally attributed to positively charged Si^+ ions. During the oxidation process, a Si atom is removed from the Si surface to react with the oxygen diffusing in through the oxide. When the oxidation process is stopped suddenly, there are unfulfilled Si ions in this region. Q_f depends on the crystal orientation and on the oxidation and annealing processes. The semiconductor to oxide interface itself is a sudden change in the structure from crystalline Si to amorphous oxide. The semiconductor surface itself will have various defects, as discussed in Chapter 1. There is some inevitable mismatch between the two structures at the interface, and consequently there are broken bonds, dangling bonds, point defects such as vacancies and Si^+, and other defects at this interface that trap charges (*e.g.*, holes). All these interface-trapped charges are represented as Q_{it} in Figure 6.67b. Q_{it} depends not only on the crystal orientation but also on the chemical composition of the interface. Both Q_f and Q_{it} overall represent a positive charge that effectively reduces the gate voltage needed for inversion. They are smaller

for the (100) surface than the (111) surface, so (100) is the preferred surface for the Si MOS device.

In addition to various charges in the oxide and at the interface shown in Figure 6.67b, there will also be a voltage difference, denoted as V_{FB}, between the semiconductor surface and the metal surface, even in the absence of an applied voltage. V_{FB} arises from the work function difference between the metal and the *p*-type semiconductor, as discussed in Chapter 4. The metal work function is generally smaller than the semiconductor work function, which means that the semiconductor surface will have an accumulation of electrons and the metal surface will have positive charges (exposed metal ions). The gate voltage needed for inversion will therefore also depend on V_{FB}. Since V_{FB} is normally positive and Q_f and Q_{it} are also positive, there may already be an inversion layer formed at the semiconductor surface even without a positive gate voltage. The fabrication of an enhancement MOSFET then requires special fabrication procedures, such as ion implantation, to obtain a positive and predictable V_{th}.

The simplest way to control the threshold gate voltage is to provide a separate electrode to the bulk of an enhancement MOSFET, as shown in Figure 6.64, and to apply a bias voltage to the bulk with respect to the source to obtain the desired V_{th} between the gate and source. This technique has the disadvantage of requiring an additional bias supply for the bulk and also adjusting the bulk to source voltage almost individually for each MOSFET.

6.13.4 ION IMPLANTED MOS TRANSISTORS AND POLY-SI GATES

The most accurate method of controlling the threshold voltage is by ion implantation, as the number of ions that are implanted into a device and their location can be closely controlled. Furthermore, ion implantation can also provide a self-alignment of the edges of the gate electrode with the source and drain regions. In the case of an *n*-channel enhancement MOSFET, it is generally desirable to keep the *p*-type doping in the bulk low to avoid small V_{DS} for reverse breakdown between the drain and the bulk (see Figure 6.64). Consequently, the surface, in practice, already has an inversion layer (without any gate voltage) due to various fixed positive charges residing in the oxide and at the interface, as shown in Figure 6.67b (positive Q_f and Q_{it} and V_{FB}). It then becomes necessary to implant the surface region under the gate with boron acceptors to remove the electrons and restore this region to a *p*-type behavior.

The ion implantation process is carried out in a vacuum chamber where the required impurity ions are generated and then accelerated toward the device. The energy of the arriving ions and hence their penetration into the device can be readily controlled. Typically, the device is implanted with *B* acceptors under the gate oxide, as shown in Figure 6.68. The distribution of implanted acceptors as a function of distance into the device from the surface of the oxide is also shown in the figure. The position of the peak depends on the energy of the ions and hence on the accelerating voltage. The peak of the concentration of implanted acceptors is made to occur just below the surface of the semiconductor. Since ion implantation involves the impact of energetic ions with the crystal structure, it results in the inevitable generation of various defects

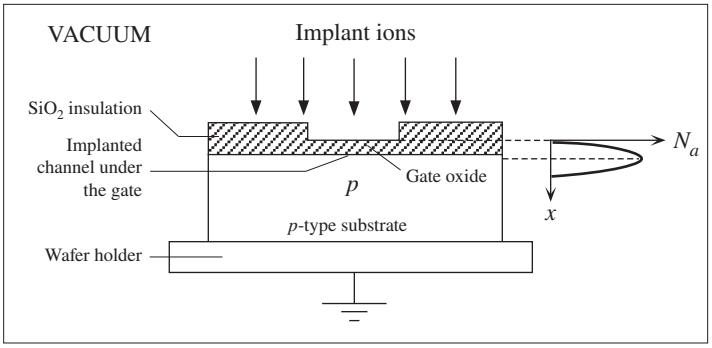


Figure 6.68 Schematic illustration of ion implantation for the control of V_{th}.

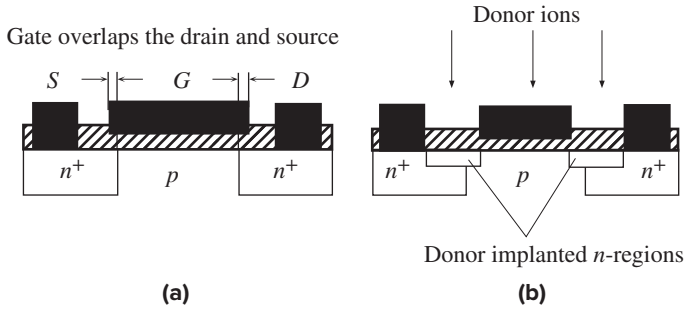


Figure 6.69 (a) There is an overlap of the gate electrode with the source and drain regions and hence additional capacitance between the gate and drain. (b) n^+-type ion implantation extends the drain and source to line up with the gate.

within the implanted region. The defects are almost totally eliminated by annealing the device at an elevated temperature. Annealing also broadens the acceptor implanted region as a result of increased diffusion of implanted acceptors.

Ion implantation also has the advantage of providing self-alignment of the drain and source with the edges of the gate electrode. In a MOS transistor, it is important that the gate electrode extends all the way from the source to the drain regions so that the channel formed under the gate can link the two regions; otherwise, an incomplete channel will be formed. To avoid the possibility of forming an incomplete channel, it is necessary to allow for some overlap, as shown in Figure 6.69a, between the gate and source and drain regions because of various tolerances and variations involved in the fabrication of a MOSFET by conventional masking and diffusional techniques. The overlap, however, results in additional capacitances between the gate and source and the gate and drain and adversely affects the high-frequency (or transient) response of the device. It is therefore desirable to align the edges of the gate electrode with the source and drain regions. Suppose that the gate electrode is made narrower so that it does not extend all the way between the source and drain regions, as shown in Figure 6.69b. If the device is now ion implanted with donors, then donor ions passing through the thin oxide will extend the n^+ regions up to the edges of the gate and thereby align the drain and source with the edges of the gate. The thick metal gate is practically impervious to the arriving donor ions.

Another method of controlling V_{th} is to use silicon instead of a metal for the gate electrode. This technique is called **silicon gate technology.** Typically, the silicon for the gate is vacuum deposited (*e.g.*, by chemical vapor deposition using silane

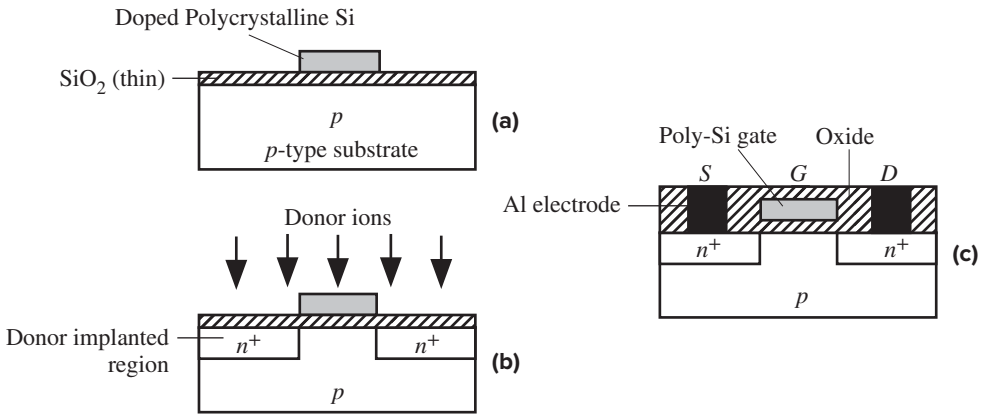


Figure 6.70 The poly-Si gate technology. (a) Poly-Si is deposited onto the oxide, and the areas outside the gate dimensions are etched away. (b) The poly-Si gate acts as a mask during ion implantation of donors to form the n^+ source and drain regions. (c) A simplified schematic sketch of the final poly-Si MOS transistor.

gas) onto the oxide, as shown in Figure 6.70. As the oxide is noncrystalline, the Si gate is polycrystalline (rather than a single crystal) and is therefore called a **poly-Si gate.** Normally it is heavily doped to ensure that it has sufficiently low resistivity to avoid *RC* time constant limitations in charging and discharging the gate capacitance during transient or ac operations. The advantage of the poly-Si gate is that its work function depends on the doping (type and concentration) and can be controlled so that V_{FB} and hence V_{th} can also be controlled. There are also additional advantages in using the poly-Si gate. For example, it can be raised to high temperatures during fabrication whereas a metal such as Al would melt at 660 °C. It can be used as a mask over the gate region of the semiconductor during the formation of the source and drain regions. If ion implantation is used to deposit donors into the semiconductor, then the n^+ source and drain regions are self-aligned with the poly-Si gate, as shown in Figure 6.70.

ADDITIONAL TOPICS

6.14 *pin* DIODES, PHOTODIODES, AND SOLAR CELLS

The *pin* Si diode is a device that has a structure with three distinct layers: a heavily doped thin p^+-type layer, a relatively thick intrinsic (*i*-Si) layer, and a heavily doped thin n^+-type layer, as shown in Figure 6.71a. For simplicity we will assume that the *i*-layer is truly intrinsic, or at least doped so lightly compared with p^+ and n^+ layers that it behaves almost as if intrinsic. The intrinsic layer is much wider than the p^+ and n^+ regions, typically 5–50 μm depending on the particular application. When the structure is first formed, holes diffuse from the p^+-side and electrons from the n^+-side into the *i*-Si layer where they recombine and disappear. This leaves behind a thin layer of exposed negatively charged acceptor ions in the p^+-side and a thin