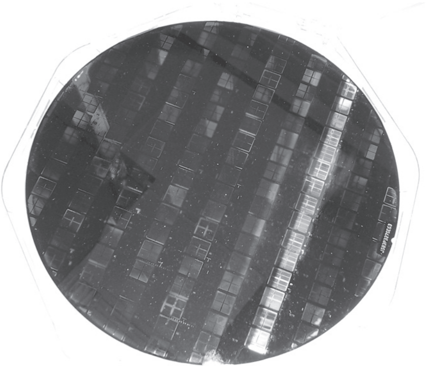
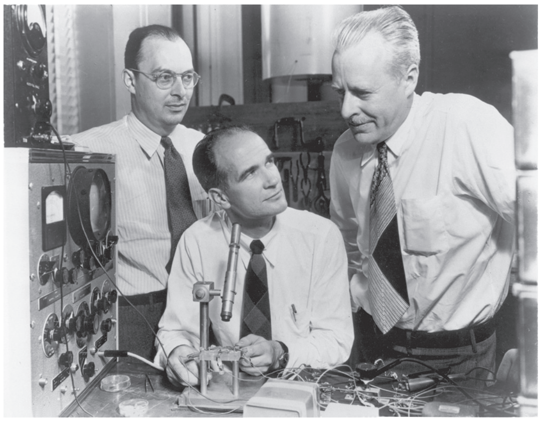
10

Semiconductor Physics: Transistors and Circuits

> The general aim of the ... research program on semiconductors initiated at the Bell Telephone Laboratories in early 1946 ... was to obtain as complete an understanding as possible of semiconductor phenomena, not in empirical terms, but on the basis of atomic theory.
>
> **John Bardeen**
> *(Nobel Lecture,* 1956*)*



Eight-inch Si wafer containing hundreds of integrated circuits, each containing millions of transistors.



John Bardeen, Walter Brattain, and William Shockley (left to right), the inventors of the first practical transistor at AT&T Bell Laboratories in 1947. (With permission of Lucent Technologies Inc./Bell Labs.)

10.1 SILICON, TRANSISTORS, AND COMPUTERS

As carbon is the building block for living things, silicon (Si) is the building block for information technology. Si-based semiconductors are the basis of most modern electronic circuits. It is abundant (from sand), and it can be purified to an extremely high degree. ***Semiconductor*** devices can be designed to carry out switching and gate operations. That is, semiconductors can be turned into conductors when needed, and then turned back into insulators, as in turning a water faucet on and off. This chapter explains how this is done. Other technologically important semiconductors are germanium, gallium arsenide, gallium phosphide, and indium arsenide. These have special uses, such as making lasers and light-emitting diodes, discussed in Chapter 12.

The understanding of semiconductors, and Si especially, in the 1940s led to a rapid development of ***solid-state electronics***—the basis of most computer and Internet technology. *Solid-state* refers to the use of crystals as electronic components in circuits. This development culminated in the development in 1947 of the first practical semiconductor transistor by John Bardeen, Walter Brattain, and William Shockley at AT&T Bell Laboratories. Although other scientists before Bardeen, Brattain, and Shockley had accomplished closely related work, these three were recognized widely as having made the big breakthrough, and were awarded the Nobel Prize in 1956 for the invention of the transistor.

As Bardeen said in his Nobel lecture, the immediate goal of the research he and his team were doing was to understand the basic physics of semiconductor crystals, on the basis of the theories of atomic physics that had been developed not long before. They also paid close attention to the practical implications of their newly found physics understanding, and, using what they had learned, built the world's first transistor.

It was immediately clear to physicists and engineers that transistors had important applications as electronic switches and as power amplifiers. Nevertheless, it took 11 years from the invention of the transistor to the making of the first miniaturized, or integrated, transistor circuit. This was accomplished by Jack Kilby at Texas Instruments (TI) in 1958 and started the digital revolution. Computer processors and memory are based on ***integrated circuits*** (ICs). ICs contain hundreds of millions of transistors packed into an area less than the size of a small coin.

In this chapter, we will explore how semiconductor crystals are modified by adding impurities, and how this leads to transistor action—the ability to control an electric voltage by applying a second, smaller voltage. This enables electronic switching and amplification and other operations useful in communication technology.

10.2 CONTROLLING THE CONDUCTIVITY OF SILICON

As we learned in the previous chapter, pure Si crystals are not good conductors. On the other hand, various elements can be put into an otherwise pure Si crystal to alter its properties. This process is called ***doping***, and the added element is called an ***impurity*** or a ***dopant***. In a typical doped crystal, there is about one "impurity atom" for every 1 million Si atoms. From what we have learned about energy bands and the principles of quantum theory, we can understand how doping changes the electrical conductivity of the crystal.

Figure 10.1 shows a schematic diagram of a crystal of pure Si, with each Si atom having four outer-shell ***electrons***. Each Si atom is surrounded by eight outer-shell electrons—four from itself and one each from its four nearest neighbors. The crystal is neutral.

Recall (Chapter 9, Section 9.7) that at very low temperature, a pure Si crystal cannot conduct electricity because its energy bands are all either empty or completely filled. At room temperature, there is very small conductivity, but let us ignore that for now. Also shown in Figure 10.1 is a Si crystal with a very small amount of phosphorous (P) added as an impurity. The P atoms replace a small fraction of the Si atoms, which would otherwise be at those locations. Each P atom has five outer electrons (see Figure 9.19). This means that surrounding each P atom that has replaced a Si atom there is one extra electron that would not be present in a pure Si crystal. This electron is indicated

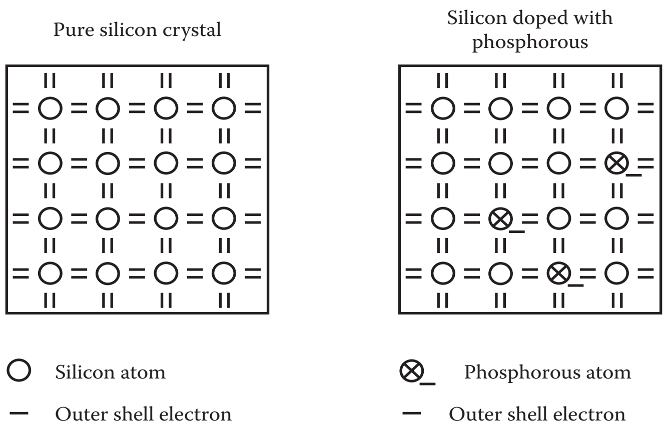


FIGURE 10.1 Pure Si crystal (on left) with four outer-shell electrons (–) for each Si atom (○). On the right is a Si crystal that is weakly doped with P atoms (⊗), having five outer-shell electrons. This creates a semiconductor.

in the drawing as an extra (–) symbol. Note that the crystal is still neutral because the P atom also has one excess proton in its nucleus, as compared with Si.

Where does the P atom's excess electron go? To answer this we must discuss both its location and its energy. Recall (Chapter 9, Section 9.7) that in a pure Si crystal there is a small energy band gap just above the highest filled energy shell. The excess electron from the P atom cannot be in the same energy band as the other electrons because that band is completely filled, and the Exclusion Principle forbids adding one more, because this would require two electrons being in the same state. So, the excess electron has no choice but to go into the next higher energy band—the conduction band, which was empty in the pure Si crystal (at low temperature). This is shown in **Figure 10.2** as the black dot in the conduction band, just above the small gap. Si that is doped with P is called an ***n-type semiconductor*** because there are extra electrons, which carry negative (*n*) charge, compared to an undoped crystal. The crystal is, nevertheless, still neutral.

The excess electrons in a phosphorous-doped (n-type) Si crystal make the crystal highly conductive. Because the excess electrons are in a partly empty energy band, they are mobile; that is, able to move. When a low voltage is applied between two ends of the crystal, the electrons in the partly filled band can easily be accelerated, gradually raising them in energy into unoccupied energy states, as indicated by the small arrow in Figure 10.2. This acceleration corresponds to electrons moving through the crystal, as shown in **Figure 10.3**. Thus, a phosphorous-doped Si crystal is a good conductor.

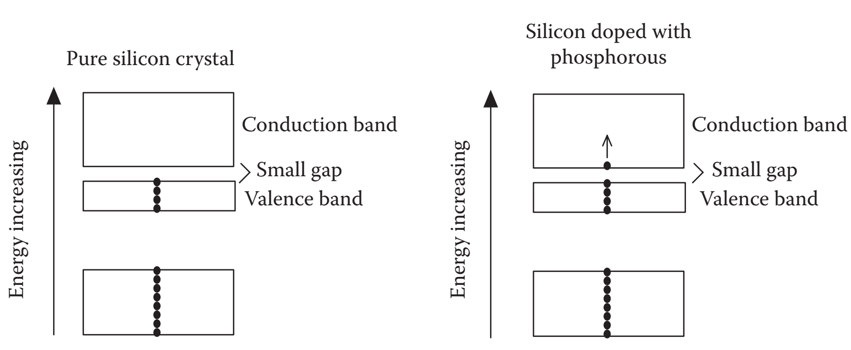


FIGURE 10.2 Energy bands and electron occupation for a pure Si crystal (left) and for a phosphorous-doped Si crystal (right). The temperature is assumed to be very low. The excess electrons from the doping go into a partly empty band, making them able to conduct electrical current. This is an n-type semiconductor.

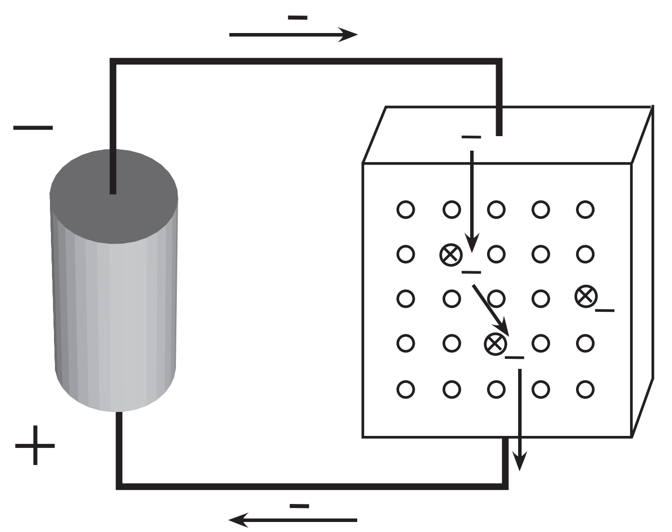


FIGURE 10.3 A phosphorous-doped Si crystal is an n-type semiconductor, which conducts current when a voltage is applied.

Notice that for every electron that moves into the crystal from the wire, one electron leaves the crystal at the other side. The arrows shown inside the crystal indicate the coordinated replacing of each electron by the electron coming behind it. Therefore, there is no charge buildup, and the crystal remains neutral.

A simple model for electron conduction in an n-type semiconductor is a ski lift with every seat occupied, taking people down a mountain (see **Figure 10.4**). As one person enters a lift seat at the top of the mountain, another person leaves a seat at the bottom. In a doped semiconductor, the partly filled band acts like the ski lift: as one electron enters one side of the crystal to occupy an energy level in the partly filled band, another electron leaves the crystal at the other side.

Doping with P atoms is not the only way to make Si conductive. By instead adding boron (B) atoms to a pure Si crystal, we also can make a Si crystal conductive. A boron atom has three outer electrons, as shown earlier in Figure 9.19. So, when a B atom replaces a Si atom in the crystal, it contributes one fewer electron in its outer shell than a Si atom would. This results in the absence of one outer-shell electron in the region near the B atom, shown as the small empty box in **Figure 10.5**. This small empty region is called a ***hole***. A hole is a place in a crystal where an electron normally would be, but is not. Boron doping creates a crystal that is deficient in outer-shell electrons. This is called a ***p-type semiconductor***, because a lack of negative charge appears like

FIGURE 10.4 Electrons traveling in a P-doped Si crystal are analogous to people traveling down a ski lift.

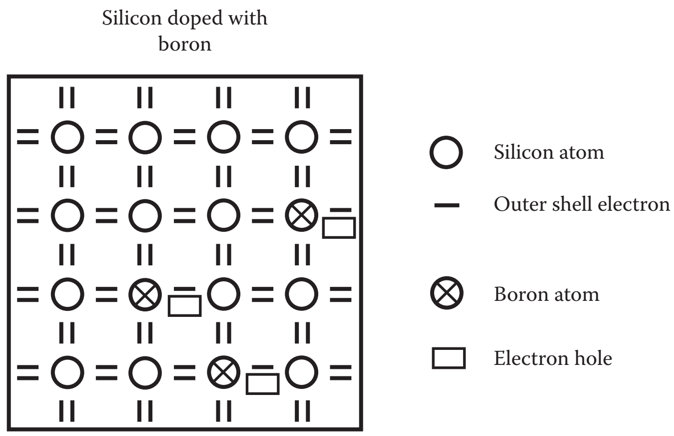


FIGURE 10.5 Si crystal doped with boron, which leaves a "hole" in the electron outer shell near the boron atoms. This creates a p-type semiconductor.

a positive charge. Again, as in the case of doping with P atoms, the crystal as a whole is still neutral, because the boron atom has one fewer proton (+ charge) in its nucleus than does a Si atom.

THINK AGAIN

Do not confuse the symbol "p" in the term *p-type* (positive-type) with the symbol P, which means the P (or phosphorous) atom.

How does this absence of an electron, or presence of a hole, affect the electrical conductivity? There is now a partly empty energy band, with a hole in it, as shown by the small empty rectangle in **Figure 10.6**. The presence of holes in the crystal's energy bands allows current to flow easily.

There is an electron in Figure 10.6 that has energy slightly below the energy of the empty hole. If a voltage is applied to the crystal, this electron can be accelerated. As it accelerates, it gains energy and moves up into the energy level that the hole formerly occupied. This in turn leaves a hole at a lower energy, as shown in **Figure 10.7**. The acceleration of electrons into holes allows electron current to flow through a p-type doped crystal. Note that if there were no energy hole, no electrons could accelerate under the influence of a voltage, because there would be no higher available energy levels for them to go into.

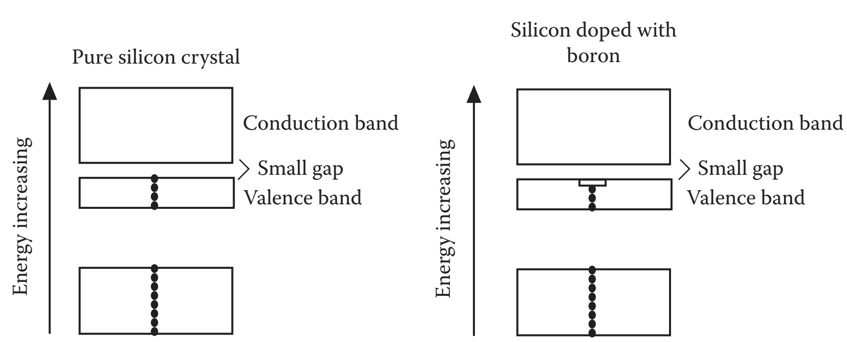


FIGURE 10.6 Energy bands and electron occupation for pure Si crystal (left) and boron-doped (p-type) Si crystal (right). Low temperature is assumed. The deficit of electrons from the doping creates a partly empty (not filled) band, making the crystal able to conduct electrical current. This is a p-type semiconductor.

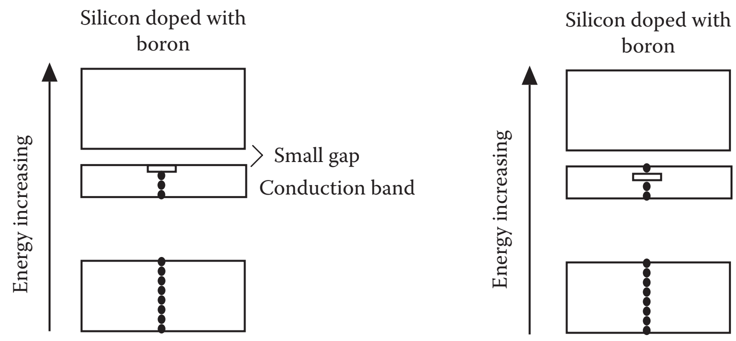


FIGURE 10.7 Electron occupation for boron-doped Si crystal (left), showing a hole in the highest-energy occupied band. Under the influence of an applied voltage, an electron gains energy and moves into the empty hole, causing a new hole just below the energy of the former hole.

10.3 P-N JUNCTIONS AND DIODES

The first solid-state devices used in electronics were diodes, used as radio receivers. A ***diode*** is similar to a one-way turnstile for electrons. Electrons can flow in one direction through a diode, but not in the other direction, as in **Figure 10.8**.

The earliest diodes were constructed by making contact between the end of a fine metal wire and a piece of semiconductor. Later, diodes were greatly improved by using all semiconductor materials in their construction. They play an important role in modern electronic circuits. Understanding diode operation will prepare us to understand the operation of transistors—the most important component in modern computer circuits.

A diode is made of two joined semiconductor crystals—one p-type and the other n-type, as in **Figure 10.9**. The region just around the contact is called a ***p-n junction***. The figure shows the two pieces before and after being joined. In the n-type piece, excess electrons, from doping, are shown as minus symbols (–), and in the p-type piece, holes are shown by the box symbols. All other symbols such as the Si, P, and B symbols are not shown, nor are the Si outer-shell electrons. Remember that all of these are present in the crystals, making each crystal neutral (zero total charge) before they are joined.

After joining, the crystals lose their neutral property in a small region around the junction. Charge builds up as some of the excess electrons flow from the n-type piece into the p-type piece, where they occupy former holes. Why do they flow that way? There are no external forces on these electrons initially to make them flow one way or the other, but they wander, or diffuse. This motion is a natural consequence of their random jiggling motion due to thermal energy at room temperature. Both pieces are initially good conductors because they are doped, so electrons can move through them.

Diffusion is the random wandering of particles throughout some region. You can think of electrons as being like salmon in a large lake, wandering randomly from place to place, as in **Figure 10.10**. If the lake has no outlet, the salmon will spread out more or

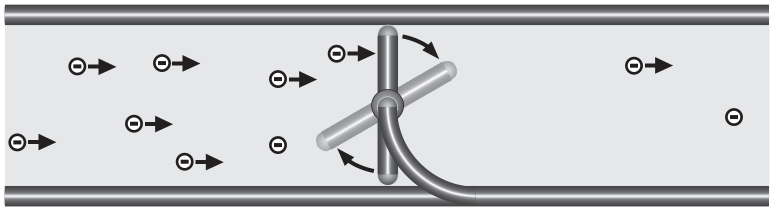


FIGURE 10.8 A good analogy for a diode is a one-way turnstile. A bar can rotate clockwise, but not counterclockwise, allowing electrons (or people) to pass only from left to right.

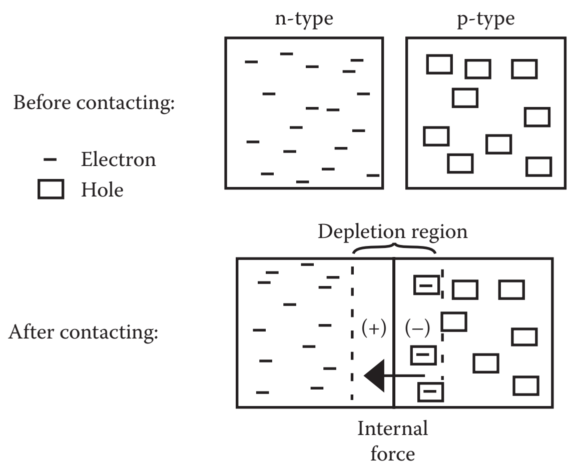


FIGURE 10.9 A p-n junction before and after contacting. A few electrons diffuse into the p-type side until the internal force on electrons, caused by the resulting charge imbalance, holds back any further net charge drift. The joined crystals comprise a diode.

less evenly throughout the lake. The average density of salmon stays about the same in each area. Then, if a channel is opened up into a second, empty lake, salmon will tend to wander at random into that second lake until they are spread out roughly evenly in both lakes. After some time, there will be on average as many salmon leaving the first lake as entering it, so that the number of salmon in each lake stays roughly constant.

Electrons are charged negatively, so if many electrons diffuse into the p-type semiconductor crystal, then the part of that crystal nearest the junction will become charged negatively, shown by the (–) symbol in Figure 10.9. Then the part of the n-type crystal nearest the junction (initially neutral), from which the electrons have departed, will become charged positively, shown by the (+) symbol. There is a thin region (about 10 μm in length) around the junction where the n-type piece is positively charged and the p-type piece is negatively charged. This region is called the ***depletion region***. In regions away from the depletion region, the crystals are neutral, as usual.

The result of the diffusion of electrons is that in the depletion region the two pieces of crystal act as a battery, with one side of this region becoming positive and the other side negative. We know that the negatively charged region (the p-type side) will repel electrons, and the positively charged side (the n-type side) will attract electrons. This effect is called the internal battery created at the junction of the p-type and n-type crystals. This internal battery creates a force on the electrons that pushes them toward the n-type side. In Figure 10.9, the arrow indicates this internal force on electrons. This causes the

QUICK QUESTION 10.1

If you put a drop of food coloring into still water, diffusion takes place. Explain.

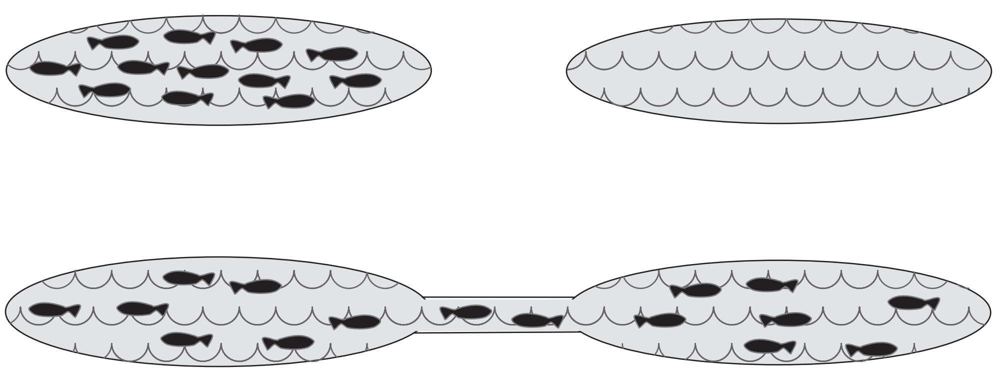


FIGURE 10.10 Salmon swimming between two lakes as an analogy of diffusing electrons.

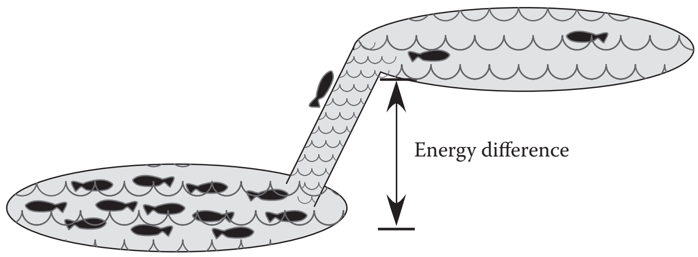


FIGURE 10.11 Salmon swimming in two lakes connected by a waterfall. Fast fish can go up. Slow fish cannot. Any fish can go down.

potential energy of electrons to be higher on the p-side of the diode. That is, it takes work to push an electron "uphill" from the n-side to the p-side, against the force.

In our salmon analogy, the effect of the internal battery is like having a waterfall between the two lakes, as in **Figure 10.11**. The higher lake is a region of higher potential energy. To reach the higher lake, salmon have to swim swiftly, energetically jumping up the waterfall. Once they reach the higher lake, they can stay there with no effort. But, if they happen to wander to the top of the falls, they may fall down into the lower lake. The waterfall is an impediment only in the upward direction. If the waterfall is high, then there are few salmon in the lower lake energetic enough to make the jump into the higher lake. On the other hand, any salmon in the higher lake can easily fall into the lower lake, regardless of how much energy it has. After some time, there will be many slow-moving salmon in the lower lake, swimming around but not able to make the jump, and a few fast salmon waiting to make the jump. In the upper lake, there will be very few salmon. Therefore, the rate of fish going down the falls is small. There will be roughly equal numbers going up and down, so that the number in each lake will not be changing. There is no net flow of fish because the tendency for upward diffusion is balanced by the tendency for downward travel between the lakes.

If we were to disturb the balance between upward diffusion of the salmon and the downward force associated with the waterfall, then a net flow, or "current," of salmon would result. Raising or lowering the height of the upper lake relative to the lower lake would disturb this balance. **Figure 10.12a** again shows the two lakes with relative

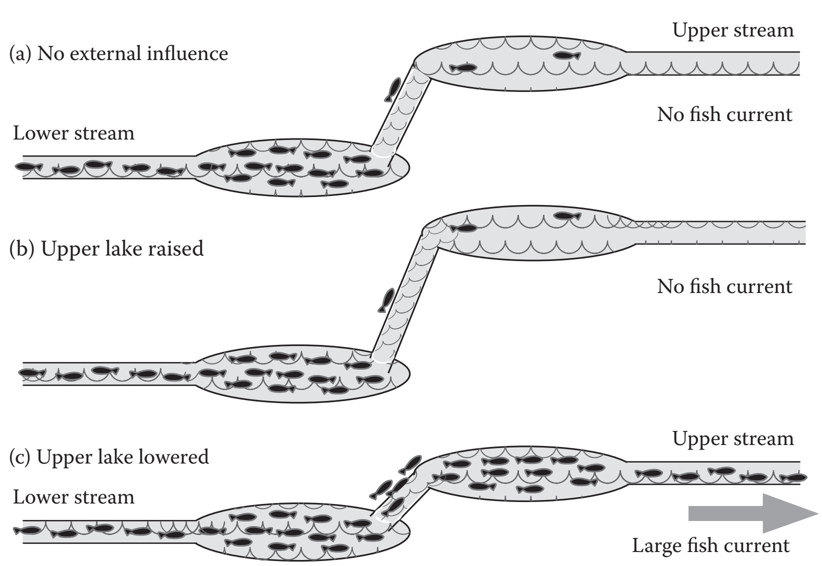


FIGURE 10.12 Analogy for the operation of a diode. Lowering the height of the upper lake allows large numbers of not-so-energetic salmon to leap up the falls and swim away in the stream to the right.

height such that there is a balance between upward diffusion and the downward force, so there is zero net fish exchange between lakes. In this case, there is zero net fish flow in the two streams going in and out of the lakes. If the upper lake is then raised to a higher level than before, as in Figure 10.12b, fish will have an even harder time jumping up the waterfall, so there is still zero or very little net fish current. On the other hand, if the upper lake were dropped to a lower level than before, as in Figure 10.12c, there is a dramatic effect: now a very large number of salmon, moving at moderate speeds, can make it up the falls, whereas they were not able to make it up when the lakes were at their original heights. Because the rate of salmon going down the falls is not much affected by lowering the upper lake, there is now a large net rate of salmon going from the lower lake into the higher lake. This results in a large net number of fish traveling from the lower stream to the upper stream. There is now a large net "current" of fish moving from left to right in the figure.

The same happens with electrons moving in a diode. In Figure 10.9, the force from the internal diode battery (likened to the waterfall) holds the slow, low-energy electrons back from diffusing "up" to the p-type side from the n-type side, where the potential energy is higher. Only a few energetic electrons can do so, depending on the temperature. Recall the discussion around Figure 9.28. Any electron in the p-type side can easily diffuse back into the n-type side. A balance gets established between electron diffusion and the internal-battery force caused by those same electrons. A brief time after joining the semiconductor crystals, these processes come into equilibrium, and after that time there is no further net flow of electrons.

If one disturbs this delicate balance between upward diffusion and the force from the internal battery, then current flows. This can be done by connecting an external battery to the two ends of the joined crystal, as in **Figure 10.13.** If, as in part (a), we connect the (–) side of the battery to the p-type side of the diode, the battery gives an increased potential energy to electrons at the p-side, analogous to raising the height of the upper lake in the salmon example. This slows down the already small rate of diffusion from the n-type to the p-type side, but does not affect the diffusion rate in the other direction. This leads to a very small net electron current from the (–) side of the battery to its (+) side. This current is so small that we will say it is zero for simplicity of discussion. We refer to this manner of attaching the battery as being in the backward direction.

On the other hand, if, as in Figure 10.13b, we connect the (+) side of the battery to the p-type side of the diode, then the battery decreases the potential energy of electrons

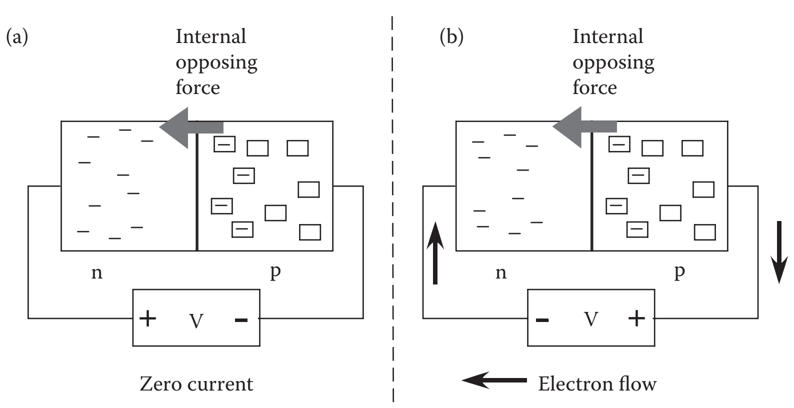


FIGURE 10.13 In a diode, the internal force opposes electron flow. An external battery adds a force on electrons pushing from the wire at the (-) side of the battery toward the wire at the (+) side of the battery. Only if the (+) side of the battery is connected to the p-type side of the diode will the battery add a force on electrons that counteracts the internal force and allows electrons to flow.

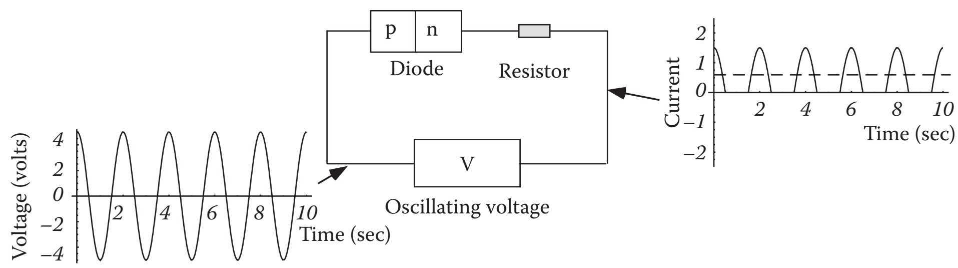


FIGURE 10.14 A p-n junction diode acting as a voltage rectifier. When the oscillating source V (the signal) creates a plus voltage on the left side of the diode, the diode conducts current from left to right, sending current through the resistor. In contrast, when the source creates minus voltage on the left side, the diode blocks the current flow through the resistor.

at the p-side, analogous to lowering the height of the upper lake in the salmon example. This allows a much larger number of the moderate-energy electrons to diffuse from the n-type side to the p-type side. This results in a large net electron flow from the (–) side of the battery to its (+) side. This explains why a diode acts as a one-way turnstile for electrons. To summarize, current flows through a diode only if the battery's plus voltage is connected to the p-type side and the minus voltage is connected to the n-type side. We call this applying voltage in the forward direction.

THINK AGAIN

Keep in mind the distinct meanings of the terms *p-type* (positive-type) and *n-type* (negative-type), which refer to types of crystals, and *plus* (+) and *minus* (-), which refer to polarities of voltages produced by batteries.

10.3.1 Rectifying an Alternating Signal

A diode acts like a one-way turnstile for electrons. The conductivity in the forward direction is large, whereas the conductivity in the backward direction is nearly zero. An application of a diode is as a ***rectifier***—a device that converts an oscillating signal (containing plus and minus voltages) into a positive-only signal. Rectifiers are a key component in radio receivers, discussed further in Real-World Example 10.1.

Figure 10.14 shows a circuit with an oscillating signal voltage, V, connected to a diode and a ***resistor*** in series. When the voltage is positive, as shown, current flows through the diode and the resistor. When the voltage is negative, the current equals zero.

REAL-WORLD EXAMPLE 10.1: A SIMPLE CRYSTAL AM RADIO RECEIVER

If the oscillating voltage shown in Figure 10.14 is replaced by an antenna, and the resistor is replaced by an audio speaker, a simple radio receiver is created, as shown in **Figure 10.15**. The other connection of the speaker is attached to a metal wire that is inserted into the ground or earth outside (or a piece of indoor plumbing that is connected to the earth). The antenna is a long metal wire, inside of which a time-varying current is produced by an incoming radio wave. The incoming wave labeled (a) in **Figure 10.16**, is a carrier

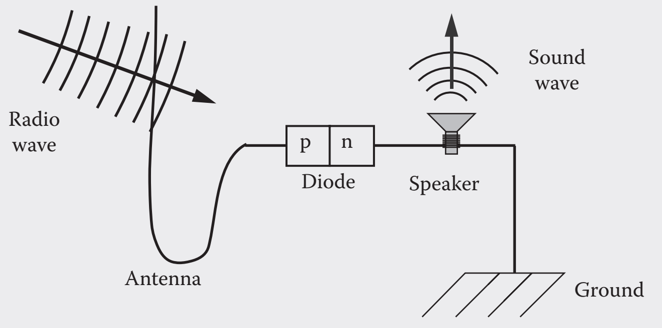


FIGURE 10.15 A simple crystal radio receiver, made of an antenna, a diode, and a small speaker.

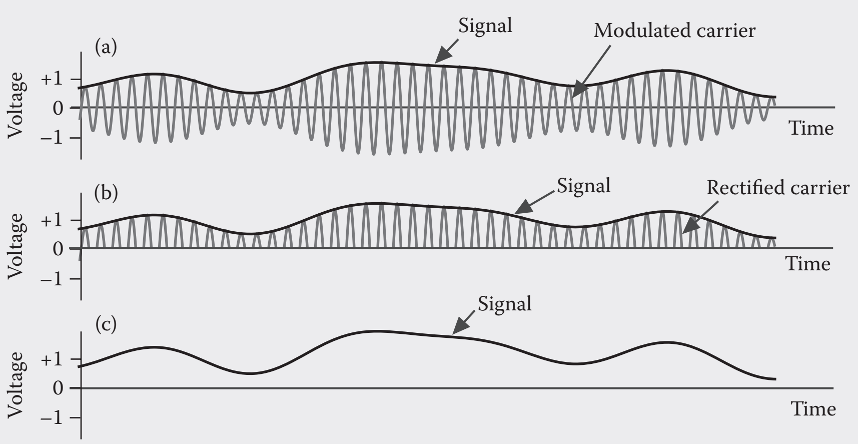


FIGURE 10.16 (a) The incoming radio wave is a carrier wave modulated by an audio signal. (b) After the diode rectifies the incoming signal, only positive voltages remain. (c) The slow response of the speaker cone follows the original audio signal.

wave (e.g., at 650 kHz), the ***amplitude*** of which is modulated by an audio signal. For details about modulation, see Chapter 8, Section 8.2. The diode rectifies the oscillating voltage induced in the wire by the radio wave to produce the wave shown in part Figure 10.16b. The electromagnet in the speaker is connected to the output voltage of the diode, which rectifies the incoming signal voltage. The rectified voltage creates a current in the speaker coil, which drives the motion of the speaker cone. The magnet and cone cannot rapidly respond to the kHz carrier wave of the radio signal; instead, they move in proportion to the slower-changing amplitude of the rectified signal. Without the diode, the slow response would average to zero, and the cone would not move. Changes of the amplitude of the radio wave cause corresponding changes in the amplitude of the rectified voltage driving the speaker. This modulates the cone position, creating a sound wave, labeled (c) in Figure 10.16. No battery or other external power source is needed, because the radio wave itself can deliver enough power to drive a small speaker such as an earphone.

10.4 TRANSISTORS

Now we can use our understanding of diodes to understand transistors—one of the most important devices in computer circuitry. A ***transistor*** is a semiconductor device that acts as a controllable valve for electrons, as a valve or faucet does for water. As in **Figure 10.17**, if a source of water is kept at a higher pressure than the drain, then when

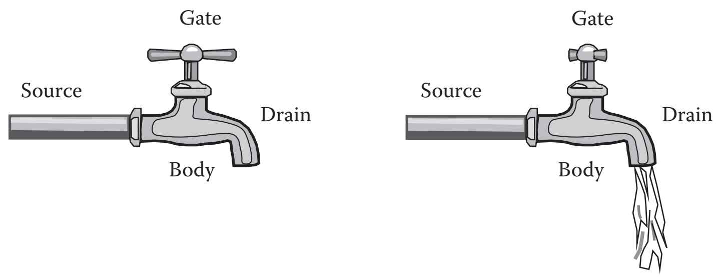


FIGURE 10.17 A transistor acts as a valve for electrons.

the control knob is activated, opening the valve, water flows from the source through the drain. If the drain has higher pressure than the source, backward flow will occur. For transistors, the control knob is called the control gate and voltage is analogous to the water pressure.

Transistors provide the low-power voltage-controlled switches needed to build practical logic circuits for use in computers. The first transistors to be put into practical use were bipolar transistors—composed of a pair of back-to-back diodes. Bipolar transistors—discussed at the end of this chapter—were initially used in semiconductor-based computer circuits, but they use a lot of power to operate. This generates heat in the circuit, and the need to remove this heat limits the speed at which computer chips can be operated reliably. Bipolar transistors were replaced in the late 1970s by the ***field-effect transistor***, or ***FET*** (pronounced f-e-t, or simply "fet"). FETs are preferred for use in computer logic circuits because they use very little electrical power when they switch.

In a ***FET***, the flow of electrons is controlled by an electric field that is applied to a narrow region in a semiconductor device. This narrow region, called an electron channel, conducts electrons only when exposed to the effect of an electric field; hence the name field-effect transistor. A schematic of a FET is shown in **Figure 10.18**. The FET contains two p-n junctions similar to those in Figure 10.9. Two n-type regions are contacted to a p-type region, and there is an ***insulator*** region, labeled *i*, adjacent to the p-region. The insulator is silicon dioxide, SiO_2, or *oxide* for short. The two n-type regions

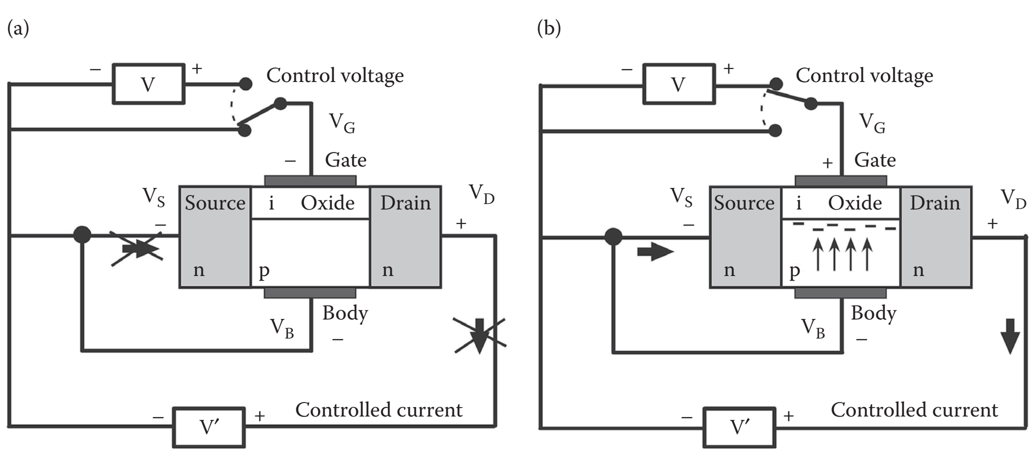


FIGURE 10.18 An n-FET. An insulating oxide layer, labeled *i*, separates the gate from the body. (a) With no voltage applied to the gate, no current can flow from the source to the drain. (b) A positive voltage V applied to the gate creates a conducting n-type channel through the p-type semiconductor, allowing electron current to flow from source to drain and through the battery V′.

are called the source and the drain, reminiscent of water plumbing. In addition, there are two conducting plates (metal or another conducting material[1]). One conducting plate, called the control gate, contacts the insulator region. The other conducting plate, called the body, contacts the p-region. The body is connected to the source through a wire, and to the minus sides of both batteries. This structure is called an n-channel metal-oxide-semiconductor (MOS) device.

Recall that a p-type semiconductor has a deficit of electrons in its outer electron shells; that is, it has electron holes (absences) in its outer shells. An n-type semiconductor has an excess of electrons in its outer shells. At each p-n junction in the FET structure, there is a thin depletion region where all holes are filled by diffusing electrons. Because the p-n junction on the left side of the insulator has zero voltage across it, according to our understanding of diode action, no current flows through this junction. The p-n junction on the right side has a voltage applied to it in the backward direction, so no current flows through it as well. When the control switch in the figure is in the down position, as in Figure 10.18a, the control gate and the body are at the same voltage; that is, there is zero voltage between them. There is no current flowing anywhere in this case.

When the control switch is moved to the up position, as shown in Figure 10.18b, a positive voltage appears at the gate, relative to the body. The electric field from the battery labeled V pushes electrons from its minus side, through the body, into the p-type region, and onto the lower surface of the oxide insulating region. The electrons cannot pass into the insulator, so a negative charge builds up near its surface, shown by minus signs. If enough electrons build up in the region of the p-type semiconductor near the surface, it becomes n-type. This occurs if all of the holes in this region are filled and some extra electrons arrive as well. These extra electrons can move under the influence of an applied voltage. Now there will be a continuous conducting path of n-type semiconductor material between the source and the drain, and current flows. The direction of flow is shown with the bold arrows. The FET shown in Figure 10.18 is called an ***n-channel FET***, or ***n-FET***, because an n-type channel is created for electrons to flow.

To make the description of FET operation more clear, **Figure 10.19** separately shows the two parts of the circuit: the controlling circuit and the controlled circuit. The voltage labeled V in the controlling circuit controls current in the controlled circuit. Notice that no current flows between the controlling circuit and the controlled circuit. This makes the FET highly efficient—almost no current is used in the controlling action.

We can describe the n-FET operation using a simplified drawing showing the FET as a controlled switch, as illustrated in **Figure 10.20**. The four voltages labeled V_G, V_B, V_S, and V_D are the voltages at the control gate, body, source, and drain. Notice that for the example given in the figures above, the body voltage is always the same as the source voltage because they are connected by a wire; that is, $V_B = V_S$.

Rule for n-FET operation: If the control gate voltage is significantly greater than the body voltage, then the n-FET is ON. Otherwise, the n-FET is OFF.

For a typical FET, "significantly greater than" means greater by about 0.5 volts (V). We can state the condition for an n-FET to be ON as: "V_G is greater than V_B by at least +0.5 V." The ON state is analogous to a conducting switch being connected: current can flow between the source and drain.

[1] The earlier-used aluminum metal was later replaced by heavily doped polycrystalline silicon (polysilicon).

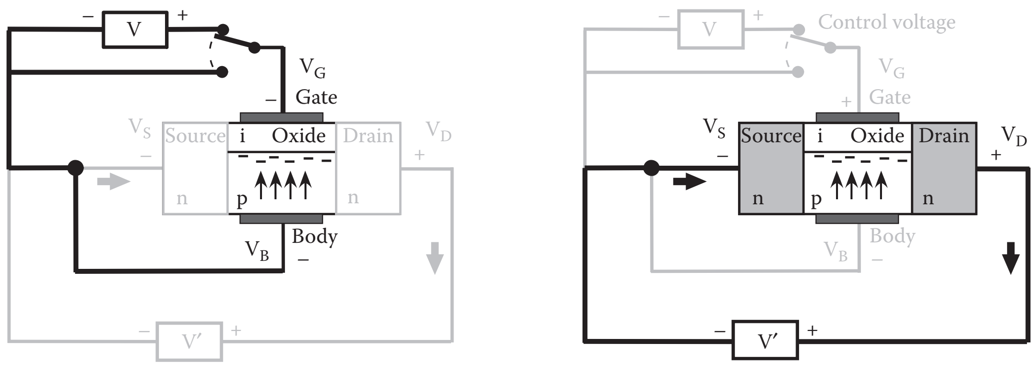


FIGURE 10.19 Highlighted controlling circuit and controlled circuit of an n-FET.

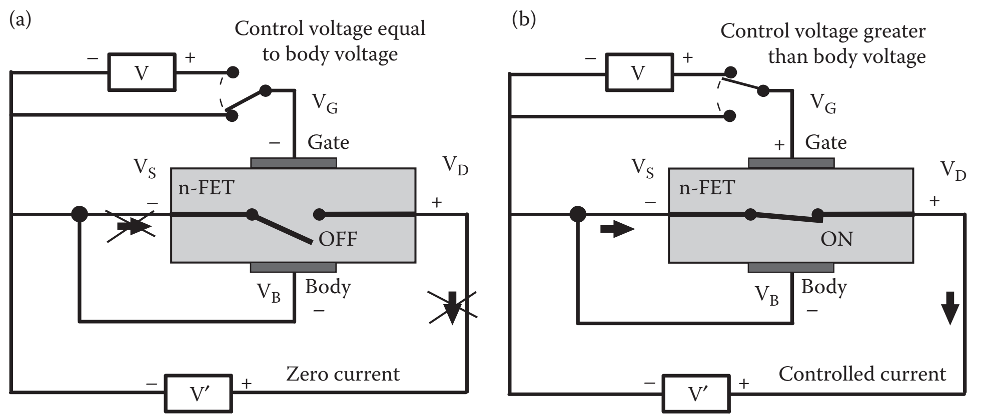


FIGURE 10.20 An n-FET represented as a controlled switch. (a) The control voltage is roughly equal to the body voltage, and the FET acts like a switch that is nonconducting, or OFF. (b) If the control voltage is at least 0.5 V more positive than the body voltage, the FET acts as a switch that is conducting, or ON.

Voltages can be defined in a relative manner (just as energy is relative). The n-FET will be ON if gate and body voltages are both positive, or if both are negative, or if the gate is positive and the body is negative, as long as the gate voltage is significantly greater than the voltage on the body. For example, the n-FET is ON if $V_G = 5$ V and $V_B = 0$ V. In addition, it is ON if $V_G = 0$ V and $V_B = -5$ V. (Note that in terms of negative numbers, –5 is less than 0.) Finally, it is ON if $V_G = 5$ V and $V_B = -5$ V. It is OFF if, for example, $V_G = 0.4$ V and $V_B = 0$ V.

Another type of FET is the ***p-channel FET***, shown in **Figure 10.21**, in which the n- and p- regions of Figure 10.20 are interchanged. The battery polarities are also reversed. A p-FET will become conducting if the voltage applied to the gate is negative relative to the body, as in Figure 10.21b. When the gate is more negative than the body, electrons are pushed away from the surface of the oxide insulator, leaving holes in that part of the n-type region. The small empty boxes depict these holes. This creates a p-type channel through the n-type region; thus the name p-FET. The presence of the holes allows electrons to flow from the drain to the source, under a force exerted by the V' battery.

Rule for p-FET operation: If the gate voltage is less than the body voltage by at least 0.5 V (i.e., V_G is less than $V_B - 0.5$ V), then the p-FET is conducting, or ON.

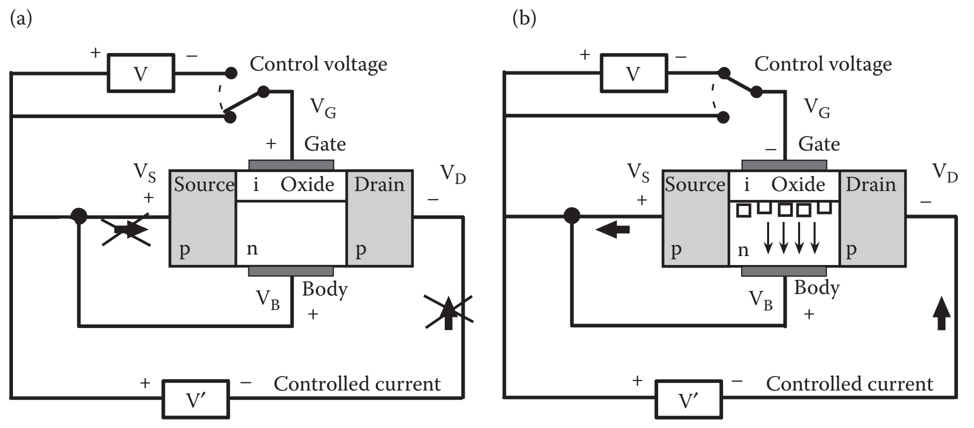


FIGURE 10.21 A p-FET. (a) With the gate voltage equal to the body voltage, no current flows from the source to the drain. (b) A negative voltage *V* applied to the gate relative to the body creates a conducting p-type channel through the n-type semiconductor, allowing electron current to flow from drain to source.

The ON state is analogous to an electrical switch being closed: current can flow between the source and drain in either direction, according to the sign of the applied voltage. Otherwise, the p-FET is OFF. This is illustrated in terms of controlled switches in **Figure 10.22**.

The p-FET can be ON if both gate and body voltages are positive, or if both are negative, or if the gate is negative and the body is positive, as long as the gate voltage is significantly less than the body voltage. For example, the p-FET is ON if $V_G = 0$ V and $V_B = 5$ V. In addition, it is ON if $V_G = –5$ V and $V_B = 0$ V. Finally, the p-FET is ON if $V_G = –5$ V and $V_B = 5$ V.

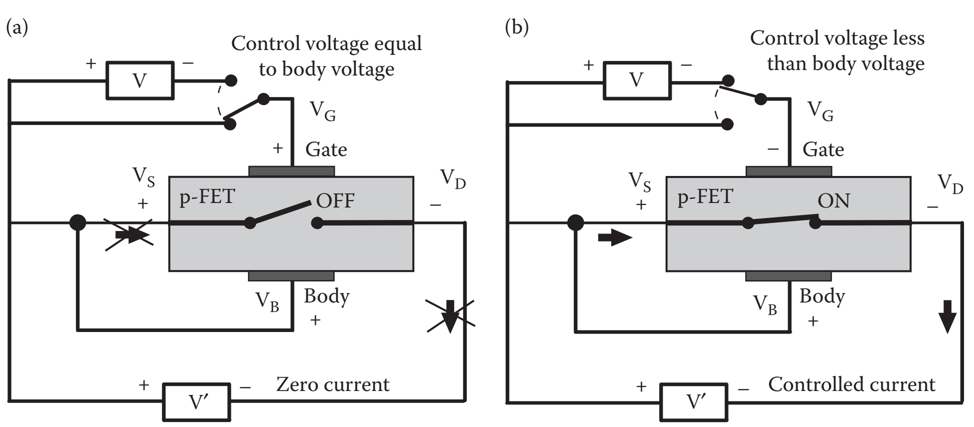
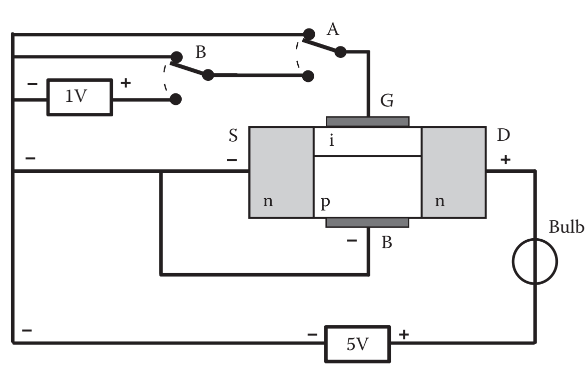


FIGURE 10.22 A p-FET represented as a controlled switch. (a) The control voltage is equal to the body voltage and the FET is OFF. (b) If the control voltage is at least 0.5 V less than the body voltage, then the FET is ON.



QUICK QUESTION 10.2

An n-type FET is connected to two mechanical push-ON switches, A and B, as well as two batteries and a light bulb, as shown to the left. If the switches are logic inputs, and the bulb is the logic output, what logic operation does this circuit perform? (See Chapter 6, Section 6.4.)

To summarize, we call an n-FET a "high-ON switch." It turns ON (conducting) when the control voltage to its gate is high relative to the body, and remains OFF otherwise. In contrast, a p-FET is a "low-ON switch." It turns ON (conducting) when the control voltage to its gate is low relative to the body and remains OFF otherwise.

10.5 CMOS COMPUTER LOGIC

Computer logic circuits are made using FETs. The FET implementation of the NOT logic gate is shown in **Figure 10.23**. A p-FET and an n-FET are connected in series as shown, with the two drains (D) connected to one another. The source (S) of the p-FET is held at a fixed high voltage, V_{HIGH}, which is typically +5 V, whereas the source of the n-FET is held at a low voltage, V_{LOW}, typically 0 V. The logic input is the input voltage, V_{IN}. The input voltage is sent to both of the FET control gates (G). The logic output is the voltage, V_{OUT}, taken from the common drain voltage.

The operation of the NOT gate can be understood from the diagrams. Recall that an n-FET is a high-ON switch and a p-FET is a low-ON switch. When the input voltage is high (+5 V), the n-FET is ON and the p-FET is OFF. This means that the output voltage is connected through the n-FET to the low voltage, $V_{LOW} = 0$ V, and the output is low. In contrast, when the input voltage is low (0 V), the n-FET is OFF and the p-FET is ON. This means that the output voltage is connected through the p-FET to the high voltage, $V_{HIGH} = +5$ V, so the output is high. In each case, the output voltage is the opposite of the input voltage. This is a logical NOT operation.

As usual, we use voltage values to represent logic values. Let us define a logical zero (0) to correspond to 0 V, and a logical one (1) to correspond to +5 V. The logic table for the circuit in Figure 10.23 is shown below. It corresponds to the logic table for the NOT operation.

V_{IN}	n-FET	p-FET	V_{OUT}
+5 V	ON	OFF	0 V
0 V	OFF	ON	+5 V

NOT

A	OUT
1	0
0	1

The type of circuit in Figure 10.23 is called ***CMOS*** (pronounced *cee-moss*), for ***complementary metal-oxide semiconductor***. *Complementary* reflects the fact that the two types of FETS in the circuit act in complementary ways—when one is ON, the other

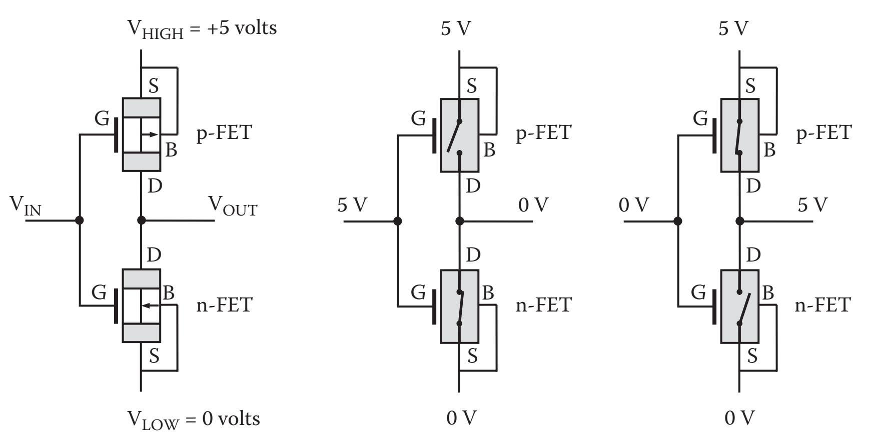


FIGURE 10.23 NOT gate constructed with an n-FET in series with a p-FET. The symbols S, G, D, and B label the source, control gate, drain, and body of each FET, respectively. When the input voltage is high (+5 V), the output is low (0 V), and vice versa.

is OFF. The advent of CMOS circuits in the 1970s was a breakthrough for computing capability, especially for microcomputers, which were introduced in the 1980s, and which later evolved into desktops and laptops. The use of back-to-back complementary FETs minimizes power consumption, because current flows only during the brief time of the switching operation. By the year 2006, the energy needed for a single switching operation reached as little as one femtojoule (10^{-15} J or fJ). After the switching takes place, no current is needed in the circuit to hold a constant output voltage indefinitely. This contrasts with the older circuit types, which used bipolar transistors and lots of current and power.

THINK AGAIN

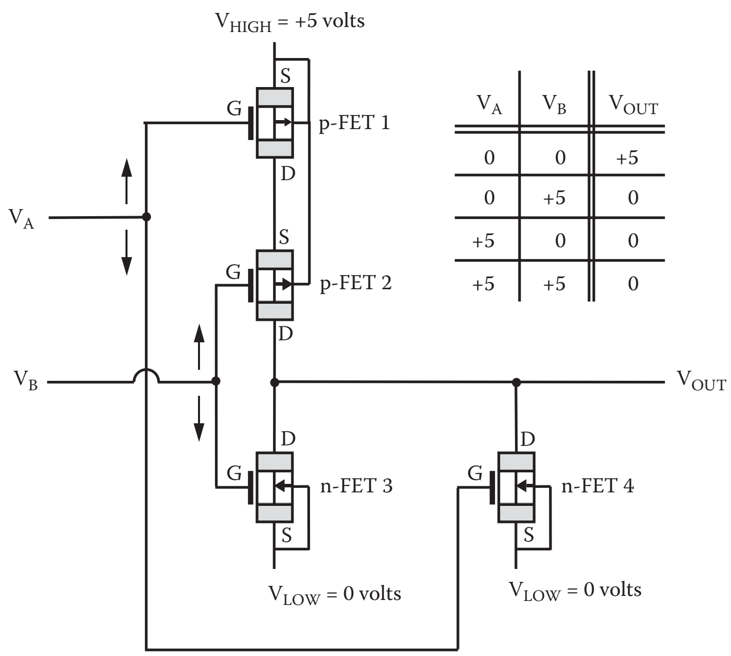
Do not confuse the terms ***logic gate*** and transistor *control gate*. This is an unfortunate double usage of the word *gate*. A logic gate is any device that performs a logic operation. A control gate refers specifically to the input side of a FET.

Recall from Chapter 6, Section 6.4 that the NOR logic operation can be used to construct any other logic gate. This makes the NOR logic gate especially useful. The logic table for the NOR operation is:

NOR

A	B	OUT
0	0	1
0	1	0
1	0	0
1	1	0

We can construct the NOR logic gate using CMOS transistors, as shown in **Figure 10.24**. The circuit consists of two n-FETs and two p-FETs, numbered 1 through 4.



V_A	V_B	V_{OUT}
0	0	+5
0	+5	0
+5	0	0
+5	+5	0

FIGURE 10.24 NOR gate constructed with two n-FETs and two p-FETs.

The logic inputs are the voltages V_A and V_B. Input V_A is sent to the control gate of p-FET-1 and simultaneously to n-FET-4. Input V_B is sent to the control gate of p-FET-2 and simultaneously to n-FET-3. The bodies of FET-1 and FET-2 are connected to the constant +5 voltage source. The p-FET-1 is ON only if V_A is less than +5 V, whereas p-FET-2 is ON only if V_B is less than +5 V. The bodies of FETs 3 and 4 are connected directly to V_{LOW}. Therefore, n-FET-3 is ON only if V_B is greater than 0, whereas n-FET-4 is ON only if V_A is greater than 0.

Because there are two inputs, there are four possible combinations of input values, listed in **Table 10.1**. The only condition that leads to high (+5 V) output V_{OUT} is the case that both inputs are low (0 V), as shown in **Figure 10.25a**. In this case, FET-3 and FET-4 are OFF, and FET-1 and FET-2 are ON. This means that the output is connected through FET-1 and FET-2 to V_{HIGH}, and disconnected from V_{LOW}. For any other combination of input voltages (e.g., Figure 10.25b), at least one of FET-1 or FET-2 is OFF, disconnecting the output from V_{HIGH}, and at least one of FET-3 or FET-4 is ON, connecting the output to V_{LOW}.

Complex logic circuits can be implemented by combining NOR gates, as discussed in Chapter 6, Section 6.4.2. Many modern computer circuits are constructed in just this way. Instead of building several types of electronic circuits—one for each logic operation (NOT, AND, OR)—it turns out to be more efficient and economical to use only one type of electronic gate to do the tasks of the others.

We can also implement the NAND logic operation by using CMOS, as shown in **Figure 10.26**. The only combination of inputs that gives $V_{OUT} = 0$ is the case that

TABLE 10.1
Logic Table for FET NOR Gate.

V_A	V_B	FET 1	FET 2	FET 3	FET 4	V_{OUT}
0	0	ON	ON	OFF	OFF	+5
0	+5	ON	OFF	ON	OFF	0
+5	0	OFF	ON	OFF	ON	0
+5	+5	OFF	OFF	ON	ON	0

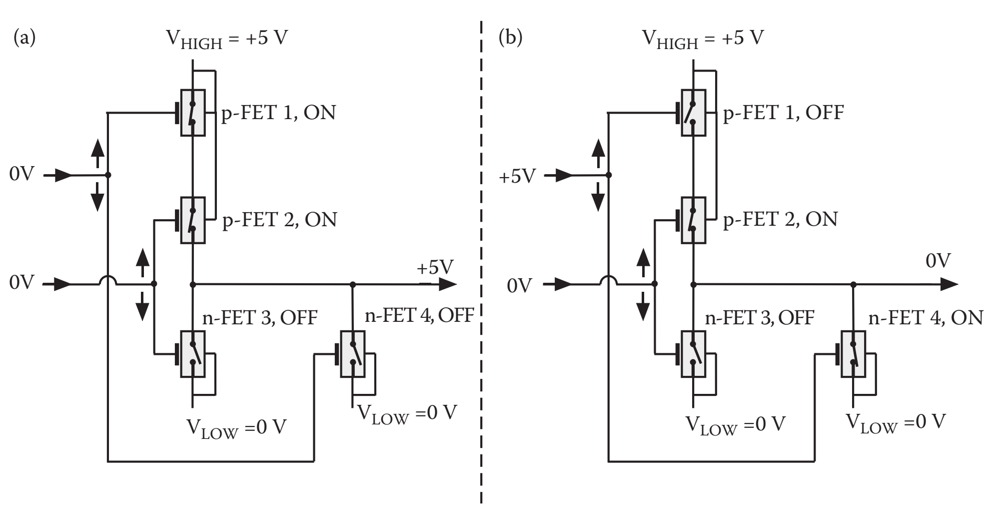
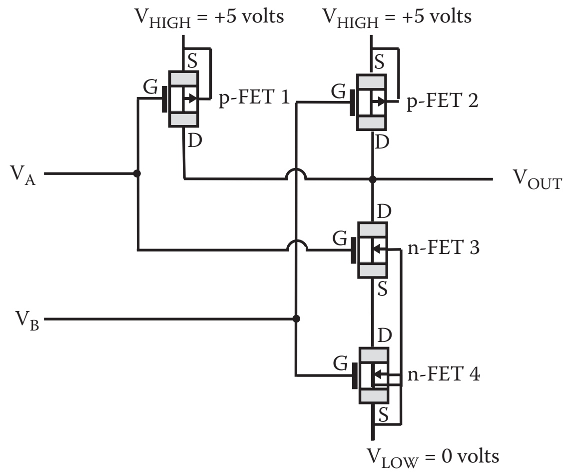


FIGURE 10.25 NOR gate constructed with two n-FETs and two p-FETs. Part (a) shows the only input condition for which the output equals +5 V, part (b) shows one of the other three possible input conditions.



		A		B		
V_A	V_B	FET1	FET3	FET2	FET4	V_{OUT}
0	0	On	Off	On	Off	+5
0	+5	On	Off	Off	On	+5
+5	0	Off	On	On	Off	+5
+5	+5	Off	On	Off	On	0

A NAND B

A	B	OUT
0	0	1
0	1	1
1	0	1
1	1	0

FIGURE 10.26 NAND gate constructed with two n-FETs and two p-FETs.

FETs 3 and 4 are both ON, and FETs 1 and 2 are both OFF. This occurs only when both inputs equal +5 V. In every other case, the point in the circuit connected to V_{OUT} is connected to one or the other of the +5 V voltage sources (and is automatically disconnected from V_{LOW}.

IN-DEPTH LOOK 10.1: WATER-EFFECT TRANSISTORS

A whimsical analogy using the pressure of water in pipes can help us understand the operation of the FET. For fun, I call this the *water-effect transistor*, or *WET*. Recall, from Chapter 5, Section 5.7, that pressure in a water tank is analogous to the voltage in a circuit. Pressure in water refers to the potential to create a fast, powerful stream of water if a small rupture is opened in the tank's wall. Pressure does not refer to the actual flow of water, only the potential to create a flow if given the chance. A FET is analogous to a pressure-activated water valve. There are two possible types of such a valve. I will call these high-ON types and low-ON types. A high-ON valve turns ON (open) when its control line pressure is high, and remains OFF (closed) otherwise. A low-ON valve turns ON when its control line is low, and remains OFF otherwise.

Models for such valves are shown in **Figure 10.27**. As in the semiconductor FET, there are source, drain, control gate, and body regions. Each region is a hollow pipe or channel where water can flow. The pressures in the body and in the source are kept equal through a small connecting pipe. For the High-ON valve, high water pressure at the gate pushes down a piston if the gate pressure is significantly greater than the pressure in the body. This opens a path in the pipe for water flow between the source and drain. If the gate and body pressures are equal, a spring keeps the piston in the up position, closing the valve.

For the low-ON valve, the piston is held down by a spring in the OFF position when the gate pressure is higher than or equal to the body pressure. When the gate pressure drops significantly below the body pressure, the body pressure pushes the piston up, opening the valve, turning it ON. This allows water flow, as indicated by the double-headed arrow.

In summary, the high-ON valve is ON only if the gate pressure is significantly greater than the source (body) pressure, while the low-ON valve is ON only if the gate pressure is significantly lower than the (body) source pressure. The high-ON valve is analogous to the n-FET, and the low-ON valve is analogous to the p-FET.

By combining both types of water valves, we could construct a NOT logic gate. **Figure 10.28** shows a high-ON valve and a low-ON valve—controlled by a common line—and two large constant-pressure tanks. The two tanks are connected to pumps (not shown) that maintain the tank pressures—one high pressure and one low pressure. Because the control lines of the two valves are connected, they have the same pressure.

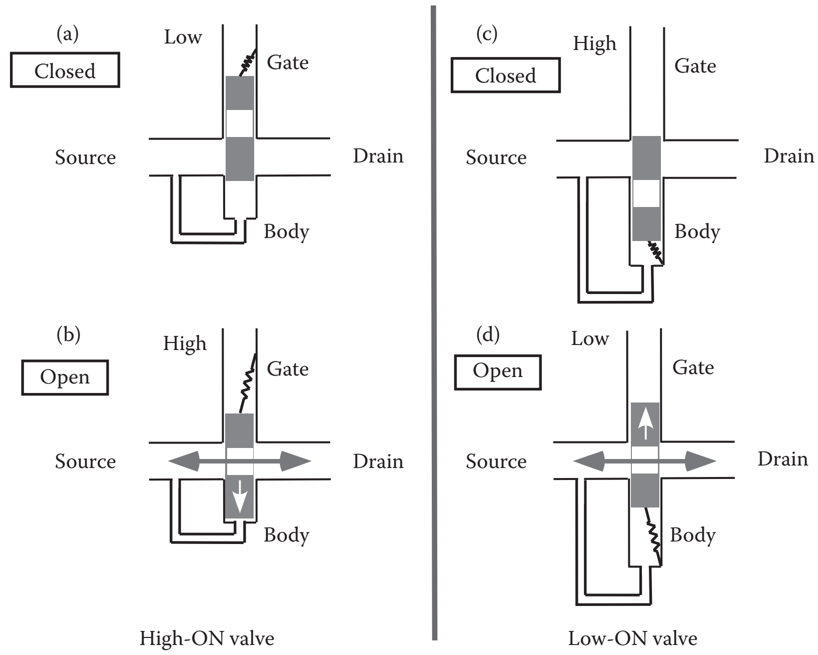


FIGURE 10.27 The water-effect transistor, or WET. The shaded regions are solid and block the flow of water. A high-ON valve is closed when the gate control pressure is lower than or equal to the body pressure (a), and is open when the gate pressure is significantly higher than the body pressure (b). A low-ON valve is closed when the gate pressure is higher than or equal to the body pressure (c), and is open when the gate pressure is significantly lower than the body pressure (d). Double-headed arrows indicate flow of water, which can occur in either direction.

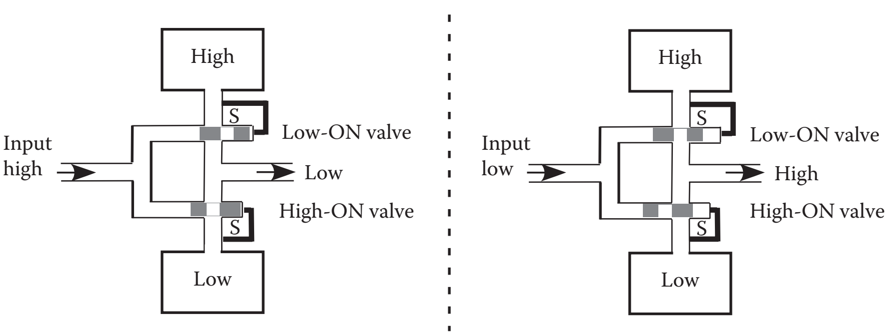


FIGURE 10.28 NOT logic gate constructed with a high-ON valve in series with a low-ON valve. The symbol S labels the source for each valve. When input pressure is high, output is low, and vice versa.

When one valve is open, the other is closed, as can be seen by considering two cases. When the input pressure is high, the output pipe is connected only to the low-pressure tank, and so it is low, as shown in the left figure. When the input pressure is low, the output pipe is connected only to the high-pressure tank, and so it is high. Bit values can be represented as 1 (high), and 0 (low). Thus, this arrangement acts as a NOT logic gate.

Notice that the only time that water flows is during or just after the switching from high to low or from low to high. If we think of water as a resource, then this is an efficient system, because it consumes little water. It operates mostly by virtue of pressure, not flow.

If several such water valves were combined in the proper manner, it would be possible to construct any type of logic gate. By using many such valves, one could construct a rudimentary computer that could carry out, for example, addition of two several-bit numbers.

10.6 MINIATURIZATION, INTEGRATED CIRCUITS, AND PHOTOLITHOGRAPHY

> Where … the ENIAC [computer] is equipped with 18,000 vacuum tubes and weighs 30 tons, computers in the future may have 1,000 vacuum tubes and perhaps weigh just 1½ tons.
>
> –*Popular Mechanics* (*March* 1949)

The transistor is the most important component in computer circuits. The first semiconductor double-junction transistor, which was constructed in 1947 by Bardeen, Brattain, and Shockley at AT&T Bell Laboratories, was a crude and large affair by today's standards. It was a ***bipolar transistor***, which operates on different principles than the FET, as described in In-Depth Look 10.2. **Figure 10.29** is a photograph of their first transistor, measuring about 1 cm on a side. Texas Instruments introduced the first commercially available transistor in 1954. Since then, the size of transistors has shrunk by more than 10,000 times. This allows circuits to be fantastically miniaturized, so that hundreds of millions of logic gates fit into less than a square centimeter (about one-sixth of a square inch). By the year 2006, the FET gate had been reduced in width to only 37 nanometers (nm), and the oxide layer that insulates the gate was only 1 nm thick. To get a feeling for these numbers, imagine that the entire metropolitan area of Greater Los Angeles could be miniaturized and placed into an area the size of a small coin, with all of its freeways, streets, buildings, cars, and people. That gives an idea of the complexity of a modern computer chip. The historical trend of miniaturization (Moore's law) that began at least as early as 1980 indicates that the number of components that can be put onto a semiconductor circuit of a fixed size doubles every 18 months. This trend is expected to continue at least until 2020.

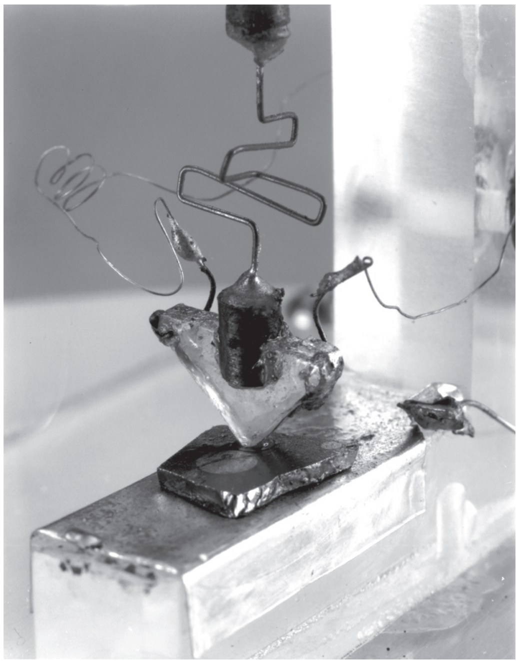


FIGURE 10.29 The first bipolar transistor was about one-half inch across. (With permission of AT&T Archives and History Center.)

Modern computer processors and memory circuits are based on ***integrated circuits*** **(*ICs*)**. An IC is a circuit in which all elements are fabricated on a single semiconductor crystal. The first IC was made by at TI in 1958 Jack Kilby and is pictured in **Figure 10.30**. It contained one double-junction transistor, three resistors, and one capacitor. These elements were connected using external wires, so the circuit is not considered monolithic ("single stone"). The monolithic IC was proposed in 1959 at Fairchild Semiconductor company by Robert Noyce, who suggested using the lithography process to fabricate wires as thin metal strips layered directly on the surface of a crystal. Noyce, a few years later, co-founded Intel Corporation with Gordon Moore. Intel made the first single-chip microprocessor—the Intel 4004—in 1971. This chip, shown in **Figure 10.31**, measured 1/8 inch by 1/16 inch and held 2,300 MOS transistors. It was as powerful in computing capability as the earlier Eniac computer, which contained 18,000 vacuum tubes and occupied 3,000 cubic feet.

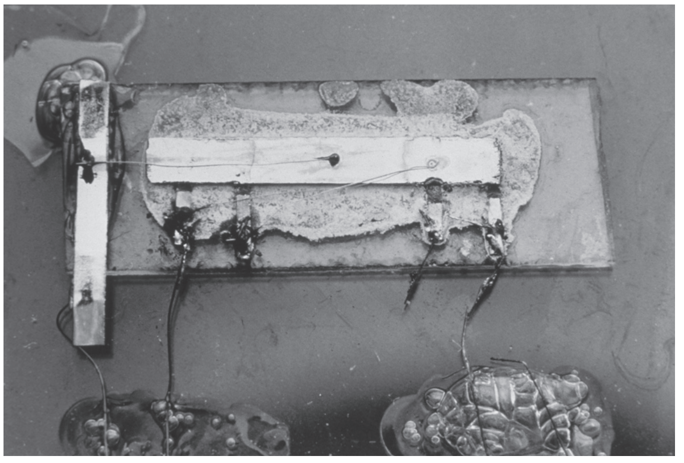


FIGURE 10.30 The first IC, built by Jack Kirby. The main part of the device is a bar of germanium (Ge) measuring 7/16 × 1/16 in., with protruding wires and glued to a glass slide. (Courtesy of Texas Instruments.)

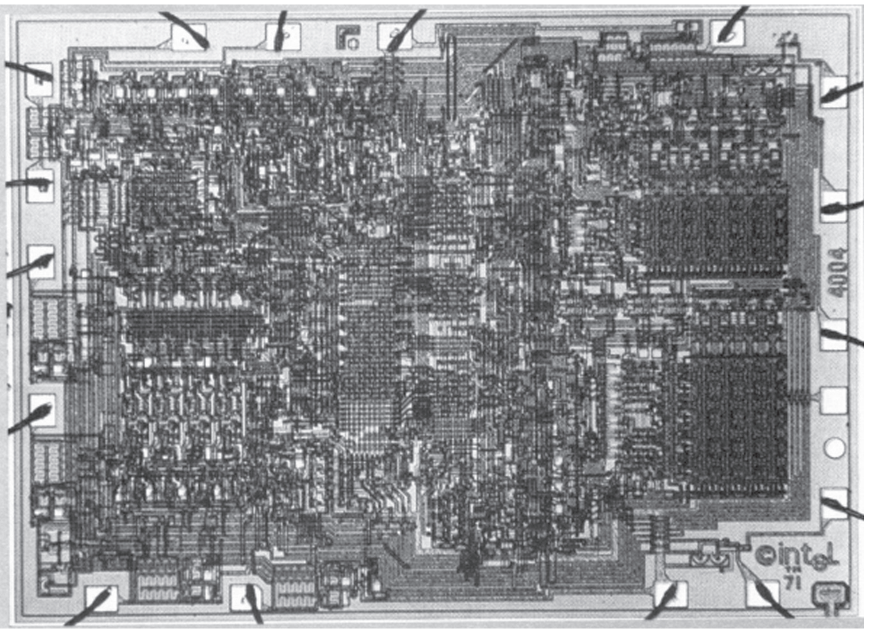


FIGURE 10.31 Photo of the first single-chip microprocessor—the Intel 4004—introduced in 1971. (With permission of Intel Corp.)

The process used to create millions of transistors on a piece of Si is analogous to a method known to artists—lithography. Lithography is a technique for transferring a pattern of ink from a flat surface, such as stone, to a piece of paper. The stone can be reused many times to create many nearly identical artworks. Lithography was discovered in 1798 in Germany, and was named from the Greek *lithos*, "stone." This process yields a one-to-one transfer of the image. ***Photolithography*** is a high-tech form of lithography developed for transferring images onto a Si crystal. Jules Andrus at Bell Laboratories first used photolithography to create semiconductor devices in 1957. In the following sections, we describe the basic steps of making ICs.

10.6.1 Silicon Crystal Preparation

A nearly perfect Si crystal is needed to begin the making of an IC. The method used for making such crystals was described in Real-World Example 4.1. After a thin Si crystal wafer, called a *substrate*, is polished, it is exposed to water vapor (H_2O) at high temperature (1,000°C). This causes *oxidation*, meaning that oxygen atoms penetrate the upper layer of crystal surface and chemically bond with the Si to create a very thin layer of a type of SiO_2. A typical film thickness is 500 nm. This silica form of SiO_2 is noncrystalline (amorphous, or random) and is an electrical insulator.

A second useful type of thin film that can be applied to the surface is a thin film of metal, which serves as a conductor. When patterned into thin strips, metal films act as wiring connecting different components on the surface. A commonly used metal for this purpose is aluminum. Copper has better conductivity properties than does aluminum, but is harder to work with in fabricating reliable circuits. Other types of thin films are often used as conductors, such as polycrystalline Si that has been heavily doped.

10.6.2 Lithography for Fabricating a p-n Junction

Because the transistor is composed of two p-n junctions (diodes), we will use the fabrication of a single p-n junction as our basic example of photolithography. The overall principles of fabrication are similar in all cases, including far more complex designs. Each technique consists of a series of distinct steps, summarized as:

1. Thin-film deposition
2. Photoresist deposition
3. Photo-exposure through a patterned mask
4. Chemical etching to remove unwanted portions of the photoresist
5. Etching away of the exposed portions of the thin film
6. Removal of remaining photoresist
7. Introduction of dopants
8. Metal film deposition to form conducting contacts (circuit wires)

Some of these steps are carried out in separate machines, with the wafer being moved from machine to machine either by hand or by robots.

Figure 10.32 illustrates each of these steps. Starting with a phosphorous-doped n-type Si crystal as substrate (a), a SiO_2 thin film is created on the surface (b) by oxidation, as mentioned above. A thin layer of a liquid polymer (plastic) material is spread evenly on this oxide surface, as shown in (c). Upon heating, the polymer dries into a thin film. This polymer is called a photoresist, and has the property that when exposed to ultraviolet (UV) light, the exposed area undergoes a chemical change and becomes stronger and resistant to being dissolved by certain chemicals called developers. This is analogous to processing chemicals used in photography.

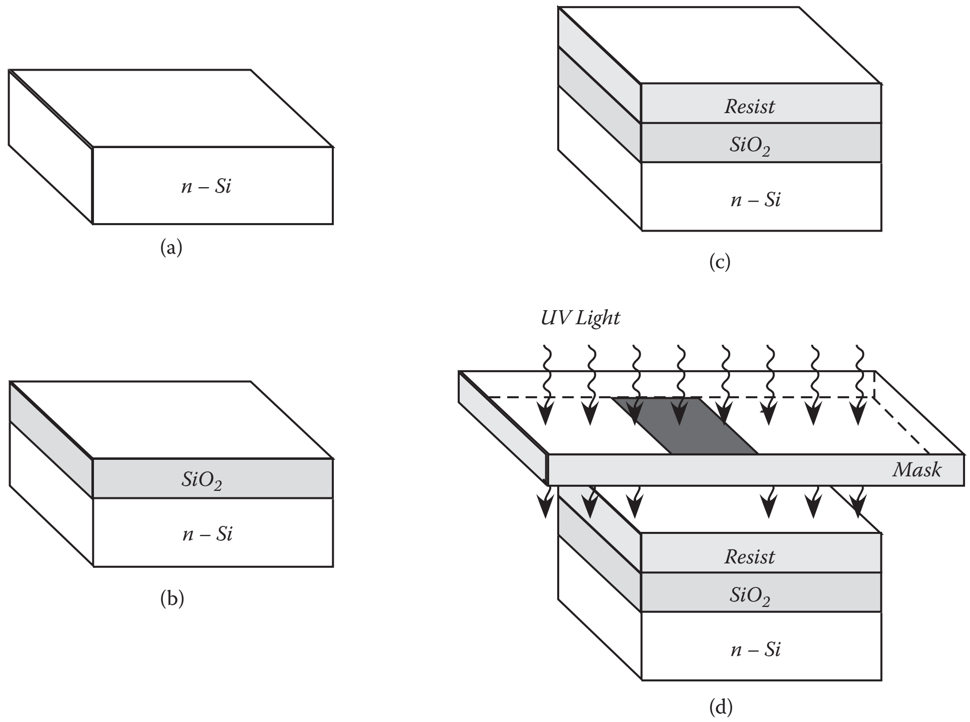


FIGURE 10.32 Fabrication of a p-n junction (a–d). In part (d), light is blocked in the dark-shaded region of the mask. (Adapted from S.M. Sze, *Semiconductor Devices, Physics and Technology,* Wiley, 2002. With permission.)

Some readers will be familiar with the idea of etching glass or metal for artistic purposes. A pattern of an acid-resistant substance (analogous to a photoresist) is painted on a glass surface. Then the whole surface is treated with an acid solution, which etches, or removes, the unprotected parts of the glass surface to a certain depth, which depends on the time of exposure. To stop the etching process, the acid is washed away using water. The pattern created in the etched glass is the negative of that pattern originally made by the photoresist. Silica (SiO_2) is a type of glass, and therefore is susceptible to etching by acid.

A similar process is used in lithography for making a p-n junction, except that the features to be created are far smaller than in the art example. Therefore, the manner for creating the pattern in the photoresist must be capable of much finer control. The method used is to first create a tiny pattern in a metal mask. This mask is a separate thin piece of metal, which is cut or machined very finely using a focused beam of electrons. The metal mask, with its patterns of holes, is placed in contact with the polymer photoresist layer on the SiO_2 layer on the substrate, as in Figure 10.32d. UV light is shone through the openings in the metal mask and into the photoresist just below it. The light causes the photoresist to harden (polymerize) only in the areas directly below the openings in the mask. In Figure 10.32d, the mask is simply a solid bar, shown shaded. The width of the bar may be 500 nm or so. Light does not penetrate this bar and so the resist there is not hardened.

In the next step, a chemical developer is spread over the surface, where it dissolves away the unhardened photoresist in the center region, leaving the pattern of photoresist shown in **Figure 10.33a.** Then a strong acid solution is applied to the surface, where it etches into the SiO_2 (silica glass) thin film, leaving the pattern shown in Figure 10.33b. Next, the remaining photoresist is removed by using a second, stronger developer chemical, leaving the result in Figure 10.33c.

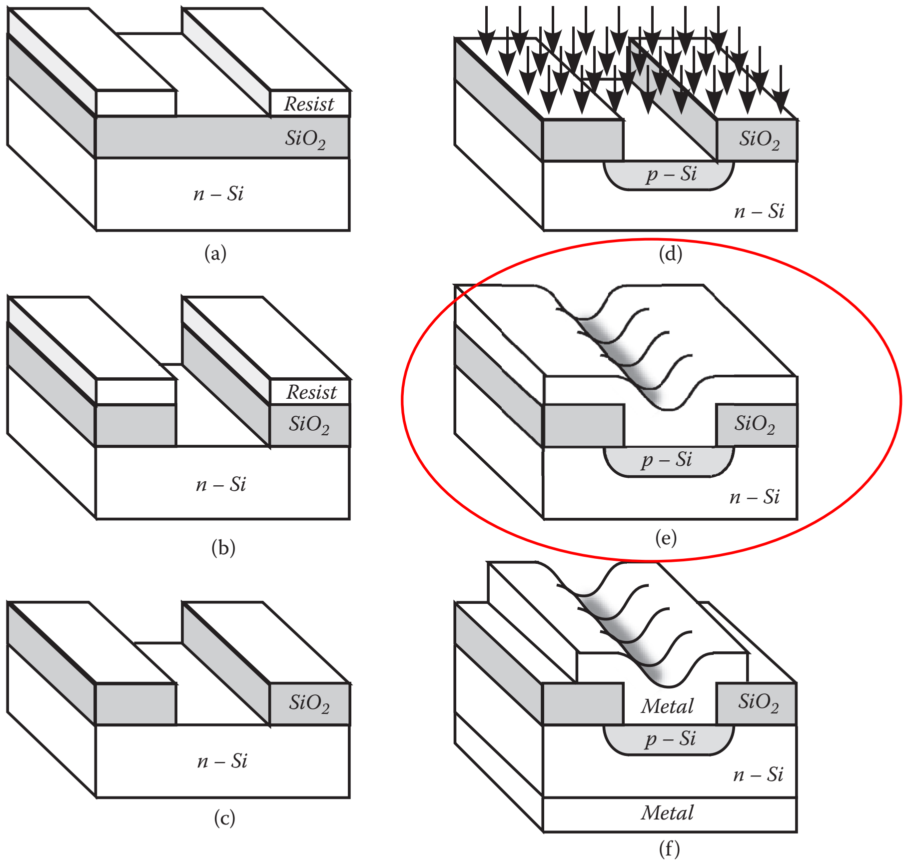


FIGURE 10.33 Fabrication of a p-n junction (e–j). (Adapted from S. M. Sze, *Semiconductor Devices, Physics and Technology,* Wiley, 2002. With permission.)

To create a p-n junction, we need, in addition to the n-type Si crystal, a region of p-type Si crystal. This is created in the region of the substrate that is now not covered by any SiO_2. This small portion of the substrate is converted from n-type to p-type by bombarding the wafer with B (boron) atoms. These enter the crystal by diffusion and, with their deficiency of electrons (compared to Si), create electron holes. This p-type material is created only in the region labeled p-Si in Figure 10.33d.

Now the p-n junction is complete, except for a means of connecting it in an electronic circuit. Rather than connecting external wires for this purpose, thin layers of aluminum metal are deposited on the top and bottom sides on the wafer, as in Figure 10.33e and f. The final diode device is shown in Figure 10.33f. Current will flow if a battery is connected with its plus side to the p-type Si (the top of the device) and the battery's minus side to the n-type (bottom). If connected in the opposite manner, no current (or very little) will flow.

When realizing the complexity of creating just one junction by the many steps described above, it is mind boggling to contemplate creating an entire IC containing millions of diodes, transistors, and other components. Fortunately, all of these components can be fabricated simultaneously, rather than one at a time. A very challenging step in this process of creating an IC is the design and electron-beam cutting of the needed masks. This is a specialty unto itself, and we will not discuss it in detail. The important point is that a single set of masks can be used repeatedly to create thousands of nearly identical circuits. This mass production capability is the essence of lithography.

IN-DEPTH LOOK 10.2: BIPOLAR TRANSISTORS

The same property that makes FETs ideal for logic circuits—low power consumption—makes them less useful as power amplifiers. Power amplifiers are used in audio and music systems such as guitar amplifiers. They are also used to boost signal levels in long-distance digital communications systems—discussed in later chapters.

The earliest and still most common transistor-based power amplifier is the bipolar junction transistor. In contrast to the FET, which is a voltage-controlled valve, a bipolar junction transistor is a current-controlled valve. It is a double diode structure, the main electron current of which can be controlled by a second, smaller current. Bipolar junction transistors come in two types—npn and pnp. An npn transistor is made by contacting a very thin (0.1–5 μm) p-type crystal, called the base, between two larger n-type crystals, called the emitter and collector, as in **Figure 10.34**. This forms two p-n junctions, back to back, labeled 1 and 2. Electron diffusion causes the two n-type crystals each to develop a positive charge near the junctions, indicated in the figure by (+). The p-type crystal develops a negative charge, indicated by (–). As in a diode, these charge buildups lead to forces on electrons in the directions shown by the large arrows at the top of the diagram. These two forces (interpreted as *internal batteries*) push electrons in two directions away from the center of the base region. This is like having a potential energy "ridge" at the middle of the base. Electrons can "fall" in either direction away from this ridge.

First, consider the case that the switch S is in the down position (shown as dashed), so the battery A is irrelevant. The conducting switch keeps the base at the same voltage (potential energy) as the emitter. For this setting of the switch, junction 1 has no external voltage, so no current flows through it. Furthermore, battery B provides a backward external voltage across junction 2 (+ sign to the n-type side), so net current does not flow through junction 2. Therefore the bulb is not lit.

Now consider what happens after the switch S is moved to the up position, connecting battery A across junction 1. Assume that the voltage V_A is much smaller than voltage V_B. Now battery A provides a forward external voltage to junction 1 (+ sign to the p-type side), counteracting the internal force that tries to hold electrons from passing through junction 1 from left to right in the diagram. Because of this counteracting force, junction 1 is now conducting, and this allows electrons to flow around the circuit labeled "small current." A few electrons move per second from battery A, through the emitter, into the base and into the (+) side of battery A.

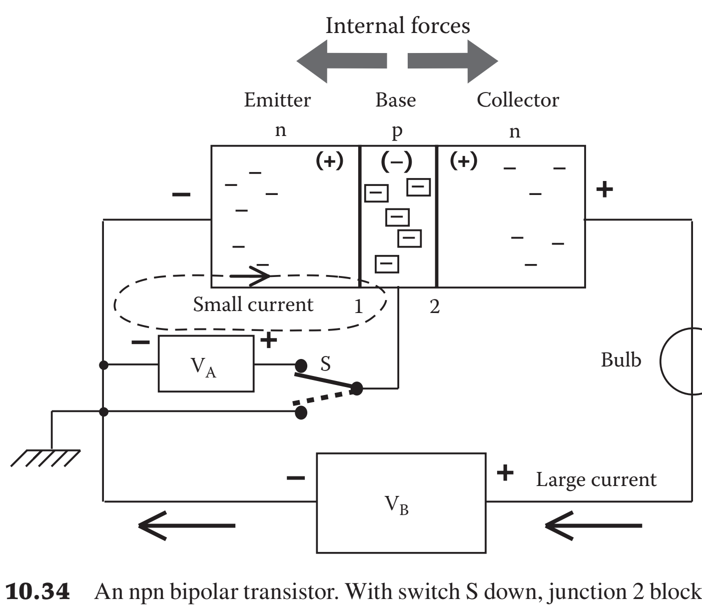


FIGURE 10.34 An npn bipolar transistor. With switch S down, junction 2 blocks current from flowing in the large outer loop. When S is up, as shown, the small current created by battery A in the inner circuit loop enables a large flow of electrons in the outer loop, driven by battery B.

There are also electrons that can be provided by battery B, whose (–) side is trying to push them into the emitter and pull them out through the collector. Battery B provides a much larger voltage than battery A, and provides far more electrons per second than battery A does. As electrons enter the base from the emitter on their way to the (+) side of battery B, many of them find themselves on the right side of the potential energy ridge, at the middle of the base. They are rapidly swept into the collector (hence its name) before they ever reach the wire leading to battery A. This opens up the possibility for another electron to enter the base from the emitter and take its place. The consequence is that a large current flows in the outer loop in the diagram, enabled by the much smaller current flowing in the inner loop. This large current lights the bulb.

A pnp transistor has the same structure as a npn transistor, except that the n-type and p-type regions are interchanged. In terms of the fish analogy used earlier to understand diode action, it is like having a small pool on a ridge of a mountaintop, with water falling on both sides, as in **Figure 10.35**. If a fish on top moves slightly to the right, it falls down on the right, whereas if instead it moves to the left, it falls down on the left. If the left lake is raised, it gives a helping boost to fish trying to go up the falls from left to right. Now many not-so-energetic fish can make it up to the small pool at the top of the ridge, at which point most of them are swept down the falls on the right side into the collector lake.

Bipolar junction transistors were used in early computer logic circuits because this technology was developed earlier than the FET technology. The disadvantage of these transistors for logic circuitry is that they carry considerable current, and this leads to heating of the circuit. Recall from Chapter 5, Section 5.7 that all conducting materials have some resistance to electron current, and this resistance is similar to friction, leading to heat. When building a computer that contains millions of transistors, heating at each transistor is the last thing we want. In fact, one of the major limitations to the number of transistors that can be fabricated on a computer chip is the amount of heat generated, because if the circuit gets too hot, the transistors cease to operate properly. The high temperature creates too many electrons in the upper, conducting energy band, and their sensitivity to the emitter-base current is lessened. For this reason, FETs are preferred and are used almost exclusively in modern computers.

Before the invention of the semiconductor-junction transistor in 1947, signal amplification was done primarily using vacuum tubes. These are familiar today mostly for their use in guitar amplifiers and some high-end stereo systems. Tube amplifiers were used by AT&T for long-distance telephone signals, but they were unreliable, bulky, and expensive. After it was found during the Second World War that semiconductors such as Si and germanium could be used as receivers for microwaves, the leaders of AT&T decided to

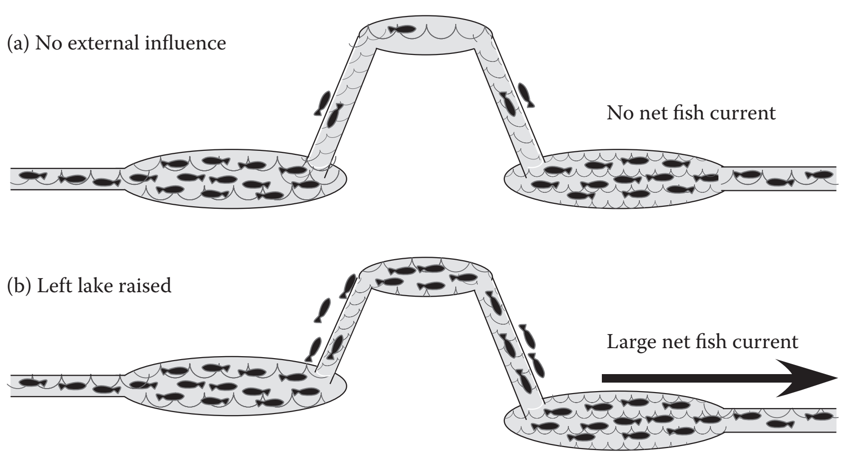


FIGURE 10.35 Fish analogy for transistor action.

research whether semiconductors could also be used as signal amplifiers. Although there had been several precursor inventions along these lines elsewhere, the work at AT&T provided the understanding to make semiconductors into a practical technology.

The team assigned to the project consisted of Shockley, Brattain, and Bardeen. After some initial ideas of Shockley's, and failures in the laboratory, Brattain and Bardeen finally were successful in creating the first pnp point-contact transistor. With Bardeens' theoretical insight and Brattain's expert laboratory technique, the two devised and made the device shown earlier in the photograph in Figure 10.29**.** As shown in the schematic in **Figure 10.36**, it consisted of a 1-centimeter triangular block made of an insulating material with two pieces of gold metal foil on the surface, separated by a 0.05 millimeter gap at the tip of the triangle. This tiny gap region was gently pressed onto the surface of a crystal of doped n-type germanium (Ge) semiconductor. With gold metal in contact with Ge, which had an excess of electrons, some of the electrons diffused into the metal, which is conducting and initially had no charge imbalance. The electron diffusion stopped when the charge separation created a strong enough internal electric force to prevent any further separation of charge. This diffusion caused small regions under the metal contacts to have a deficit of electrons, and thus they were p-type semiconductors.

Figure 10.36c shows the gap and the small regions below the metal contacts that became p-type because of the electron diffusion into the metal. There is an n-type region between two p-type regions. A pnp double junction transistor was created. Compare to Figure 10.34, which shows an npn transistor. The gold metal on the left side connects to the emitter (p-type region), the gold metal on the right side connects to the collector (p-type region), and the metal at the bottom connects to the base (n-type). When Brattain connected wires with negative voltage to the base and collector, and a wire with positive voltage to the emitter, he found that a small current flowing through the emitter-base loop created a large current flowing through the emitter-collector loop. This was the first example of amplification by a pnp transistor.

The three shared the 1956 Nobel Prize for their contributions to physics. The race was on to use this breakthrough, and many others that quickly followed, for improving communication systems—telephone, radio, and television—as well as computing machines.

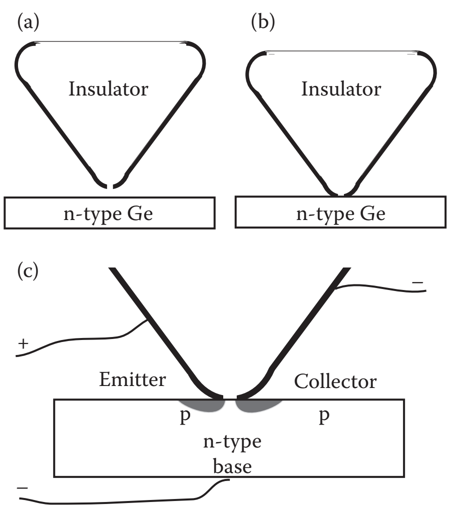


FIGURE 10.36 The first bipolar transistor, shown earlier in Figure 10.29. (a) Triangular insulator material with gold metal foil strips (thick lines) on edges. (b) Making contact with an n-type germanium crystal. (c) Close-up of contact region, showing resulting p-type regions shaded. Metal wires are connected with plus or minus voltages as shown.

SUMMARY AND LOOK FORWARD

In this chapter, we studied the important physics principles governing semiconductor behavior that were discovered in the 1940s and 1950s. Knowledge of this area of physics allowed scientists and engineers in the 1960s and 1970s to create electronic devices that are essential to the operation of computers and the Internet—miniaturized logic and switching circuits.

The keys to understanding semiconductor physics are as follows:

1. Doping—The addition of impurity atoms into a pure Si crystal to make it conducting. Dopants such as P (phosphorous), which bring extra electrons into the crystal, make an n-type crystal, whereas dopants such as B (boron), which are electron-deficient, create electron holes and make a p-type crystal.
2. Diffusion—Electrons move freely and randomly through a conducting material (a doped semiconductor or a metal).
3. p-n junction—If two materials having different concentrations of excess electrons are joined, diffusion of electrons takes place, and a separation of charge occurs. This charge separation causes an internal electric field to build up, which prevents the diffusion from causing any further charge separation.

On the basis of these principles, useful electronic components were created.

1. Diode—A p-n junction, in which current can flow only if the applied voltage is more positive at the p-type side of the diode than at the n-type side. This can be used to rectify a signal (i.e., remove its negative components) leaving a positive-only signal.
2. Transistor—The workhorse of the Information Age is the transistor, in particular the MOS FET, or MOSFET. A voltage applied between the gate and the body of the transistor turns the transistor ON, connecting the source and drain through a conducting channel. By using one n-FET and one p-FET in a complementary arrangement, voltage switching can be achieved with very little current flow. Lower current allows circuits to use a minimum of power and therefore not overheat.

We have reached an important milestone in our understanding of the physics behind computer operation. From the principles of quantum physics and electromagnetism, we are able to understand the workings of transistors. The CMOS-based electronic circuits we have studied can be used to construct any logic operation desired. In Chapter 6, we discussed how logic operations are used to accomplish any computer information-processing task. We are close to understanding all of the elements needed for the operation of computer hardware, but one element has not been covered—computer memory. We will discuss that in Chapter 11. Another application of semiconductors is based on how they emit or absorb light, which is the basis of photocells, photodetectors, light-emitting diodes (LEDs), and lasers. We will study the physics of semiconductors and light in the next chapter.

SOCIAL IMPACTS: LABELING EVERY OBJECT IN THE WORLD

Unlike a UPC bar code, Electronic Product Code provides for the unique identification of any physical object in the world.

Steve Meloan
(Sun Developer Network, 2003*)*

Most people think of a personal computer or a PDA as things connected to the network. But here we are connecting trees, race cars, and astronauts to the network. In the future, everything of value will be on the network in one form of another. And once they're on the network, we can aggregate data from those diverse devices, and then deliver that data to equally diverse devices.

John Fowler
(Software CTO of Sun Microsystems)

Radio-Frequency ID is about to take on a life of its own.

"RFID Goes Mainstream: Alien Technology Whitepaper" (2007)

Supermarkets use black-and-white printed bar codes to identify items for pricing and inventory. To read its code, a worker moves it close to an optical reader. Imagine a new system, in which tiny microchips, smaller than an ant's head, are attached or inserted into every item in a store, and could be read electronically from a distance of several meters. Now imagine that these chips could be implanted under the skin of dogs, cats, children, husbands, wives, and parents. Finally, imagine a system where every person, pet, car, bike, skateboard, guitar, compact disk, book, pair of pants, gun, bullet, flag, etc., had not only an electronic "bar code" identifying the type of item, but also a unique bar code identifying every individual item. The basis of such a system is already developed and is being implemented increasingly in the commercial world. As with any new technology, such a system has wonderfully positive potential, and frightening potential for abuse.

Before discussing the social issues, let us briefly consider how the technology works. First, how many items could be labeled with unique codes? Using a binary number having N bits to label each item, we can create 2^N unique codes (see Chapter 2). For example, with $N = 4$, we can create $2^4 = 2 \times 2 \times 2 \times 2 = 16$ unique codes: 0000, 0001, 0010, etc., up to 1111. If, instead, we used 33 bits to create each number, we could create $2^{33} = 8{,}589{,}934{,}592$ unique codes, one for every human on Earth. For example, your personal code could be

$$011000111000100001110010000101010.$$

If we wanted to expand this labeling scheme to every single object on Earth, this could be done, at least in theory. If we used 90 bits to create them, we could create $2^{90} = 1.2 \times 10^{27}$ unique codes. These are enough codes to label uniquely every cubic millimeter of material (including rocks, dirt, and ocean) within 1 kilometer of the Earth's surface.

The technology used for labeling is radio frequency identification (RFID) tags. Each tag contains a tiny antenna, which is used for receiving and transmitting, and a Si chip used for data storage. A tag reading device, or reader, transmits a low-power radio signal to the tag, which responds by broadcasting a message back, containing information stored in the tag's memory. There are two types of RFID tags. Passive tags have no battery and must receive power from the radio waves sent by the reader. Therefore, these tags work only at close range, up to a couple of meters. Active tags contain their own battery for power and can therefore be read from greater distances.

Two significant commercial developments in this area are the Electronic Product Code (EPC) Network and the subdermal RFID implant. The EPC Network is based on inserting tags into each item produced by a company so that it can be tracked during manufacturing, distribution, sales, and (potentially) after being purchased by a customer. Subdermal RFID implants are inserted under the skin, often in the back of the upper arm, or triceps muscle. These can be used either to monitor people, such as Alzheimer's patients who might wander off and get lost, or workers in banks or other high-security positions, where their identity can be easily verified. Note that a person carrying a typical implanted chip cannot be tracked from a distance more than a few hundred meters.

Great benefits can result from using these labeling and tracking technologies. Companies can put resources when and where they are needed and potentially save costs, as well as do a better job of conserving energy. The worldwide market size for RFID products and services was about $4.5 billion in 2008. People with health risks can "get chipped," and then if they end up unconscious in an emergency ward, doctors can identify them by their code and look up their medical records in an online database. Companies can make more money by better identifying what types of products are bought where and in what combinations. On the lighter side, news agencies reported that the Baja Beach Club in Barcelona, Spain is the first business to offer RFID subdermal implants in customers to give them access to VIP areas and provide an easy payment option. On a darker note, some people might be in favor of implanting chips under the skin of sex offenders or other criminals, although no program of this kind is in force as of 2008.

It is easy to see why some people are concerned about potential abuses of this technology. It could raise the specter of a dictatorial government (called Big Brother by George Orwell [1]) monitoring our every move [2]. There have been studies that suggest subdermal implants might cause cancer in laboratory mice [3,4]. Those standing to benefit financially from the industry point to counter reports arguing that the mice studies should not be seen as conclusive [5]. There are concerns about people being forced to accept chip implants by governments (security workers or "troublemakers") or companies (employees). Wisconsin, North Dakota, and California have passed laws prohibiting employers and others from forcing anyone to have an RFID device implanted under their skin [5]. There are concerns that a person's RFID implant could be read by others without permission. There are also privacy concerns about RFID chips placed into passports, driver's licenses, or student identification cards. Also, hackers could possibly alter the data in someone else's implanted chip, for their own purposes.

So which is it: *Brave New World*, or *Better-Living-Through-Technology World*? Or both? Or neither?

REFERENCES

1. Orwell, George. *Nineteen-Eighty-Four*. London: Harcourt, Brace & Co., 1949.
2. Albrecht, Katherine, and Liz McIntyre. *Spychips: How Major Corporations and Government Plan to Track Your Every Move with RFID*. New York: Plume, 2006.
3. Albrecht, Katherine. "Microchip-Induced Tumors in Laboratory Rodents and Dogs: A Review of the Literature 1990–2006." CASPIAN Consumer Privacy, November 19, 2007. http://www.antichips.com/cancer.

4. Le Calvez, Sophie, Marie-France Perron-Lepage, and Roger Burnett. "Subcutaneous Microchip-Associated Tumours in B6C3F1 Mice: A Retrospective Study to Attempt to Determine their Histogenesis," *Experimental and Toxicologic Pathology* 57 (2006): 255.
5. Wustenberg, William. "Effective Carcinogenicity Assessment of Permanent Implantable Medical Devices: Lessons from 60 Years of Research Comparing Rodents with Other Species." AlterNetMD Consulting, Farmington, MN, 2007.

SOURCES FOR SOCIAL IMPACT

How Stuff Works http://electronics.howstuffworks.com/rfid1.htm.
RFID Journal. http://www.rfidjournal.com.

SUGGESTED READING

See the general physics references given at the end of Chapter 1.

Bierman, Alan W. *Great Ideas in Computer Science—A Gentle Introduction.* Cambridge, MA: MIT Press, 1997.
Brinkman, William, Douglas Haggan, and William Troutman. "A History of the Invention of the Transistor and Where It Will Lead Us," *IEEE Journal of Solid-State Circuits* 32 (1997). http://www.sscs.org/AdCom/transistorhistory.pdf.
Hepher, M. "The Photoresist Story," *Journal of Photographic Science* 12 (1964).
Riordan, Michael, and Lillian Hoddeson. *Crystal Fire*. New York: W. W. Norton & Company, 1997.
Seitz, Frederick, and Norman G. Einspruch. *Electronic Genie: The Tangled History of Silicon*. Urbana: University of Illinois Press, 1998.
"Transistorized!" ScienCentral, Inc. and The American Institute of Physics, 1997. http://www.pbs.org.

For instructors:
There are few references with elementary discussions of FET operation. Books that are helpful:

Ferry, David K., and Jonathon P. Bird. *Electronic Materials and Devices*. San Diego: Academic, 2001 (see pp. 249–282).
Melissinos, Adrian C. *Principles of Modern Technology*. Cambridge, UK: Cambridge, 1990. (Sections 1.7–1.9 describe FET operation, including depletion FETs. Section 2.7 describes DRAM.)
Neamen, Donald. *An Introduction to Semiconductor Devices*. New York: McGraw-Hill, 2006.
Pierret, Robert F. "Field Effect Devices," in *Modular Series on Solid State Devices*, edited by R. F. Pierret and G. W. Neudeck, Vol. IV. Reading, MA: Addison-Wesley, 1983. (Sections 2.1, 2.2, and 5.1 give qualitative discussions of FET operation.)
Sze, S.M. *Semiconductor Devices, Physics and Technology*. Hoboken, NJ: Wiley, 2002.
Taur, Yaun, and Tak H. Ning. *Fundamentals of VLSI Technology*. Cambridge, UK: Cambridge, 1998 (Sections 2.3.1, 2.3.2, 5.1 and Chapter 3 Introduction).

KEY TERMS

Amplitude
Bipolar transistor
CMOS
Depletion region
Diffusion
Diode
Dopant or impurity
Doping
Electron
Field-effect transistor (FET)
Hole
Insulator
Integrated circuit (ICs)
Logic gate
n-channel FET, or n-FET
n-type semiconductor
p-channel FET, or p-FET
p-n junction

p-type semiconductor
Photolithography
Rectifier
Resistor
Solid-state electronics
Transistor

ANSWERS TO QUICK QUESTIONS

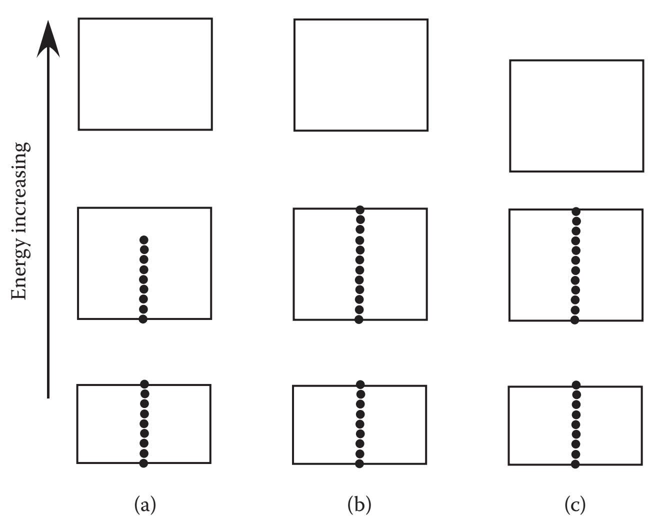
Q10.1 The molecules of dye feel random forces exerted on them from the nearby water molecules colliding with them. Each collision pushes a given dye molecule in some random direction. After many such collisions, different dye molecules will find themselves at different locations in the water. As a whole, the dye molecules eventually spread out to fill the water more or less uniformly.

Q10.2 If both switches are pushed ON (logic value = 1), the FET's gate becomes more positive than its body, and current flows through the source and drain, lighting the bulb. This performs the AND logic operation.

EXERCISES AND PROBLEMS

Exercises

E10.1 For some elements, when several atoms of the same element are brought together, they form a crystal. The energy-level diagram of the allowed energies of electrons for an individual atom changes dramatically when this occurs. Below are three energy-level diagrams for three different crystals. Label each diagram as a conductor, insulator, or semiconductor and explain what distinguishes them. (In the pictures here, dots indicate full or partially full bands, but the particular numbers of dots do not have any significance.)



E10.2 Explain why the electrical conductivity of a pure Si crystal changes when it is warmed from 80ºC below freezing to room temperature (20ºC). Does the conductivity increase or decrease? How does the resistance change?

E10.3 If the band gap in Si were equal to 1.8 aJ instead of 0.18 aJ, would this make it a better or worse conductor at room temperature? Why? *Hint*: You may need to look at some information given in the previous chapter.

E10.4 In 2006, hyperpure Si cost as much as \$3/g (about \$100/oz). One cubic centimeter of Si has a mass of 2.3 g, meaning that the cost per cubic centimeter equals about \$7.00. Consider a cylinder (boule) of hyperpure Si that is 20 in. (51 cm) in length and has diameter of 8 in. (20 cm). The boule is sliced into 510 1-mm-thick disks or wafers, and 100 ICs are made on each wafer.

(a) What is the volume (in cm^3) and cost of the Si for each wafer? *Hint*: The volume of a wafer equals its thickness times its surface area, and the surface area equals πr^2, where r is the radius (one-half the diameter) of the wafer.
(c) What is the total cost of the boule?
(d) What is the Si cost per IC?

E10.5 (a) Explain how adding small amounts of P to a pure Si crystal greatly increases its conductivity.
(b) Predict, using the Periodic Table, at least one other element besides P that would likely have a similar effect when doped into Si.
(c) Explain how adding small amounts of B to a pure Si crystal greatly increases its conductivity. On the basis of the Periodic Table, predict at least one other element besides B that would likely have a similar effect when doped into Si.

E10.6 Invent and explain a situation from everyday life that is analogous to the statement, "The presence of electron holes allows current to flow easily in a crystal that would otherwise be nonconducting."

E10.7 Use the following analogy to explain how a semiconductor diode works to allow electrical current to flow in one direction, but not in the other. Two boxes, called box A and box B, sit on a table and contain a few hundred small marbles each. One box (box A) is continuously shaken side to side randomly, so the marbles in it never stop rolling around, bouncing off the walls and each other. A small opening is cut into the side of each box, down to the level of the table, and a piece of cardboard is folded into a rectangular channel, allowing marbles to freely move from one box into the other. Continue to develop this analogy to illustrate diode operation. (Read the explanation in terms of swimming fish.) What do you need to do with box B to make the diode analogy complete? *Hint*: Think about potential energy and the two lakes in the fish analogy.

E10.8 A fun model for a p-n junction can made by imagining a rock concert, where the area in front of the stage is separated into two zones by a curtain. The zone to the left is the mosh pit—a region filled with closely packed people (moshers), with zero space between them. The mosh pit is "doped" with a few crowd surfers, who have enough energy to move above the heads of the moshers. (An amusing animation of crowd surfing is at http://members.aol.com/rik0lar/moshing/mosh.htm.) In the zone to the right of the curtain, the closeness of standing people is not great enough to allow crowd surfing—a surfer would fall to the floor between people. Elaborate this model further to explain what happens when the curtain is suddenly opened. Invent a mechanism analogous to the internal battery that develops at a p-n junction, which prevents the charge separation from increasing indefinitely.

E10.9 (a) Extend the mosh-pit analogy discussed in the previous exercise to model the operation of an n-FET. As a joke, I call this a MOSHFET. Consider the insulator region to be represented by the stage, which is too high for crowd surfers to reach.
(b) Do the same for the p-FET.

E10.10 (a) When a material is classified as a conductor, insulator, or semiconductor, what is meant by these classifications in terms of the electrical current flow?
(b) Explain why a ceramic object (made of randomly packed crystallites) might not be a good conductor.
(c) In an n-type semiconductor, it is the excess electrons (which are in the higher-energy conducting band) that are free to move. We say such electrons are mobile. In a p-type semiconductor, in which energy band do the mobile electrons reside? Make an energy-band drawing to explain your answer.

E10.11 (a) In a FET, does the voltage between the source and drain determine whether it is ON or OFF? Explain.
(b) For each of the following cases, would an n-FET be ON or OFF? (i) V_G = 0 V and V_B = –5 V; (ii) V_G = 0.3 V and V_B = 0 V; (iii) V_G = 1 V and V_B = 0 V; (iv) V_G = 0 V and V_B = –5 V; (v) V_G = 0 V and V_B = –1 V.
(c) For each of the following cases, would a p-FET be ON or OFF? (i) V_G = 0 V and V_B = 1 V; (ii) V_G = 4 V and V_B = 5 V; (iii) V_G = –1 V and V_B = 0 V; (iv) V_G = 0 V and V_B = –5 V; (v) V_G = 0 V and V_B = –0.3 V.

E10.12 A rectangular IC chip has sides of length 1 cm and 2 cm. This chip is completely covered with transistors each having dimensions 1×1 µm. Calculate the maximum number of such transistors that can fit on this chip.

E10.13 Using state-of-the-art transistors in 2006, a single switching operation could be achieved using as little as 1 fJ (10^{-15} J). The smallest switching time achieved was 10 ps.

(a) What power does this correspond to?
(b) How many such switching operations can be performed in 1 sec?

E10.14 Do some research online and/or in the book references given at the end of this chapter to learn more about the history of science leading to the development of semiconductor electronic devices. Present the history in the form of a timeline, showing 10 to 20 major events occurring between about 1900 and 1960. Provide commentary on each event. Give references.

Problems

P10.1 A pure Si crystal contains about 5×10^{22} atoms/cm^3.

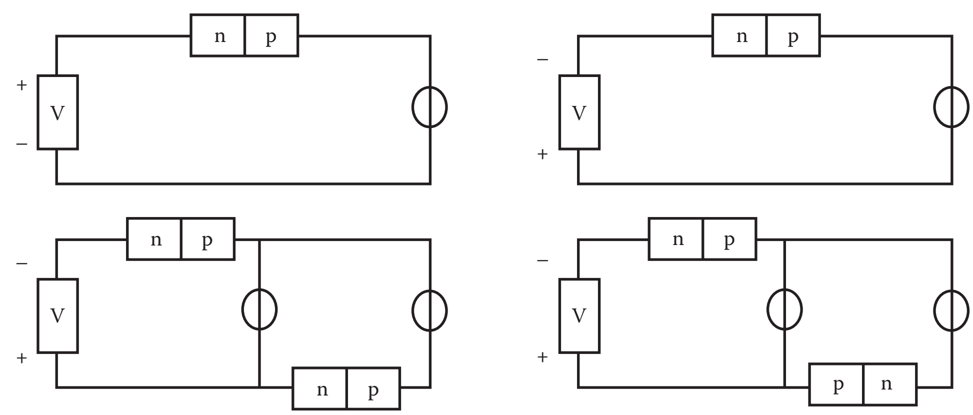
(a) How many protons per cubic centimeter does a Si crystal contain?
(b) How many electrons per cubic centimeter does a Si crystal contain?
(c) In a pure Si crystal at temperature 200°C (very hot), there are about 1.5×10^{14} electrons/cm^3 in the conducting band. Which would be more conducting—a pure Si crystal at 200°C, or a Si crystal doped with one part per million of phosphorus at room temperature (about 22°C)?

P10.2 The discussion of p-n junctions in connection with Figure 10.9 assumes that some of the excess electrons in the n-type region diffuse the to the p-type region; that is, only electrons in the highest energy band diffuse. This is a simplification of the actual situation in real crystals. Redraw Figure 10.9 allowing also for holes to move (diffuse) from the p-type region to the n-type region. When we say a hole diffuses, this means that some other physical objects move to fill in a hole, leaving a new hole nearby. In this case, what physical object actually moves, and which energy band is this object in?

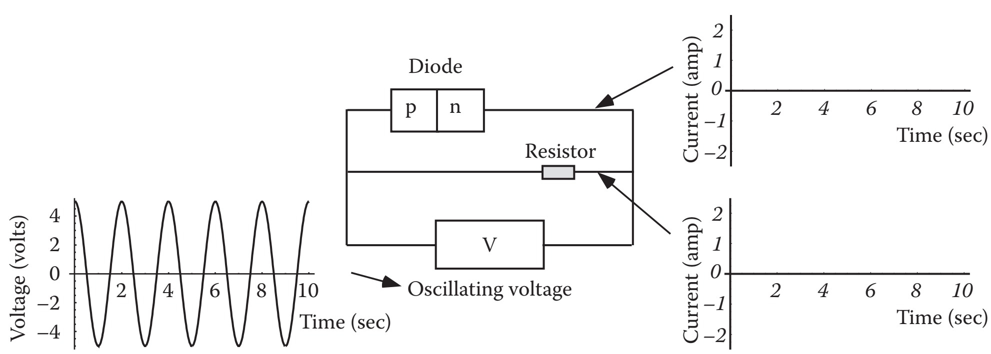
P10.3 A certain p-n junction is made of Si doped with P at a density 5×10^{16} atoms/cm^3 and Si doped with B at a density 5×10^{16} atoms/cm^3. The depletion region is 10 μm thick and 100×100 nm in area.

(a) What is the volume of the depletion region, in cubic centimeters?
(b) How many excess electrons from the n-type side have diffused over to the p-type side?

P10.4 In the four drawings below, indicate which light bulbs (indicated by ovals) would light.



P10.5 A diode and a resistor in series are connected to an oscillating voltage, as shown below. The graph of voltage indicates the voltage at the left side of the voltage source (V) relative to the voltage at the right side. Make graphs of the current in the resistor and in the diode, assuming that current flowing left to right corresponds to a positive current, and that the maximum current in each equals 2 amp.

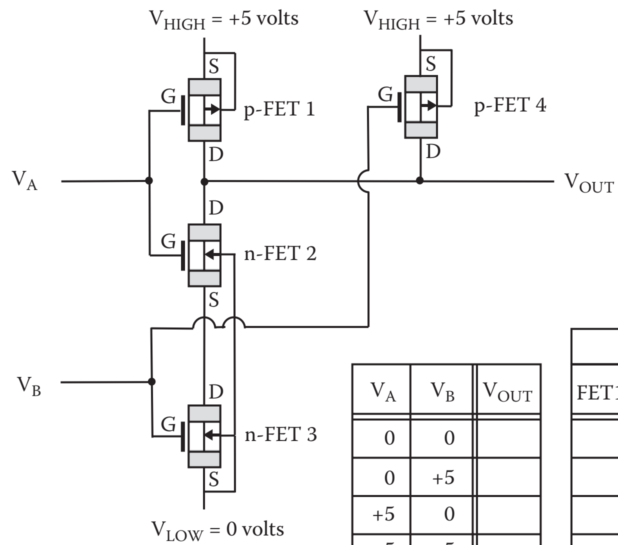


P10.6 Recall from Chapter 6 that NOR logic gates can be used to construct any other logic gate. By combining NOR gates of the form shown in Figure 10.24, how many n-FETs and p-FETs are needed to construct each of the following logic devices?

(a) AND
(b) OR
(c) half-adder
(d) full-adder

P10.7 Construct a diagram showing how the NOR logic operation could be implemented using the whimsical water-effect, or WET devices, described in In-Depth Look 10.1.

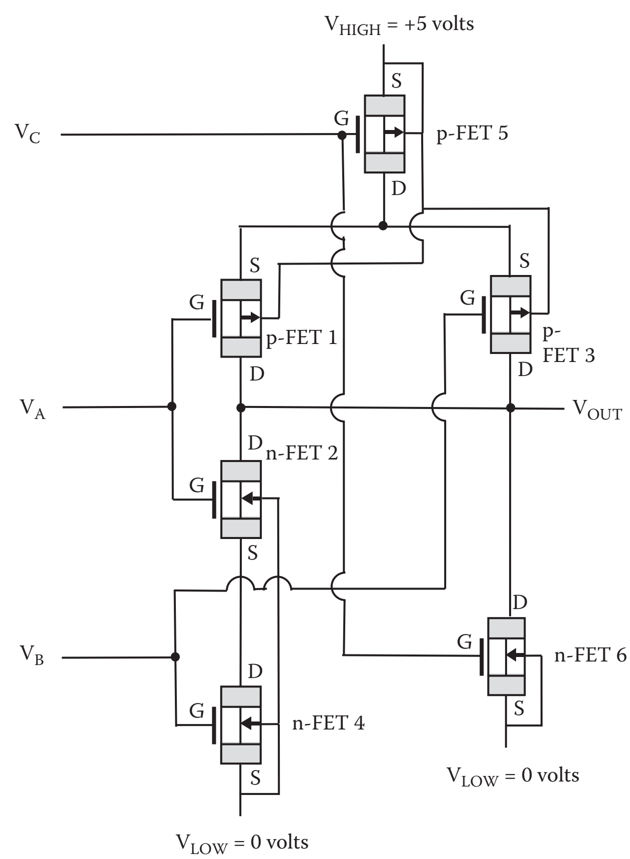
P10.8 Prepare the voltage (logic) tables for voltage inputs and outputs for the following CMOS circuit. Fill in all of the blanks, and explain in words the overall operation. *Hint*: This is a NAND gate.



V_A	V_B	V_{OUT}
0	0	
0	+5	
+5	0	
+5	+5	

A		B	
FET1	FET2	FET3	FET4

P10.9 Work out the voltage (logic) tables for voltage inputs and outputs for the following CMOS circuit. Fill in all of the blanks in the logic table below and explain in words the overall operation. *Hint*: This gate operation can be summarized as: [NOT(V_C = +5)] AND [NOT(V_A = +5) OR NOT(V_B = +5)]



				A		B		C	
V_A	V_B	V_C	V_{out}	p-1	n-2	p-3	n-4	p-5	n-6
0	0	0							
0	0	+5							
0	+5	0							
0	+5	+5							
+5	0	0							
+5	0	+5							
+5	+5	0							
+5	+5	+5							

P10.10 (This problem depends on In-Depth Look 10.2)

(a) Describe the operation of a bipolar-junction transistor that has a thin n-doped region sandwiched between two p-doped regions. This is a pnp transistor. Draw a circuit diagram showing how the batteries should be wired (including their polarities; i.e., location of plus and minus terminals) to obtain the same valvelike action discussed in this chapter for npn transistors.

(b) Draw the corresponding fish-transistor diagram, analogous to the diagram in Figure 10.35 in In-Depth Look 10.2.

P10.11 (This problem depends on In-Depth Look 10.2)

Further develop the boxes and marbles analogy in E10.7 to explain how a bipolar semiconductor transistor works.