CHAPTER

3

TRANSISTOR PHYSICS

The active devices in electronic systems are called **transistors**. Their function is to control the flow of current from one node based on the potential at another node. In terms of our analog of Figure 2.1 (p. 14), we construct a valve that is operated by the level in another tank. Although it is easy to imagine mechanical linkages that could cause a valve to operate, we will not succumb to the temptation to use that model. Rather, we will describe a somewhat less familiar but still intuitive model that accurately embodies all the necessary physics of the transistor.

BOLTZIAN HYDRAULICS

We saw in Chapter 2 that the density of gas molecules in the atmosphere decreases exponentially with height. For the planet Earth, the gravitational attraction, temperature, and molecular weight are such that the atmospheric density decreases by a factor of e for each approximately 20-kilometer increase in elevation. It is easy to imagine a planet with much higher gravitational attraction, lower temperature, and heavier molecules, such that the change in elevation required to decrease the density by a factor of e would be about 1 meter. Because such a planet would not be a hospitable place to live, we will travel to it only to illustrate the principles on which transistors operate. We will call this imaginary planet Boltzo.

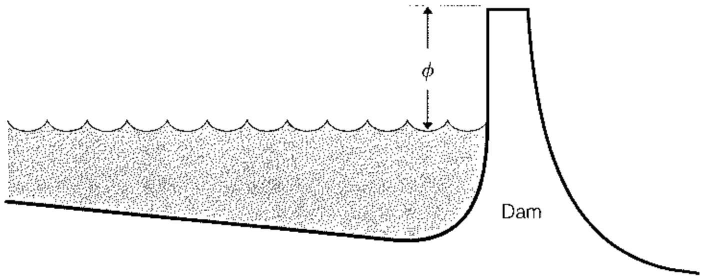

FIGURE 3.1 Cross-section of reservoir showing water level a distance ϕ below the top of the dam. Any water vapor must rise to this height to spill over the dam.

We will use a hydraulic power system similar to that shown in Figure 2.1 (p. 14). If we locate our Boltzian reservoir about 200 meters above sea level, the potential energy of a water molecule in the reservoir (measured in units of kT) will be about equal to that of an electron in the 5-volt power supply of an electronic circuit. (Had we located our reservoir at an altitude of 25 meters, we would have approximated the operation of a neural system, but that story is a bit more complicated. We will content ourselves for the moment with creating a hydraulic transistor.)

Early Boltzian engineers constructed a dam in the bottom of a deep canyon, as shown in Figure 3.1. The purpose of the dam was to create a reservoir to store water. When the rains were heavy and the reservoir nearly full, however, the Boltzians noticed that the water level did not stay constant, even if no water was withdrawn from the reservoir. They made an exhaustive search, but found no sign of leakage. Finally, they sought the advice of a wise and venerable Boltzian philosopher. After surveying the situation, he gave his reply: "Contemplate atop the dam in the quiet of the night."

This pronouncement instantly became the subject of much discussion in learned circles—what could he possibly mean by such a reply? One young engineering student named Lily Field grew tired of the endless arguments. Against prevailing sentiment, she undertook the long journey to the dam site. Sitting atop the massive structure as the noises of the day faded into darkness, she contemplated the meaning of the philosopher's words. All wind had ceased; not a blade of grass or a leaf stirred. Still she had the strong sense of a cold and heavy force against her back. Holding out her silk scarf like a sail, she noticed that it billowed out, as if pressed by some invisible force. When she raised it to eye level, the billowing was considerably weaker. She found a long stick, attached the scarf to its tip, and held it aloft; the billowing effect was scarcely detectable.

The mystery was solved. The water was not escaping in liquid form at all it was *evaporating*, and the vapor was pouring over the dam in enormous quantities. There was a source of water behind the dam, and no source on the opposite side of the dam. There was thus a difference in density of water vapor on the two sides of the dam. Water vapor was *diffusing* from the region of high vapor density to the region of low density. The density gradient, and hence the diffusion rate, was proportional to the density at the top of the dam. Because the density of the vapor decreased exponentially with height above the water level, the effect was noticeable only when the reservoir was very nearly full.

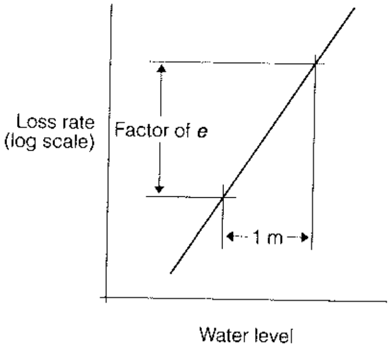

FIGURE 3.2 Loss rate of water vapor over the dam of Figure 3.1 as a function of water level in the reservoir. The distance ϕ decreases as the water level rises. Because the vapor density decreases exponentially with height above the water level, the loss rate increases exponentially with the water level.

When Field returned to the city, she looked up data on the water loss rate and level over the several years that the dam had been operating. She plotted the log of the loss rate as a function of the height of the water level in the reservoir. The result was a straight line, as shown in Figure 3.2. For each meter that the water level rose, the loss rate increased by a factor of e.

Field presented her findings in a poster session at the next International Hydraulics Conference (IHC); her study soon became the talk of the entire meeting. A special evening discussion session was organized. Field showed the data, and pointed out that the engineers could greatly reduce the vapor loss by extending the height of the dam only a few meters. The additional structure would be required to support not the weight of liquid water, but only the density gradient of water *vapor*, so it could be constructed of light material such as wood or plastic, rather than of massive reinforced concrete.

In an invited paper at the IHC the following year, Dr. Field presented a mechanism for controlling the flow of vapor over the dam, shown in Figure 3.3. A barrier of light material is allowed to slide vertically in a slot in the middle of the dam. The barrier is supported on a series of floats. The floats are buoyed up by water from another reservoir or tank. If the level in the second tank is high, the barrier will be high and the flow of vapor over the dam will be low. If the level in the second tank is low, the barrier will be low, and the flow of vapor over the dam will be large. The flow rate I across the barrier will be directly proportional to the density gradient in the horizontal direction. That gradient will, in turn, be proportional to the density of water vapor at height ϕ above the liquid surface, given in Equation 2.14 (p. 25):

$$I = I_0 e^{-\frac{w\phi}{kT}} \tag{3.1}$$

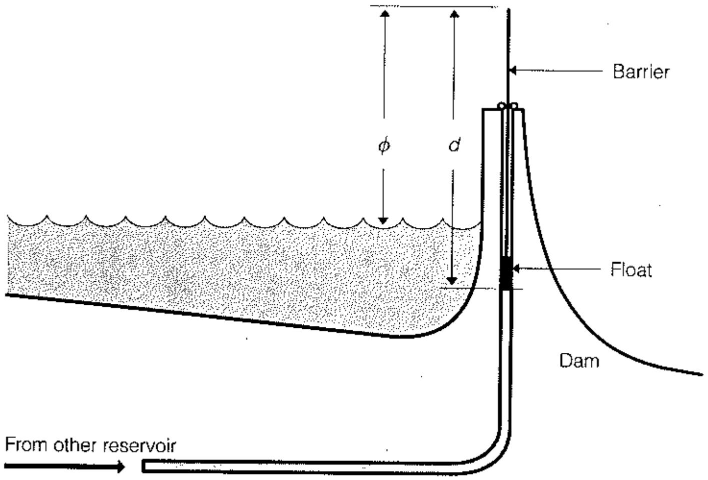

FIGURE 3.3 Hydraulic transistor. The barrier floats on water in a chamber filled from another reservoir. The water-vapor flow rate is an exponentially decreasing function of the water level in the second reservoir.

From the basic dimensions of the structure, Equation 3.1 can be expressed as

$$I = I_0 e^{-\frac{w}{kT}(h_2 - h_1 + d)} \tag{3.2}$$

where h_1 is the elevation of the main reservoir and h_2 is the elevation in the control reservoir. The physical height d of the barrier can be factored into a new preexponential constant,

$$I_1 = I_0 e^{-\frac{wd}{kT}}$$

The product wh of the molecular weight w and the reservoir level h is just the potential energy per molecule P. Equation 3.2 can thus be expressed as

$$I = I_1 e^{\frac{P_1 - P_2}{kT}} \tag{3.3}$$

The current decreases exponentially with the energy difference between the control reservoir and source reservoir. Equation 3.3 is the basic governing equation for the operation of a transistor. The only difference between Boltzo and silicon is that, in the latter, gravitational forces have been replaced by electrostatic ones. The Boltzian landscape is the energy profile of our electronic structures. Cross-sections of the energy terrain are called **energy diagrams**, and appear in all discussions of electron device physics. We will see the energy diagram for an MOS transistor later in this chapter.

SEMICONDUCTORS

All physical structures—neural, electronic, or mechanical—are built out of atoms. Before proceeding, we will briefly review the properties of these basic building blocks out of which transistors and neurons are made.

Atoms and All That

Atoms can be viewed as a swarm of electrons circulating around a nucleus containing protons and neutrons. The negatively charged electrons are, of course, attracted to the positively charged nucleus, and will orbit as close to it as the laws of quantum mechanics allow. An **element** is a type of atom; each atom of a given element will have the same number of electrons. The total number of electrons in orbit around an atom of a particular element is called the **atomic number** of that element. The electrons are arranged in quantum-mechanical orbits or shells. There is a maximum number of electrons each shell can hold. We can thus classify elements according to the number of electrons in the outermost shell. Such a classification is called a **periodic table** of the elements. A simplified periodic table showing the elements with which we will be concerned is given in Table 3.1.

Hydrogen, the element with atomic number 1, has one lone proton for a nucleus, and one electron in the innermost shell. It is a Group I element; the group number refers to the number of electrons in the outermost shell. The first shell can contain at most two electrons. Helium, a Group Zero element, has two electrons in its inner shell, which is therefore full. After one shell is full, additional electrons are forced to populate the next larger shell. The element lithium, atomic number 3, has the first shell full with two electrons, and one electron in the second

TABLE 3.1 Simplified periodic table of the elements, showing the valence and position of elements that form semiconductor crystals. The Group IV elements shown form a diamond lattice. Silicon is by far the most commonly used semiconductor. Boron, aluminum, and gallium are acceptor impurities in silicon; phosphorus and arsenic are donors. In addition, Group III elements can combine with Group V elements to form diamondlike crystals in which alternate lattice sites are occupied by atoms of each element. The best known of these Group III–V semiconductors is gallium arsenide, which is used for microwave transistors and light-emitting diodes. Group II–VI crystals also are semiconductors. Zinc sulfide is a common phosphor in television display tubes, and cadmium sulfide was the earliest widely used photosensitive material.

I	II	III	IV	V	VI	VII	Zero
H							He
Li	Be	B	C	N	O	F	Ne
Na	Mg	Al	Si	P	S	Cl	Ar
K	Zn	Ga	Ge	As	Se	Br	Kr
Rb	Cd	In	Sn	Sb	Te	I	Xe

shell. It is therefore a Group I element. The second, third, and fourth shells can each hold eight electrons. Neon, with atomic number 10, has both its first and second shells full.

In the world of atoms, having a full outer shell is a happy circumstance. So happy, in fact, that atoms fortunate enough to attain this condition have absolutely no desire to interact with other atoms. Such elements as helium, neon, and argon are **inert gases**, the ultimate snobs of the atomic pecking order. Other, less fortunate atoms strive mightily to emulate their austere brethren by forming alliances with similar atoms. These alliances, called **covalent bonds**, are based on sharing electrons such that all parties to the charade can make believe they have a full shell, even though some of the electrons that fill the outer shell are shared with neighbors. Small communal aggregates of this sort are called **molecules**. An example is methane, CH_4. Carbon, being a Group IV atom, has four electrons with which to play, but is desperately seeking four more to fill its outer shell. Each hydrogen has only one electron with which to play, but needs only one more to fill its outer shell. The ultimately blissful arrangement results when each of the four hydrogens shares its electron with the carbon, and the carbon shares one electron with each hydrogen. Not quite neon, but not too bad!

Crystals

Molecules can satisfy a social need for a few atoms, but for regimentation on a massive scale there is no substitute for a **crystal**. In the simplest crystal, every atom is an equivalent member of a vast army, arrayed in three dimensions as far as the eye can see. Three elements in Group IV of the periodic table crystallize naturally into a remarkable structure called the **diamond lattice**: carbon, silicon, and germanium. Each atom is covalently bonded to four neighbors arranged at the corners of a regular tetrahedron. Group IV is unique in chemistry. Eight electrons are required to complete an atomic shell. Atoms with four electrons can team up with four neighbors, sharing one electron with each. Such an arrangement fills the shell for everyone; no electrons are left over, and no bond is missing an electron. A pure crystal formed this way is called an **intrinsic** semiconductor; it is an electrical insulator, because there are no charged particles free to move around and to carry current.

Conduction

If we alter the crystal by replacing a small fraction of its atoms with impurity atoms of Group V, the crystal becomes conductive. The addition of impurities is called **doping**. Group V elements have one *more* electron than the four needed for the covalent bonds. This extra electron is only weakly bound to the impurity site in the lattice; at room temperature, it is free to move about and to carry current. Such atoms are called **donors**, because they donate a free electron to the crystal. The free electrons are negative, and a semiconductor crystal doped with donors is said to be **n-type**. The entire crystal is charge neutral, because it is made of atoms that have as many positively charged protons in their nuclei as they have electrons in their shells. When an electron leaves its donor, the donor is said to be **ionized**. An ionized donor has a positive charge because it has lost one electron.

It also is possible to dope a Group IV semiconductor with atoms of Group III. These impurities have one *less* electron than is required for the four covalent bonds. Group III dopants are therefore called **acceptors**. The absence of one electron in a bond is called a **hole**. We can think of the hole as mobile at room temperature, moving about the crystal. (It is actually electrons that move, and the "motion" of the hole is in the opposite direction. We think of the hole as a "bubble" in the electronic fluid.) The hole, being the absence of an electron, carries a *positive* charge. Once the hole has been filled, the acceptor acquires a negative charge, and is said to be **ionized**. Doping a semiconductor with acceptors renders it conductive, the current being carried by positive holes. Such a crystal is called **p-type**.

It is thus possible to provide either positive or negative charge carriers by doping the crystal appropriately. The concentration of dopants can be controlled precisely over many orders of magnitude, from lightly doped (approximately 10^{15} atoms per cubic centimeter) to heavily doped (approximately 10^{19} atoms per cubic centimeter). Heavily doped n-type material is called $n+$, and heavily doped p-type material is called $p+$. The density of impurity atoms is always small compared to the approximately 5×10^{22} atoms per cubic centimeter in the crystal itself.

Because electrons and holes are both charged, and are both used to carry current, we refer to them generically as **charge carriers**.

MOS TRANSISTORS

A cross-section of the simplest transistor structure is illustrated in Figure 3.4. It shows an intrinsic substrate into which two highly doped regions have been fabricated. Consistent with our hydraulic analogy, one of the highly doped regions is called the **source**, and the other is called the **drain**. The entire surface is covered with a very thin layer of SiO_2 (quartz), which is an excellent electrical insulator. On top of the insulator is a metallic control electrode, called the **gate**, that spans the intrinsic region between source and drain. Current flows from source to drain in the region just under the gate oxide called the **channel**. This structure was first described by a freelance inventor named Lilienfeld in a patent issued in 1933 [Lilienfeld, 1928]. It is called an MOS transistor because the active region consists of a metallic gate, an oxide insulator, and a semiconductor channel. In today's technology, the metallic gate often is made of heavily doped polycrystalline silicon, called **polysilicon** or **poly** for short. The details of how transistors are fabricated are the subject of Appendix A. At this point, we will consider the electrical operation of such a transistor.

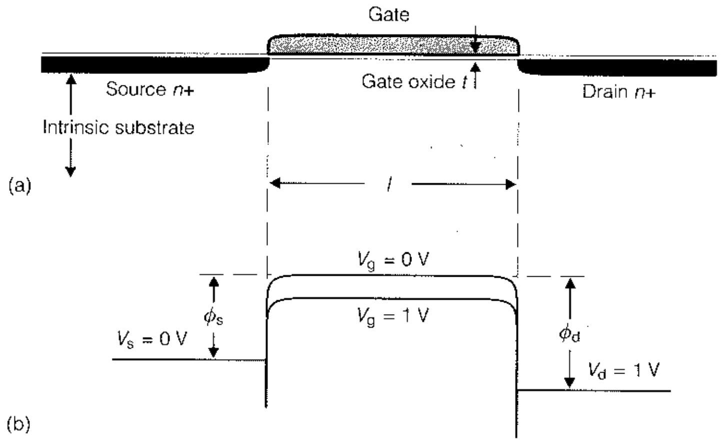

FIGURE 3.4 Cross-section (a) and energy diagram (b) of an *n*-channel transistor. In a typical 1988 process, the gate-oxide thickness is approximately 400 angstroms (0.04 micron), and the minimum channel length *l* is approximately 1.5 microns. When the circuit is in operation, the drain is biased positively; hence, the barrier for electrons is greater at the drain than at the source. Applying a positive voltage at the gate lowers the electron barrier at both source and drain, allowing electrons to diffuse from source to drain.

In Figure 3.4, an energy diagram of the cross-section of part (a) is shown in part (b). The energy of the charge carrier is plotted upward. The surface of the electronic "fluid" is called the **Fermi level.** The drain Fermi level is lower than that of the source by qV_{ds}, where V_{ds} is the drain voltage relative to the source, and q is the electronic charge.

The gate oxide is very thin compared to the source-drain spacing. In a typical 1988 process, the source–drain spacing l is 1.5 microns and the oxide thickness t is 400 angstroms (100 angstroms is equal to 0.01 micron). For this reason, the potential in the channel at the surface of the silicon, just under the gate, is dominated by the gate voltage. There are fringe effects of the source and drain voltages, which will be discussed in Appendix B. For the moment, we will ignore those effects. As a first approximation, then, we assume that any change in the potential on the gate will be reflected in an equal change in potential in the channel. As the gate potential is lowered, the lowest potential will be at the silicon surface. As this barrier between source and drain is lowered, more charges can flow along the surface from source to drain. The barrier for two gate voltages is shown in Figure 3.4(b). The device is exactly analogous to Dr. Field's hydraulic transistor, described at the beginning of the chapter.

The barrier ϕ_{s} from source to channel is lower than ϕ_{d} from drain to channel, so more charges will be able to surmount ϕ_{s} than can climb over ϕ_{d}. The charge-carrier density at the source end of the channel thus will be larger than that at the drain end of the channel. Current flows through the channel by diffusion, from the region of high density at the source to that of low density at the drain. We will now compute the channel current.

The carrier density N_{s} at the source end of the channel is given by Equation 2.16 (p. 25):

$$N_{\text{s}} = N_0 e^{-\frac{\phi_{\text{s}}}{kT}} \tag{3.4}$$

where N_0 is the carrier density at the Fermi level. A similar relation holds for N_{d}, the carrier density at the drain end of the channel:

$$N_{\text{d}} = N_0 e^{-\frac{\phi_{\text{d}}}{kT}} \tag{3.5}$$

When the transistor was fabricated, there was a built-in barrier ϕ_0 between source and channel. The control of this barrier is the most crucial element of a high-quality processing line. As the gate potential is lowered, the barrier will be lowered accordingly:

$$\phi_{\text{s}} = \phi_0 + q(V_{\text{g}} - V_{\text{s}}) \tag{3.6}$$

$$\phi_{\text{d}} = \phi_0 + q(V_{\text{g}} - V_{\text{d}}) \tag{3.7}$$

We can now write the barrier energies in Equations 3.4 and 3.5 in terms of the source, gate, and drain voltages:

$$N_{\text{s}} = N_0 e^{-\frac{\phi_0 + q(V_{\text{g}} - V_{\text{s}})}{kT}} \tag{3.8}$$

$$N_{\text{d}} = N_0 e^{-\frac{\phi_0 + q(V_{\text{g}} - V_{\text{d}})}{kT}} \tag{3.9}$$

From Equations 3.8 and 3.9, we can compute the gradient of carrier density with respect to the distance z along the channel (z is equal to zero at the source). Because no carriers are lost as they travel from source to drain, the current is the same at any z, and the gradient will not depend on z. The density thus will be a linear function of z:

$$\frac{dN}{dz} = \frac{N_{\text{d}} - N_{\text{s}}}{l} = \frac{N_1}{l} e^{-\frac{qV_{\text{g}}}{kT}} \left(e^{\frac{qV_{\text{d}}}{kT}} - e^{\frac{qV_{\text{s}}}{kT}} \right) \tag{3.10}$$

where $N_1 = N_0 e^{-\phi_0/(kT)}$.

The electrical current is just the total number of charges times the average diffusion velocity given in Equation 2.10 (p. 23). The current per unit channel width w is thus

$$\frac{I}{w} = qNv_{\text{diff}} = -qD\frac{dN}{dz} \tag{3.11}$$

where D is the diffusion constant of carriers in the channel. Substituting Equation 3.10 into Equation 3.11, we obtain the general form of the MOS transistor current:

$$I = I_0 e^{-\frac{qV_g}{kT}}\left(e^{\frac{qV_s}{kT}} - e^{\frac{qV_d}{kT}}\right) \tag{3.12}$$

The accumulated preexponential constants have been absorbed into one giant constant I_0.

For a transistor with its source connected to the power supply rail, V_s is equal to zero, and Equation 3.12 becomes:

$$I = I_0 e^{-\frac{qV_{gs}}{kT}}\left(1 - e^{\frac{qV_{ds}}{kT}}\right) \tag{3.13}$$

where V_{gs} and V_{ds} are the gate-to-source and drain-to-source voltages, respectively.

Because there are charge carriers with positive as well as negative charge, there are two kinds of MOS transistor: Those using electrons as their charge carriers are called **n-channel**, whereas those using holes are called **p-channel**; the technology is thus called **complementary MOS**, or CMOS. For positive q (p-channel device), the current increases as the gate voltage is made negative with respect to the source; for negative q (n-channel device), the opposite occurs. kT is the thermal energy per charge carrier, so the quantity kT/q has the units of potential; it is called the **thermal voltage**, and its magnitude is equal to 0.025 volt at room temperature. A carrier must slide down a potential barrier of kT/q to raise its energy by kT. As we noted in Chapter 2, electrochemists and biologists write RT/F in place of kT/q. The way it is written does not change its value.

We have made a number of simplifying assumptions, which will be addressed in Appendix B. Equation 3.12, however, captures all the essential quantitative principles of transistor operation. Notice that, in Equation 3.10, the roles of the source and of the drain are completely symmetrical; therefore, if we interchange them, the magnitude of the current given by Equation 3.12 is identical, with the current flowing in the opposite direction.

The energy diagrams for both types of transistors are identical to that shown in Figure 3.4, but the energy axis has a different meaning. For a p-channel device, upward means higher energy for positive charges, or positive voltage. For an n-channel device, upward means higher energy for negative charges, or negative voltage. Opposite charges attract. An n-channel device requires positive gate voltages to attract negative electrons out of its source into its channel; a p-channel device requires negative gate voltages to attract positive holes out of its source into its channel.

CIRCUIT PROPERTIES OF TRANSISTORS

The symbols used in schematic diagrams for both n- and p-channel transistors are given in Figure 3.5, which shows the source, gate, and drain terminals. We put a bubble on the gate of the p-channel symbol to remind us that the transistor turns on as we make the gate more negative relative to the source. We normally will draw the positive supply at the top of the diagram, and the most negative supply at the bottom. For this reason, the *sources* of p-channel devices usually are located at the *top*, whereas those of n-channel devices normally are at the *bottom*. Implicit in the schematic is a shadow of the Boltzian landscape, with upward meaning positive voltage, as in the energy diagram for p-channel transistors. Positive current (the flow of positive charges) is from high to low. Ground, the reference level, is the most negative supply, or sea level, for positive charges.

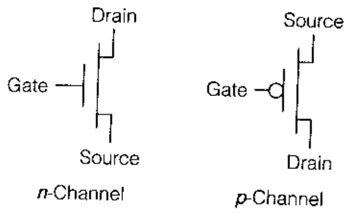

FIGURE 3.5 Circuit symbols for n- and p-channel transistors. These abstractions are used in schematic diagrams of transistor circuits. Note that the assignment of source and drain depends on the voltages to which these terminals are connected, because the physical device structure is symmetrical.

The measured current–voltage characteristics of a typical transistor are shown in Figure 3.6. The drain current is zero for $V_{ds} = 0$, as expected. For a given gate voltage, the drain current increases with V_{ds} and then saturates after a few kT/q, as predicted by Equation 3.13. The current in the flat part of the curves is nearly independent of V_{ds} and is called the **saturation current**, I_{sat}. A plot of the saturation current as a function of gate voltage V_{gs} is shown in Figure 3.7. I_{sat} increases exponentially with V_{gs} as predicted by Equation 3.13,

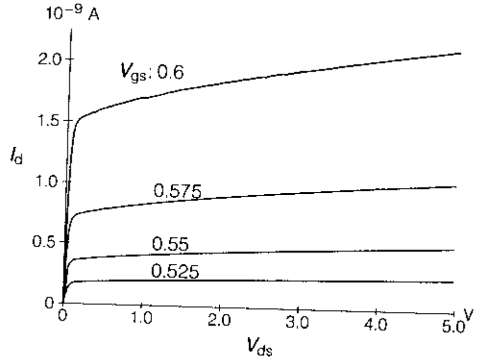

FIGURE 3.6 Drain current as a function of drain–source voltage for several values of gate–source voltage. The channel length of this transistor is 6 microns. The drain current increases rapidly with drain voltage, and saturates within a few kT/q to a gently sloping region of nearly constant current. The slope in the saturation region is due to the change in channel length with drain voltage. The slope in this region is proportional to drain current.

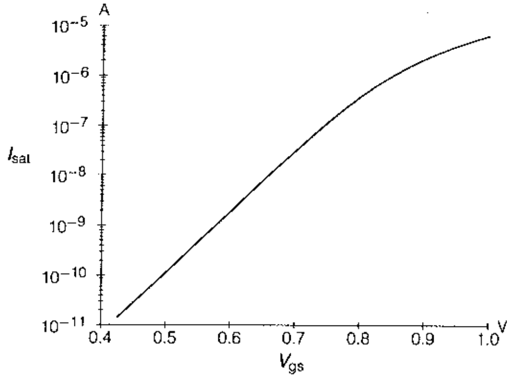

FIGURE 3.7 Saturation current of the transistor of Figure 3.6 as a function of gate voltage. The drain voltage was fixed at 2 volts. The current is an exponential function of gate voltage over many orders of magnitude. In this region, the current increases by a factor of e every 37 millivolts, corresponding to $\kappa = 0.676$. Above approximately 0.8 volt, the limiting effects of mobile charge on the current are evident. The nominal "threshold" of the device is approximately 0.9 volt.

but the voltage required for a factor of e increase in I_{sat} is 37 millivolts, rather than the 25 millivolts we expected.

In our simplified derivation, we assumed that the gate voltage was 100-percent effective in reducing the barrier potential. This assumption is valid for a structure built on an intrinsic substrate, as shown in Figure 3.4. Real transistors, such as those from which the data of Figure 3.7 were taken, are not built on intrinsic substrates. The n-channel transistors are fabricated on p-type substrates, and vice versa. Charges from the ionized donors or acceptors in the substrate under the channel reduce the effectiveness of the gate at controlling the barrier energy (see Appendix B). For our purposes, the effect can be taken into account by replacing kT/q by $kT/(q\kappa)$ in the gate term in Equation 3.12, and rescaling I_0. Equation 3.12 can thus be written

$$I = I_0 e^{-\frac{q\kappa V_{\text{g}}}{kT}} \left(e^{\frac{qV_{\text{s}}}{kT}} - e^{\frac{qV_{\text{d}}}{kT}} \right) \tag{3.14}$$

The transistor shown has a value of κ approximately equal to 0.7. The values of κ can vary considerably among processes, but are reasonably constant among transistors in a single fabrication batch. Throughout this book, we will treat kT/q as the unit of voltage. Because this quantity appears in nearly every expression, we have developed a shorthand notation for it. If the magnitude of kT/q is used to scale all voltages in an expression—as, for example, when a voltage appears as the argument of an exponential—we often will write the voltage as though it were dimensionless. Using this notation, we can write Equation 3.14 for an n-channel transistor as follows:

$$I = I_0 e^{\kappa V_{\text{g}}} \left(e^{-V_{\text{s}}} - e^{-V_{\text{d}}} \right) = I_{\text{sat}} \left(1 - e^{-V_{\text{ds}}} \right) \tag{3.15}$$

Where V_{ds} is the drain–source voltage. For a p-channel device, the signs of all voltages are reversed.

At the upper end of the current range of Figure 3.7, the current increases less rapidly than does the exponential predicted by Equation 3.15. This deviation from the exponential behavior occurs when the charge on the mobile carriers becomes comparable to the total charge on the gate. The gate voltage at which the mobile charge begins to limit the flow of current is called the **threshold voltage**. For gate voltages higher than threshold, the saturation current increases as the square of the gate voltage. Most circuits described in this book operate in **subthreshold**—their gate voltages are well below the threshold voltage. Typical digital circuits operate well above threshold. The detailed model described in Appendix B describes transistor characteristics over the entire range of operation.

Subthreshold operation has many advantages, three of which we are now in a position to appreciate:

1. Power dissipation is extremely low—from 10^{-12} to 10^{-6} watt for a typical circuit
2. The drain current saturates in a few kT/q, allowing the transistor to operate as a current source over most of the voltage range from near ground to V_{DD}
3. The exponential nonlinearity is an ideal computation primitive for many applications

We will encounter many more beneficial properties, and some limitations, of subthreshold operation as we proceed.

CURRENT MIRRORS

An often-used circuit configuration is shown in Figure 3.8. Here, each transistor is **diode-connected**; that is, its gate is connected to its drain. For a typical process, values of I_0 are such that even the smallest drain current used (10^{-12} amp) requires V_{gs} approximately equal to 0.4 volt. Higher drain currents

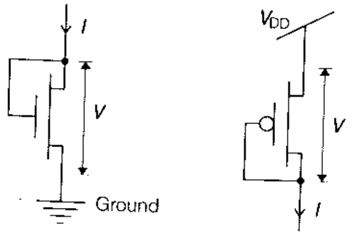

FIGURE 3.8 Diode connected n- and p-channel transistors. Because the drain-source voltage is always a few hundred millivolts, a device in this configuration is guaranteed to be saturated. The current-voltage characteristic is thus exponential, like that of Figure 3.7.

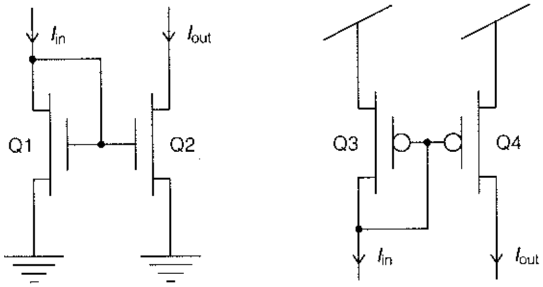

FIGURE 3.9 Current mirror connected *n*- and *p*-channel transistors. This configuration turns a source of current (the input) into an equal *sink* of current (the output). Reflecting currents in this manner is the most common operation in analog circuit design.

require higher values. Thus, the drain curves of Figure 3.6 are well into saturation for any useful V_{gs}. For this reason, the current through these diode-connected transistors has the same exponential dependence on voltage as that shown for the saturation current in Figure 3.7.

It is common to have a current of a certain sign—for example, a source of positive charges—and an input that requires an equal but opposite current—for example, a source of negative charges. A simple circuit that performs this inversion of current polarity is shown in Figure 3.9. The input current I_{in} biases a diode-connected transistor Q1. The resulting V_{gs} is just sufficient to bias the second transistor Q2 to a saturation current I_{out} equal to I_{in}. The value of I_{out} will be nearly independent of the drain voltage of Q2 as long as Q2 stays in saturation. A similar arrangement is shown for currents of the opposite sign using p-channel transistors Q3 and Q4. The p-channel circuit *reflects* a current to ground into a current from V_{DD}, so it is called a **current mirror**. The n-channel current mirror reflects a current from V_{DD} into a current to ground. We will use these circuit configurations in nearly every example in this book.

SUMMARY

We have seen how we can construct a physical structure that allows a voltage on one terminal to control the flow of current into another terminal. Although the first proposal for a device of this type was made in the 1930s [Lilienfeld, 1926], it took 3 decades to reduce the ideas to a production process. By the 1960s, MOS technology came into its own—today, it is the major technology on which the computer revolution has been built. For our purposes, MOS transistors are controlled sources of both positive and negative current. Their control terminals do not draw current from the nodes to which they are connected. MOS transistors are, in that sense, the most ideal active devices extant. The exponential dependence of drain current on gate voltage allows us to control current levels over many orders of magnitude. We will develop increasingly complex configurations of these simple elements, culminating in complete neural subsystems for vision and hearing.

REFERENCES

Lilienfeld, J.E. Method and apparatus for controlling electric currents. Patent (1,745,175: January 28, 1930): October 8, 1926.

Lilienfeld, J.E. Device for controlling electric current. Patent (1,900,018: March 7, 1933): March 28, 1928.