How Transistors Work

We begin by discussing the so-called bipolar transistor such as those shown in Figures 10-44 and 10-45. In normal operation, the transistor's emitter-base junction is forward biased and the base-collector junction is reverse biased, as shown in Figure 10-44*a*. The heavily doped *p*-type emitter emits holes that flow across the emitter-base junction into the base. Because the base is very thin, most of these holes flow across the base into the collector. This flow constitutes a current I_c from the emitter to the collector. However, some of the holes recombine in the base, producing a positive charge that inhibits the further flow of current. To prevent this, some of the holes that do not reach the collector are drawn off the base as a base current I_b in a circuit connected to the base. In Figure 10-44, therefore, I_c is almost but not quite equal to I_e, and I_b is much smaller than either I_c or I_e. It is customary to express I_c as

$$I_c = \beta I_b \qquad \textbf{10-50}$$

where β is called the *current gain* of the transistor. Transistors can be designed to have values of β as low as 10 or as high as several hundred.

Figure 10-44*b* shows a simple *pnp* transistor used as an amplifier. A small time-varying input voltage v_s is connected in series with a bias voltage V_{eb}. The base current is then the sum of a steady current I_b produced by the bias voltage V_{eb} and a varying current i_b due to the signal voltage v_s. Because v_s may at any instant be either positive or negative, the bias voltage V_{eb} must be large enough to ensure that there will always

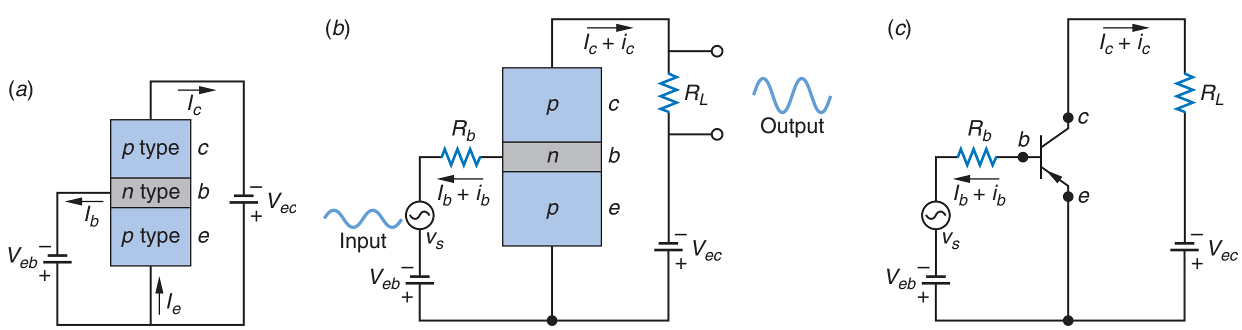

FIGURE 10-44 (*a*) A *pnp* transistor biased for normal operation. Holes from the emitter can easily diffuse across the base, which is only tens of nanometers thick. Most of the holes flow to the collector, producing the current I_c. (*b*) A simple *pnp* transistor used as an amplifier. (*c*) A typical CD or DVD player will have many such transistors connected in series.

be a forward bias on the emitter-base junction. The collector current will consist of two parts: a direct current $I_c = \beta I_b$ and an alternating current $i_c = \beta i_b$. We thus have a current amplifier in which the time-varying output current i_c is β times the input current i_b. In such an amplifier, the steady currents I_c and I_b, although essential to the operation of the transistor, are usually not of interest. The input signal voltage v_s is related to the base current by Ohm's law:

$$i_b = \frac{v_s}{R_b + r_b} \qquad \textbf{10-51}$$

where r_b is the internal resistance of the transistor between the base and emitter. Similarly, the collector current i_c produces a voltage v_L across the output or load resistance R_L given by

$$v_L = i_c R_L \qquad \textbf{10-52}$$

Using Equation 10-50, we have

$$i_c = \beta i_b = \beta \frac{v_s}{R_b + r_b}$$

The output voltage is thus related to the input voltage by

$$v_L = \beta \frac{R_L}{R_b + r_b} v_s \qquad \textbf{10-53}$$

The ratio of the output voltage to the input voltage is the *voltage gain* of the amplifier:

$$\text{Voltage gain} = \frac{v_L}{v_s} = \beta \frac{R_L}{R_b + r_b} \qquad \textbf{10-54}$$

In a practical case β may be 100, and the ratio $R_L/(R_b + r_b)$ may be $1/2$. Then the voltage gain is 50. A more detailed derivation shows that v_L and v_s are 180° *out of phase*; that is, when v_s has its most positive value, v_L has its most negative value. For a simple amplifier this phase shift is not important because all input voltages, regardless of frequency, are affected identically. This simple voltage amplifier is the basis of many circuits that are components of communication systems.

A typical amplifier, such as that in a tape or CD player, has several transistors similar to the one in Figure 10-44 connected in series so that the output of one transistor serves as the input for the next. Thus, the very small voltage produced by the passage of the magnetized tape past the pickup heads controls the large amounts of power required to drive the loudspeakers. The power delivered to the speakers is supplied by the sources of direct voltage connected to each transistor. The resistance of a semiconductor material depends on the impurity concentration. Reverse-biased diodes have capacitance that can be controlled by controlling the bias voltage. These two facts are exploited in the manufacture of *integrated circuits*, in which millions of transistors along with associated resistors and capacitors are interconnected on a single tiny piece, or "chip," of silicon or germanium.

EXAMPLE 10-11 Currents Across a *pn* Junction Compare the currents through a *pn* junction diode when the forward bias is changed from 0.25 V to 0.75 V. How does the current under the 0.75 V forward bias compare with that through the diode if the bias voltage is reversed to −0.75 V? Assume that the diode is operating at 297 K.

SOLUTION

The net current is given by Equation 10-49. Writing $k/e = 8.63 \times 10^{-5}$ eV/K, we have for the 0.50V change in forward bias that

$$\frac{I_{\text{net}}(0.75\ \text{V})}{I_{\text{net}}(0.25\ \text{V})} = \frac{I_0\left(e^{0.75/8.63\times10^{-5}\times297} - 1\right)}{I_0\left(e^{0.25/8.63\times10^{-5}\times297} - 1\right)} = \frac{e^{29.3} - 1}{e^{9.76} - 1} = 3.0 \times 10^8$$

Thus, the increase of 0.50 V in forward bias increases the current through the diode more than a hundred million times!

The effect on the current of reversing the bias is equally dramatic:

$$\frac{I_{\text{net}}(+0.75\ \text{V})}{I_{\text{net}}(-0.75\ \text{V})} = \frac{e^{29.3} - 1}{e^{-29.3} - 1} = -5.2 \times 10^{12}$$

Limits to Transistor Size

Integrated circuits (called ICs or chips) combine “active” electronic devices (transistors and diodes) with “passive” ones (resistors and capacitors) on a single semiconductor crystal wafer. A single chip may contain several million transistors of a type called metal oxide semiconductor field-effect transistors (MOSFET). These differ from the bipolar transistors described above in that the current through the transistor is controlled by a voltage (hence electric field) across the “gate” rather than the current through the base. It has been the experience of the electronics industry for over three decades that transistor performance (switching speed) and density (number of transistors per unit area on the chip) double every three years. So consistent has this rate of improvement been that it is referred to as “Moore’s law” after G. Moore, who first made the observation. However, continued improvements in performance and density are faced with serious physical limitations.

The common electronic device is the silicon-based metal oxide semiconductor (MOS) transistor, consisting of a source (≈ emitter), a drain (≈ collector), and a gate (≈ base) (see Figure 10-45). The source and drain are separated by an oppositely charged channel, as Figure 45 illustrates. If a voltage applied via the gate across the insulating oxide layer separating the gate from the channel repels the charges in the channel and attracts those in the source and drain, a conducting layer is formed in the doped silicon of the channel and current flows—the transistor is “on.” If the applied voltage attracts channel charges and repels those in the source and drain, no conducting layer forms and the transistor is “off.”

The increases in switching speed have been obtained by making the individual transistors smaller and smaller but doing so without reducing the free charge in the source, channel, and drain regions so that the device’s resistance remains low and, therefore, power consumption remains low. Maintaining the free charge means increasing the concentration of dopant atoms in the silicon lattice (see Figures 10-27 and 10-28). The concentration of dopant atoms in state-of-the-art technology MOS transistors is currently of the order of 1 percent, which is very near the thermodynamic maximum for silicon. Larger concentrations result in the dopant atoms “clumping” together rather than participating in the covalent bonding of the silicon

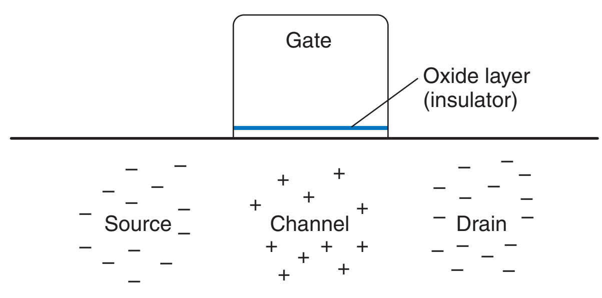

FIGURE 10-45 Schematic of a metal oxide semiconductor (MOS) transistor. Voltage applied to the gate, depending on its sign, attracts or repels charges in the channel, thereby turning the transistor “off” and “on.”

crystal. This results in reduction of the free charge and a corresponding increase in the resistance. Searches for new dopant atoms that create higher free charge concentrations have thus far been unsuccessful.

Concomitant reduction of the thickness of the gate oxide insulating layer faces an equally daunting quantum-mechanical limitation. Reducing the thickness of the oxide layer increases the electric field across the layer for a given applied voltage. (Think of the oxide layer as a parallel-plate capacitor.) This increases the charge density in the channel, thereby lowering the resistance, a desirable effect. However, oxide layer thicknesses are now in the 1.5 to 2.0 nm range (about 3 to 4 atomic layers). Tunneling of electrons across the potential barrier, that is, through the oxide layer, leads to a sizable leakage current through the gate. Recall that the tunneling probability is exponentially dependent on the thickness of the barrier (see Equation 6-76). While the leakage current does not seem to damage the oxide, it does result in excessive power consumption and potential circuit failure. Suitable materials with much higher dielectric constants would make possible increased oxide layer thicknesses and hence reduced tunneling and continued size reduction, but here again no suitable materials have yet been found. These and related problems must be solved if the performance increases that have driven the electronics industry over the past 35 years are to continue.