

# **3**

**PART**

# **Basic Electronics**

*This page intentionally left blank*

# 19

CHAPTER

# Introduction to Semiconductors

SINCE THE 1960S, WHEN THE TRANSISTOR BECAME COMMON IN CONSUMER DEVICES, *SEMICONDUCTORS* have acquired a dominating role in electronics. The term *semiconductor* arises from the ability of these materials to conduct some of the time, but not all the time. The conductivity can be controlled to produce effects such as amplification, rectification, oscillation, signal mixing, and switching.

## The Semiconductor Revolution

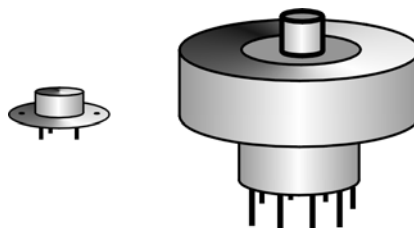
Decades ago, *vacuum tubes*, also known as *electron tubes*, were the only devices available for use as amplifiers, oscillators, detectors, and other electronic circuits and systems. A typical tube (called a *valve* in England) ranged from the size of your thumb to the size of your fist. They are still used in some power amplifiers, microwave oscillators, and video display units.

Tubes generally require high voltage. Even in modest radio receivers, 100 V to 200 V dc was required when tubes were employed. This mandated bulky power supplies, and created an electrical shock hazard. Nowadays, a transistor of microscopic dimensions can perform the functions of a tube in most situations. The power supply can be a couple of AA cells or a 9-V transistor battery.

Even in high-power applications, transistors are smaller and lighter than tubes. Figure 19-1 is a size comparison drawing between a transistor and a vacuum tube for use in an AF or RF power amplifier.

*Integrated circuits* (ICs), hardly larger than individual transistors, can do the work of hundreds or even thousands of vacuum tubes. An excellent example of this technology is found in personal computers and the peripheral devices used with them.

**19-1** A power-amplifier transistor (at left) is much smaller than a vacuum tube of comparable power-handling capacity (right).



## Semiconductor Materials

Various elements, compounds, and mixtures can function as semiconductors. The two most common materials are *silicon* and a compound of gallium and arsenic known as *gallium arsenide* (often abbreviated GaAs). In the early years of semiconductor technology, *germanium* formed the basis for many semiconductors; today it is seen occasionally, but not often. Other substances that work as semiconductors are *selenium*, *cadmium* compounds, *indium* compounds, and the oxides of certain metals.

### Silicon

Silicon (chemical symbol Si) is widely used in diodes, transistors, and integrated circuits. Generally, other substances, or *impurities*, must be added to silicon to give it the desired properties. The best quality silicon is obtained by growing crystals in a laboratory. The silicon is then fabricated into *wafers* or *chips*.

### Gallium Arsenide

Another common semiconductor is the compound gallium arsenide. Engineers and technicians call this material by its acronym-like chemical symbol, GaAs, pronounced “gas.” If you hear about “gas-fets” and “gas ICs,” you’re hearing about gallium-arsenide technology.

GaAs devices require little voltage, and will function at higher frequencies than silicon devices because the charge carriers move faster through the semiconductor material. GaAs devices are relatively immune to the effects of ionizing radiation such as X rays and gamma rays. GaAs is used in light-emitting diodes (LEDs), infrared-emitting diodes (IREDS), laser diodes, visible-light and infrared (IR) detectors, ultra-high-frequency (UHF) amplifying devices, and a variety of integrated circuits.

### Selenium

Selenium exhibits conductivity that varies depending on the intensity of visible light or IR radiation that strikes it. All semiconductor materials exhibit this property, known as *photoconductivity*, to some degree; but in selenium the effect is especially pronounced. For this reason, selenium is useful for making *photocells*. Selenium is also used in certain types of *rectifiers*. A rectifier is a component or circuit that converts ac to pulsating dc.

A significant advantage of selenium is the fact that it is electrically rugged. Selenium-based components can withstand brief *transients*, or spikes, of abnormally high voltage, better than components made with most other semiconductor materials.

### Germanium

Pure elemental germanium is a poor electrical conductor. It becomes a semiconductor only when *impurities* are added. Germanium was used extensively in the early years of semiconductor technology. Some diodes and transistors still use it.

A germanium diode has a low voltage drop (0.3 V, compared with 0.6 V for silicon and 1 V for selenium) when it conducts, and this makes it useful in some situations. But germanium is easily destroyed by heat. Extreme care must be used when soldering the leads of a germanium component.

## Metal Oxides

Certain metal oxides have properties that make them useful in the manufacture of semiconductor devices. When you hear about MOS (pronounced “moss”) or CMOS (pronounced “sea moss”) technology, you are hearing about *metal-oxide semiconductor* and *complementary metal-oxide semiconductor* devices, respectively.

An advantage of MOS and CMOS devices is the fact that they need almost no power to function. They draw so little current that a battery in a MOS or CMOS device lasts just about as long as it would on the shelf. Another advantage is high speed. This allows operation at high frequencies in RF equipment, and makes it possible to perform many switching operations per second for use in computers.

Certain types of transistors, and many kinds of ICs, make use of this technology. In integrated circuits, MOS and CMOS allow for a large number of discrete diodes and transistors on a single chip. Engineers would say that MOS/CMOS has *high component density*.

The biggest problem with MOS and CMOS technology is the fact that the devices are easily damaged by static electricity. Care must be used when handling components of this type. Technicians working with MOS and CMOS components must literally ground themselves by wearing a metal wrist strap connected to a good earth ground. Otherwise, the electrostatic charges that normally build up on their bodies can destroy MOS and CMOS components when equipment is constructed or serviced.

## Doping and Charge Carriers

For a semiconductor material to have the properties necessary in order to function as electronic components, impurities are usually added. The impurities cause the material to conduct currents in certain ways. The addition of an impurity to a semiconductor is called *doping*. Sometimes the impurity is called a *dopant*.

### Donor Impurities

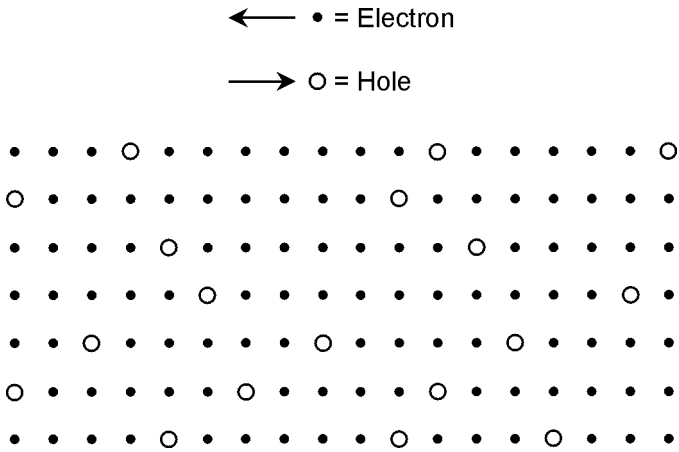
When an impurity contains an excess of electrons, the dopant is called a *donor impurity*. Adding such a substance causes conduction mainly by means of electron flow, as in an ordinary metal such as copper or aluminum. The excess electrons are passed from atom to atom when a voltage exists across the material. Elements that serve as donor impurities include antimony, arsenic, bismuth, and phosphorus. A material with a donor impurity is called an *N-type semiconductor*, because electrons have negative (N) charge.

### Acceptor Impurities

If an impurity has a deficiency of electrons, the dopant is called an *acceptor impurity*. When a substance such as aluminum, boron, gallium, or indium is added to a semiconductor, the material conducts by means of *hole flow*. A *hole* is a missing electron—or more precisely, a place in an atom where an electron should be, but isn't. A semiconductor with an acceptor impurity is called a *P-type semiconductor*, because holes have, in effect, a positive (P) charge.

### Majority and Minority Carriers

Charge carriers in semiconductor materials are either electrons, each of which has a unit negative charge, or holes, each of which has a unit positive charge. In any semiconductor substance, some



**19-2** Pictorial representation of hole flow. Solid black dots represent electrons, moving in one direction. Open circles represent holes, moving in the opposite direction.

of the current takes the form of electrons passed from atom to atom in a negative-to-positive direction, and some of the current occurs as holes that move from atom to atom in a positive-to-negative direction.

Sometimes electrons account for most of the current in a semiconductor. This is the case if the material has donor impurities, that is, if it is of the N type. In other cases, holes account for most of the current. This happens when the material has acceptor impurities, and is thus of the P type. The dominating charge carriers (either electrons or holes) are called the *majority carriers*. The less abundant ones are called the *minority carriers*. The ratio of majority to minority carriers can vary, depending on the way in which the semiconductor material has been manufactured.

Figure 19-2 is a simplified illustration of electron flow versus hole flow in a sample of N-type semiconductor material, where the majority carriers are electrons and the minority carriers are holes. The solid black dots represent electrons. Imagine them moving from right to left in this illustration as they are passed from atom to atom. Small open circles represent holes. Imagine them moving from left to right in the illustration. In this particular example, the positive battery or power-supply terminal (or “source of holes”) would be out of the picture toward the left, and the negative battery or power-supply terminal (or “source of electrons”) would be out of the picture toward the right.

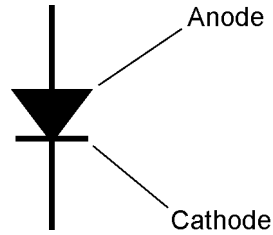
## The P-N Junction

Merely connecting up a piece of semiconducting material, either P or N type, to a source of current can be interesting, and a good subject for science experiments. But when the two types of material are brought together, the boundary between them, called the *P-N junction*, behaves in ways that make semiconductor materials truly useful in electronic components.

### The Semiconductor Diode

Figure 19-3 shows the schematic symbol for a *semiconductor diode*, formed by joining a piece of P-type material to a piece of N-type material. The N-type semiconductor is represented by the short, straight line in the symbol, and is called the *cathode*. The P-type semiconductor is represented by the arrow, and is called the *anode*.

**19-3** Schematic symbol for a semiconductor diode.

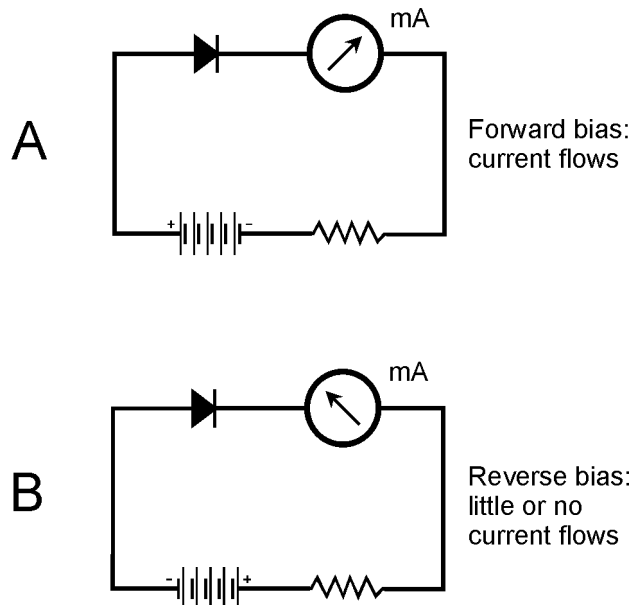


In the diode as shown in Figure 19-3, electrons can move easily in the direction opposite the arrow, and holes can move easily in the direction in which the arrow points. But current cannot, under most conditions, flow the other way. Electrons normally do not move with the arrow, and holes normally do not move against the arrow.

If you connect a battery and a resistor in series with the diode, you'll get a current to flow if the negative terminal of the battery is connected to the cathode and the positive terminal is connected to the anode, as shown in Fig. 19-4A. No current will flow if the battery is reversed, as shown in Fig. 19-4B. (The resistor is included in the circuit to prevent destruction of the diode by excessive current.)

It takes a specific, well-defined minimum applied voltage for conduction to occur through a semiconductor diode. This is called the *forward breakover voltage*. Depending on the type of material, the forward breakover voltage varies from about 0.3 V to 1 V. If the voltage across the junction is not at least as great as the forward breakover voltage, the diode will not conduct, even when it is connected as shown in Fig. 19-4A. This effect, known as the *forward breakover effect* or the *P-N junction threshold effect*, can be of use in circuits designed to limit the positive and/or negative peak voltages that signals can attain. The effect can also be used in a device called a *threshold detector*, in which a signal must be stronger than a certain amplitude in order to pass through.

**19-4** Series connection of a battery, a resistor, a current meter, and a diode. At A, forward bias results in a flow of current. At B, reverse bias results in no current.



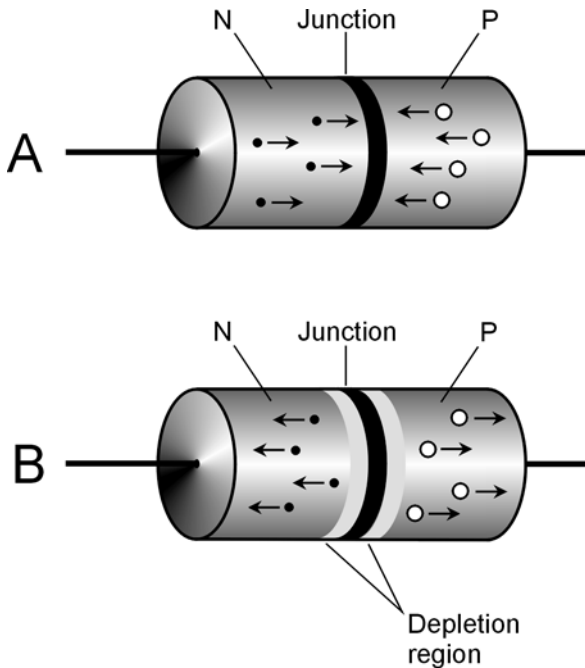
### How the Junction Works

When the N-type material is negative with respect to the P type, as in Fig. 19-4A, electrons flow easily from N to P. The N-type semiconductor, which already has an excess of electrons, receives more; the P-type semiconductor, with a shortage of electrons, has some more taken away. The N-type material constantly feeds electrons to the P type in an attempt to create an electron balance, and the battery or power supply keeps robbing electrons from the P-type material. This condition is illustrated in Fig. 19-5A, and is known as *forward bias*. Current can flow through the diode easily under these circumstances.

When the battery or dc power-supply polarity is switched so the N-type material is positive with respect to the P type, the situation is called *reverse bias*. Electrons in the N-type material are pulled toward the positive charge pole, away from the P-N junction. In the P-type material, holes are pulled toward the negative charge pole, also away from the P-N junction. The electrons are the majority carriers in the N-type material, and the holes are the majority carriers in the P-type material. The charge therefore becomes depleted in the vicinity of the P-N junction, and on both sides of it, as shown in Fig. 19-5B. This zone, where majority carriers are deficient, is called the *depletion region*. A shortage of majority carriers in any semiconductor substance means that the substance cannot conduct well. Thus, the depletion region acts like an electrical insulator. This is why a semiconductor diode will not normally conduct when it is reverse-biased. A diode is, in effect, a one-way current gate—usually!

### Junction Capacitance

Some P-N junctions can alternate between conduction (in forward bias) and nonconduction (in reverse bias) millions or billions of times per second. Other junctions are slower. The main limiting



**19-5** At A, forward bias of a P-N junction. At B, reverse bias of the same junction. Solid black dots represent electrons. White dots represent holes. Arrows indicate direction of charge-carrier movement.



factor is the capacitance at the P-N junction during conditions of reverse bias. As the *junction capacitance* of a diode increases, maximum frequency at which it can alternate between the conducting state and the nonconducting state decreases.

The junction capacitance of a diode depends on several factors, including the operating voltage, the type of semiconductor material, and the cross-sectional area of the P-N junction. If you examine Fig. 19-5B, you might get the idea that the depletion region, sandwiched between two semiconducting sections, can play a role similar to that of the dielectric in a capacitor. This is true! In fact, a reverse-biased P-N junction actually is a capacitor. Some semiconductor components, called *varactor diodes*, are manufactured with this property specifically in mind.

The junction capacitance of a diode can be varied by changing the reverse-bias voltage, because this voltage affects the width of the depletion region. The greater the reverse voltage, the wider the depletion region gets, and the smaller the capacitance becomes.

## Avalanche Effect

Sometimes, a diode conducts when it is reverse-biased. The greater the reverse-bias voltage, the more like an electrical insulator a P-N junction gets—up to a point. But if the reverse bias rises past a specific critical value, the voltage overcomes the ability of the junction to prevent the flow of current, and the junction conducts as if it were forward-biased. This phenomenon is called the *avalanche effect* because conduction occurs in a sudden and massive way, something like a snow avalanche on a mountainside.

The avalanche effect does not damage a P-N junction (unless the voltage is extreme). It's a temporary thing. When the voltage drops back below the critical value, the junction behaves normally again.

Some components are designed to take advantage of the avalanche effect. In other cases, the avalanche effect limits the performance of a circuit. In a device designed for voltage regulation, called a *Zener diode*, you'll hear about the *avalanche voltage* or *Zener voltage* specification. This can range from a couple of volts to well over 100 V. Zener diodes are often used in voltage-regulating circuits.

For *rectifier diodes* in power supplies, you'll hear or read about the *peak inverse voltage* (PIV) or *peak reverse voltage* (PRV) specification. It's important that rectifier diodes have PIV ratings great enough so that the avalanche effect will not occur (or even come close to happening) during any part of the ac cycle.

## Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. The term *semiconductor* arises from
  - (a) resistor-like properties of metal oxides.
  - (b) variable conductive properties of some materials.
  - (c) the fact that electrons conduct better than holes.
  - (d) insulating properties of silicon and GaAs.
2. Which of the following is *not* an advantage of semiconductor devices over vacuum tubes?
  - (a) Smaller size
  - (b) Lower working voltage

- (c) Lighter weight
  - (d) Ability to withstand high voltage spikes
3. Of the following substances, which is the most commonly used semiconductor?
- (a) Germanium
  - (b) Galena
  - (c) Silicon
  - (d) Copper
4. GaAs is
- (a) a compound.
  - (b) an element.
  - (c) a mixture.
  - (d) a gas.
5. A disadvantage of MOS devices is the fact that
- (a) the charge carriers move fast.
  - (b) the material does not react to ionizing radiation.
  - (c) they can be damaged by electrostatic discharges.
  - (d) they must always be used at high frequencies.
6. Selenium works especially well in
- (a) photocells.
  - (b) high-frequency detectors.
  - (c) RF power amplifiers.
  - (d) voltage regulators.
7. Of the following, which material allows the lowest forward voltage drop in a diode?
- (a) Selenium
  - (b) Silicon
  - (c) Copper
  - (d) Germanium
8. A CMOS integrated circuit
- (a) can only work at low frequencies.
  - (b) requires very little power to function.
  - (c) requires considerable power to function.
  - (d) can only work at high frequencies.
9. The purpose of doping is to
- (a) make the charge carriers move faster.
  - (b) cause holes to flow.

- (c) give a semiconductor material specific properties.
  - (d) protect devices from damage in case of transients.
10. A semiconductor material is made into N type by
- (a) adding an acceptor impurity.
  - (b) adding a donor impurity.
  - (c) injecting protons.
  - (d) taking neutrons away.
11. Which of the following does not result from adding an acceptor impurity?
- (a) The material becomes P type.
  - (b) Current flows mainly in the form of holes.
  - (c) Most of the carriers have positive electric charge.
  - (d) The substance acquires an electron surplus.
12. In a P-type material, electrons are
- (a) the majority carriers.
  - (b) the minority carriers.
  - (c) positively charged.
  - (d) entirely absent.
13. Holes move from
- (a) minus to plus.
  - (b) plus to minus.
  - (c) P-type to N-type material.
  - (d) N-type to P-type material.
14. When a P-N junction does not conduct even though a voltage is applied, the junction is
- (a) reverse-biased at a voltage less than the avalanche voltage.
  - (b) overdriven.
  - (c) biased past the breaker voltage.
  - (d) in a state of avalanche effect.
15. Holes flow the opposite way from electrons because
- (a) charge carriers flow continuously.
  - (b) they have opposite electric charge.
  - (c) they have the same electric charge.
  - (d) Forget it! Holes flow in the same direction as electrons.
16. If an electron is considered to have a charge of  $-1$  unit, then a hole can be considered to have
- (a) a charge of  $-1$  unit.
  - (b) no charge.

- (c) a charge of +1 unit.
  - (d) a charge that depends on the semiconductor type.
17. When a P-N junction is forward-biased, conduction will not occur unless
- (a) the applied voltage exceeds the forward breakover voltage.
  - (b) the applied voltage is less than the forward breakover voltage.
  - (c) the junction capacitance is high enough.
  - (d) the depletion region is wide enough.
18. If the reverse bias exceeds the avalanche voltage in a P-N junction,
- (a) the junction will be destroyed.
  - (b) the junction will insulate; no current will flow.
  - (c) the junction will conduct current.
  - (d) the capacitance will become extremely low.
19. Avalanche voltage is routinely exceeded when a P-N junction acts as a
- (a) current rectifier.
  - (b) variable resistor.
  - (c) variable capacitor.
  - (d) voltage regulator.
20. Which of the following does not affect the junction capacitance of a diode?
- (a) the cross-sectional area of the P-N junction
  - (b) the width of the depletion region
  - (c) the phase of an applied ac signal
  - (d) the reverse-bias voltage

## 20 CHAPTER

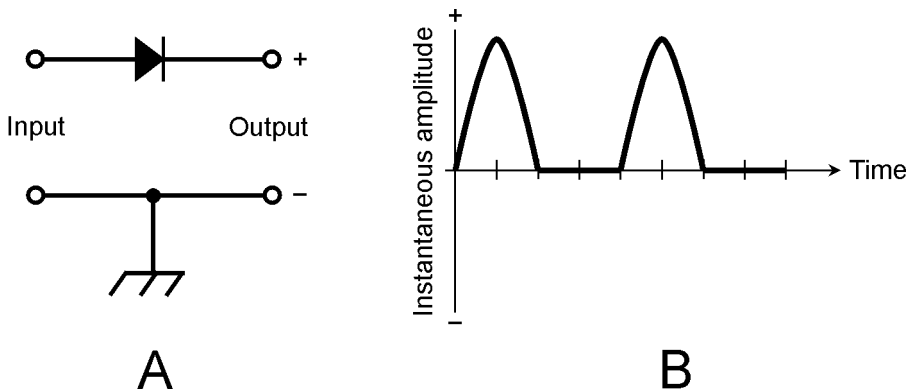
# How Diodes Are Used

IN THE EARLY YEARS OF ELECTRONICS, NEARLY ALL DIODES WERE VACUUM TUBES. TODAY, MOST ARE made from semiconductors. Contemporary diodes can do almost everything that the old ones could, and also some things that people in the tube era could only dream about.

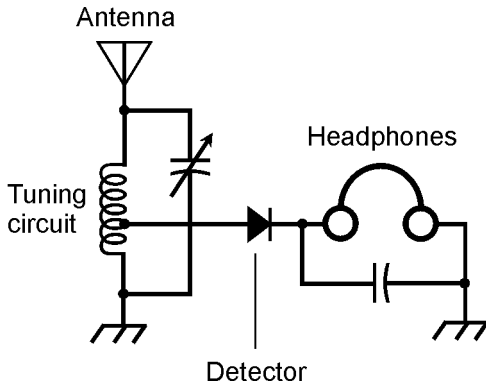
## Rectification

The hallmark of a *rectifier diode* is that it passes current in only one direction. This makes it useful for changing ac to dc. Generally speaking, when the cathode is negative with respect to the anode, current flows; when the cathode is positive relative to the anode, there is no current. The constraints on this behavior are the forward breakover and avalanche voltages, as you learned about in Chap. 19.

Examine the circuit shown at A in Fig. 20-1. Suppose a 60-Hz ac sine wave is applied to the input. During half the cycle, the diode conducts, and during the other half, it doesn't. This cuts off half of every cycle. Depending on which way the diode is hooked up, either the positive half or the negative half of the ac cycle will be removed. Drawing B in Fig. 20-1 shows a graph of the output



**20-1** At A, a half-wave rectifier circuit. At B, the output of the circuit shown at A when an ac sine wave is applied to the input.



**20-2** Schematic diagram of a crystal-set radio receiver.

of the circuit at A. Remember that electrons flow from negative to positive, against the arrow in the diode symbol.

The circuit and wave diagram of Fig. 20-1 show a *half-wave rectifier* circuit. This is the simplest possible rectifier. That's its chief advantage over other, more complicated rectifier circuits. You'll learn about the various types of rectifier diodes and circuits in Chap. 21.

## Detection

One of the earliest diodes, existing even before vacuum tubes, was actually a primitive semiconductor device. Known as a *cat whisker*, it consisted of a fine piece of wire in contact with a small piece of the mineral *galena*. This strange-looking contraption had the ability to act as a rectifier for extremely weak RF currents. When the cat whisker was connected in a circuit such as the one shown in Fig. 20-2, the result was a device capable of picking up amplitude-modulated (AM) radio signals and producing audio output that could be heard in the headset.

The galena, sometimes called a “crystal,” gave rise to the nickname *crystal set* for this primitive radio receiver. You can still build a crystal set today, using a simple RF diode, a coil, a tuning capacitor, a headset, and a long-wire antenna. Notice that there's no battery! The audio is provided by the received signal alone.

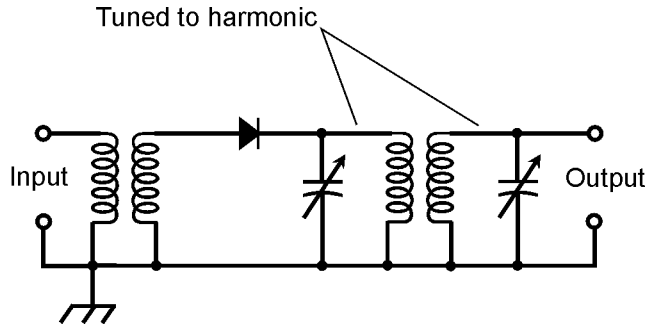
The diode in Fig. 20-2 acts to recover the audio from the radio signal. This process is called *detection*; the circuit is called a *detector* or *demodulator*. If the detector is to be effective, the diode must be of the proper type. It must have low junction capacitance, so that it can work as a rectifier (and not as a capacitor) at radio frequencies. Some modern RF diodes are microscopic versions of the old cat whisker, enclosed in a glass case with axial leads.

## Frequency Multiplication

When current passes through a diode, half of the cycle is cut off, as shown in Fig. 20-1B. This occurs no matter what the frequency, from 60-Hz utility current through RF, as long as the diode capacitance is not too great.

The output wave from the diode looks much different than the input wave. This condition is known as *nonlinearity*. Whenever there is nonlinearity of any kind in a circuit—that is, whenever the output waveform is shaped differently from the input waveform—there are harmonics in the output. These are waves at integer multiples of the input frequency. (If you've forgotten what harmonics are, refer to Chap. 9.)

**20-3** A frequency-multiplier circuit using a semiconductor diode.



Often, nonlinearity is undesirable. Then engineers strive to make the circuit *linear*, so the output waveform has exactly the same shape as the input waveform. But sometimes harmonics are desired. Then nonlinearity is introduced deliberately to produce *frequency multiplication*. Diodes are ideal for this purpose. A simple frequency-multiplier circuit is shown in Fig. 20-3. The output *LC* circuit is tuned to the desired  $n$ th harmonic frequency,  $nf_0$ , rather than to the input or fundamental frequency,  $f_0$ .

For a diode to work as a frequency multiplier, it must be of a type that would also work well as a detector at the same frequencies. This means that the component should act like a rectifier, but not like a capacitor.

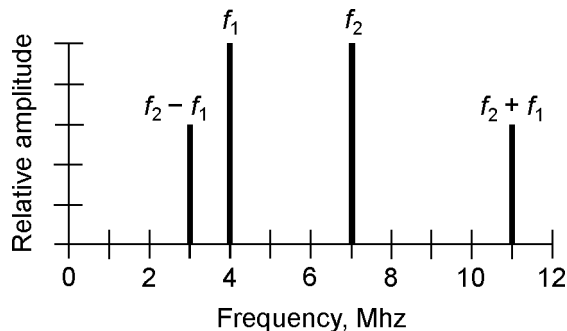
## Signal Mixing

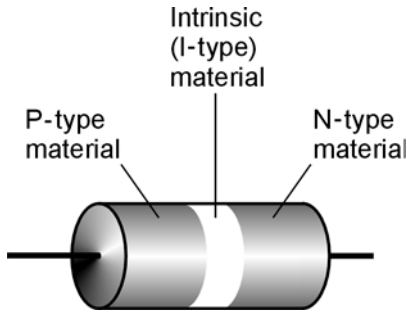
When two waves having different frequencies are combined in a nonlinear circuit, new waves are produced at frequencies equal to the sum and difference of the frequencies of the input waves. Diodes can provide this nonlinearity.

Suppose there are two signals with frequencies  $f_1$  and  $f_2$ . For mathematical convenience, let's assign  $f_2$  to the wave with the higher frequency, and  $f_1$  to the wave with the lower frequency. If these signals are combined in a nonlinear circuit, new waves result. One of them has a frequency of  $f_2 + f_1$ , and the other has a frequency of  $f_2 - f_1$ . These sum and difference frequencies are known as *beat frequencies*. The signals themselves are called *mixing products* or *heterodynes* (Fig. 20-4).

Figure 20-4, incidentally, is an illustration of a *frequency domain* display. The amplitude (on the vertical scale or axis) is shown as a function of the frequency (on the horizontal scale or axis). This sort of display is what engineers see when they look at the screen of a lab instrument known as a *spectrum analyzer*. In contrast, an ordinary oscilloscope displays amplitude (on the vertical scale or axis) as a function of time (on the horizontal scale or axis). The oscilloscope provides a *time domain* display.

**20-4** Spectral (frequency-domain) illustration of signal mixing.





**20-5** The PIN diode has a layer of intrinsic (I type) semiconductor material at the P-N junction.

## Switching

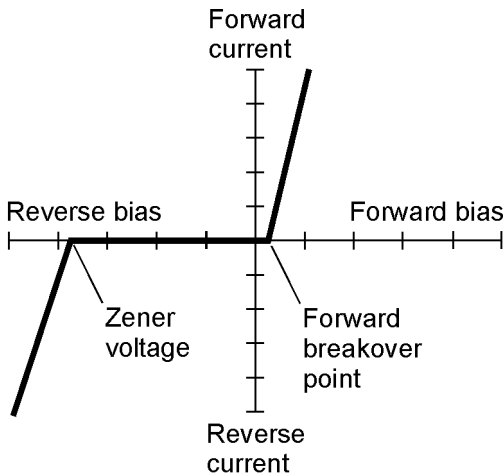
The ability of diodes to conduct with forward bias, and to insulate with reverse bias, makes them useful for switching in some electronic applications. Diodes can perform switching operations much faster than any mechanical device.

One type of diode, made for use as an RF switch, has a special semiconductor layer sandwiched in between the P-type and N-type material. The material in this layer is called an *intrinsic* (or *I-type*) *semiconductor*. The *intrinsic layer* (or *I layer*) reduces the capacitance of the diode, so that it can work at higher frequencies than an ordinary diode. A diode with an I-type semiconductor layer sandwiched in between the P- and N-type layers is called a *PIN diode* (Fig. 20-5).

Direct-current bias, applied to one or more PIN diodes, allows RF currents to be effectively channeled without using relays and cables. A PIN diode also makes a good RF detector, especially at very high frequencies.

## Voltage Regulation

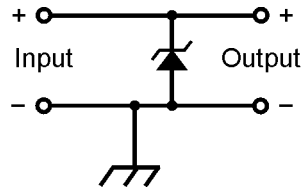
Most diodes have an avalanche breakdown voltage that is much higher than the reverse bias ever gets. The value of the avalanche voltage depends on how a diode is manufactured. *Zener diodes* are specially made so they exhibit well-defined, constant avalanche voltages.



**20-6** Current through a Zener diode as a function of the bias voltage.



**20-7** Connection of a Zener diode for voltage regulation.



Suppose a certain Zener diode has an avalanche voltage, also called the *Zener voltage*, of 50 V. If reverse bias is applied to the P-N junction, the diode acts as an open circuit as long as the bias is less than 50 V. But if the reverse-bias voltage reaches 50 V—even for a brief instant of time—the diode conducts. This effectively prevents the reverse-bias voltage from exceeding 50 V.

The current through a Zener diode, as a function of the voltage, is shown in Fig. 20-6. The Zener voltage is indicated by the abrupt rise in reverse current as the reverse-bias voltage increases. A simple Zener-diode voltage-limiting circuit is shown in Fig. 20-7. Note the polarity of the diode: the cathode is connected to the positive pole, and the anode is connected to the negative pole.

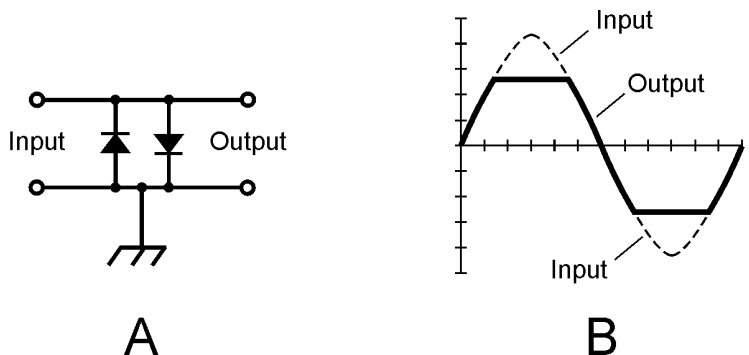
## Amplitude Limiting

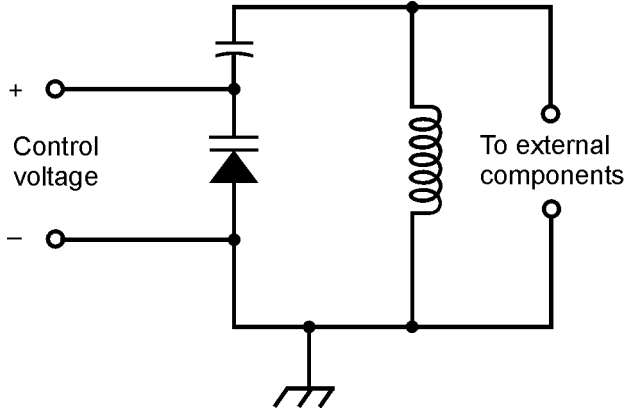
In Chap. 19, you learned that a diode will not conduct until the forward-bias voltage is at least as great as the forward breakover voltage. There's a corollary to this: a diode will always conduct when the forward-bias voltage reaches or exceeds the forward breakover voltage, when the device is conducting current in the forward direction. In the case of silicon diodes this is approximately 0.6 V. For germanium diodes it is about 0.3 V, and for selenium diodes it is about 1 V.

This phenomenon can be used to advantage when it is necessary to limit the amplitude of a signal, as shown in Fig. 20-8. By connecting two identical diodes back-to-back in parallel with the signal path (A), the maximum peak amplitude is limited, or *clipped*, to the forward breakover voltage of the diodes. The input and output waveforms of a clipped signal are illustrated at B. This scheme is sometimes used in radio receivers to prevent “blasting” when a strong signal comes in.

The downside of the *diode limiter* circuit, such as the one shown in Fig. 20-8, is the fact that it introduces distortion when clipping occurs. This might not be a problem for reception of digital signals, for frequency-modulated signals, or for analog signals that rarely reach the limiting voltage. But for amplitude-modulated signals with peaks that rise past the limiting voltage, it can cause trouble.

**20-8** At A, connection of two diodes to act as an ac limiter. At B, illustration of sine-wave peaks cut off by the action of the diodes in an ac limiter.





**20-9** Connection of a varactor diode in a tuned circuit.

## Frequency Control

When a diode is reverse-biased, there is a region at the P-N junction with dielectric (insulating) properties. As you know from Chap. 19, this is called the depletion region, because it has a shortage of majority charge carriers. The width of this zone depends on several things, including the reverse-bias voltage.

As long as the reverse bias is less than the avalanche voltage, varying the bias affects the width of the depletion region. This in turn varies the junction capacitance. This capacitance, which is always small (on the order of picofarads), varies inversely with the square root of the reverse-bias voltage, as long as the reverse bias remains less than the avalanche voltage. Thus, for example, if the reverse-bias voltage is quadrupled, the junction capacitance drops to one-half; if the reverse-bias voltage is decreased by a factor of 9, then the junction capacitance increases by a factor of 3.

Some diodes are manufactured especially for use as variable capacitors. Such a device is known as varactor diode, as you learned in Chap. 19. Varactors are used in a special type of circuit called a *voltage-controlled oscillator* (VCO). Figure 20-9 is a simple example of the LC circuit in a VCO, using a coil, a fixed capacitor, and a varactor. This is a parallel-tuned circuit. The fixed capacitor, whose value is large compared with that of the varactor, serves to keep the coil from short-circuiting the control voltage across the varactor. Notice that the symbol for the varactor has two lines on the cathode side.

## Oscillation and Amplification

Under certain conditions, diodes can be made to produce microwave RF signals. Three types of diodes that can do this are *Gunn diodes*, *IMPATT diodes*, and *tunnel diodes*.

### Gunn Diodes

A Gunn diode can produce up to 1 W of RF power output, but more commonly it works at levels of about 0.1 W. Gunn diodes are usually made from gallium arsenide. A Gunn diode oscillates because of the *Gunn effect*, named after J. Gunn of International Business Machines (IBM), who first observed it in the 1960s. A Gunn diode doesn't work like a rectifier, detector, or mixer. Instead, the oscillation takes place as a result of a quirk called *negative resistance*.

Gunn-diode oscillators are often tuned using varactor diodes. A Gunn-diode oscillator, connected directly to a microwave horn antenna, is known as a *Gunnplexer*. These devices are popular with amateur-radio experimenters at frequencies of 10 GHz and above.

### IMPATT Diodes

The acronym *IMPATT* comes from the words *impact avalanche transit time*. This, like negative resistance, is a rather esoteric phenomenon. An *IMPATT diode* is a microwave oscillating device like a Gunn diode, except that it uses silicon rather than gallium arsenide.

An IMPATT diode can be used as an amplifier for a microwave transmitter that employs a Gunn-diode oscillator. As an oscillator, an IMPATT diode produces about the same amount of output power, at comparable frequencies, as a Gunn diode.

### Tunnel Diodes

Another type of diode that will oscillate at microwave frequencies is the *tunnel diode*, also known as the *Esaki diode*. It produces enough power so it can be used as a local oscillator in a microwave radio receiver, but not much more.

Tunnel diodes work well as amplifiers in microwave receivers, because they generate very little unwanted noise. This is especially true of gallium arsenide devices.

## Energy Emission

Some semiconductor diodes emit radiant energy when a current passes through the P-N junction in a forward direction. This phenomenon occurs as electrons fall from higher to lower energy states within atoms.

### LEDs and IREDS

Depending on the exact mixture of semiconductors used in manufacture, visible light of almost any color can be produced by diodes when bias is applied to them in the forward direction. Infrared-emitting devices also exist. The most common color for a *light-emitting diode* (LED) is bright red. An *infrared-emitting diode* (IRED) produces energy at wavelengths slightly longer than those of visible red light.

The intensity of the radiant energy from an LED or IRED depends to some extent on the forward current. As the current rises, the brightness increases, but only up to a certain point. If the current continues to rise, no further increase in brilliance takes place. The LED or IRED is then said to be in a state of *saturation*.

### Digital Displays

Because LEDs can be made in various different shapes and sizes, they are ideal for use in digital displays. You've seen digital clock radios that use them. They are common in car radios. They make good indicators for "on/off," "a.m./p.m.," "battery low," and other conditions.

In recent years, LED displays have been largely replaced by *liquid crystal displays* (LCDs). The LCD technology has advantages over LED technology, including lower power consumption and better visibility in direct sunlight. However, LCDs require backlighting when the ambient illumination is low.

## Communications

Both LEDs and IREDS are useful in communications because their intensity can be modulated to carry information. When the current through the device is sufficient to produce output, but not enough to cause saturation, the LED or IRED output follows along with rapid current changes. Analog and digital signals can be conveyed over light beams in this way. Some modern telephone systems make use of modulated light, transmitted through clear fibers. This is known as *fiber-optic* technology.

Special LEDs and IREDS produce *coherent radiation*. These are called *laser diodes*. The rays from these diodes aren't the intense, parallel beams that most people imagine when they think about lasers. A laser LED or IRED generates a cone-shaped beam of low intensity. But it can be focused into a parallel beam, and the resulting rays have some of the same advantages found in larger lasers, including the ability to travel long distances with little decrease in their intensity.

## Photosensitive Diodes

Virtually all P-N junctions exhibit conductivity that varies with exposure to radiant electromagnetic energy such as IR, visible light, and UV. The reason that conventional diodes are not affected by these rays is that they are enclosed in opaque packages. Some *photosensitive diodes* have variable dc resistance that depends on the intensity of the electromagnetic rays. Other types of diodes produce their own dc in the presence of radiant energy.

### Silicon Photodiodes

A silicon diode, housed in a transparent case and constructed in such a way that visible light can strike the barrier between the P-type and N-type materials, forms a *silicon photodiode*. A reverse-bias voltage is applied to the device. When radiant energy strikes the junction, current flows. The current is proportional to the intensity of the radiant energy, within certain limits.

Silicon photodiodes are more sensitive at some wavelengths than at others. The greatest sensitivity is in the *near infrared* part of the spectrum, at wavelengths just a little bit longer than the wavelength of visible red light.

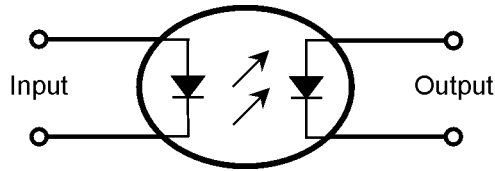
When radiant energy of variable intensity strikes the P-N junction of a reverse-biased silicon photodiode, the output current follows the light-intensity variations. This makes silicon photodiodes useful for receiving modulated-light signals of the kind used in fiber-optic communications systems.

### The Optoisolator

An LED or IRED and a photodiode can be combined in a single package to get a component called an *optoisolator*. This device, the schematic symbol for which is shown in Fig. 20-10, creates a modulated-light signal and sends it over a small, clear gap to a receptor. An LED or IRED converts an electrical signal to visible light or IR; a photodiode changes the visible light or infrared back into an electrical signal.

When a signal is electrically coupled from one circuit to another, the two stages interact. The input impedance of a given stage, such as an amplifier, can affect the behavior of the circuits that feed power to it. This can lead to various sorts of trouble. Optoisolators overcome this effect, because the coupling is not done electrically. If the input impedance of the second circuit changes, the impedance that the first circuit sees is not affected, because it is simply the impedance of the LED

- 20-10** An optoisolator has an LED or IRED at the input and a photodiode at the output.



or IRED. That is where the “isolator” in “optoisolator” comes from. The circuits can be *electronically coupled*, and yet at the same time remain *electrically isolated*.

## Photovoltaic Cells

A silicon diode, with no bias voltage applied, can generate dc all by itself if enough electromagnetic radiation hits its P-N junction. This is known as the *photovoltaic effect*. It is the principle by which solar cells work.

*Photovoltaic cells* are specially manufactured to have the greatest possible P-N junction surface area. This maximizes the amount of light that strikes the junction. A single silicon photovoltaic cell can produce about 0.6 V of dc electricity. The amount of current that it can deliver, and thus the amount of power it can provide, depends on the surface area of the junction.

Photovoltaic cells can be connected in series-parallel combinations to provide power for solid-state electronic devices such as portable radios. These arrays can also be used to charge batteries, allowing for use of the electronic devices when radiant energy is not available (for example, at night!). A large assembly of solar cells, connected in series-parallel, is called a *solar panel*. The power produced by a solar panel depends on the intensity of the light that strikes it, the sum total of the surface areas of all the cells, and the angle at which the light strikes the cells. Some solar panels can produce several kilowatts of electrical power in direct sunlight that shines in such a way that the sun's rays arrive perpendicular to the surfaces of all the cells.

## Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

- When a diode is forward-biased, the anode voltage
  - is negative relative to the cathode voltage.
  - is positive relative to the cathode voltage.
  - is the same as the cathode voltage.
  - alternates between positive and negative relative to the cathode voltage.
- If a diode is connected in series with the secondary winding of an ac transformer, and if the peak voltage across the diode never exceeds the avalanche voltage, then the output of the complete transformer-diode circuit is
  - ac with half the frequency of the input.
  - ac with the same frequency as the input.
  - ac with twice the frequency of the input.
  - none of the above.

3. A crystal set
  - (a) can be used to transmit radio signals.
  - (b) requires a battery with long life.
  - (c) requires no battery.
  - (d) is used for rectifying 60-Hz ac.
4. A diode detector
  - (a) is used in power supplies.
  - (b) is employed in some radio receivers.
  - (c) is used to generate microwave RF signals.
  - (d) changes dc into ac.
5. If the output wave in a circuit has the same shape as the input wave, then
  - (a) the circuit is operating in a linear manner.
  - (b) the circuit is operating as a frequency multiplier.
  - (c) the circuit is operating as a mixer.
  - (d) the circuit is operating as a rectifier.
6. Suppose the two input signal frequencies to a mixer circuit are 3.522 MHz and 3.977 MHz. At which of the following frequencies can we expect a signal to exist at the output?
  - (a) 455 kHz
  - (b) 886 kHz
  - (c) 14.00 MHz
  - (d) 1.129 MHz
7. Fill in the blanks to make the following sentence correct: "A spectrum analyzer provides a display of \_\_\_\_ as a function of \_\_\_\_."
  - (a) amplitude/time
  - (b) time/frequency
  - (c) frequency/time
  - (d) amplitude/frequency
8. Zener voltage is a specialized manifestation of
  - (a) forward breakover voltage.
  - (b) peak forward voltage.
  - (c) avalanche voltage.
  - (d) reverse bias.
9. The forward breakover voltage of a silicon diode is
  - (a) about 0.3 V.
  - (b) about 0.6 V.
  - (c) about 1.0 V.
  - (d) dependent on the avalanche voltage.

10. A diode audio limiter circuit
  - (a) is useful for voltage regulation.
  - (b) always uses Zener diodes.
  - (c) rectifies the audio to reduce distortion.
  - (d) can cause distortion under some conditions.
11. The capacitance of a varactor varies with the
  - (a) forward voltage.
  - (b) reverse voltage.
  - (c) avalanche voltage.
  - (d) forward breakover voltage.
12. The purpose of the I layer in a PIN diode is to
  - (a) minimize the junction capacitance.
  - (b) optimize the avalanche voltage.
  - (c) reduce the forward breakover voltage.
  - (d) increase the current through the diode.
13. Which of these diode types can be used as the key element in the oscillator circuit of a microwave radio transmitter?
  - (a) A rectifier diode
  - (b) A PIN diode
  - (c) An IMPATT diode
  - (d) None of the above
14. A Gunnplexer is often used as a
  - (a) microwave communications device.
  - (b) low-frequency RF detector.
  - (c) high-voltage rectifier.
  - (d) signal mixer or frequency divider.
15. The most likely place you would find an LED would be in
  - (a) a rectifier circuit.
  - (b) a mixer circuit.
  - (c) a digital frequency display.
  - (d) an oscillator circuit.
16. Coherent electromagnetic radiation is produced by a
  - (a) Gunn diode.
  - (b) varactor diode.
  - (c) rectifier diode.
  - (d) laser diode.

17. Suppose you want a circuit to operate in a stable manner when the load impedance varies. You might consider a coupling method that employs

- (a) a Gunn diode.
- (b) an optoisolator.
- (c) a photovoltaic cell.
- (d) a PIN diode.

18. The electrical power that a solar panel can provide in direct sunlight depends on all of the following factors except

- (a) the ac voltage applied to the panel.
- (b) the total surface area of all the cells in the panel.
- (c) the angle at which the sunlight strikes the cells.
- (d) the intensity of the sunlight that strikes the cells.

19. Emission of energy in an IRED is caused by

- (a) high-frequency radio waves.
- (b) rectification.
- (c) changes in electron energy within atoms.
- (d) none of the above.

20. A photodiode, when not used as a photovoltaic cell, has

- (a) reverse bias.
- (b) no bias.
- (c) forward bias.
- (d) negative resistance.



## 21 CHAPTER

# Power Supplies

A *POWER SUPPLY* CONVERTS UTILITY AC TO DC FOR USE WITH CERTAIN ELECTRICAL AND ELECTRONIC devices. In this chapter, we'll examine the components of a typical power supply.

## Power Transformers

Power transformers can be categorized as step-down or step-up. As you remember, the output, or secondary, voltage of a step-down unit is lower than the input, or primary, voltage. The reverse is true for a step-up transformer.

### Step-down

Most solid-state electronic devices, such as radios, need only a few volts. The power supplies for such equipment use step-down power transformers. The physical size of the transformer depends on the current. Some devices need only a small current and a low voltage. The transformer in a radio receiver, for example, can be physically small. A ham radio transmitter or hi-fi amplifier needs more current. This means that the secondary winding of the transformer must consist of heavy-gauge wire, and the core must be bulky to contain the magnetic flux.

### Step-up

Some circuits need high voltage. The cathode-ray tube (CRT) in a conventional home television set needs several hundred volts. Some ham radio power amplifiers use vacuum tubes working at more than 1 kV dc. The transformers in these appliances are step-up types. They are moderate to large in size, because of the number of turns in the secondary, and also because high voltages can spark, or *arc*, between wire turns if the windings are too tight. If a step-up transformer needs to supply only a small amount of current, it need not be big. But for ham radio transmitters and radio or television broadcast amplifiers, the transformers are large, heavy, and expensive.

### Transformer Ratings

Transformers are rated according to output voltage and current. For a given unit, the *volt-ampere* (VA) capacity is often specified. This is the product of the voltage and current.

A transformer with 12-V output, capable of delivering 10 A, has  $12\text{ V} \times 10\text{ A} = 120\text{ VA}$  of capacity. The nature of power-supply filtering, to be discussed later in this chapter, makes it necessary for the power-transformer VA rating to be greater than the wattage consumed by the load.

A high-quality, rugged power transformer, capable of providing the necessary currents and/or voltages, is crucial in any power supply. The transformer is usually the most expensive component to replace.

## Rectifier Diodes

*Rectifier diodes* are available in various sizes, intended for different purposes. Most rectifier diodes are made of silicon, and are known as *silicon rectifiers*. Some are fabricated from selenium, and are called *selenium rectifiers*. Two important features of a power-supply diode are the *average forward current* ( $I_o$ ) rating and the *peak inverse voltage* (PIV) rating.

### Average Forward Current

Electric current produces heat. If the current through a diode is too great, the heat will destroy the P-N junction. When designing a power supply, it is wise to use diodes with an  $I_o$  rating of at least 1.5 times the expected average dc forward current. If this current is 4.0 A, for example, the rectifier diodes should be rated at  $I_o = 6.0\text{ A}$  or more.

Note that  $I_o$  flows through the *diodes*. The current drawn by the *load* is often different from this. Also, note that  $I_o$  is an *average* figure. The *instantaneous* forward current is another thing, and can be 15 or 20 times the  $I_o$ , depending on the nature of the filtering circuit.

Some diodes have *heatsinks* to help carry heat away from the P-N junction. A selenium diode can be recognized by the appearance of its heatsink, which looks something like a baseboard radiator built around a steam pipe.

Diodes can be connected in parallel to increase the current rating over that of an individual diode. When this is done, small-value resistors should be placed in series with each diode in the set to equalize the current. Each resistor should have a value such that the voltage drop across it is about 1 V under normal operating conditions.

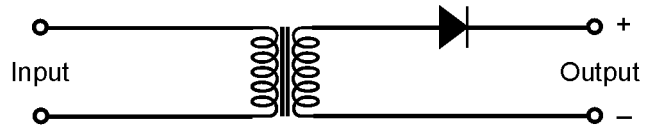
### Peak Inverse Voltage

The PIV rating of a diode is the instantaneous reverse-bias voltage that it can withstand without the avalanche effect taking place. A good power supply has diodes whose PIV ratings are significantly greater than the peak ac input voltage. If the PIV rating is not great enough, the diode or diodes in a supply conduct for part of the reverse cycle. This degrades the efficiency of the supply because the reverse current bucks the forward current.

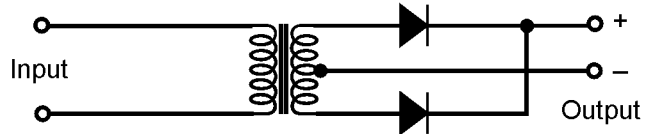
Diodes can be connected in series to get a higher PIV capacity than a single diode alone. This scheme is sometimes seen in high-voltage supplies, such as those needed for tube-type power amplifiers. High-value resistors, of about  $500\ \Omega$  for each peak-inverse volt, are placed across each diode in the set to distribute the reverse bias equally among the diodes. In addition, each diode is shunted by (that is, connected in parallel with) a capacitor of  $0.005\ \mu\text{F}$  or  $0.1\ \mu\text{F}$ .

## Half-Wave Circuit

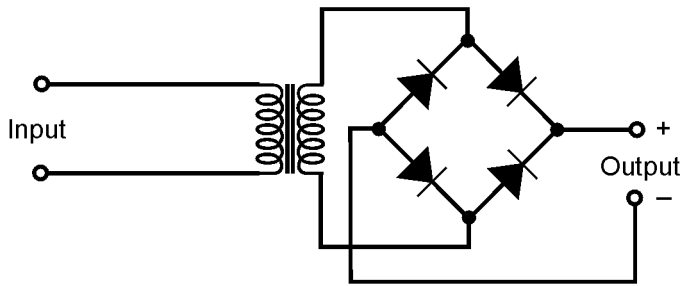
The simplest rectifier circuit, called the *half-wave rectifier* (Fig. 21-1A), has a single diode that chops off half of the ac cycle. The effective (eff) output voltage from a power supply that uses a



A



B



C

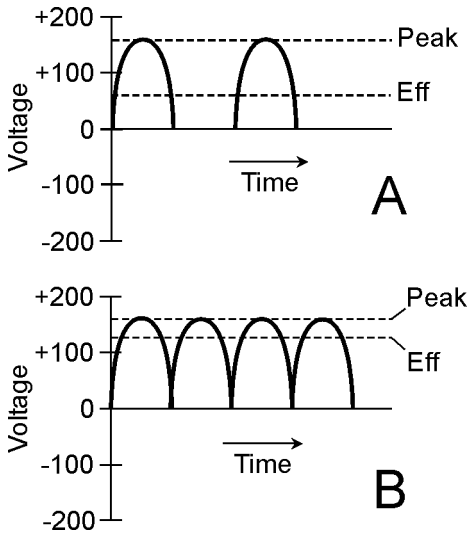
- 21-1 At A, a half-wave rectifier circuit. At B, a full-wave center-tap rectifier circuit. At C, a full-wave bridge rectifier circuit.

half-wave rectifier is much less than the peak transformer output voltage, as shown in Fig. 21-2A. The peak voltage across the diode in the reverse direction can be as much as 2.8 times the applied rms ac voltage.

Most engineers like to use diodes whose PIV ratings are at least 1.5 times the maximum expected peak reverse voltage. Therefore, in a half-wave rectifier circuit, the diodes should be rated for at least  $2.8 \times 1.5$ , or 4.2, times the rms ac voltage that appears across the secondary winding of the power transformer.

Half-wave rectification has shortcomings. First, the output is difficult to filter. Second, the output voltage can drop considerably when the supply is required to deliver high current. Third, half-wave rectification puts a strain on the transformer and diodes because it *pumps* them. The circuit works the diodes hard during half the ac cycle, and lets them loaf during the other half.

Half-wave rectification is usually adequate for use in a power supply that is not required to deliver much current, or when the voltage can vary without affecting the behavior of the equipment connected to it. The main advantage of a half-wave circuit is that it costs less than more sophisticated circuits.



**21-2** At A, the output of a half-wave rectifier. At B, the output of a full-wave rectifier. Note the difference in how the effective (eff) voltages compare with the peak voltages.

## Full-Wave Center-Tap Circuit

A better scheme for changing ac to dc takes advantage of both halves of the ac cycle. A *full-wave center-tap rectifier* has a transformer with a tapped secondary (Fig. 21-1B). The center tap is connected to *electrical ground*, also called *chassis ground*. This produces voltages and currents at the ends of the winding that are in phase opposition with respect to each other. These two ac waves can be individually half-wave rectified, cutting off one half of the cycle and then the other, over and over.

The effective output voltage from a power supply that uses a full-wave center-tap rectifier is greater, relative to the peak voltage, than is the case with the half-wave rectifier (Fig. 21-2B). The PIV across the diodes can, nevertheless, be as much as 2.8 times the applied rms ac voltage. Therefore, the diodes should have a PIV rating of at least 4.2 times the applied rms ac voltage to ensure that they won't break down.

The output of a full-wave center-tap rectifier is easier to filter than that of a half-wave rectifier because the frequency of the pulsations in the dc (known as the *ripple frequency*) from a full-wave rectifier is twice the ripple frequency of the pulsating dc from a half-wave rectifier, assuming identical ac input frequency in either situation. If you compare Fig. 21-2B with Fig. 21-2A, you will see that the full-wave-rectifier output is closer to pure dc than the half-wave rectifier output. Another advantage of a full-wave center-tap rectifier is the fact that it's gentler with the transformer and diodes than a half-wave rectifier. Yet another asset: When a load is applied to the output of a power supply that uses a full-wave center-tap rectifier circuit, the voltage drops less than is the case with a half-wave supply. But because the transformer is more sophisticated, the full-wave center-tap circuit costs more than a half-wave circuit that delivers the same output voltage at the same rated maximum current.

## Full-Wave Bridge Circuit

Another way to get full-wave rectification is the *full-wave bridge rectifier*, often called simply a *bridge*. It is diagrammed in Fig. 21-1C. The output waveform is similar to that of the full-wave center-tap circuit (Fig. 21-2B).

The effective output voltage from a power supply that uses a full-wave bridge rectifier is somewhat less than the peak transformer output voltage, as shown in Fig. 21-2B. The peak voltage across the diodes in the reverse direction is about 1.4 times the applied rms ac voltage. Therefore, each diode needs to have a PIV rating of at least  $1.4 \times 1.5$ , or 2.1, times the rms ac voltage that appears at the transformer secondary.

The bridge circuit does not require a center-tapped transformer secondary. It uses the entire secondary winding on both halves of the wave cycle, so it makes even more efficient use of the transformer than the full-wave center-tap circuit. The bridge is also easier on the diodes than half-wave or full-wave center-tap circuits.

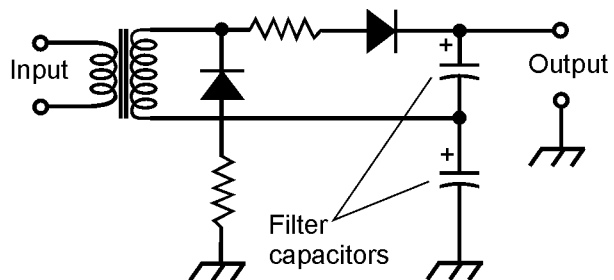
## Voltage-Doubler Circuit

Diodes and capacitors can be interconnected to deliver a dc output that is approximately twice the positive or negative peak ac input voltage. This is called a *voltage-doubler power supply*. This circuit works well only when the load draws low current. Otherwise, the *voltage regulation* is poor; the voltage drops a lot when the current demand is significant.

The best way to build a high-voltage power supply is to use a step-up transformer, not a voltage-doubling scheme. Nevertheless, a voltage-doubler power supply can be, and sometimes is, used when the cost of the circuit must be minimized and the demands placed on it are expected to be modest.

Figure 21-3 is a simplified diagram of a voltage-doubler power supply. It works on the entire ac cycle, so it is called a *full-wave voltage doubler*. This circuit subjects the diodes to voltage peaks in the reverse direction that are 2.8 times the applied rms ac voltage. Therefore, the diodes should be rated for PIV of at least 4.2 times the rms ac voltage that appears across the transformer secondary. When the current drawn is low, the dc output voltage of this type of power supply is approximately 2.8 times the rms ac input voltage.

Proper operation of a voltage-doubler power supply depends on the ability of the capacitors to hold a charge under maximum load. The capacitors must have large values, as well as be capable of handling high voltages. The capacitors serve two purposes: to boost the voltage and to filter the out-



**21-3** A full-wave voltage-doubler power supply.

Connections to  
equipment chassis  
(for electrical grounding)

put. The resistors, which have low ohmic values and are connected in series with the diodes, protect the diodes against *surge currents* that occur when the power supply is first switched on.

## Filtering

Most dc-powered devices need something better—more pure—than the rough, pulsating dc that comes right out of a rectifier circuit. The pulsations (ripple) in the rectifier output can be eliminated by a *filter*.

### Capacitors Alone

The simplest power-supply filter consists of one or more large-value capacitors, connected in parallel with the rectifier output (Fig. 21-4). A good component for this purpose is known as an *electrolytic capacitor*. This type of capacitor is *polarized*, meaning that it must be connected in the correct direction in the circuit. Each capacitor is also rated for a certain maximum voltage. Pay attention to these ratings if you ever work with electrolytic capacitors!

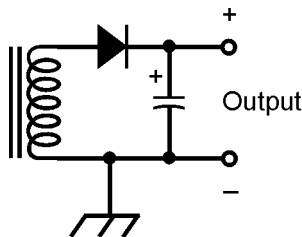
Filter capacitors work by trying to maintain the dc voltage at its peak level, as shown in Fig. 21-5. This is easier to do with the output of a full-wave rectifier (drawing A) than with the output of a half-wave rectifier (drawing B). With a full-wave rectifier receiving a 60-Hz ac electrical input, the ripple frequency is 120 Hz, but with a half-wave rectifier it is 60 Hz. The filter capacitors are thus recharged twice as often with a full-wave rectifier, as compared with a half-wave rectifier. This is why full-wave rectifier circuits produce more pure dc than half-wave rectifier circuits.

### Capacitors and Chokes

Another way to smooth out the dc from a rectifier is to place a large-value inductor in series with the output, and a large-value capacitor in parallel. The inductor is called a *filter choke*.

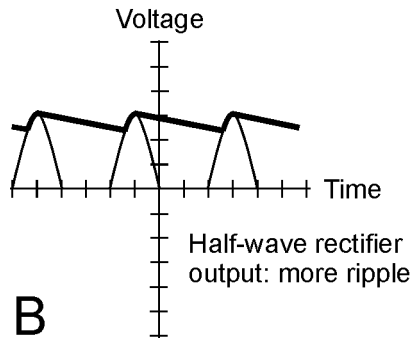
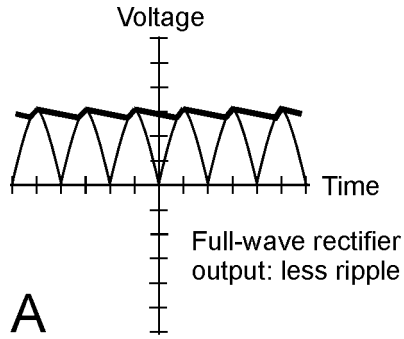
In a filter that uses a capacitor and an inductor, the capacitor can be placed on the rectifier side of the choke. This is a *capacitor-input filter* (Fig. 21-6A). If the filter choke is placed on the rectifier side of the capacitor, the circuit is a *choke-input filter* (Fig. 21-6B). Capacitor-input filtering can be used when a power supply is not required to deliver much current. The output voltage, when the load is light (not much current is drawn), is higher with a capacitor-input filter than with a choke-input filter having identical input. If the supply needs to deliver large or variable amounts of current, a choke-input filter is a better choice, because the output voltage is more stable.

If the output of a power supply must have an absolute minimum of ripple, two or three capacitor/choke pairs can be connected in *cascade* (Fig. 21-7). Each pair constitutes a *section* of the filter. Multisection filters can consist of capacitor-input or choke-input sections, but the two types are never

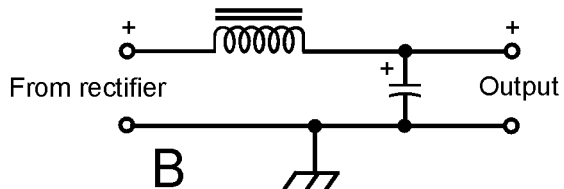
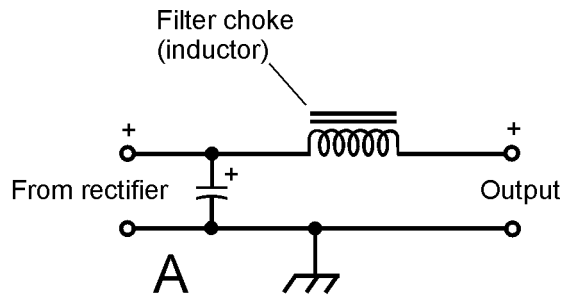


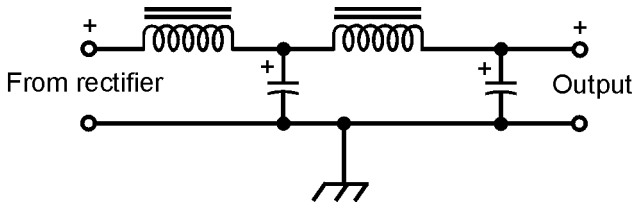
**21-4** A large-value capacitor can be used all by itself as a power-supply filter.

21-5 Filtering of ripple from a full-wave rectifier (A) and from a half-wave rectifier (B).



21-6 At A, a capacitor-input filter. At B, a choke-input filter.





21-7 Two choke-input filter sections in cascade.

mixed. In the example of Fig. 21-7, both capacitor/choke pairs are called *L sections*. If the second capacitor is omitted, the filter becomes a *T section*. If the second choke is omitted, the filter becomes a *pi section*. These sections are named because their schematic diagrams look something like the uppercase English L, the uppercase English T, and the uppercase Greek  $\Pi$ , respectively.

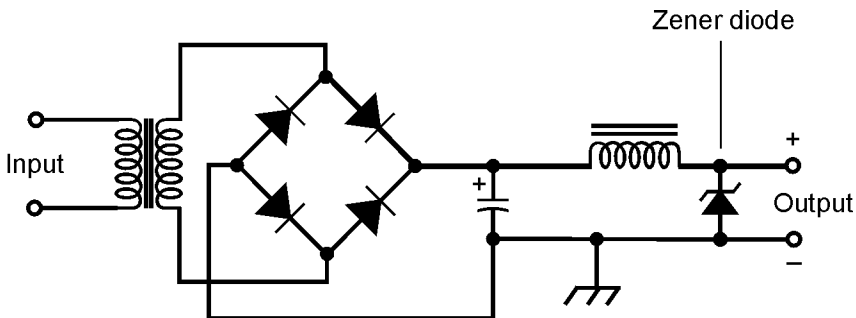
## Voltage Regulation

If a special diode called a *Zener diode* is connected in parallel with the output of a power supply, the diode limits the output voltage. The diode must have an adequate power rating to prevent it from burning out. The limiting voltage depends on the particular Zener diode used. Zener diodes are available for any reasonable power-supply voltage.

Figure 21-8 is a diagram of a full-wave bridge dc power supply including a Zener diode for voltage regulation. Note the direction in which the Zener diode is connected in this application: with the arrow pointing from minus to plus. This is contrary to the polarity used for rectifier diodes. It's important that the polarity be correct with a Zener diode, or it will burn out.

A Zener-diode voltage regulator is inefficient when the supply is used with equipment that draws high current. When a supply must deliver a lot of current, a *power transistor* is used along with the Zener diode to obtain regulation. Figure 21-9 shows such a circuit.

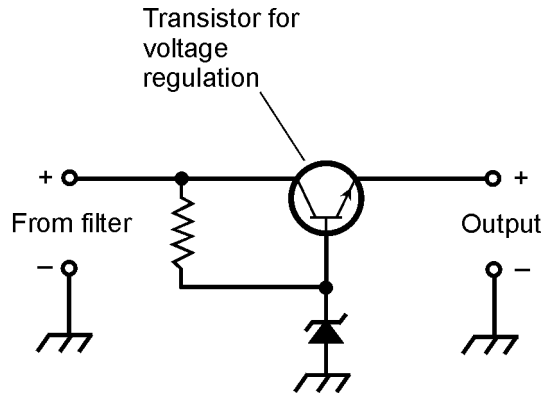
Voltage regulators are available in *integrated-circuit* (IC) form. The *regulator IC*, also called a *regulator chip*, is installed in the power-supply circuit at the output of the filter. In high-voltage power supplies, *electron tubes* are sometimes used as voltage regulators. These are particularly rugged, and can withstand much higher temporary overloads than Zener diodes, transistors, or chips. However, some engineers consider such *regulator tubes* archaic.



21-8 A power supply with a Zener-diode voltage regulator in the output.



21-9 A voltage-regulator circuit using a Zener diode and an NPN transistor.



## Protection of Equipment

The output of a power supply should be free of sudden changes that can damage equipment or components, or interfere with their proper performance. It is also important that voltages not appear on the external surfaces of a power supply, or on the external surfaces of any equipment connected to it.

### Grounding

The best electrical ground for a power supply is the third-wire ground provided in up-to-date ac utility circuits. In an ac outlet, this connection appears as a hole shaped like an uppercase letter D turned on its side. The contacts inside this hole should be connected to a wire that ultimately terminates in a metal rod driven into the earth at the point where the electrical wiring enters the building. That constitutes an *earth ground*.

In older buildings, *two-wire ac systems* are common. These can be recognized by the presence of two slots in the utility outlets, but no ground hole. Some of these systems employ reasonable grounding by means of a scheme called *polarization*, where one slot is longer than the other, the longer slot being connected to electrical ground. But this is not as good as a *three-wire ac system*, in which the ground connection is independent of both the outlet slots.

Unfortunately, the presence of a three-wire or polarized outlet system does not always mean that an appliance connected to an outlet is well grounded. If the appliance design is faulty, or if the ground holes at the outlets were not grounded by the people who installed the electrical system, a power supply can deliver unwanted voltages to the external surfaces of appliances and electronic devices. This can present an electrocution hazard, and can also hinder the performance of equipment connected to the supply.

- **Warning:** All exposed metal surfaces of power supplies should be connected to the grounded wire of a three-wire electrical cord. The third prong of the plug should never be defeated or cut off. Some means should be found to ensure that the electrical system in the building has been properly installed, so you don't work under the illusion that your system has a good ground when it actually does not. If you are in doubt about this, consult a professional electrician.

## Surge Currents

At the instant a power supply is switched on, a surge of current occurs, even with nothing connected to the supply output. This is because the filter capacitors need an initial charge, so they draw a large current for a short time. The *surge current* is far greater than the normal operating current. An extreme current surge of this sort can destroy the rectifier diodes if they are not sufficiently rated and/or protected. The phenomenon is worst in high-voltage supplies and voltage-multiplier circuits. Diode failure as a result of current surges can be prevented in at least three ways:

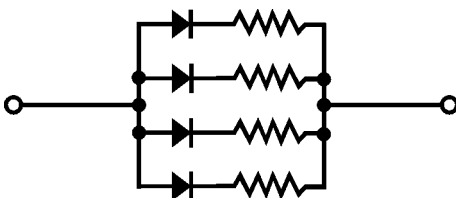
- Use diodes with a current rating of many times the normal operating level.
- Connect several diodes in parallel wherever a diode is called for in the circuit. *Current-equalizing resistors* are necessary (Fig. 21-10). The resistors should have small, identical ohmic values. The diodes should all be identical.
- Use an *automatic switching* circuit in the transformer primary. This type of circuit applies a reduced ac voltage to the transformer for a second or two, and then applies the full input voltage.

## Transients

The ac that appears at utility outlets is a sine wave with a constant voltage near 117 V rms or 234 V rms. But there are often *voltage spikes*, known as *transients*, that can attain positive or negative peak values of several thousand volts. Transients are caused by sudden changes in the load in a utility circuit. A thundershower can produce transients throughout an entire town. Unless they are suppressed, transients can destroy the diodes in a power supply. Transients can also cause problems with sensitive electronic equipment such as computers or microcomputer-controlled appliances.

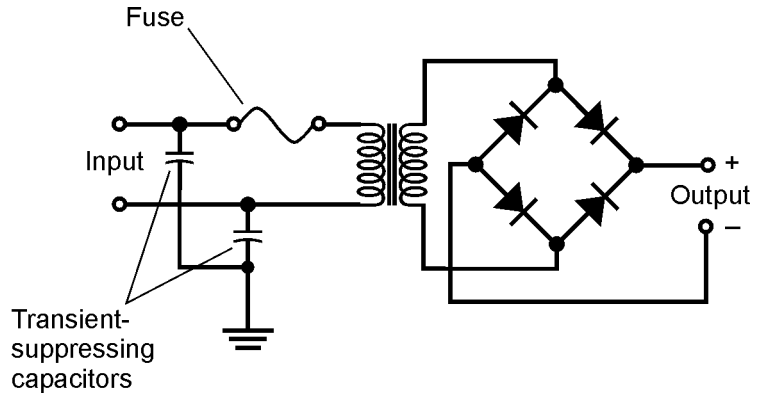
The simplest way to get rid of common transients is to place a small capacitor of about 0.01  $\mu\text{F}$ , rated for 600 V or more, between each side of the transformer primary and electrical ground, as shown in Fig. 21-11. A good component for this purpose is a *disk ceramic capacitor* (not an electrolytic capacitor). Disk ceramic capacitors have no polarity issues. They can be connected in either direction to work equally well.

Commercially made *transient suppressors* are available. These devices, often mistakenly called “surge protectors,” use sophisticated methods to prevent sudden voltage spikes from reaching levels where they can cause problems. It is a good idea to use transient suppressors with all sensitive electronic devices, including computers, hi-fi stereo systems, and television sets. In the event of a thundershower, the best way to protect such equipment is to physically unplug it from the wall outlets until the event has passed.



**21-10** Diodes in parallel, with current-equalizing resistors in series with each diode.

**21-11** A full-wave bridge rectifier with transient-suppression capacitors and a fuse in the transformer primary circuit.



## Fuses

A *fuse* is a piece of soft wire that melts, breaking a circuit if the current exceeds a certain level. A fuse is placed in series with the transformer primary, as shown in Fig. 21-11. A short circuit or overload anywhere in the power supply, or in equipment connected to it, will burn the fuse out. If a fuse blows out, it must be replaced with another of the same rating. Fuses are rated in amperes (A). Thus, a 5-A fuse will carry up to 5 A before blowing out, and a 20-A fuse will carry up to 20 A.

Fuses are available in two types: the *quick-break fuse* and the *slow-blow fuse*. A quick-break fuse is a straight length of wire or a metal strip. A slow-blow fuse usually has a spring inside along with the wire or strip. It's best to replace blown-out fuses with new ones of the same type. Quick-break fuses in slow-blow situations can burn out needlessly, causing inconvenience. Slow-blow fuses in quick-break environments might not provide adequate protection to the equipment, letting excessive current flow for too long before blowing out.

## Circuit Breakers

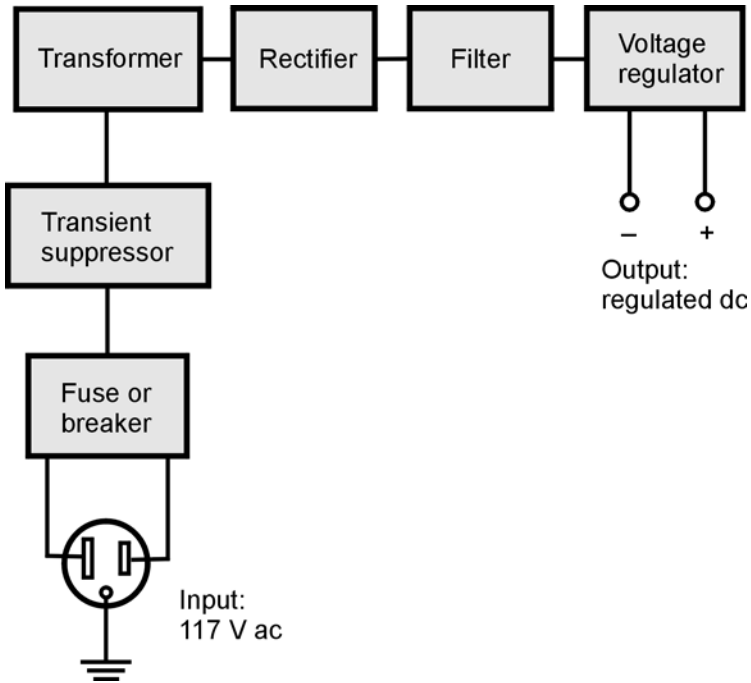
A *circuit breaker* performs the same function as a fuse, except that a breaker can be reset by turning off the power supply, waiting a moment, and then pressing a button or flipping a switch. Some breakers reset automatically when the equipment has been shut off for a certain length of time. Circuit breakers are rated in amperes, just like fuses.

If a fuse or breaker keeps blowing out or tripping, or if it blows or trips immediately after it has been replaced or reset, then something is wrong with the power supply or with the equipment connected to it. Burned-out diodes, a bad transformer, and shorted filter capacitors in the supply can all cause this trouble. A short circuit in the equipment connected to the supply, or the connection of a device in the wrong direction (polarity), can cause repeated fuse blowing or circuit-breaker tripping.

*Never* replace a fuse or breaker with a larger-capacity unit to overcome the inconvenience of repeated fuse/breaker blowing/tripping. Find the cause of the trouble, and repair the equipment as needed. The "penny in the fuse box" scheme can endanger equipment and personnel, and it increases the risk of fire in the event of a short circuit.

## The Complete System

Figure 21-12 is a block diagram of a complete power supply. Note the sequence in which the portions of the system, called *stages*, are connected. A final note of warning is in order here:



**21-12** Block diagram of a complete power supply that can deliver high-quality dc output with ac input.

- **High-voltage power supplies can retain deadly voltages after they have been switched off and unplugged. This is because the filter capacitors retain their charge for some time. If you have any doubt about your ability to safely build or work with a power supply, leave it to a professional.**

## Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

- The output of a rectifier circuit without a filter included is
  - 60 Hz ac.
  - smooth dc.
  - pulsating dc.
  - 120 Hz ac.
- Which of the following components is not necessarily required in a power supply designed to produce 12-V dc output with 117-V rms ac input?
  - The transformer
  - The filter
  - The rectifier
  - All of the above components are required.

3. Of the following appliances, which would need the biggest transformer?
  - (a) A clock radio
  - (b) A television broadcast transmitter
  - (c) A shortwave radio receiver
  - (d) A home television set
4. An advantage of full-wave bridge rectification is the fact that
  - (a) it uses the whole transformer secondary for the entire ac input cycle.
  - (b) it costs less than other rectifier types.
  - (c) it cuts off half of the ac wave cycle.
  - (d) it never needs a filter.
5. In a power supply designed to provide high power at low voltage, the best rectifier circuit would probably be the
  - (a) half-wave arrangement.
  - (b) full-wave, center-tap arrangement.
  - (c) quarter-wave arrangement.
  - (d) voltage doubler arrangement.
6. The part of a power supply immediately preceding the regulator is
  - (a) the transformer.
  - (b) the rectifier.
  - (c) the filter.
  - (d) the ac input.
7. If a half-wave rectifier is used with 165-V pk ac input, the effective dc output voltage is
  - (a) considerably less than 165 V.
  - (b) slightly less than 165 V.
  - (c) exactly 165 V.
  - (d) slightly more than 165 V.
8. If a full-wave bridge circuit is used with a transformer whose secondary provides 50 V rms, the peak voltage that occurs across the diodes in the reverse direction is approximately
  - (a) 50 V pk.
  - (b) 70 V pk.
  - (c) 100 V pk.
  - (d) 140 V pk.
9. What is the principal disadvantage of a voltage-doubler power supply circuit?
  - (a) Excessive current
  - (b) Excessive voltage
  - (c) Insufficient rectification
  - (d) Poor regulation under heavy loads

10. Suppose a transformer secondary provides 10-V rms ac to a voltage-doubler circuit. What is the approximate dc output voltage with no load?
- (a) 14 V
  - (b) 20 V
  - (c) 28 V
  - (d) 36 V
11. The ripple frequency from a full-wave rectifier is
- (a) twice that from a half-wave circuit.
  - (b) the same as that from a half-wave circuit.
  - (c) half that from a half-wave circuit.
  - (d)  $\frac{1}{4}$  that from a half-wave circuit.
12. Which of the following would make the best filter for a power supply?
- (a) A capacitor in series
  - (b) A choke in series
  - (c) A capacitor in series and a choke in parallel
  - (d) A capacitor in parallel and a choke in series
13. If you need exceptionally good ripple filtering for a power supply, which of the following alternatives will yield the best results?
- (a) Connect several capacitors in parallel.
  - (b) Use a choke-input filter.
  - (c) Connect several chokes in series.
  - (d) Use two capacitor/choke filtering sections in cascade.
14. Voltage regulation can be accomplished by a Zener diode connected in
- (a) parallel with the filter output, forward-biased.
  - (b) parallel with the filter output, reverse-biased.
  - (c) series with the filter output, forward-biased.
  - (d) series with the filter output, reverse-biased.
15. A current surge takes place when a power supply is first turned on because
- (a) the transformer core is suddenly magnetized.
  - (b) the diodes suddenly start to conduct.
  - (c) the filter capacitor(s) must be initially charged.
  - (d) arcing takes place in the power switch.
16. Transient suppression is of importance mainly because it minimizes the risk of
- (a) diode failure.
  - (b) transformer imbalance.
  - (c) filter capacitor overcharging.
  - (d) poor voltage regulation.

17. If a fuse blows, and it is replaced with one having a lower current rating, there is a good chance that
- (a) the power supply will be severely damaged.
  - (b) the diodes will not rectify.
  - (c) the fuse will blow out right away.
  - (d) transient suppressors won't work.
18. Suppose you see a fuse with nothing but a straight wire inside. You can assume that this fuse
- (a) is a slow-blow type.
  - (b) is a quick-break type.
  - (c) has a low current rating.
  - (d) has a high current rating.
19. In order to minimize the risk of diode destruction as a result of surge currents that can occur when a power supply is first switched on, which of the following techniques can be useful?
- (a) Connecting multiple diodes in parallel, with low-value resistors in series with each diode
  - (b) Connecting multiple diodes in parallel, with low-value capacitors in series with each diode
  - (c) Connecting multiple diodes in series, with low-value chokes across each diode
  - (d) Connecting multiple diodes in series, with low-value resistors across each diode
20. To repair a damaged power supply with which you are not completely familiar, you should
- (a) install bleeder resistors before beginning your work.
  - (b) remove the fuse before beginning your work.
  - (c) leave it alone and have a professional work on it.
  - (d) short out all the diodes before beginning your work.

## 22 CHAPTER

# The Bipolar Transistor

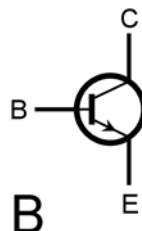
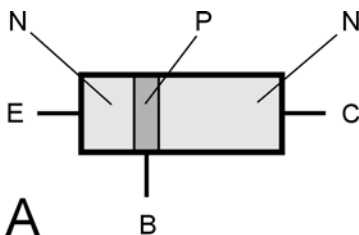
THE WORD *TRANSISTOR* IS A CONTRACTION OF “CURRENT-TRANSFERRING RESISTOR.” A *BIPOLAR transistor* has two P-N junctions. There are two configurations: a P-type layer sandwiched between two N-type layers, or an N-type layer between two P-type layers.

## NPN versus PNP

A simplified drawing of an *NPN bipolar transistor* is shown in Fig. 22-1A, and the schematic symbol is shown in Fig. 22-1B. The P-type, or center, layer is called the *base*. One of the N-type semiconductor layers is the *emitter*, and the other is the *collector*. Sometimes these are labeled *B*, *E*, and *C* in schematic diagrams. A *PNP bipolar transistor* has two P-type layers, one on either side of a thin N-type layer (Fig. 22-2A). The schematic symbol is shown in Fig. 2-22B.

It's easy to tell whether a bipolar transistor in a diagram is NPN or PNP. If the device is NPN, the arrow at the emitter points outward. If the device is PNP, the arrow at the emitter points inward.

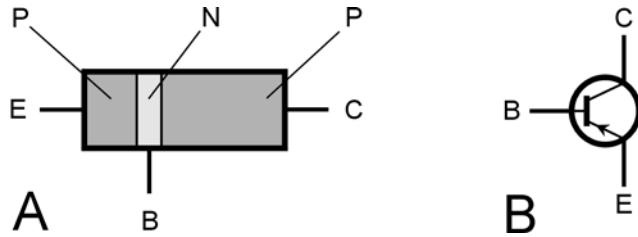
Generally, PNP and NPN transistors can perform the same functions. The differences are the polarities of the voltages and the directions of the resulting currents. In most applications, an NPN device can be replaced with a PNP device or vice versa, the power-supply polarity can be reversed, and the circuit will work in the same way—as long as the new device has the appropriate specifications.



22-1 At A, pictorial diagram of an NPN transistor. At B, the schematic symbol. Electrodes are E = emitter, B = base, and C = collector.



**22-2** At A, pictorial diagram of a PNP transistor. At B, the schematic symbol. Electrodes are E = emitter, B = base, and C = collector.



## Biasing

Imagine a bipolar transistor as consisting of two diodes in reverse series. You can't normally connect two diodes this way and get a working transistor, but the analogy is good for *modeling* the behavior of bipolar transistors. A dual-diode NPN transistor model is shown in Fig. 22-3A. The base is formed by the connection of the two anodes. The emitter is one of the cathodes, and the collector is the other cathode. Figure 22-3B shows the equivalent real-world NPN transistor circuit.

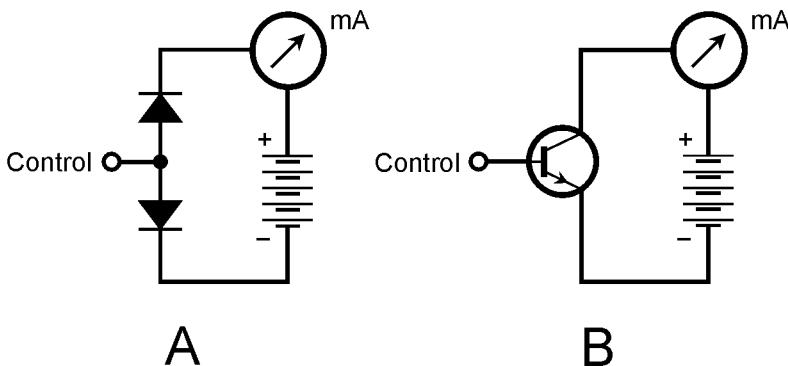
### The NPN Case

The normal method of biasing an NPN transistor is to have the collector voltage positive with respect to the emitter. This is shown by the connection of the battery in Figs. 22-3A and 22-3B. Typical dc voltages for a transistor power supply range between 3 V and about 50 V. A typical voltage is 12 V.

In the model and also in the real-world transistor circuit, the base is labeled “control,” because the flow of current through the transistor depends critically on what happens at this electrode.

### Zero Bias for NPN

Suppose the base of a transistor is at the same voltage as the emitter. This is known as *zero bias*. When the forward bias is zero, the emitter-base current, often called simply *base current* and de-



**22-3** At A, the dual-diode model of a simple NPN circuit. At B, the actual transistor circuit.

noted  $I_B$ , is zero, and the emitter-base (E-B) junction does not conduct. This prevents current from flowing between the emitter and collector, unless a signal is injected at the base to change the situation. Such a signal must, at least momentarily, attain a positive voltage equal to or greater than the forward breakover voltage of the E-B junction.

### Reverse Bias for NPN

Now imagine that a second battery is connected between the base and the emitter in the circuit of Fig. 22-3B, with the polarity such that  $E_B$  becomes negative with respect to the emitter. The addition of this new battery will cause the E-B junction to be *reverse-biased*. No current flows through the E-B junction in this situation (as long as the new battery voltage is not so great that avalanche breakdown occurs). A signal might be injected at the base to cause a flow of current, but such a signal must attain, at least momentarily, a positive voltage high enough to overcome both the reverse bias and the forward breakover voltage of the junction.

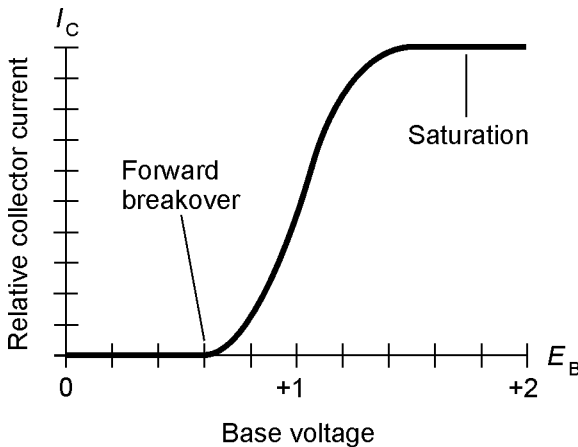
### Forward Bias for NPN

Now suppose that  $E_B$  is made positive with respect to the emitter, starting at small voltages and gradually increasing. This is *forward bias*. If the forward bias is less than the forward breakover voltage, no current will flow. But as the base voltage  $E_B$  reaches the breakover point, the E-B junction will start to conduct.

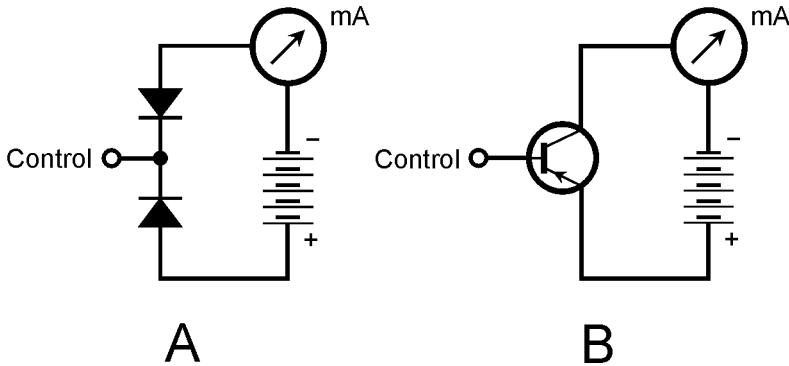
The base-collector (B-C) junction of a bipolar transistor is normally reverse-biased. It will remain reverse-biased as long as  $E_B$  is less than the supply voltage (in this case 12 V). In practical transistor circuits, it is common for  $E_B$  to be set at a fraction of the supply voltage. Despite the reverse bias of the B-C junction, a significant emitter-collector current, called *collector current* and denoted  $I_C$ , will flow once the E-B junction conducts.

In a real transistor circuit such as the one shown in Fig. 22-3B, the meter reading will jump when the forward breakover voltage of the E-B junction is reached. Then even a small rise in  $E_B$ , attended by a rise in  $I_B$ , will cause a large increase in  $I_C$ , as shown in Fig. 22-4.

If  $E_B$  continues to rise, a point will eventually be reached where the  $I_C$  versus  $E_B$  curve levels off. The transistor is then said to be *saturated* or *in saturation*. It is wide open, conducting as much as it can.



22-4 Relative collector current ( $I_C$ ) as a function of base voltage ( $E_B$ ) for a hypothetical NPN silicon transistor.



22-5 At A, the dual-diode model of a simple PNP circuit. At B, the actual transistor circuit.

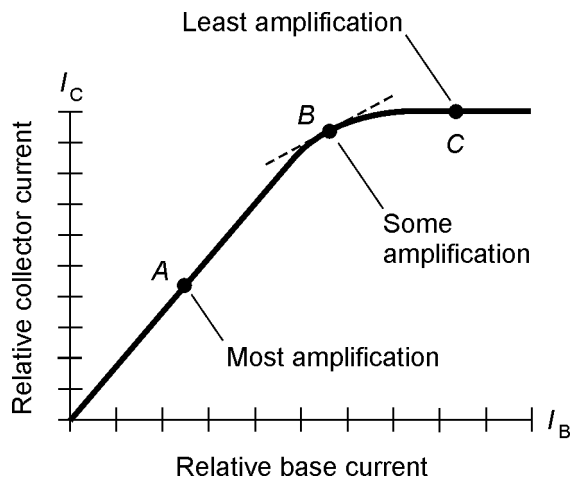
### PNP Biasing

For a PNP transistor, the situation is a mirror image of the case for an NPN device. The diodes are reversed, the arrow points inward rather than outward in the transistor symbol, and all the polarities are reversed. The dual-diode PNP model, along with the real-world transistor circuit, are shown in Fig. 22-5. In the preceding discussion, replace every occurrence of the word “positive” with the word “negative.” Qualitatively, the same things happen: small changes in  $E_B$  cause small changes in  $I_B$ , which in turn produce large fluctuations in  $I_C$ .

## Biasing for Amplification

Because a small change in  $I_B$  causes a large variation in  $I_C$  when the bias is just right, a transistor can operate as a *current amplifier*. If you look at Fig. 22-6, you’ll see that there are some bias values at which a transistor won’t provide any current amplification. If the transistor is in saturation, the  $I_C$

22-6 Three different transistor bias points. The most amplification is obtained when the bias is near the middle of the straight-line portion of the curve.



versus  $I_B$  curve is horizontal. A small change in  $I_B$ , in these portions of the curve, causes little or no change in  $I_C$ . But if the transistor is biased near the middle of the straight-line part of the curve in Fig. 22-6, the transistor will work as a current amplifier.

The same situation holds for the curve in Fig. 22-4. At some bias points, a small change in  $E_B$  does not produce much, if any, change in  $I_C$ ; at other points, a small change in  $E_B$  produces a dramatic change in  $I_C$ . Whenever we want a transistor to amplify a signal, it's important that it be biased in such a way that a small change in the base current or voltage will result in a large change in the collector current.

### Static Current Amplification

Current amplification is often called *beta* by engineers. It can range from a factor of just a few times up to hundreds of times. The beta of a transistor can be expressed as the *static forward current transfer ratio*, abbreviated  $H_{FE}$ . Mathematically, this is the collector current divided by the base current:

$$H_{FE} = I_C / I_B$$

For example, if a base current,  $I_B$ , of 1 mA results in a collector current,  $I_C$ , of 35 mA, then  $H_{FE} = 35/1 = 35$ . If  $I_B = 0.5$  mA and  $I_C = 35$  mA, then  $H_{FE} = 35/0.5 = 70$ . The  $H_{FE}$  specification for a particular transistor represents the greatest amount of current amplification that can be obtained with it.

### Dynamic Current Amplification

A more practical way to define current amplification is as the ratio of the difference in  $I_C$  to the difference in  $I_B$  that occurs when a small signal is applied to the base of a transistor. Abbreviate the words "the difference in" by  $d$ . Then, according to this second definition:

$$\text{Current amplification} = dI_C / dI_B$$

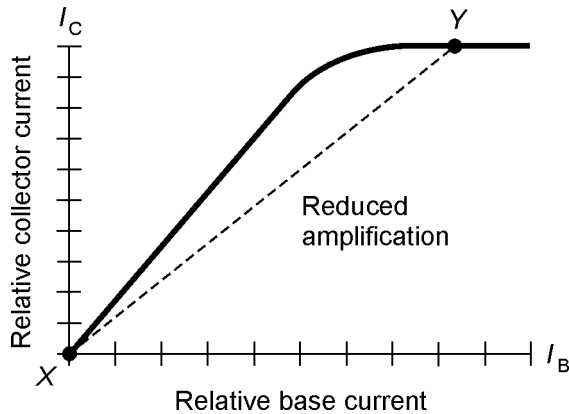
Figure 22-6 is a graph of the collector current as a function of the base current ( $I_C$  versus  $I_B$ ) for a hypothetical transistor. Three different points are shown, corresponding to three different bias scenarios. The ratio  $dI_C/dI_B$  is different for each of the points in this graph. Geometrically,  $dI_C/dI_B$  at a given point is the slope of a line tangent to the curve at that point. The tangent line for point  $B$  in Fig. 22-6 is a dashed straight line; the tangent lines for points  $A$  and  $C$  lie right along the curve and are therefore not shown. The steeper the slope of the line, the greater is  $dI_C/dI_B$ . Point  $A$  provides the highest value of  $dI_C/dI_B$ , provided the input signal is not too strong. This value is very close to  $H_{FE}$ .

For small-signal amplification, point  $A$  in Fig. 22-6 represents a good bias level. Engineers would say that it's a good *operating point*. At point  $B$ ,  $dI_C/dI_B$  is smaller than at point  $A$ , so point  $B$  is not as good for small-signal amplification. At point  $C$ ,  $dI_C/dI_B$  is practically zero. The transistor won't amplify much, if at all, if it is biased at this point.

### Overdrive

Even when a transistor is biased for the greatest possible current amplification (at or near point  $A$  in Fig. 22-6), a strong ac input signal can drive it to point  $B$  or beyond during part of the signal cycle.

22-7 Excessive input reduces amplification.



Then,  $dI_C/dI_B$  is reduced, as shown in Fig. 22-7. Points X and Y in the graph represent the instantaneous current extremes during the signal cycle in this particular case.

When conditions are like those in Fig. 22-7, a transistor amplifier will cause *distortion* in the signal. This means that the output wave will not have the same shape as the input wave. This phenomenon is known as *nonlinearity*. It can sometimes be tolerated, but often it is undesirable. When the input signal to a transistor amplifier is too strong, the condition is called *overdrive*, and the amplifier is said to be *overdriven*.

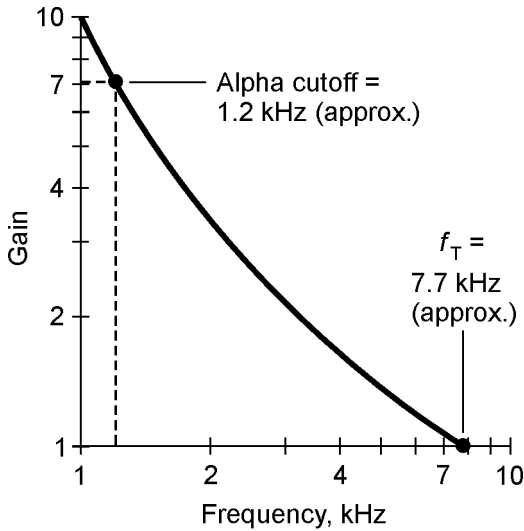
Overdrive can cause problems other than signal distortion. An overdriven transistor is in or near saturation during part of the input signal cycle. This reduces circuit efficiency, causes excessive collector current, and can overheat the base-collector (B-C) junction. Sometimes overdrive can destroy a transistor.

## Gain versus Frequency

Another important specification for a transistor is the range of frequencies over which it can be used as an amplifier. All transistors have an *amplification factor*, or *gain*, that decreases as the signal frequency increases. Some devices work well only up to a few megahertz; others can be used to several gigahertz.

Gain can be expressed in various ways. In the preceding discussion, you learned a little about *current gain*, expressed as a ratio. You will also hear about *voltage gain* or *power gain* in amplifier circuits. These, too, can be expressed as ratios. For example, if the voltage gain of a circuit is 15, then the output signal voltage (rms, peak, or peak-to-peak) is 15 times the input signal voltage. If the power gain of a circuit is 25, then the output signal power is 25 times the input signal power.

Two expressions are commonly used for the gain-versus-frequency behavior of a bipolar transistor. The *gain bandwidth product*, abbreviated  $f_T$ , is the frequency at which the gain becomes equal to 1 with the emitter connected to ground. If you try to make an amplifier using a transistor at a frequency higher than its  $f_T$  specification, you are bound to fail. The *alpha cutoff frequency* of a transistor is the frequency at which the gain becomes 0.707 times its value when the input signal frequency is 1 kHz. A transistor can have considerable gain at its alpha cutoff frequency. By looking at this



**22-8** Alpha cutoff and gain bandwidth product for a hypothetical transistor.

specification for a particular transistor, you can get an idea of how rapidly it loses its ability to amplify as the frequency goes up. Some devices die off faster than others.

Figure 22-8 shows the gain bandwidth product and alpha cutoff frequency for a hypothetical transistor, on a graph of gain versus frequency. Note that the scales of this graph are not linear; that is, the divisions are not evenly spaced. This type of graph is called a *log-log* graph because both scales are *logarithmic* rather than linear.

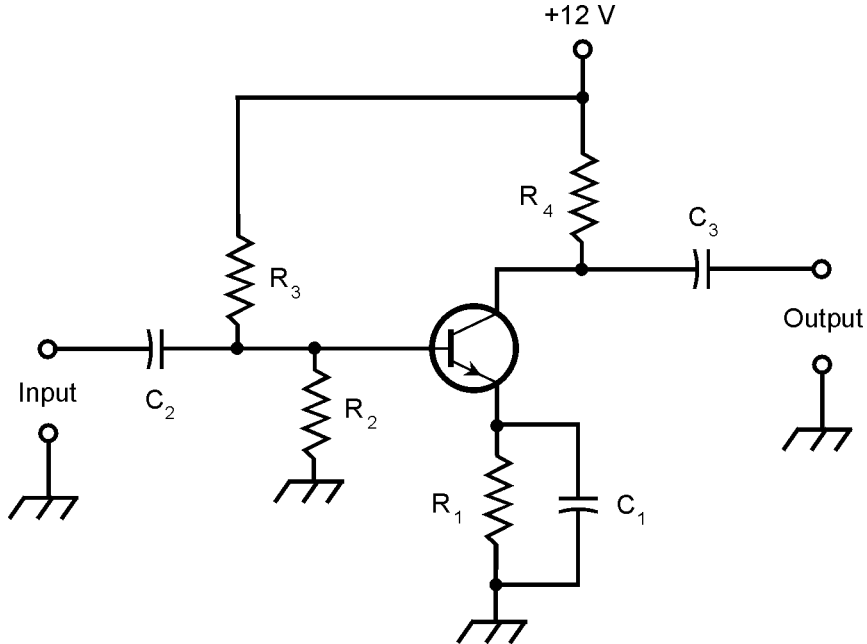
## Common Emitter Circuit

A transistor can be hooked up in three general ways. The emitter can be grounded for signal, the base can be grounded for signal, or the collector can be grounded for signal. An often-used arrangement is the *common emitter circuit*. “Common” means “grounded for the signal.” The basic configuration is shown in Fig. 22-9.

A terminal can be at ground potential for a signal, and yet have a significant dc voltage. In the circuit shown, capacitor  $C_1$  appears as a short circuit to the ac signal, so the emitter is at *signal ground*. But resistor  $R_1$  causes the emitter to have a certain positive dc voltage with respect to ground (or a negative voltage, if a PNP transistor is used). The exact dc voltage at the emitter depends on the resistance of  $R_1$ , and on the bias. The bias is set by the ratio of the values of resistors  $R_2$  and  $R_3$ . The bias can be anything from zero, or ground potential, to +12 V, the supply voltage. Normally it is a couple of volts.

Capacitors  $C_2$  and  $C_3$  block dc to or from the input and output circuitry (whatever that might be) while letting the ac signal pass. Resistor  $R_4$  keeps the output signal from being shorted out through the power supply. A signal enters the common emitter circuit through  $C_2$ , where it causes the base current,  $I_B$ , to vary. The small fluctuations in  $I_B$  cause large changes in the collector current,  $I_C$ . This current passes through resistor  $R_4$ , causing a fluctuating dc voltage to appear across this resistor. The ac part of this passes unhindered through capacitor  $C_3$  to the output.

The circuit of Fig. 22-9 is the basis for many amplifiers, from audio frequencies through ultra-high radio frequencies. The common emitter configuration produces the largest gain of any arrangement. The output wave is  $180^\circ$  out of phase with respect to the input wave.



**22-9** Common emitter configuration. This diagram shows an NPN transistor circuit.

## Common Base Circuit

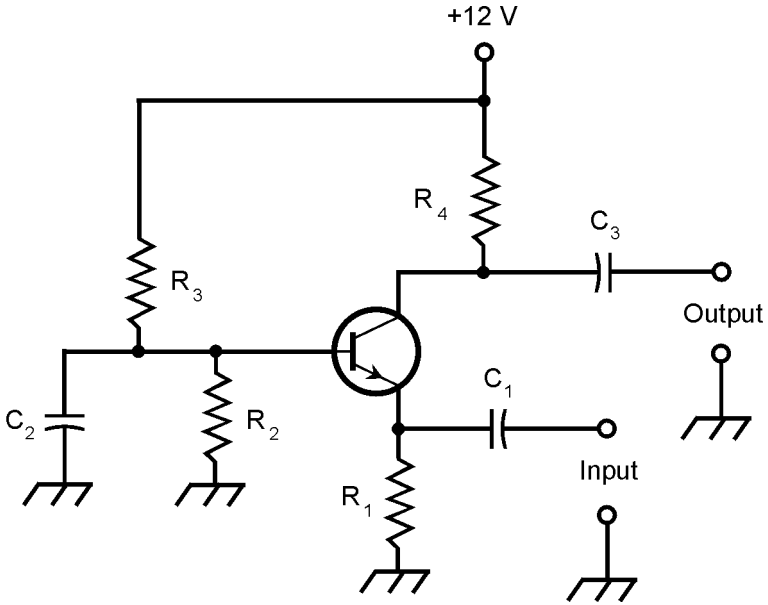
As its name implies, the *common base circuit* (Fig. 22-10) has the base at signal ground. The dc bias is the same as for the common emitter circuit, but the input signal is applied at the emitter, instead of at the base. This causes fluctuations in the voltage across  $R_1$ , causing variations in  $I_B$ . The result of these small current fluctuations is a large change in the current through  $R_4$ . Therefore, amplification occurs. The output wave is in phase with the input wave.

The signal enters through capacitor  $C_1$ . Resistor  $R_1$  keeps the input signal from being shorted to ground. Bias is provided by  $R_2$  and  $R_3$ . Capacitor  $C_2$  keeps the base at signal ground. Resistor  $R_4$  keeps the signal from being shorted out through the power supply. The output is taken through  $C_3$ .

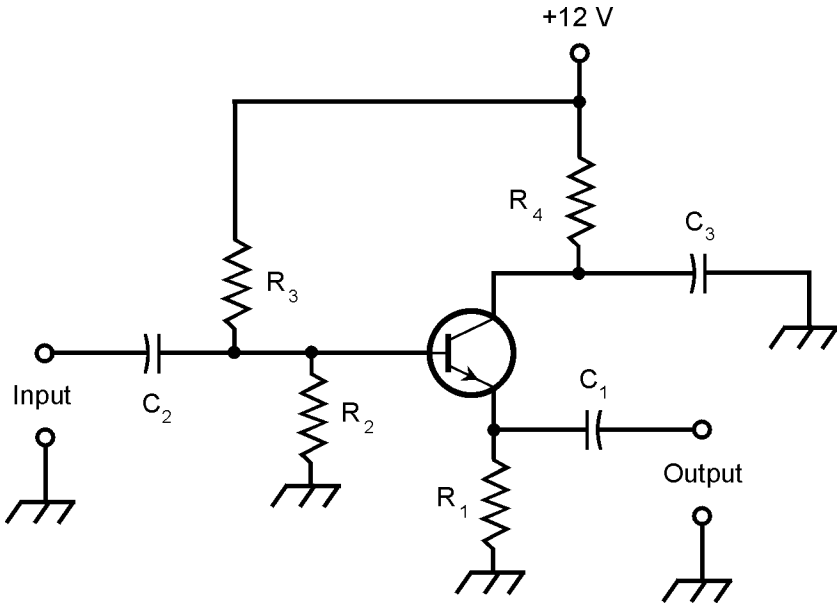
The common base circuit provides somewhat less gain than a common emitter circuit. However, it is more stable than the common emitter configuration in some applications, especially RF power amplifiers.

## Common Collector Circuit

A *common collector circuit* (Fig. 22-11) operates with the collector at signal ground. The input is applied at the base, just as it is with the common emitter circuit. The signal passes through  $C_2$  onto the base of the transistor. Resistors  $R_2$  and  $R_3$  provide the correct bias for the base. Resistor  $R_4$  limits the current through the transistor. Capacitor  $C_3$  keeps the collector at signal ground. A fluctuating direct current flows through  $R_1$ , and a fluctuating dc voltage therefore appears across it. The ac



**22-10** Common base configuration. This diagram shows an NPN transistor circuit.



**22-11** Common collector configuration, also known as an emitter follower. This diagram shows an NPN transistor circuit.



part of this voltage passes through  $C_1$  to the output. Because the output follows the emitter current, this circuit is sometimes called an *emitter follower circuit*.

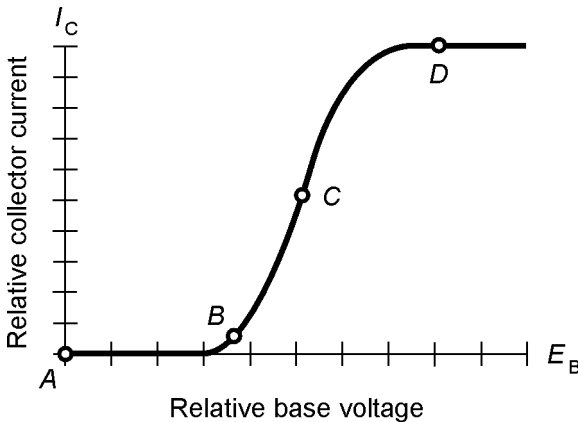
The output wave of a common collector circuit is in phase with the input wave. This circuit is unique because its input impedance is high, while its output impedance is low. For this reason, the common collector circuit can be used to match high impedances to low impedances. When well designed, an emitter follower works over a wide range of frequencies, and is a low-cost alternative to a broadband impedance-matching transformer.

## Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. In a PNP circuit, the collector
  - (a) has an arrow pointing inward.
  - (b) is positive with respect to the emitter.
  - (c) is biased at a small fraction of the base bias.
  - (d) is negative with respect to the emitter.
2. In many cases, a PNP transistor can be replaced with an NPN device and the circuit will do the same thing, provided that
  - (a) the power supply or battery polarity is reversed.
  - (b) the collector and emitter leads are interchanged.
  - (c) the arrow is pointing inward.
  - (d) Forget it! A PNP transistor can never be replaced with an NPN transistor.
3. A bipolar transistor has
  - (a) three P-N junctions.
  - (b) three semiconductor layers.
  - (c) two N-type layers around a P-type layer.
  - (d) a low avalanche voltage.
4. In the dual-diode model of an NPN transistor, the emitter corresponds to
  - (a) the point where the cathodes are connected together.
  - (b) the point where the cathode of one diode is connected to the anode of the other.
  - (c) the point where the anodes are connected together.
  - (d) either of the diode cathodes.
5. The current through a transistor depends on
  - (a)  $E_C$ .
  - (b)  $E_B$  relative to  $E_C$ .
  - (c)  $I_B$ .
  - (d) more than one of the above.

6. With no signal input, a bipolar transistor would have the least  $I_C$  when
  - (a) the emitter is grounded.
  - (b) the E-B junction is forward-biased.
  - (c) the E-B junction is reverse-biased.
  - (d) the E-B current is high.
7. When a transistor is conducting as much as it can, it is said to be
  - (a) in a state of cutoff.
  - (b) in a state of saturation.
  - (c) in a state of reverse bias.
  - (d) in a state of avalanche breakdown.
8. Refer to the curve shown in Fig. 22-12. Which operating point is best if a large amplification factor is desired with a weak signal input?
  - (a) Point *A*
  - (b) Point *B*
  - (c) Point *C*
  - (d) Point *D*
9. In Fig. 22-12, the forward breakover point for the E-B junction is nearest to
  - (a) no point on this graph.
  - (b) point *B*.
  - (c) point *C*.
  - (d) point *D*.
10. In Fig. 22-12, saturation is nearest to
  - (a) point *A*.
  - (b) point *B*.
  - (c) point *C*.
  - (d) point *D*.



22-12 Illustration for Quiz Questions 8 through 11.

11. In Fig. 22-12, the greatest gain occurs at
  - (a) point *A*.
  - (b) point *B*.
  - (c) point *C*.
  - (d) point *D*.
12. In a common emitter circuit, the gain bandwidth product is
  - (a) the frequency at which the gain is 1.
  - (b) the frequency at which the gain is 0.707 times its value at 1 MHz.
  - (c) the frequency at which the gain is greatest.
  - (d) the difference between the frequency at which the gain is greatest, and the frequency at which the gain is 1.
13. The bipolar-transistor configuration most often used for matching a high input impedance to a low output impedance puts signal ground at
  - (a) the emitter.
  - (b) the base.
  - (c) the collector.
  - (d) any point; it doesn't matter.
14. The output is in phase with the input in
  - (a) a common emitter circuit.
  - (b) a common base circuit.
  - (c) a common collector circuit.
  - (d) more than one of the above.
15. The greatest possible amplification is obtained in
  - (a) a common emitter circuit.
  - (b) a common base circuit.
  - (c) a common collector circuit.
  - (d) more than one of the above.
16. The input is applied to the collector in
  - (a) a common emitter circuit.
  - (b) a common base circuit.
  - (c) a common collector circuit.
  - (d) none of the above.
17. The configuration noted for its stability in RF power amplifiers is the
  - (a) common emitter circuit.
  - (b) common base circuit.
  - (c) common collector circuit.
  - (d) emitter follower circuit.

18. In a common base circuit, the output is taken from
- (a) the emitter.
  - (b) the base.
  - (c) the collector.
  - (d) more than one of the above.
19. Suppose that the input signal to a transistor amplifier results in saturation during part of the cycle. This produces
- (a) the greatest possible amplification.
  - (b) reduced efficiency.
  - (c) avalanche effect.
  - (d) nonlinear output impedance.
20. Suppose that the gain of a transistor in a common emitter circuit is 100 at a frequency of 1 kHz, and the gain is 70.7 at 335 kHz. The gain drops to 1 at 210 MHz. The alpha cutoff frequency is
- (a) 1 kHz.
  - (b) 335 kHz.
  - (c) 210 MHz.
  - (d) impossible to define based on this data.

## 23 CHAPTER

# The Field Effect Transistor

THE BIPOLAR TRANSISTOR ISN'T THE ONLY SEMICONDUCTOR TRANSISTOR THAT CAN AMPLIFY. THE other major category of transistor is the *field effect transistor* (FET). There are two main types of FET: the *junction FET* (JFET) and the *metal-oxide FET* (MOSFET).

## Principle of the JFET

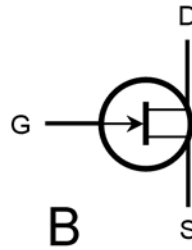
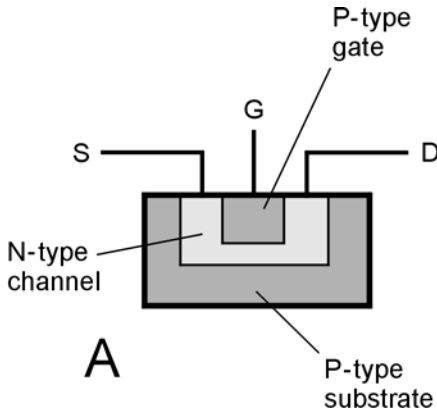
In a JFET, the current varies because of the effects of an electric field within the device. Charge carriers (electrons or holes) flow from the *source* (S) electrode to the *drain* (D) electrode. This results in a *drain current*,  $I_D$ , that is normally the same as the *source current*,  $I_S$ . The rate of flow of charge carriers—that is, the current—depends on the voltage at a control electrode called the *gate* (G). Fluctuations in *gate voltage*,  $E_G$ , cause changes in the current through the *channel*, which is the path between the source and the drain. The current through the channel is normally equal to  $I_D$ . Small fluctuations in  $E_G$  can cause large variations in  $I_D$ . This fluctuating drain current can produce significant fluctuations in the voltage across an output resistance.

### N-Channel versus P-Channel

A simplified drawing of an *N-channel JFET*, and its schematic symbol, are shown in Fig. 23-1. The N-type material forms the channel, or the path for charge carriers. The majority carriers are electrons. The drain is placed at a positive dc voltage with respect to the source.

In an N-channel device, the gate consists of P-type material. Another section of P-type material, called the *substrate*, forms a boundary on the side of the channel opposite the gate. The voltage on the gate produces an electric field that interferes with the flow of charge carriers through the channel. The more negative  $E_G$  becomes, the more the electric field chokes off the current through the channel, and the smaller  $I_D$  becomes.

A *P-channel JFET* (Fig. 23-2) has a channel of P-type semiconductor. The majority charge carriers are holes. The drain is negative with respect to the source. The more positive  $E_G$  gets, the more the electric field chokes off the current through the channel, and the smaller  $I_D$  becomes.

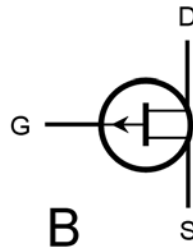
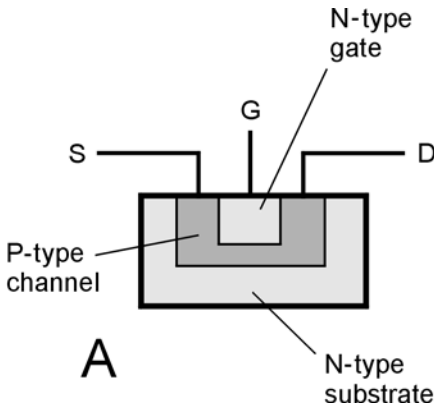


**23-1** At A, pictorial diagram of an N-channel JFET. At B, the schematic symbol. Electrodes are S = source, G = gate, and D = drain.

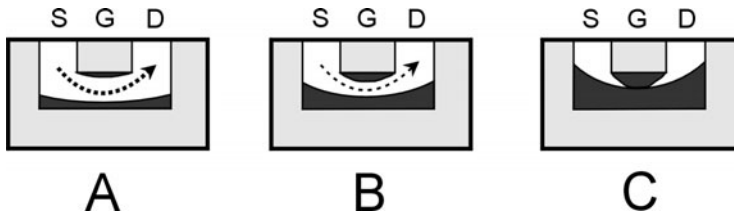
You can recognize the N-channel JFET in schematic diagrams by the arrow pointing inward at the gate, and the P-channel JFET by the arrow pointing outward. Also, you can tell which is which (sometimes arrows are not included in schematic diagrams) by the power-supply polarity. A positive drain indicates an N-channel JFET, and a negative drain indicates a P-channel JFET.

In electronic circuits, N-channel and P-channel devices can do the same kinds of things. The main difference is the polarity. An N-channel device can almost always be replaced with a P-channel JFET, and the power-supply polarity reversed, and the circuit will still work if the new device has the right specifications. Just as there are different kinds of bipolar transistors, there are various types of JFETs, each suited to a particular application. Some JFETs work well as weak-signal amplifiers and oscillators; others are made for power amplification.

Field effect transistors have certain advantages over bipolar transistors. Perhaps the most important is that FETs are available that generate less internal noise than bipolar transistors. This makes them excellent for use as weak-signal amplifiers at very high or ultrahigh frequencies. Field effect transistors have high input impedance, which can also be an advantage in weak-signal amplifiers.



**23-2** At A, pictorial diagram of a P-channel JFET. At B, the schematic symbol. Electrodes are S = source, G = gate, and D = drain.



**23-3** At A, the depletion region (darkest area) is narrow, the channel (white area) is wide, and many charge carriers (heavy dashed line) flow. At B, the depletion region is wider, the channel is narrower, and fewer charge carriers flow. At C, the depletion region obstructs the channel, and no charge carriers flow.

## Depletion and Pinchoff

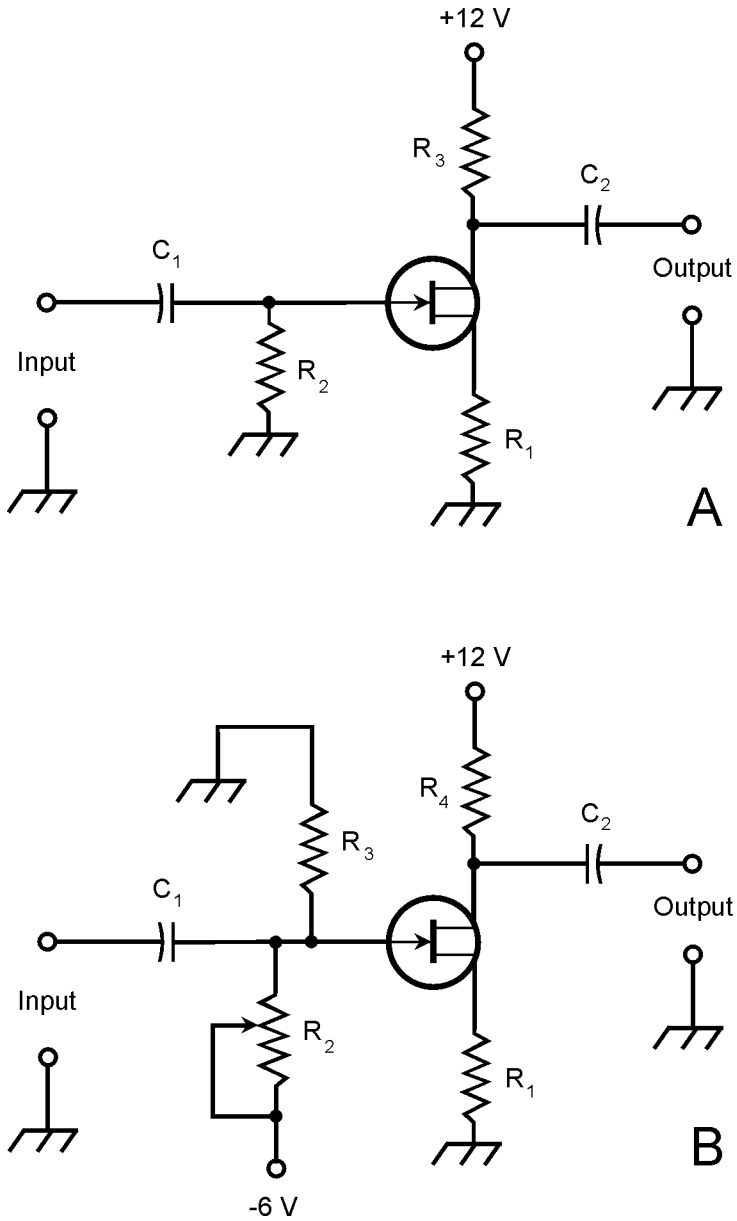
The JFET works because the voltage at the gate causes an electric field that interferes, more or less, with the flow of charge carriers along the channel. A simplified drawing of the situation for an N-channel device is shown in Fig. 23-3.

As the drain voltage  $E_D$  increases, so does the drain current  $I_D$ , up to a certain level-off value. This is true as long as the gate voltage  $E_G$  is constant, and is not too large negatively. But as  $E_G$  becomes increasingly negative (Fig. 23-3A), a *depletion region* (shown as a solid dark area) begins to form in the channel. Charge carriers cannot flow in this region; they must pass through a narrowed channel. The more negative  $E_G$  becomes, the wider the depletion region gets, as shown in drawing B. Ultimately, if the gate becomes negative enough, the depletion region completely obstructs the flow of charge carriers. This condition is called *pinchoff*, and is illustrated at C.

## JFET Biasing

Two biasing methods for N-channel JFET circuits are shown in Fig. 23-4. In Fig. 23-4A, the gate is grounded through resistor  $R_2$ . The source resistor,  $R_1$ , limits the current through the JFET. The drain current,  $I_D$ , flows through  $R_3$ , producing a voltage across this resistor. The ac output signal passes through  $C_2$ . In Fig. 23-4B, the gate is connected through potentiometer  $R_2$  to a voltage that is negative with respect to ground. Adjusting this potentiometer results in a variable negative  $E_G$  between  $R_2$  and  $R_3$ . Resistor  $R_1$  limits the current through the JFET. The drain current,  $I_D$ , flows through  $R_4$ , producing a voltage across it. The ac output signal passes through  $C_2$ .

In both of these circuits, the drain is positive relative to ground. For a P-channel JFET, reverse the polarities in Fig. 23-4. Typical power-supply voltages in JFET circuits are comparable to those for bipolar transistor circuits. The voltage between the source and drain, abbreviated  $E_D$ , can range from about 3 V to 150 V dc; most often it is 6 to 12 V dc. The biasing arrangement in Fig. 23-4A is preferred for weak-signal amplifiers, low-level amplifiers, and oscillators. The scheme at B is more often employed in power amplifiers having substantial input signal amplitudes.



**23-4** Two methods of biasing an N-channel JFET. At A, fixed gate bias; at B, variable gate bias.



## Amplification

The graph of Fig. 23-5 shows  $I_D$  as a function of  $E_G$  for a hypothetical N-channel JFET. The drain voltage,  $E_D$ , is assumed to be constant. When  $E_G$  is fairly large and negative, the JFET is pinched off, and no current flows through the channel. As  $E_G$  gets less negative, the channel opens up, and  $I_D$  begins flowing. As  $E_G$  gets still less negative, the channel gets wider and  $I_D$  increases. As  $E_G$  approaches the point where the S-G junction is at forward breakover, the channel conducts as well as it possibly can. If  $E_G$  becomes positive enough so the S-G junction conducts, some of the current in the channel leaks out through the gate. This is usually an unwanted phenomenon.

### The FET Amplifies Voltage

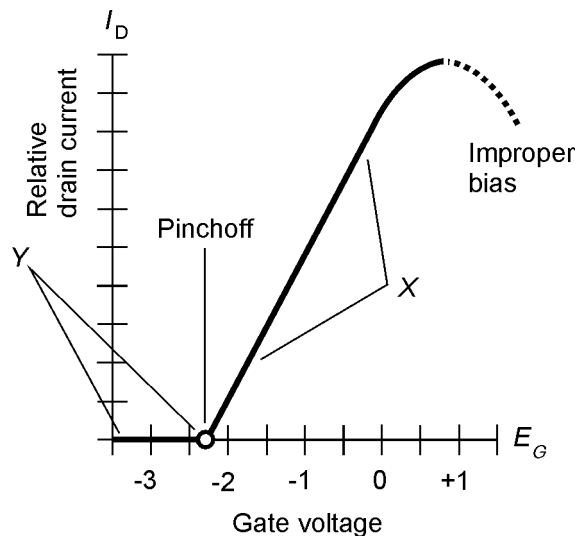
The best amplification for weak signals is obtained when  $E_G$  is such that the slope of the curve in Fig. 23-5 is the greatest. This is shown roughly by the range marked X. For power amplification, however, results are often best when the JFET is biased at or beyond pinchoff, in the range marked Y.

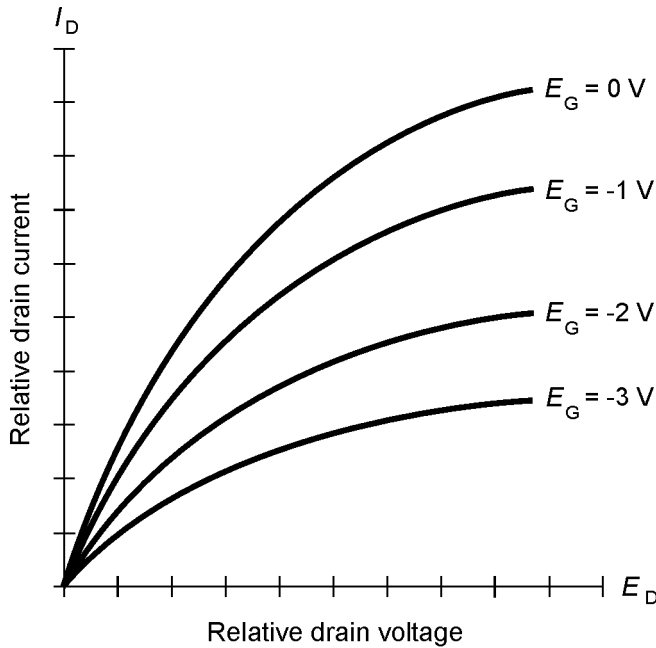
In either circuit shown in Fig. 23-4,  $I_D$  passes through the drain resistor. Small fluctuations in  $E_G$  cause large changes in  $I_D$ , and these variations in turn produce wide swings in the dc voltage across  $R_3$  (in the circuit at A) or  $R_4$  (in the circuit at B). The ac part of this voltage goes through capacitor  $C_2$ , and appears at the output as a signal of much greater ac voltage than that of the input signal at the gate.

### Drain Current versus Drain Voltage

Do you suspect that the current  $I_D$ , passing through the channel of a JFET, increases in a linear manner with increasing drain voltage  $E_D$ ? This seems reasonable, but it is not what usually happens. Instead,  $I_D$  rises for awhile as  $E_D$  increases steadily, and then  $I_D$  starts to level off. The current  $I_D$  can be plotted graphically as a function of  $E_D$  for various values of  $E_G$ . When this is done, the result is a *family of characteristic curves* for the JFET. The graph of Fig. 23-6 shows a family of characteristic

**23-5** Relative drain current ( $I_D$ ) as a function of gate voltage ( $E_G$ ) for a hypothetical N-channel JFET.





**23-6** A family of characteristic curves for a hypothetical N-channel JFET.

curves for a hypothetical N-channel device. The graph of  $I_D$  versus  $E_G$ , one example of which is shown in Fig. 23-5, is also an important specification that engineers consider when choosing a JFET for a particular application.

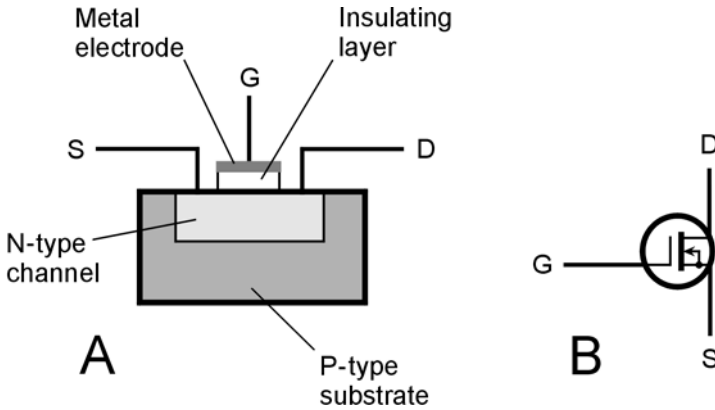
### Transconductance

Recall the discussion of *dynamic current amplification* from the last chapter. This is a measure of how well a bipolar transistor amplifies a signal. The JFET equivalent of this is called *dynamic mutual conductance* or *transconductance*.

Refer again to Fig. 23-5. Suppose that  $E_G$  is a certain value, resulting in a certain current  $I_D$ . If the gate voltage changes by a small amount  $dE_G$ , then the drain current will change by a certain increment  $dI_D$ . The transconductance is the ratio  $dI_D/dE_G$ . Geometrically, this translates to the slope of a line tangent to the curve of Fig. 23-5. The value of  $dI_D/dE_G$  is not the same at every point along the curve. When the JFET is biased beyond pinchoff, in the region marked *Y*, the slope of the curve is zero. Then there is no fluctuation in  $I_D$  when  $E_G$  changes by small amounts. There can be a change in  $I_D$  when there is a change in  $E_G$  only when the channel conducts current. The region where the transconductance,  $dI_D/dE_G$ , is the greatest is the region marked *X*, where the slope of the curve is steepest. This region of the curve represents conditions where the most gain can be obtained from the device.

## The MOSFET

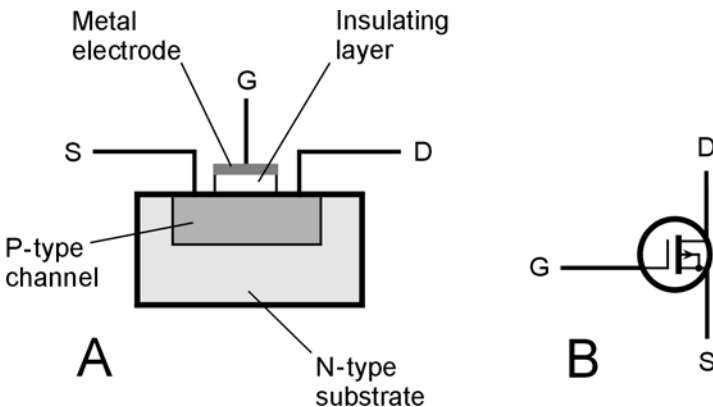
The acronym *MOSFET* (pronounced “*moss-fet*”) stands for *metal-oxide-semiconductor field effect transistor*. A simplified cross-sectional drawing of an N-channel MOSFET, along with the schematic symbol, is shown in Fig. 23-7. The P-channel device is shown in the drawings of Fig. 23-8.



23-7 At A, pictorial diagram of an N-channel MOSFET. At B, the schematic symbol. Electrodes are S = source, G = gate, and D = drain.

### Extremely High Input Impedance

When the MOSFET was first developed, it was called an *insulated-gate FET* or *IGFET*. That's still a good name for it. The gate electrode is insulated from the channel by a thin layer of dielectric material. As a result, the input impedance is even higher than that of a JFET. The gate-to-source resistance of a typical MOSFET is, as a matter of fact, comparable to that of a typical capacitor! This means that a MOSFET draws essentially no current, and therefore no power, from the signal source. This makes the device ideal for low-level and weak-signal amplifiers. But MOS devices aren't quite perfect. They have an Achilles heel. They are electrically fragile.



23-8 At A, pictorial diagram of a P-channel MOSFET. At B, the schematic symbol. Electrodes are S = source, G = gate, and D = drain.

## Beware of Static!

The trouble with MOSFETs is that they can be easily damaged by electrostatic discharges. When building or servicing circuits containing MOS devices, technicians must use special equipment to ensure that their hands don't carry electrostatic charges that might ruin the components. If a discharge occurs through the dielectric of a MOS device, the component is permanently destroyed. Warm and humid climates do not offer total protection against the hazard. This author learned that fact by ruining several MOSFETs while designing circuits in the summertime—in Miami, Florida!

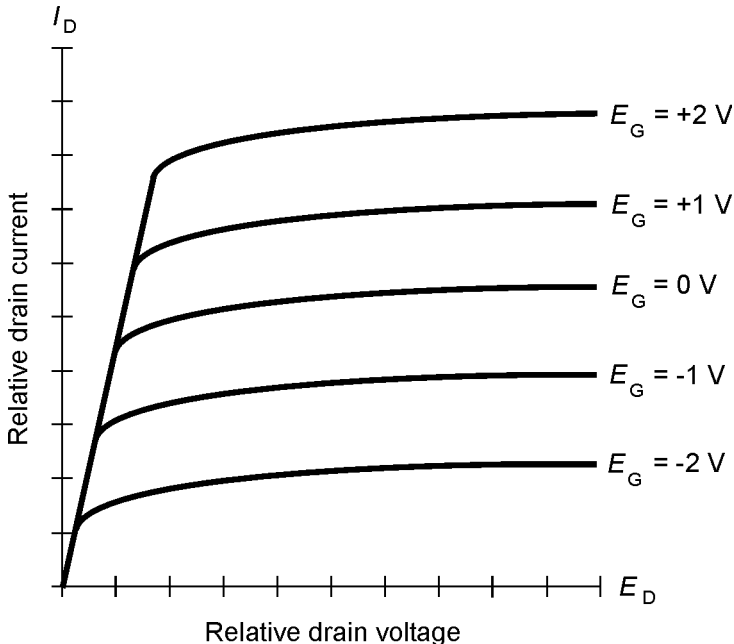
## Flexibility in Biasing

In electronic circuits, an N-channel JFET can sometimes be replaced directly with an N-channel MOSFET, and P-channel devices can be similarly interchanged. But the characteristic curves for MOSFETs are not the same as those for JFETs. The main difference is that the S-G junction in a MOSFET is not a P-N junction. Therefore, forward breakover cannot occur. A gate bias voltage,  $E_G$ , more positive than +0.6 V can be applied to an N-channel MOSFET, or an  $E_G$  more negative than -0.6 V to a P-channel device, without a current leak taking place.

A family of characteristic curves for a hypothetical N-channel MOSFET is shown in the graph of Fig. 23-9.

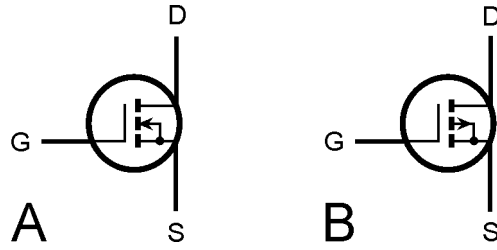
## Depletion Mode versus Enhancement Mode

Normally the channel in a JFET is wide open; as the depletion region gets wider and wider, choking off the channel, the charge carriers are forced to pass through a narrower and narrower path. This is known as the *depletion mode* of operation for a field effect transistor. A MOSFET can also



23-9 A family of characteristic curves for a hypothetical N-channel MOSFET.

- 23-10** Schematic symbols for enhancement mode MOSFETs. At A, the N-channel device; at B, the P-channel device.

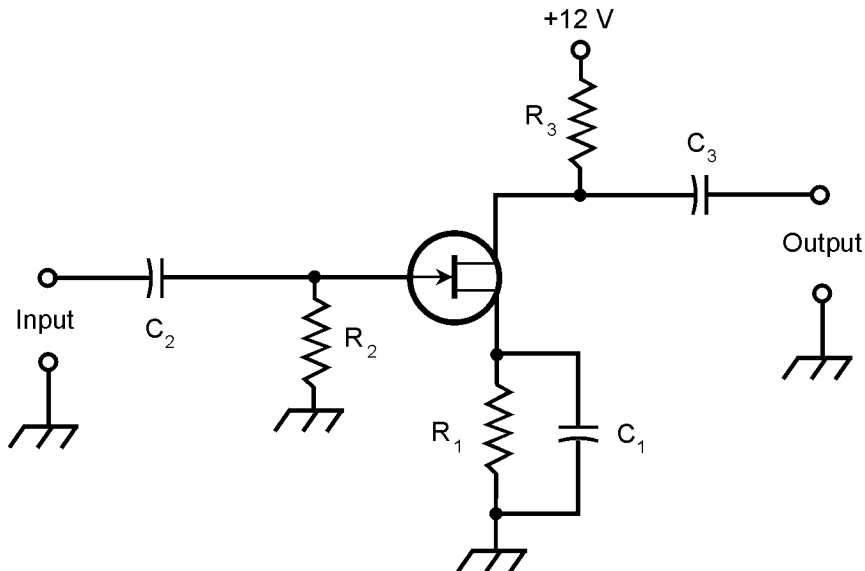


be made to work in the depletion mode. The drawings and schematic symbols of Figs. 23-7 and 23-8 show depletion-mode MOSFETs.

Metal-oxide semiconductor technology also allows an entirely different means of operation. An *enhancement-mode* MOSFET normally has a pinched-off channel. It is necessary to apply a bias voltage,  $E_G$ , to the gate so that a channel will form. If  $E_G = 0$  in an enhancement-mode MOSFET, then  $I_D = 0$  when there is no signal input. The schematic symbols for N-channel and P-channel enhancement-mode devices are shown in Fig. 23-10. Note that the vertical line is broken. This is how you can recognize an enhancement-mode device in circuit diagrams.

## Common Source Circuit

There are three different circuit hookups for FETs, just as there are for bipolar transistors. These three arrangements have the source, the gate, or the drain at signal ground. In the *common source circuit*, the source is placed at signal ground. Signal input is applied to the gate. The general configuration is shown in Fig. 23-11. An N-channel JFET is used here, but the device could be an



- 23-11** Common source configuration. This diagram shows an N-channel JFET circuit.

N-channel, depletion-mode MOSFET and the circuit diagram would be the same. For an N-channel enhancement-mode device, an extra resistor would be necessary, running from the gate to the positive power supply terminal. For P-channel devices, the supply would provide a negative, rather than a positive, voltage.

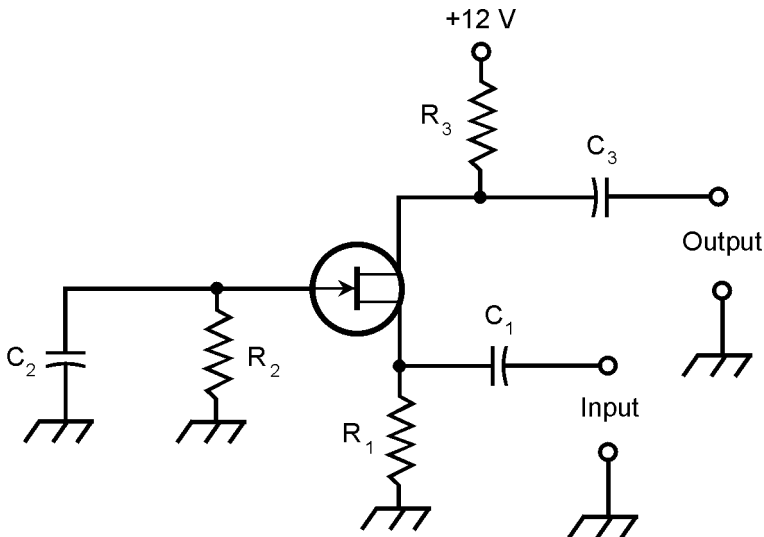
Capacitor  $C_1$  and resistor  $R_1$  place the source at signal ground while elevating the source above ground for dc. The ac signal enters through  $C_2$ . Resistor  $R_2$  adjusts the input impedance and provides bias for the gate. The ac signal passes out of the circuit through  $C_3$ . Resistor  $R_3$  keeps the output signal from being shorted out through the power supply. The circuit of Fig. 23-11 is the basis for low-level RF amplifiers and oscillators.

The common source arrangement provides the greatest gain of the three FET circuit configurations. The output is  $180^\circ$  out of phase with the input.

## Common Gate Circuit

In the *common gate circuit* (Fig. 23-12), the gate is placed at signal ground. The input is applied to the source. The illustration shows an N-channel JFET. For other types of FETs, the same considerations apply as previously described for the common source circuit. Enhancement-mode devices would require a resistor between the gate and the positive supply terminal (or the negative terminal if the MOSFET is P-channel).

The dc bias for the common gate circuit is basically the same as that for the common source arrangement. But the signal follows a different path. The ac input signal enters through  $C_1$ . Resistor  $R_1$  keeps the input from being shorted to ground. Gate bias is provided by  $R_1$  and  $R_2$ . Capacitor



**23-12** Common gate configuration. This diagram shows an N-channel JFET circuit.

$C_2$  places the gate at signal ground. In some common gate circuits, the gate is directly grounded, and  $R_2$  and  $C_2$  are not necessary. The output signal leaves the circuit through  $C_3$ . Resistor  $R_3$  keeps the output signal from being shorted through the power supply.

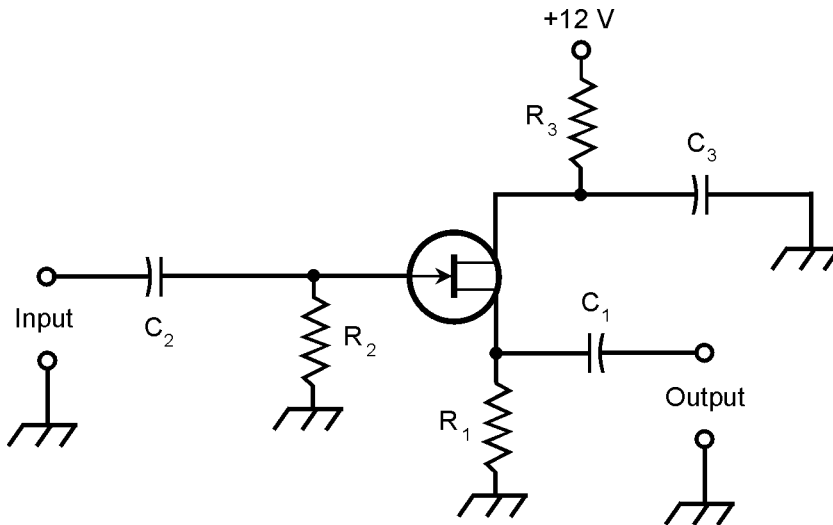
The common gate arrangement produces less gain than its common source counterpart. But a common gate amplifier is not likely to break into unwanted oscillation, making it a good choice for power-amplifier circuits, especially at RF. The output is in phase with the input.

## Common Drain Circuit

A *common drain circuit* is shown in Fig. 23-13. In this circuit, the collector is at signal ground. It is sometimes called a *source follower*. The FET is biased in the same way as for the common source and common gate circuits. In the illustration, an N-channel JFET is shown, but any other kind of FET could be used, reversing the polarity for P-channel devices. Enhancement-mode MOSFETs would need a resistor between the gate and the positive supply terminal (or the negative terminal if the MOSFET is P-channel).

The input signal passes through  $C_2$  to the gate. Resistors  $R_1$  and  $R_2$  provide gate bias. Resistor  $R_3$  limits the current. Capacitor  $C_3$  keeps the drain at signal ground. Fluctuating dc (the channel current) flows through  $R_1$  as a result of the input signal; this causes a fluctuating dc voltage to appear across  $R_1$ . The output is taken from the source, and its ac component passes through  $C_1$ .

The output of the common drain circuit is in phase with the input. This scheme is the FET analog of the bipolar common collector arrangement. The output impedance is rather low, making this circuit a good choice for broadband impedance matching.



**23-13** Common drain configuration, also known as a source follower. This diagram shows an N-channel JFET circuit.

## Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. The current through the channel of a JFET is directly affected by all of the following, except the
  - (a) drain voltage.
  - (b) transconductance.
  - (c) gate voltage.
  - (d) gate bias.
2. In an N-channel JFET, pinchoff occurs when the gate bias is
  - (a) small and positive.
  - (b) zero.
  - (c) small and negative.
  - (d) large and negative.
3. The current consists mainly of holes when a JFET
  - (a) has a P-type channel.
  - (b) is forward-biased.
  - (c) is zero-biased.
  - (d) is reverse-biased.
4. A JFET might work better than a bipolar transistor in
  - (a) a high-voltage rectifier.
  - (b) a weak-signal RF amplifier.
  - (c) a power-supply filter.
  - (d) a power transformer.
5. In a P-channel JFET,
  - (a) the drain is forward-biased.
  - (b) the source-gate junction is forward-biased.
  - (c) the drain is negative relative to the source.
  - (d) the gate must be at dc ground.
6. A JFET is sometimes biased at or beyond pinchoff in
  - (a) a power amplifier.
  - (b) a rectifier.
  - (c) a filter.
  - (d) a weak-signal amplifier.
7. The gate of a JFET exhibits a
  - (a) forward bias.
  - (b) high impedance.



- (c) low reverse resistance.
  - (d) low avalanche voltage.
8. Which of the following conditions is not normally desirable in a JFET?
- (a) A conducting channel
  - (b) Holes as the majority carriers
  - (c) A forward-biased P-N junction
  - (d) A high input impedance
9. When a JFET is pinched off,
- (a) the value of  $dI_D/dE_G$  is very large with no signal input.
  - (b) the value of  $dI_D/dE_G$  might vary considerably with no signal input.
  - (c) the value of  $dI_D/dE_G$  is negative with no signal input.
  - (d) the value of  $dI_D/dE_G$  is zero with no signal input.
10. Transconductance is the ratio of
- (a) a change in drain voltage to a change in source voltage.
  - (b) a change in drain current to a change in gate voltage.
  - (c) a change in gate current to a change in source voltage.
  - (d) a change in drain current to a change in drain voltage.
11. Characteristic curves for JFETs generally show
- (a) drain voltage as a function of source current.
  - (b) drain current as a function of gate current.
  - (c) drain current as a function of drain voltage.
  - (d) drain voltage as a function of gate current.
12. A disadvantage of MOS components is the fact that
- (a) they can be easily damaged by static electricity.
  - (b) they need a high input voltage in order to amplify.
  - (c) they draw large amounts of current.
  - (d) they produce a great deal of heat.
13. The input impedance of a MOSFET is
- (a) lower than that of a JFET.
  - (b) lower than that of a bipolar transistor.
  - (c) between that of a bipolar transistor and a JFET.
  - (d) extremely high.
14. A significant difference between MOSFETs and JFETs is the fact that
- (a) MOSFETs can handle a wider range of gate bias voltages.
  - (b) MOSFETs can deliver greater output power.
  - (c) MOSFETs are more rugged.
  - (d) MOSFETs last longer.

15. The channel in a zero-biased JFET is normally
  - (a) pinched off.
  - (b) in a state of avalanche breakdown.
  - (c) in a conducting state.
  - (d) made of P-type semiconductor material.
16. When an enhancement-mode MOSFET is at zero bias,
  - (a) the drain current is high with no signal.
  - (b) the drain current fluctuates with no signal.
  - (c) the drain current is low with no signal.
  - (d) the drain current is zero with no signal.
17. An enhancement-mode MOSFET can be recognized in schematic diagrams by the presence of
  - (a) an arrow pointing inward.
  - (b) a broken vertical line inside the circle.
  - (c) an arrow pointing outward.
  - (d) a solid vertical line inside the circle.
18. In a source follower, which of the electrodes receives the input signal?
  - (a) Any of them (doesn't matter)
  - (b) The source
  - (c) The gate
  - (d) The drain
19. Which of the following circuits produces an output signal wave that is  $180^\circ$  out of phase with the input signal wave?
  - (a) The common source circuit
  - (b) The common gate circuit
  - (c) The common drain circuit
  - (d) All of the above
20. Which of the following circuits can produce the greatest signal gain (amplification factor)?
  - (a) The common source circuit
  - (b) The common gate circuit
  - (c) The common drain circuit
  - (d) All of the above circuits can amplify to the same extent.

## 24 CHAPTER

# Amplifiers and Oscillators

NOW THAT YOU'VE LEARNED THE BASICS OF BIPOLAR AND FIELD EFFECT TRANSISTORS, YOU'RE READY to learn about amplifiers and oscillators that use these devices. As a prelude to that, let's define the most common unit that engineers use to evaluate changes in signal strength.

## The Decibel

The human senses of hearing and vision don't perceive the intensity of sound and light in a linear manner. Instead, our perception of these phenomena is *logarithmic*. Electronic circuits behave in a similar way when subjected to signals that change in amplitude. Because of this, scientists and engineers invented a unit called the *decibel* (symbolized dB), in which amplitude changes are expressed according to the *base-10 logarithm* of the actual change.

### More or Less?

In the decibel scheme, increases in amplitude are assigned *positive gain* values, and decreases in amplitude are assigned *negative gain* values. A negative gain figure is sometimes expressed as a *loss* by removing the minus sign.

If the output signal amplitude from a circuit is +6 dB relative to the input signal amplitude, then the output is stronger than the input. If the output is at -14 dB relative to the input, then the output is weaker than the input. In the first case, the circuit has a gain of +6 dB. In the second case, we can say that the circuit has a gain of -14 dB or a loss of 14 dB.

How large a unit is the decibel, anyway? The answer is that it has no absolute size; it is a relative unit. The decibel has meaning only when two or more signals are compared. An amplitude increase or decrease of plus or minus one decibel ( $\pm 1$  dB) is roughly the smallest change a listener can detect if the change is expected. If the change is not expected, the smallest difference a listener can notice is about  $\pm 3$  dB.

**For Voltage**

Consider a circuit with an rms ac input voltage of  $E_{in}$  and an rms ac output voltage of  $E_{out}$ , specified in the same units as  $E_{in}$ . Then the *voltage gain* of the circuit, in decibels, is given by this formula:

$$\text{Gain (dB)} = 20 \log(E_{out}/E_{in})$$

The base-10 logarithm of a value  $x$  is written  $\log x$ . Log functions of base 10 and base  $e$  (another type of logarithm sometimes used in physics and engineering) can be determined easily using scientific calculators. From now on, when we say “logarithm,” we mean the base-10 logarithm unless otherwise specified.

**Problem 24-1**

Suppose a circuit has an rms ac input of 1.00 V and an rms ac output of 14.0 V. What is the gain in decibels?

First, find the ratio  $E_{out}/E_{in}$ . Because  $E_{out} = 14.0$  V rms and  $E_{in} = 1.00$  V rms, the ratio is  $14.0/1.00$ , or 14.0. Next, find the logarithm of 14.0. This is about 1.146128036. Finally, multiply this number by 20, getting something like 22.92256071. Round this off to three significant figures, because that’s all you’re entitled to, getting 22.9 dB.

**Problem 24-2**

Suppose a circuit has an rms ac input voltage of 24.2 V and an rms ac output voltage of 19.9 V. What is the gain in decibels?

Find the ratio  $E_{out}/E_{in} = 19.9/24.2 = 0.822 \dots$ . (The three dots indicate extra digits introduced by the calculator. You can leave them in until the final roundoff.) Find the logarithm of this:  $\log 0.822 \dots = -0.0849 \dots$ . The gain is 20 times this, or  $-1.699 \dots$  dB, which rounds off to  $-1.70$  dB.

**For Current**

Current gain or loss is calculated in the same way as voltage gain or loss. If  $I_{in}$  is the rms ac input current and  $I_{out}$  is the rms ac output current specified in the same units as  $I_{in}$ , then:

$$\text{Gain (dB)} = 20 \log(I_{out}/I_{in})$$

**For Power**

The *power gain* of a circuit, in decibels, is calculated according to a slightly different formula. If  $P_{in}$  is the input signal power and  $P_{out}$  is the output signal power expressed in the same units as  $P_{in}$ , then:

$$\text{Gain (dB)} = 10 \log (P_{out}/P_{in})$$

The *coefficient* (that is, the factor by which the logarithm is to be multiplied) in the formula for power gain is 10, whereas for voltage and current gain it is 20.

**Problem 24-3**

Suppose a power amplifier has an input of 5.03 W and an output of 125 W. What is the gain in decibels?

Find the ratio  $P_{\text{out}}/P_{\text{in}} = 125/5.03 = 24.85 \dots$ . Then find the logarithm:  $\log 24.85 \dots = 1.395 \dots$ . Finally, multiply by 10 and round off. The gain is thus  $10 \times 1.395 \dots = 14.0$  dB.

### Problem 24-4

Suppose an *attenuator* (a circuit designed deliberately to produce power loss) provides 10 dB power reduction. The input power is 94 W. What is the output power?

An attenuation of 10 dB represents a gain of  $-10$  dB. We know that  $P_{\text{in}} = 94$  W, so the unknown in the power gain formula is  $P_{\text{out}}$ . We must solve for  $P_{\text{out}}$  in this formula:

$$-10 = 10 \log (P_{\text{out}}/94)$$

First, divide each side by 10, getting:

$$-1 = \log (P_{\text{out}}/94)$$

To solve this, we must take the *base-10 antilogarithm*, also known as the *antilog*, or the *inverse log*, of each side. The antilog function undoes the log function. The antilog of a value  $x$  is written antilog  $x$ . It can also be denoted as  $\log^{-1}x$  or  $10^x$ . Antilogarithms can be determined with any good scientific calculator. The solution process goes like this:

$$\begin{aligned} \text{antilog}(-1) &= \text{antilog}[\log (P_{\text{out}}/94)] \\ 0.1 &= P_{\text{out}}/94 \\ 94 \times 0.1 &= P_{\text{out}} \\ P_{\text{out}} &= 9.4 \text{ W} \end{aligned}$$

### Decibels and Impedance

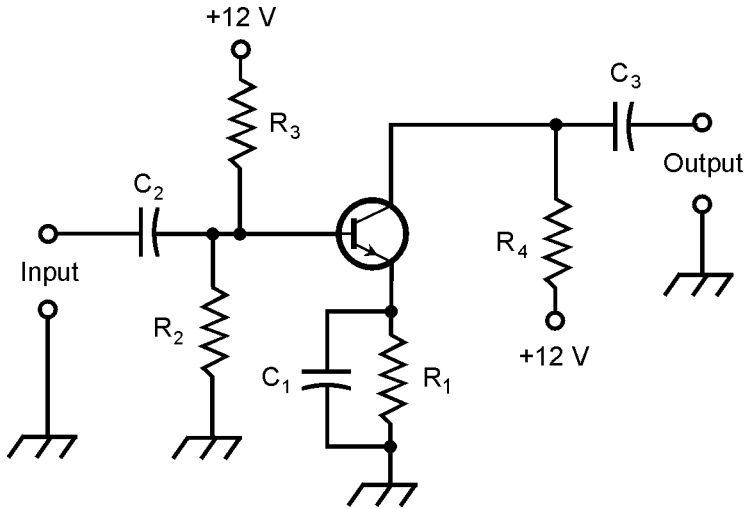
When determining the voltage gain (or loss) and the current gain (or loss) for a circuit in decibels, you should expect to get the same figure for both parameters only when the input impedance is identical to the output impedance. If the input and output impedances differ, the voltage gain or loss is generally not the same as the current gain or loss.

Consider how transformers work. A step-up transformer, in theory, has voltage gain, but this alone doesn't make a signal more powerful. A step-down transformer can exhibit theoretical current gain, but again, this alone does not make a signal more powerful. In order to make a signal more *powerful*, a circuit must increase the signal *power*—the *product* of the voltage and the current!

When determining power gain (or loss) for a particular circuit in decibels, the input and output impedances don't matter. In this sense, positive power gain always represents a real-world increase in signal strength. Similarly, negative power gain (or power loss) always represents a true decrease in signal strength.

## Basic Bipolar Transistor Amplifier

In the previous chapters, you saw some circuits that use bipolar and field effect transistors. A signal can be applied to some control point, causing a much greater signal to appear at the output. This is the principle by which all amplifiers work.



**24-1** An amplifier circuit with a bipolar transistor. Component designators are discussed in the text.

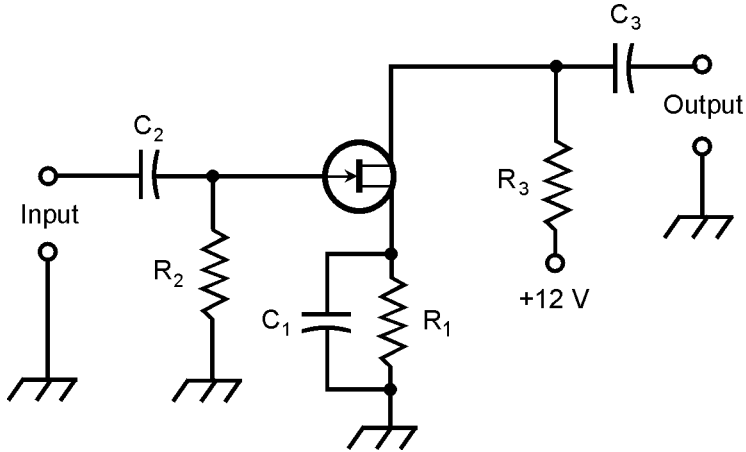
In Fig. 24-1, an NPN bipolar transistor is connected as a *common emitter amplifier*. The input signal passes through  $C_2$  to the base. Resistors  $R_2$  and  $R_3$  provide the base bias. Resistor  $R_1$  and capacitor  $C_1$  allow for the emitter to have a dc voltage relative to ground, while keeping it grounded for signals. Resistor  $R_1$  also limits the current through the transistor. The ac output signal goes through capacitor  $C_3$ . Resistor  $R_4$  keeps the ac output signal from being short-circuited through the power supply.

In this amplifier, the optimum capacitance values depend on the design frequency of the amplifier, and also on the impedances at the input and output. In general, as the frequency and/or circuit impedance increase, less capacitance is needed. At audio frequencies and low impedances, the capacitors might be as large as 100  $\mu\text{F}$ . At radio frequencies and high impedances, values will be only a fraction of a microfarad, down to picofarads at the highest frequencies and impedances. The optimum resistor values also depend on the application. In the case of a weak-signal amplifier, typical values are 470  $\Omega$  for  $R_1$ , 4.7 k $\Omega$  for  $R_2$ , 10 k $\Omega$  for  $R_3$ , and 4.7 k $\Omega$  for  $R_4$ .

## Basic JFET Amplifier

Figure 24-2 shows an N-channel JFET hooked up as a *common source amplifier*. The input signal passes through  $C_2$  to the gate. Resistor  $R_2$  provides the bias. Resistor  $R_1$  and capacitor  $C_1$  give the source a dc voltage relative to ground, while grounding it for signals. The output signal goes through  $C_3$ . Resistor  $R_3$  keeps the output signal from being short-circuited through the power supply.

A JFET has a high input impedance, and therefore the value of  $C_2$  should usually be small. If the device is a MOSFET, the input impedance is higher still, and  $C_2$  will be smaller yet, sometimes 1 pF or less. The resistor values depend on the application. In some instances,  $R_1$  and  $C_1$  are not used, and the source is grounded directly. If  $R_1$  is used, its optimum value will depend on the input



24-2 An amplifier circuit with an FET. Component designators are discussed in the text.

impedance and the bias needed for the FET. For a weak-signal amplifier, typical values are  $680\ \Omega$  for  $R_1$ ,  $10\ \text{k}\Omega$  for  $R_2$ , and  $100\ \Omega$  for  $R_3$ .

## Amplifier Classes

Amplifier circuits can be categorized as *class A*, *class AB*, *class B*, and *class C*. Each class has its own special characteristics, and works best in its own unique set of applications.

### The Class A Amplifier

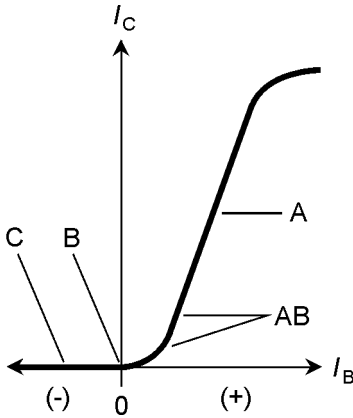
With the previously mentioned component values, the amplifier circuits in Figs. 24-1 and 24-2 operate in class A. This type of amplifier is *linear*, meaning that the output waveform has the same shape as (although a much greater amplitude than) the input waveform.

For class A operation with a bipolar transistor, the bias must be such that, with no signal input, the device is near the middle of the straight-line portion of the  $I_C$  versus  $I_B$  (collector current versus base current) curve. This is shown for an NPN transistor in Fig. 24-3. For PNP, reverse the polarity signs. With a JFET or MOSFET, the bias must be such that, with no signal input, the device is near the middle of the straight-line part of the  $I_D$  versus  $E_G$  (drain current versus gate voltage) curve. This is shown in Fig. 24-4 for an N-channel JFET. For P channel, reverse the polarity signs.

In a class A amplifier, it is important that the input signal not be too strong. An excessively strong input signal will drive the device out of the straight-line part of the characteristic curve during part of the cycle. When this occurs, the output waveshape will not be a faithful reproduction of the input waveshape, and the amplifier will become *nonlinear*. Class A amplifiers are supposed to operate in a linear fashion at all times.

### The Class AB Amplifier

Class A operation is inefficient because the transistor draws current whether there is a signal input or not. For weak-signal work, efficiency is not too critical; the things that matter are the gain and



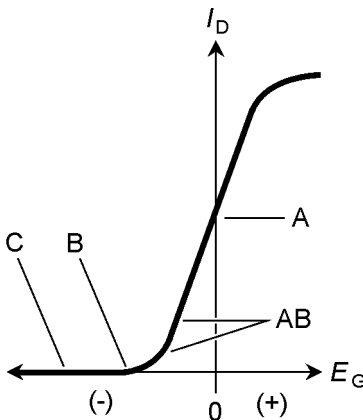
24-3 Various classes of amplifier operation for an NPN bipolar transistor.

the sensitivity. In power amplifiers, efficiency is more important, and gain and sensitivity don't matter as much.

When a bipolar transistor is biased close to cutoff under no-signal conditions, or when an FET is near pinchoff, the input signal always drives the device into the nonlinear part of the operating curve. Typical bias zones for class AB are shown in Figs. 24-3 and 24-4. A small collector or drain current flows when there is no input, but it is less than the no-signal current that flows in a class A amplifier. This is called *class AB operation*. It's more efficient than class A, but the gain and sensitivity are not as high.

There are two modes of class AB amplification. If the bipolar transistor or FET is never driven into cutoff/pinchoff during any part of the signal cycle, the amplifier is working in *class AB<sub>1</sub>*. If the device goes into cutoff/pinchoff for any part of the cycle (up to almost half), the amplifier is working in *class AB<sub>2</sub>*.

In a class AB amplifier, the output signal waveform is not identical with the input signal waveform. But if the signal wave is *modulated*, such as in a voice radio transmitter, the waveform of the *modulating signal* comes out undistorted anyway. Thus, class AB operation is useful in RF power amplifiers.



24-4 Various classes of amplifier operation for an N-channel JFET.



## The Class B Amplifier

When a bipolar transistor is biased exactly at cutoff, or an FET is biased exactly at pinchoff under zero-input-signal conditions, an amplifier is working in *class B*. These operating points are labeled on the curves in Figs. 24-3 and 24-4. The class B scheme lends itself well to RF power amplification.

In class B operation, there is no collector or drain current when there is no signal. This saves energy, because the circuit does not consume power unless there is a signal going into it. (Class A and class AB amplifiers consume some power even when the input is zero.) When there is an input signal, current flows in the device during exactly half of the cycle. The output signal waveform is greatly different from the input waveshape in a class B amplifier. In fact, it is half-wave rectified.

You'll sometimes hear of class AB or class B "linear amplifiers," especially in ham radio. The term "linear" refers to the fact that the *modulation waveform* is not distorted by such an amplifier, even though the *carrier waveform* is distorted because the transistor is not biased in the straight-line part of the operating curve.

Class AB<sub>2</sub> and class B amplifiers draw power from the input signal source. Engineers say that such amplifiers require a certain amount of *drive* or *driving power* to function. Class A and class AB<sub>1</sub> amplifiers theoretically need no driving power, although there must be an input voltage.

## The Class B Push-Pull Amplifier

Sometimes two bipolar transistors or FETs are used in a class B circuit, one for the positive half of the cycle and the other for the negative half. In this way, signal waveform distortion is eliminated. This is called a *class B push-pull amplifier*. This type of circuit, using two NPN bipolar transistors, is illustrated in Fig. 24-5. Resistor R<sub>1</sub> limits the current through the transistors. Capacitor C<sub>1</sub> keeps the input transformer center tap at signal ground, while allowing for some dc base bias. Resistors R<sub>2</sub> and R<sub>3</sub> bias the transistors precisely at their cutoff points. The two transistors must be identical. Not only should their part numbers be the same, but ideally they should be chosen *by experiment* to ensure that their characteristic curves are as closely matched as possible.

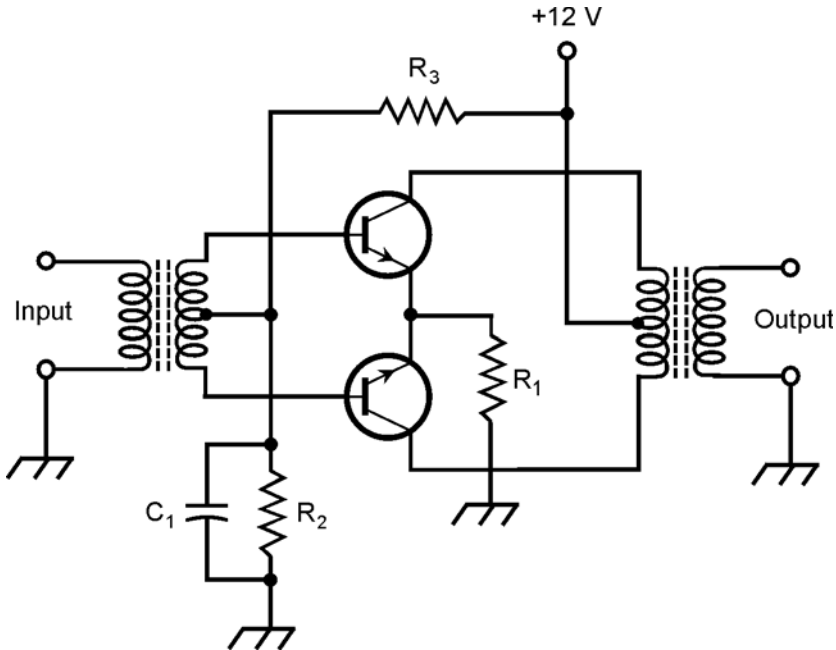
Class B push-pull is a popular arrangement for audio-frequency (AF) power amplification. It combines the efficiency of class B with the low distortion of class A. Its main disadvantage is that it needs two center-tapped transformers, one at the input and the other at the output. This makes push-pull amplifiers rather bulky and expensive compared to other types.

## The Class C Amplifier

A bipolar transistor or FET can be biased past cutoff or pinchoff, and it will still work as a power amplifier (PA), provided that the drive is sufficient to overcome the bias during part of the cycle. This is known as *class C* operation. Bias points for class C are labeled in Figs. 24-3 and 24-4.

Class C amplifiers are nonlinear, even for amplitude modulation waveforms. Because of this, a class C circuit is useful only for signals that are either full-on or full-off. Such signals include old-fashioned Morse code, and digital schemes in which the frequency or phase (but not the amplitude) of the signal is varied.

A class C amplifier needs a lot of driving power. The gain is low. For example, it might take 300 W of signal drive to get 1 kW of signal power output. However, the efficiency is better than that of class A, AB, or B amplifiers. Let's take a closer look, now, at what *amplifier efficiency* is all about.



**24-5** A class B push-pull amplifier using NPN bipolar transistors. Component designators are discussed in the text.

## Efficiency in Power Amplifiers

In power amplification, efficiency is important. It not only provides optimum output power with minimum heat generation and minimum strain on the transistors, but it conserves energy as well. This translates into reduced cost, reduced size and weight, and longer equipment life compared with inefficient power amplifiers.

### Determining dc Power Input

Suppose you connect an ammeter or milliammeter in series with the collector or drain of an amplifier and the power supply. While the amplifier is in operation, this meter will have a certain reading. The reading might appear constant, or it might fluctuate with changes in the input signal level. The *dc collector power input* to a bipolar-transistor amplifier circuit is the product of the collector current ( $I_C$ ) and the collector voltage ( $E_C$ ). Similarly, for an FET, the *dc drain power input* is the product of the drain current ( $I_D$ ) and the drain voltage ( $E_D$ ). These power figures can be further categorized as *average* or *peak* values. This discussion involves only average power.

The dc collector or drain power input can be high even when there is no input signal applied to an amplifier. A class A circuit operates this way. In fact, when a signal is applied to a class A amplifier, the meter reading, and therefore the dc collector or drain power input, does not change compared to the value under conditions of no input signal. In class AB<sub>1</sub> or class AB<sub>2</sub>, there is low current (and therefore low dc collector or drain power input) with zero input signal, and a higher current

(and therefore a higher dc power input) with an input signal. In class B and class C, there is no current (and therefore zero dc collector or drain power input) when there is no input signal. The current, and therefore the dc power input, increases with increasing signal input. The dc collector or drain power input is usually measured in watts, the product of amperes and volts. It can be indicated in milliwatts for low-power amplifiers, or kilowatts for high-power amplifiers.

## Power Output

The *power output* of an amplifier must be measured by means of a specialized ac wattmeter. The design of AF and RF wattmeters is a sophisticated specialty in engineering.

When there is no signal input to an amplifier, there is no signal output, and therefore the power output is zero. This is true no matter what the class of amplification. The greater the signal input, in general, the greater the power output of a power amplifier, up to a certain point.

Power output, like dc input, is measured in watts. For very low power circuits, it can be in milliwatts; for high-power circuits, it is sometimes given in kilowatts.

## Definition of Efficiency

The *efficiency* of a power amplifier is the ratio of the ac power output to the dc collector or drain power input.

In a bipolar-transistor amplifier, let  $P_C$  be the dc collector power input, and let  $P_{out}$  be the ac power output. For an FET amplifier, let  $P_D$  be the dc drain power input, and let  $P_{out}$  be the ac power output. Then the efficiency,  $eff$ , of the bipolar transistor amplifier is given by:

$$eff = P_{out}/P_C$$

For the FET circuit, the efficiency is:

$$eff = P_{out}/P_D$$

These are ratios, and they are always between 0 and 1. Efficiency is often expressed as a percentage instead of a ratio, so the preceding formulas are modified as follows:

$$\begin{aligned} eff_{\%} &= 100 P_{out}/P_C \\ &\text{and} \\ eff_{\%} &= 100 P_{out}/P_D \end{aligned}$$

### Problem 24-5

Suppose a bipolar-transistor amplifier has a dc collector input of 115 W and an ac power output of 65.0 W. What is the efficiency in percent?

Use the formula for the efficiency of a bipolar transistor amplifier expressed as a percentage:  
 $eff_{\%} = 100 P_{out}/P_C = 100 \times 65/115 = 100 \times 0.565 = 56.5\%$ .

### Problem 24-6

Suppose an FET amplifier is 60 percent efficient. If the power output is 3.5 W, what is the dc drain power input?

Plug in values to the formula for the efficiency of an FET amplifier expressed as a percentage. The resulting equation is solved as follows:

$$\begin{aligned} 60 &= 100 \times 3.5/P_D \\ 60 &= 350/P_D \\ 60/350 &= 1/P_D \\ P_D &= 350/60 \\ &= 5.8 \text{ W} \end{aligned}$$

### Efficiency versus Class

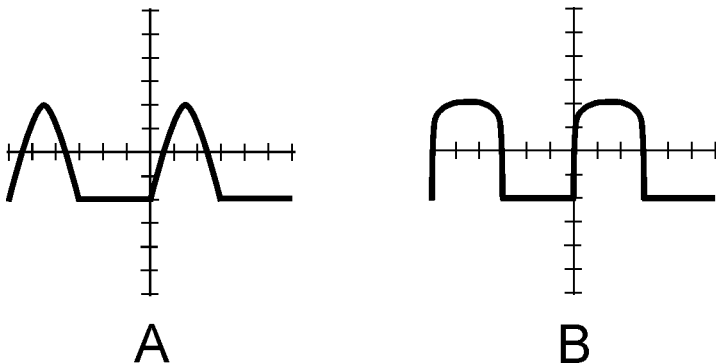
Class A amplifiers have efficiency figures from 25 percent to 40 percent, depending on the nature of the input signal and the type of transistor used. A good class AB<sub>1</sub> amplifier is 35 percent to 45 percent efficient. A class AB<sub>2</sub> amplifier, if well designed and properly operated, can be up to about 50 percent efficient. Class B amplifiers are typically 50 percent to 65 percent efficient. Class C amplifiers can have efficiency levels as high as 75 percent.

## Drive and Overdrive

Class A and AB<sub>1</sub> power amplifiers do not, in theory, take any power from the signal source to produce significant output power. This is one of the advantages of these classes of operation. It is only necessary that a certain voltage be present at the control electrode (the base, gate, emitter, or source) for these circuits to produce useful output signal power. Class AB<sub>2</sub> amplifiers need some driving power to produce ac power output. Class B amplifiers require more drive than class AB<sub>2</sub>, and class C amplifiers need still more drive.

Whatever kind of PA is used in a given situation, it is important that the driving signal not be too strong. If *overdrive* takes place, distortion occurs in the output signal. An oscilloscope can be used to determine whether or not an amplifier is being overdriven. The scope is connected to the amplifier output terminals, and the waveform of the output signal is examined. The output waveform for a particular class of amplifier always has a characteristic shape. Overdrive is indicated by a form of distortion known as *flat topping*.

In Fig. 24-6A, the output signal waveshape for a properly operating class B amplifier is shown. In Fig. 24-6B, the output of an overdriven class B amplifier is shown. Note that the peaks are blunted or truncated. The result of this can be distortion in the modulation on a radio signal, and



**24-6** At A, an oscilloscope display of the signal output waveform from a properly operating class B power amplifier. At B, a display showing distortion in the waveform caused by overdrive.

also an excessive amount of signal output at harmonic frequencies. The efficiency of the circuit can be degraded, as well. The flat tops of the distorted waves don't contribute anything to the strength of the signal at the desired frequency, but they cause a higher-than-normal dc power input, which translates into a lower-than-normal efficiency.

## Audio Amplification

The circuits you've seen so far have been general, not application-specific. With capacitors of several microfarads, and when biased for class A, these circuits are representative of audio amplifiers.

### Frequency Response

High-fidelity audio amplifiers, of the kind used in music systems, must have more or less constant gain from 20 Hz to 20 kHz. This is a frequency range of 1000:1. Audio amplifiers for voice communications must work from 300 Hz to 3 kHz, a 10:1 span of frequencies. In digital communications, audio amplifiers are designed to work over a narrow range of frequencies, sometimes less than 100 Hz wide.

Hi-fi amplifiers are usually equipped with resistor-capacitor (RC) networks that tailor the frequency response. These are *tone controls*, also called *bass* and *treble* controls. The simplest hi-fi amplifiers use a single knob to control the tone. More sophisticated amplifiers have separate controls, one for bass and the other for treble. The most advanced hi-fi systems make use of *graphic equalizers*, having controls that affect the amplifier gain over several different frequency spans.

### Volume Control

Audio amplifier systems usually consist of two or more *stages*. A stage is one bipolar transistor or FET (or a push-pull combination), plus peripheral resistors and capacitors. Stages can be cascaded one after the other to get high gain. In one of the stages in an audio system, a *volume control* is used. This control can be as simple as a potentiometer that allows the gain of a stage to be adjusted without affecting its linearity.

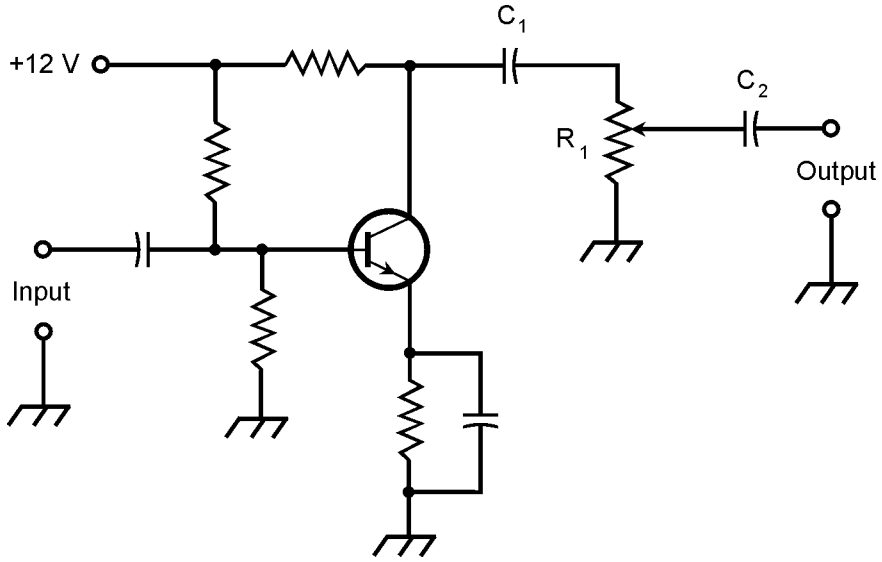
An example of a basic volume control is shown in Fig. 24-7. In this amplifier, the gain through the transistor is constant. The ac output signal passes through  $C_1$  and appears across  $R_1$ , a potentiometer. The *wiper* (indicated by the arrow) of the potentiometer picks off more or less of the ac output signal, depending on the position of the control shaft. Capacitor  $C_2$  isolates the potentiometer from the dc bias of the following stage.

A volume control should normally be placed in a stage where the audio power level is low. This allows the use of a low-wattage, low-cost potentiometer.

### Transformer Coupling

Transformers can be used to transfer (or *couple*) signals from one stage to the next in a cascaded amplifier system (also known as an *amplifier chain*). An example of *transformer coupling* is shown in Fig. 24-8. Capacitors  $C_1$  and  $C_2$  keep one end of the transformer primary and secondary at signal ground. Resistor  $R_1$  limits the current through the first transistor,  $Q_1$ . Resistors  $R_2$  and  $R_3$  provide the proper base bias for transistor  $Q_2$ .

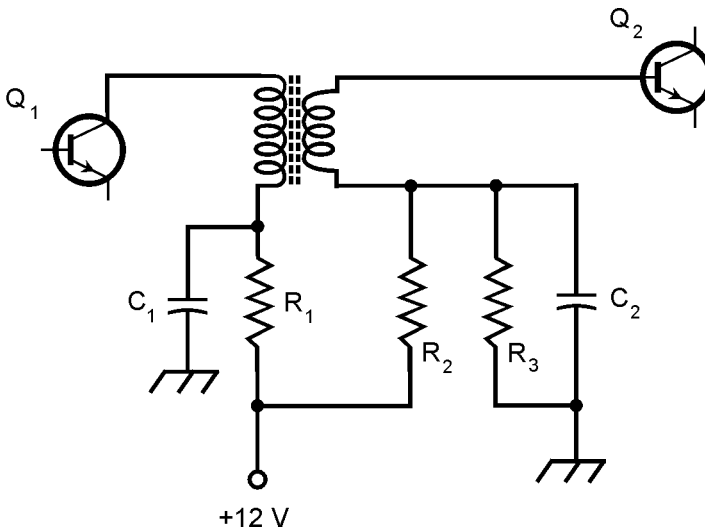
The main disadvantage of this scheme is that it costs more than *capacitive coupling*. But transformer coupling can provide an optimum signal transfer between amplifier stages. By selecting a transformer with the correct turns ratio, the output impedance of  $Q_1$  can be perfectly matched to the input impedance of  $Q_2$ .



24-7 A basic volume control (potentiometer  $R_1$ ) can be used to vary the gain in a low-power audio amplifier.

### Tuned-Circuit Coupling

In some amplifier systems that employ transformer coupling, capacitors are added across the primary and/or secondary of the transformer. This results in resonance at a frequency determined by the capacitance and the transformer winding inductance. If the set of amplifiers is intended for use at only one frequency (and this is often the case in RF systems), this method of coupling, called



24-8 An example of transformer coupling between amplifier stages. Component designators are discussed in the text.

*tuned-circuit coupling*, enhances the system efficiency. But care must be taken to be sure that the amplifier chain doesn't oscillate at the resonant frequency of the tuned circuits!

## Radio-Frequency Amplification

The *RF spectrum* extends upward in frequency to well over 300 GHz. The exact lower limit is a matter of disagreement in the literature. Some texts put it at 3 kHz, some at 9 kHz, some at 10 kHz, and some at the upper end of the AF range, which is normally considered to be 20 kHz.

### Weak-Signal Amplifiers versus Power Amplifiers

Some RF amplifiers are designed for weak-signal work. The *front end*, or first amplifying stage, of a radio receiver requires the most sensitive possible amplifier. Sensitivity is determined by two factors: the *gain*, which has already been discussed here, and the *noise figure*, a measure of how well a circuit can amplify desired signals while generating a minimum of electronic noise.

All bipolar transistors or FETs create some *white noise* because of the movement of the charge carriers among the atoms. In general, JFETs produce less noise than bipolar transistors. Gallium arsenide FETs, also called *GaAsFETs* (pronounced "gasfets"), are the least noisy of all.

The higher the frequency at which a weak-signal amplifier is designed, the more important the noise figure gets. This is because there is less *atmospheric noise* at the higher radio frequencies, as compared with the lower frequencies. At 1.8 MHz, for example, the airwaves contain much atmospheric noise, and it doesn't make any difference if the receiver introduces a little noise itself. But at 1.8 GHz, atmospheric noise is almost nonexistent, and receiver performance depends critically on the amount of internally generated noise.

Weak-signal amplifiers almost always use resonant circuits. This optimizes the amplification at the desired frequency, while helping to cut out noise on unwanted frequencies. A typical tuned GaAsFET weak-signal RF amplifier is diagrammed in Fig. 24-9. It is designed for operation at about 10 MHz.

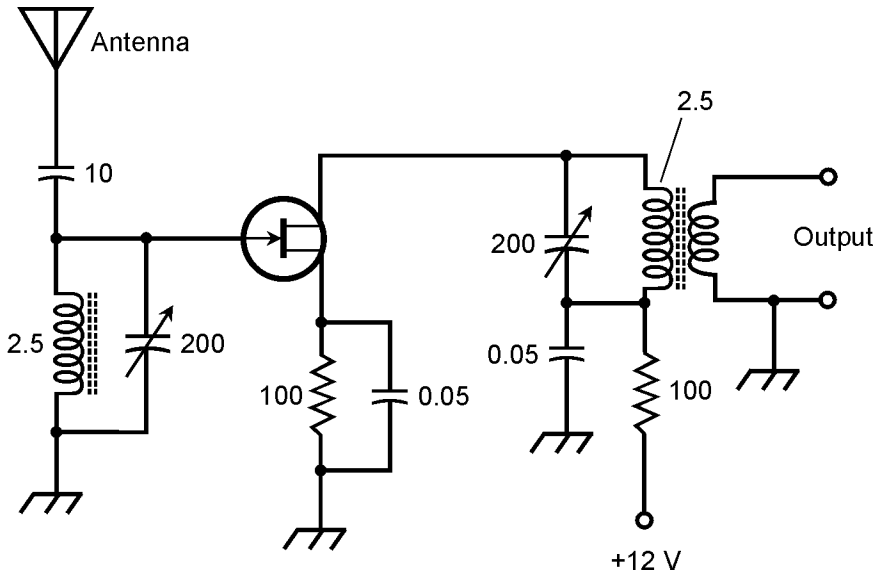
### Broadband PAs

At RF, a PA can be either *broadband* or *tuned*. The main advantage of a broadband PA is ease of operation, because it does not need tuning. The operator need not worry about critical adjustments, nor bother to change them when changing the frequency. However, broadband PAs are slightly less efficient than tuned PAs. Another disadvantage of broadband PAs is the fact that they will amplify any signal in the design frequency range, whether or not this is desired. For example, if some earlier stage in a radio transmitter is oscillating at a frequency different from the intended signal frequency, and if this undesired energy falls within the design frequency range of the broadband PA, it will be amplified. The result will be unintended RF emission from the radio transmitter. Such unwanted signals are called *spurious emissions*.

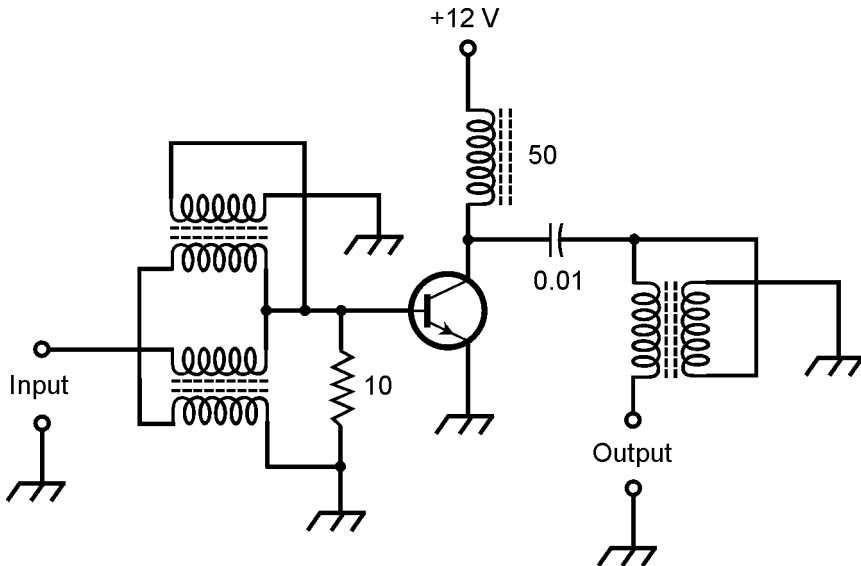
Figure 24-10 is a schematic diagram of a typical broadband PA. The NPN bipolar transistor is a power transistor. It will reliably provide several watts of continuous RF power output over a range of frequencies from 1.5 MHz through 15 MHz. The transformers are a critical part of this circuit. They must be designed to work efficiently over a 10:1 range of frequencies.

### Tuned PAs

A tuned RF power amplifier offers improved efficiency compared with broadband designs. Also, the tuning helps to reduce the chances of spurious signals being amplified and transmitted over the air. Another advantage of tuned PAs is that they can work into a wide range of load impedances. In ad-

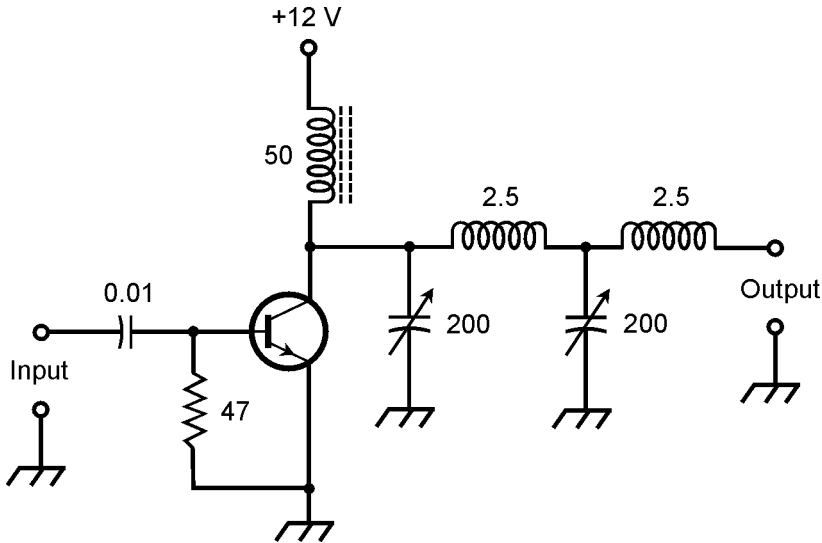


**24-9** A tuned RF amplifier for use at about 10 MHz. Resistances are in ohms. Capacitances are in microfarads ( $\mu\text{F}$ ) if less than 1, and in picofarads (pF) if more than 1. Inductances are in microhenrys ( $\mu\text{H}$ ).



**24-10** A broadband RF power amplifier, capable of producing a few watts output. Resistances are in ohms. Capacitances are in microfarads ( $\mu\text{F}$ ). Inductances are in microhenrys ( $\mu\text{H}$ ).





**24-11** A tuned RF power amplifier, capable of producing a few watts output. Resistances are in ohms. Capacitances are in microfarads ( $\mu\text{F}$ ) if less than 1, and in picofarads ( $\text{pF}$ ) if more than 1. Inductances are in microhenrys ( $\mu\text{H}$ ).

dition to a *tuning control*, or resonant circuit that adjusts the output of the amplifier to the operating frequency, there is a *loading control* that optimizes the signal transfer between the amplifier and the load (usually an antenna).

The main drawback of a tuned PA is that the adjustment takes time, and improper adjustment can result in damage to the transistor. If the tuning and/or loading controls are out of kilter, the efficiency of the amplifier will be extremely low (sometimes practically zero) while the dc collector or drain power input is high. Solid-state devices overheat quickly under these conditions.

A tuned RF PA, providing a few watts' output at 10 MHz or so, is shown in Fig. 24-11. The transistor is the same type as for the broadband amplifier discussed previously. The tuning and loading controls (left-hand and right-hand variable capacitors, respectively) should be adjusted for maximum RF power output as indicated by an RF wattmeter.

## How Oscillators Work

Once you know how amplifiers work, it's easy to understand oscillators. All oscillators are amplifiers with positive feedback. In radio communications, oscillators generate the waves, or signals, that are ultimately sent over the air. Audio-frequency oscillators find applications in such devices as music synthesizers, modems, doorbells, sirens, alarms, and electronic toys.

### Positive Feedback

Feedback can be in phase or out of phase. For a circuit to oscillate, the feedback must be in phase (positive). Out-of-phase (negative) feedback reduces the gain of an amplifier. In fact, negative feedback is used in some amplifiers to *prevent* oscillation.

The output of a common emitter or common source amplifier is out of phase from the input. If you couple the collector to the base through a capacitor, you won't get oscillation. It is necessary to reverse the phase in the feedback process in order for oscillation to occur. In addition, the amplifier gain must be high, and the coupling from the output to the input must be good. The positive feedback path must be easy for a signal to follow. Most oscillators are common emitter or common source amplifier circuits with positive feedback.

The output of a common base or common gate amplifier is in phase with the input. But these circuits have limited gain, and it's hard to make them oscillate. Common collector and common drain circuits don't have enough gain to make oscillators.

### Feedback at a Single Frequency

The frequency of an oscillator is controlled by means of tuned, or resonant, circuits. These are usually inductance-capacitance ( $LC$ ) or resistance-capacitance ( $RC$ ) combinations. The  $LC$  scheme is common in radio transmitters and receivers; the  $RC$  method is more often used in audio work. The tuned circuit makes the feedback path easy for a signal to follow at one frequency, but hard to follow at all other frequencies. As a result, oscillation takes place at a predictable and stable frequency, determined by the inductance and capacitance, or by the resistance and capacitance.

## Common Oscillator Circuits

Many circuit arrangements can be used to produce oscillation. The following several circuits are all known as *variable-frequency oscillators* (VFOs), because their frequencies are adjustable over a wide range.

### The Armstrong Circuit

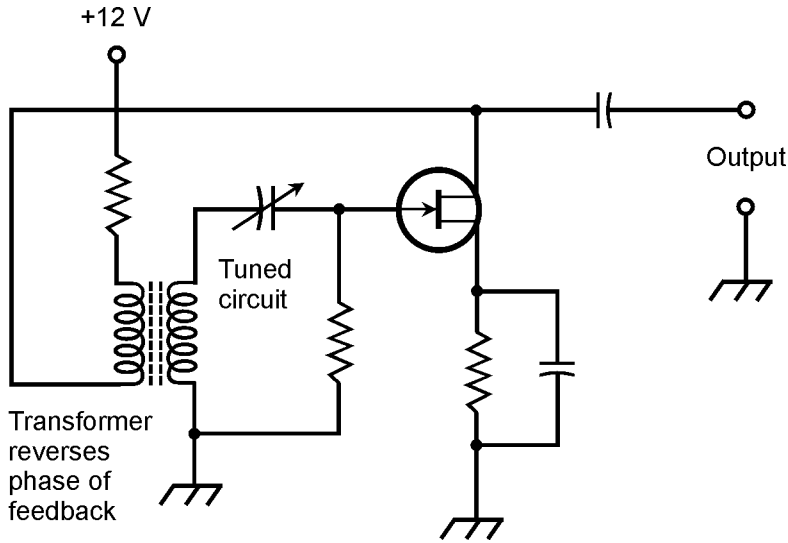
A common emitter or common source class A amplifier can be made to oscillate by coupling the output back to the input through a transformer that reverses the phase of the feedback signal. The schematic diagram of Fig. 24-12 shows a common-source amplifier whose drain circuit is coupled to the gate circuit by means of a transformer. The frequency is controlled by a capacitor in series with the secondary winding. The inductance of the secondary, along with the capacitance, forms a resonant circuit that passes energy easily at one frequency while attenuating the energy at other frequencies. This circuit is known as an *Armstrong oscillator*. A bipolar transistor can be used in place of the JFET, as long as the device is biased for class A amplification.

### The Hartley Circuit

Another method of obtaining controlled feedback at RF is shown in Fig. 24-13. In this circuit, a PNP bipolar transistor is used. The circuit uses a single coil with a tap on the windings to provide the feedback. A variable capacitor in parallel with the coil determines the oscillating frequency, and allows for frequency adjustment. This circuit is called a *Hartley oscillator*.

In the Hartley circuit, as well as in most other RF oscillator circuits, it is important to use only the minimum amount of feedback necessary to get oscillation. The amount of feedback is controlled by the position of the coil tap. The circuit shown in Fig. 24-13 uses about 25 percent of its amplifier power to produce feedback. The other 75 percent of the power can be used as output.

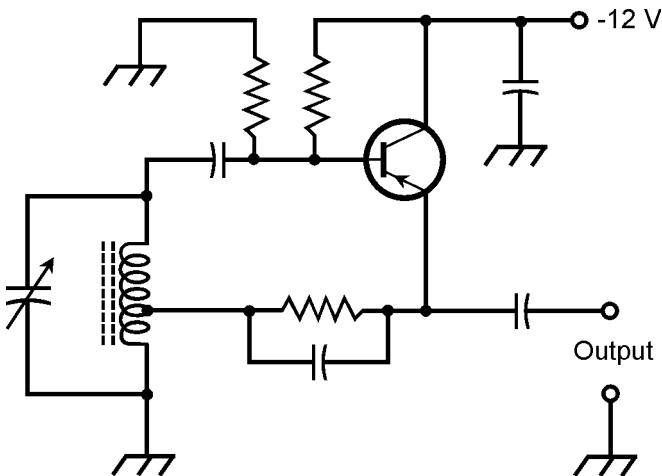
Oscillators usually produce less than 1 W of RF power output. If more power is needed, the signal can be boosted by one or more stages of amplification.



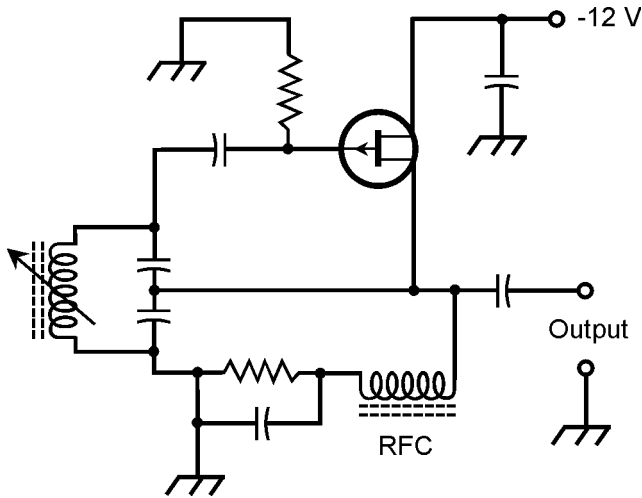
**24-12** An Armstrong oscillator using an N-channel JFET. This is a common-source amplifier with positive feedback through a tuned circuit.

### The Colpitts Circuit

The capacitance can be tapped, instead of the inductance, in the tuned circuit of an RF oscillator. Such a circuit is called a *Colpitts oscillator*, and a P-channel JFET version is diagrammed in Fig. 24-14. The amount of feedback is controlled by the ratio of the capacitances. A variable inductor provides for frequency adjustment. This is a matter of convenience, because it can be difficult to find a dual



**24-13** A Hartley oscillator using a PNP bipolar transistor. The Hartley circuit can be recognized by the tapped inductor in the tuned LC circuit.



**24-14** A Colpitts oscillator using a P-channel JFET. The Colpitts circuit can be recognized by the split capacitance in the tuned  $LC$  circuit.

variable capacitor that maintains the correct ratio of capacitances throughout its tuning range. Using fixed capacitors eliminates this problem, and it costs less, too!

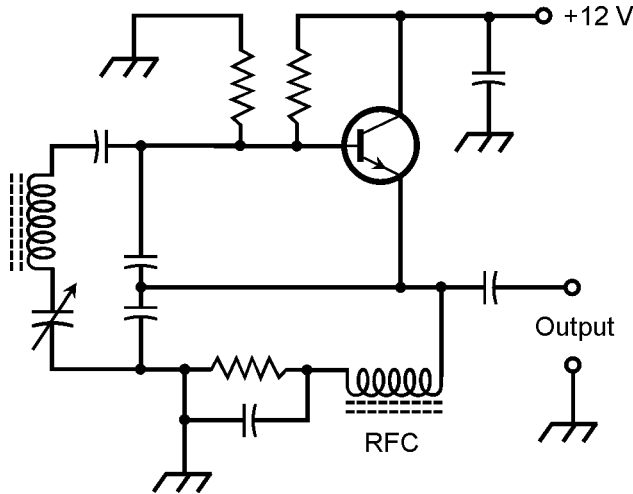
Unfortunately, finding a good variable inductor for use in a Colpitts oscillator can be just about as hard as obtaining a suitable dual-gang variable capacitor. A *permeability-tuned* coil can be used, but ferromagnetic cores impair the frequency stability of an RF oscillator. A *roller inductor* can be employed, but these are bulky and expensive. An inductor with several switch-selectable taps can be used, but this does not allow for continuous frequency adjustment. Despite these shortcomings, the Colpitts circuit offers exceptional stability and reliability when properly designed, and is preferred by some engineers for this reason.

## The Clapp Circuit

A variation of the Colpitts oscillator makes use of series resonance, instead of parallel resonance, in the tuned circuit. Otherwise, the circuit is basically the same as the parallel-tuned Colpitts oscillator. Figure 24-15 is a schematic diagram of a *series-tuned Colpitts oscillator* circuit that uses an NPN bipolar transistor. This circuit is also known as a *Clapp oscillator*. Its frequency won't change much when high-quality components are used. The Clapp oscillator is a reliable circuit. It isn't hard to get it to oscillate and keep it going. Another advantage of the Clapp circuit is that it allows the use of a variable capacitor for frequency control, while accomplishing feedback through a capacitive voltage divider.

## Getting the Output

In the Hartley, Colpitts, and Clapp oscillators just described and shown in Figs. 24-13 through 24-15, the output is taken from the emitter or source, not from the collector or drain. There's a reason for this. The output of an oscillator can be taken from the collector or drain, just as is done in a common emitter or common-source amplifier to get maximum gain. But in an oscillator, stabil-



**24-15** A series-tuned Colpitts oscillator, also known as a Clapp oscillator. This circuit uses an NPN bipolar transistor.

ity is more important than gain. The stability of an oscillator is better when the output is taken from the emitter or source, as compared with taking it from the collector or drain. Variations in the load impedance are less likely to affect the frequency of oscillation, and a sudden decrease in load impedance is less likely to cause the circuit to stop oscillating.

To prevent the output signal from being short-circuited to ground, an *RF choke* (RFC) is connected in series with the emitter or source in the Colpitts and Clapp oscillator circuits. The choke lets dc pass while blocking ac (just the opposite of a *blocking capacitor*). Typical values for RF chokes range from about 100  $\mu\text{H}$  at high frequencies, such as 15 MHz, to 10 mH at low frequencies, such as 150 kHz.

## The Voltage-Controlled Oscillator

The frequency of a VFO can be adjusted by means of a varactor diode in the tuned *LC* circuit. Recall that a varactor, also called a *varicap*, is a semiconductor diode that works as a variable capacitor when it is reverse-biased. The capacitance depends on the reverse-bias voltage. The greater this voltage, the lower the value of the capacitance.

The Hartley and Clapp oscillator circuits lend themselves well to varactor-diode frequency control. The varactor is placed in series or parallel with the tuning capacitor, and is isolated for dc by blocking capacitors. In Chap. 20, we saw an example of how a varactor can be connected in a tuned circuit (Fig. 20-9). The resulting oscillator is called a *voltage-controlled oscillator* (VCO).

Varactors are cheaper than variable capacitors or inductors. They're also less bulky. These are the chief advantages of a VCO over an old-fashioned *LC* tuned VFO.

## Diode Oscillators

At ultrahigh frequencies (UHF) and microwave radio frequencies, certain types of diodes can be used as oscillators. You learned about these diodes in Chap. 20.

## Crystal-Controlled Oscillators

*Quartz crystals* can be used in place of tuned *LC* circuits in RF oscillators, as long as it isn't necessary to change the frequency often. Crystal oscillators offer frequency stability far superior to that of *LC* tuned VFOs.

There are several ways that crystals can be connected in bipolar or FET circuits to get oscillation. One common circuit is the *Pierce oscillator*. An N-channel JFET and quartz crystal are connected in a Pierce configuration as shown in the schematic diagram of Fig. 24-16. The crystal frequency can be varied somewhat (by about 0.1 percent, or 1 part in 1000) by means of an inductor or capacitor in parallel with the crystal. But the frequency is determined mainly by the thickness of the quartz wafer, and by the angle at which it is cut from the original quartz sample.

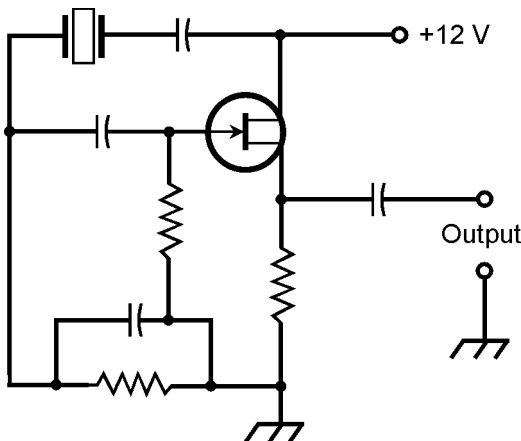
Crystals change in frequency as the temperature changes. But they are far more stable than *LC* circuits, most of the time. Some crystal oscillators are housed in temperature-controlled chambers called *crystal ovens*. In this environment, crystals maintain their frequency so well that they are sometimes used as *frequency standards* against which other oscillators are calibrated.

## The Phase-Locked Loop

One type of oscillator that combines the flexibility of a VFO with the stability of a crystal oscillator is known as a *phase-locked loop* (PLL). This makes use of a circuit called a *frequency synthesizer*.

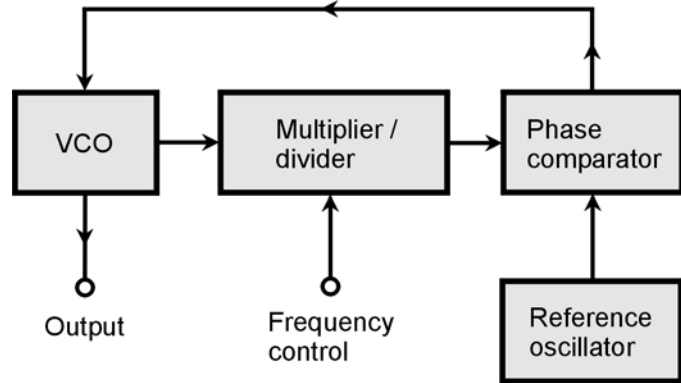
The heart of the PLL is a VCO. The output of this oscillator passes through a *programmable multiplier/divider*, a digital circuit that divides and/or multiplies the VCO frequency by integral (whole-number) values chosen by the operator. As a result, the output frequency can be any rational-number multiple of the crystal frequency. A well-designed PLL circuit can be tuned in small digital increments over a wide range of frequencies.

The output frequency of the multiplier/divider is locked, by means of a *phase comparator*, to the signal from a crystal-controlled *reference oscillator*. As long as the output from the multiplier/divider is exactly on the reference oscillator frequency, the two signals are in phase, and the output of the phase comparator is 0 V dc. If the VCO frequency begins to drift, the output frequency of the multiplier/divider will drift, too (although at a different rate). But even a frequency change of less than 1 Hz causes the phase comparator to produce a dc *error voltage*. This error voltage is either pos-



**24-16** A Pierce oscillator circuit using an N-channel JFET.

24-17 Block diagram of a phase-locked loop (PLL).



itive or negative, depending on whether the VCO has drifted higher or lower in frequency. The error voltage is applied to a varactor, causing the VCO frequency to change in a direction opposite to that of the drift. This forms a *dc feedback* circuit that maintains the VCO frequency at a precise value. It is a *loop* circuit that *locks* the VCO onto a particular frequency by means of *phase* sensing, hence the term *phase-locked loop* (PLL).

The key to the stability of the PLL lies in the fact that the reference oscillator is crystal-controlled. A block diagram of a PLL circuit is shown in Fig. 24-17. When you hear that a radio receiver, transmitter, or transceiver is synthesized, it usually means that the frequency is determined by a PLL.

The stability of a synthesizer can be enhanced by using an amplified signal from the shortwave time-and-frequency broadcast station WWV at 2.5, 5, 10, or 15 MHz, directly as the reference oscillator. These signals are frequency-exact to a minuscule fraction of 1 Hz, because they are controlled by atomic clocks. Most people don't need precision of this caliber, so you won't see consumer devices like ham radios and shortwave receivers with *primary-standard* PLL frequency synthesis. But it is employed by some corporations and government agencies.

## Oscillator Stability

In an oscillator, the term *stability* has two meanings: constancy of frequency (or minimal *frequency drift*), and reliability of performance.

### Constancy of Frequency

When designing a VFO of any kind, it's essential that the components maintain constant values, as much as possible, under all anticipated conditions. Some types of capacitors maintain their values better than others as the temperature rises or falls. Among the best are polystyrene capacitors. Silver-mica capacitors also work well when polystyrene units can't be found. Inductors are most temperature-stable when they have air cores. They should be wound, when possible, from stiff wire with strips of plastic to keep the windings in place. Some air-core coils are wound on hollow cylindrical cores, made of ceramic or phenolic material. Ferromagnetic solenoidal or toroidal cores aren't very good for VFO coils, because these materials change their permeability as the temperature varies. This changes the inductance, in turn affecting the oscillator frequency.

The best oscillators, in terms of frequency stability, are crystal-controlled. This includes circuits that oscillate at the fundamental frequency of the quartz crystal, circuits that oscillate at one of the crystal harmonic frequencies, or circuits that oscillate at frequencies derived from the crystal frequency by multiplier/dividers.

## Reliability

An oscillator should always start working as soon as power is supplied. It should keep oscillating under all normal conditions. The failure of a single oscillator can cause an entire receiver, transmitter, or transceiver to stop working.

When an oscillator is built and put to use in a radio receiver, transmitter, or audio device, *debugging* is always necessary. This is a trial-and-error process of getting the flaws, or bugs, out of the circuit. Rarely can an engineer build something straight from the drawing board and have it work just right the first time. In fact, if two oscillators are built from the same diagram, with the same component types and values in the same geometric arrangement, one circuit might work fine, and the other might not. This usually happens because of differences in the quality of components that don't show up until the acid test.

Oscillators are designed to work into a certain range of load impedances. It's important that the load impedance not be too low. (You need never be concerned that it might be too high. In general, the higher the load impedance, the better.) If the load impedance is too low, the load will draw significant power from an oscillator. Then, even a well-designed oscillator might become unstable. Oscillators aren't meant to produce powerful signals. High power can be obtained using amplification after the oscillator.

## Audio Oscillators

Audio oscillators are used in myriad electronic devices including doorbells, ambulance sirens, electronic games, telephone sets, and toys that play musical tunes. All AF oscillators are, in effect, AF amplifiers with positive feedback.

### Audio Waveforms

At AF, oscillators can use *RC* or *LC* combinations to determine frequency. If *LC* circuits are used, the inductances must be large, and ferromagnetic cores are necessary.

At RF, oscillators are usually designed to produce a sine wave output. A pure sine wave represents energy at one and only one frequency. Audio oscillators, by contrast, don't always concentrate all their energy at a single frequency. (A pure AF sine wave, especially if it is continuous and frequency-constant, can be annoying.) The various musical instruments in a band or orchestra all sound different from each other, even when they play the same note (such as middle C). The reason for this is that each instrument has its own unique waveform. A clarinet sounds different than a trumpet, which in turn sounds different than a cello or piano.

Suppose you were to use an oscilloscope to look at the waveforms of musical instruments. This can be done using a high-fidelity microphone, a sensitive, low-distortion audio amplifier, and an oscilloscope. You'd see that each instrument has its own signature. Thus, each instrument's unique sound qualities can be reproduced using AF oscillators whose waveform outputs match those of the instrument. Electronic music synthesizers use audio oscillators to generate the tones you hear.



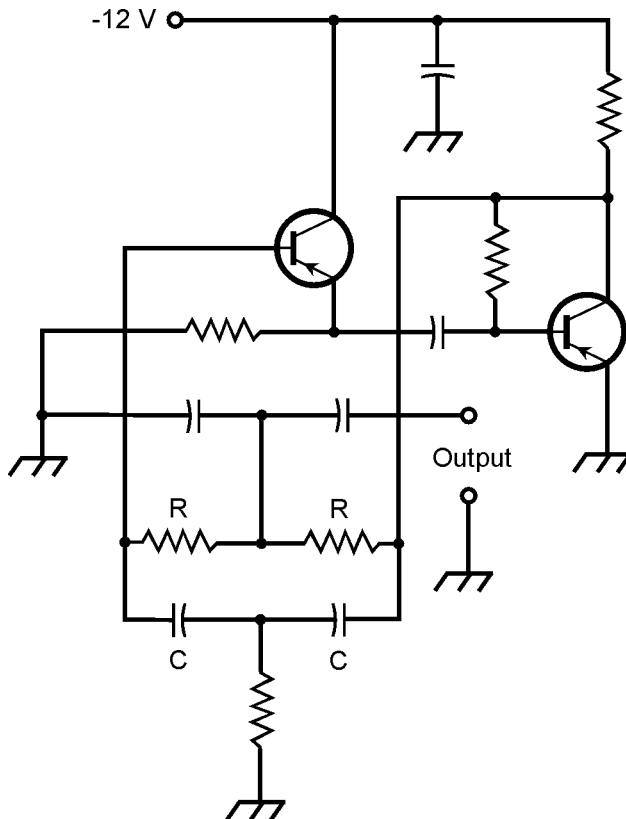
## The Twin T Oscillator

An audio oscillator circuit that is popular for general-purpose use is the *twin T oscillator* (Fig. 24-18). The frequency is determined by the values of the resistors  $R$  and capacitors  $C$ . The output is a near-perfect sine wave. The small amount of distortion helps to alleviate the irritation produced by an absolutely pure sinusoid. The circuit shown in this example uses two PNP bipolar transistors. They are biased for class A amplification.

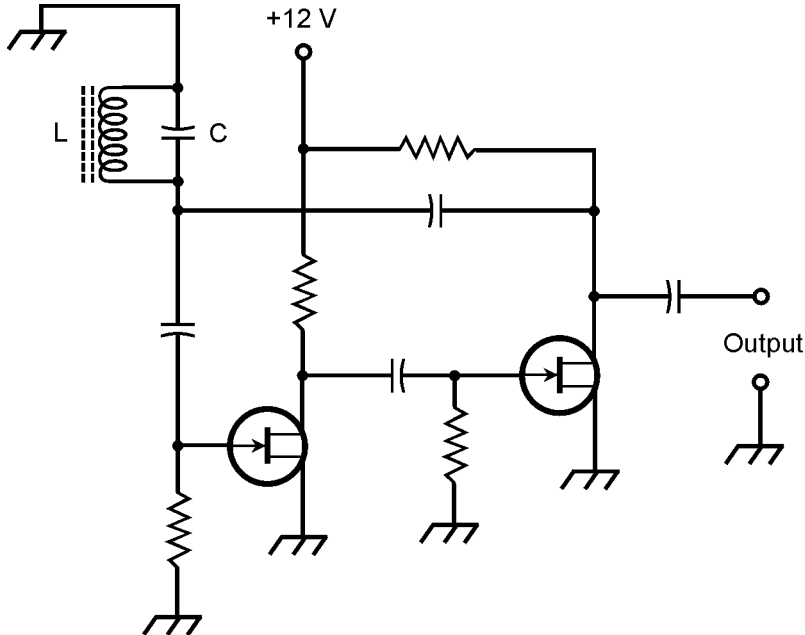
## The Multivibrator

Another popular AF oscillator circuit makes use of two identical common emitter or common source amplifier circuits, hooked up so that the signal goes around and around between them. This is sometimes called a *multivibrator* circuit, although that is technically a misnomer, the term being more appropriate to various digital signal-generating circuits.

In the example of Fig. 24-19, two N-channel JFETs are connected to form a multivibrator for use at AF. Each stage amplifies the signal in class A, and reverses the phase by  $180^\circ$ . Therefore, the signal goes through a  $360^\circ$  phase shift each time it gets back to any particular point. A  $360^\circ$  phase shift is equivalent to no phase shift at all, so it results in positive feedback.



**24-18** A twin T audio oscillator using two PNP bipolar transistors. The frequency is determined by the values of the resistors  $R$  and the capacitors  $C$ .



**24-19** A multivibrator audio oscillator using two N-channel JFETs. The frequency is determined by the value of the inductor  $L$  and the capacitor  $C$ .

The frequency of the circuit shown in Fig. 24-19 is set by means of an  $LC$  circuit. The coil uses a ferromagnetic core, because stability is not of great concern and because such a core is necessary to obtain the large inductance needed for resonance at AF. Toroidal cores or pot cores are excellent in this application. The value of  $L$  can range from about 10 mH to as much as 1 H. The capacitance is chosen according to the formula for resonant circuits, to obtain an audio tone at the frequency desired.

## IC Oscillators

In recent years, solid-state technology has advanced to the point that whole circuits can be etched onto silicon chips. Such devices are called *integrated circuits* (ICs). The *operational amplifier*, also called an *op amp*, is a type of IC that is especially useful as an audio oscillator because it has high gain, and it can easily be connected to produce positive feedback.

## Quiz

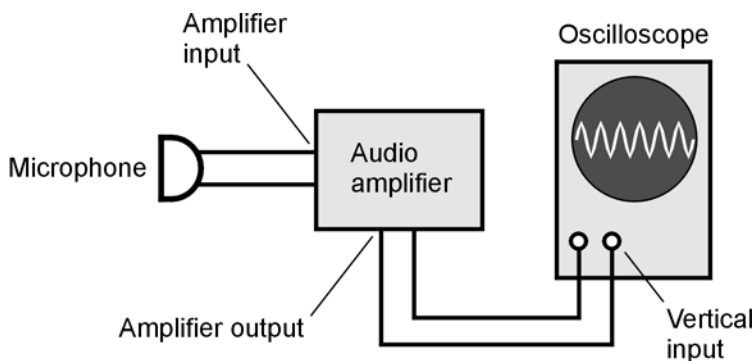
Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. The decibel is a unit of
  - (a) relative signal strength.
  - (b) voltage.

- (c) power.
  - (d) current.
2. An oscillator at RF requires the use of
- (a) a common drain or common collector circuit.
  - (b) a stage with gain.
  - (c) a tapped coil.
  - (d) a quartz crystal.
3. Suppose a circuit is found to have a gain figure of  $-15$  dB. Which of the following statements is true?
- (a) The output signal is stronger than the input signal.
  - (b) The input signal is stronger than the output signal.
  - (c) The input signal is 15 times as strong as the output signal.
  - (d) The output signal is 15 times as strong as the input signal.
4. In an oscillator circuit, the feedback should be
- (a) as great as possible.
  - (b) kept to a minimum.
  - (c) just enough to sustain oscillation.
  - (d) done through a transformer whose wires can be switched easily.
5. A power gain of 44 dB is equivalent to which of the following output/input power ratios?
- (a) 44:1
  - (b) 160:1
  - (c) 440:1
  - (d) 25,000:1
6. An RF choke
- (a) passes RF signals but blocks dc.
  - (b) passes both RF signals and dc.
  - (c) passes dc but blocks RF signals.
  - (d) blocks both dc and RF signals.
7. The optimum capacitance values in an amplifier circuit depend on
- (a) the power-supply voltage.
  - (b) the power-supply polarity.
  - (c) the input signal strength.
  - (d) the input signal frequency.
8. An oscillator might fail to start for any of the following reasons, except
- (a) low battery voltage.
  - (b) low stage gain.

- (c) in-phase feedback.
  - (d) a high-impedance load.
9. In which of the following FET amplifier types does drain current flow for exactly 50 percent of the signal cycle?
- (a) Class A
  - (b) Class AB<sub>1</sub>
  - (c) Class AB<sub>2</sub>
  - (d) Class B
10. The frequency at which a quartz crystal oscillator produces energy is largely dependent on
- (a) the load impedance.
  - (b) the physical thickness of the quartz wafer.
  - (c) the amount of resistance through the crystal.
  - (d) the power-supply voltage.
11. Which bipolar amplifier type has some distortion in the signal wave, with collector current during most, but not all, of the cycle?
- (a) Class A
  - (b) Class AB<sub>1</sub>
  - (c) Class AB<sub>2</sub>
  - (d) Class B
12. An RF oscillator usually
- (a) produces an output signal with an irregular waveshape.
  - (b) has most or all of its energy at a single frequency.
  - (c) produces a sound that depends on its waveform.
  - (d) employs an RC circuit to determine the output amplitude.
13. A class C amplifier can be made linear by
- (a) reducing the bias.
  - (b) increasing the drive.
  - (c) using two transistors in push-pull.
  - (d) no means; a class C amplifier is always nonlinear.
14. A frequency synthesizer has
- (a) high power output.
  - (b) high frequency drift rate.
  - (c) exceptional stability.
  - (d) an adjustable waveshape.
15. A graphic equalizer is a form of
- (a) bias control for an NPN bipolar transistor.

- (b) gain control for an RF oscillator.
  - (c) tone control that can be used in an audio amplifier.
  - (d) circuit for adjusting the waveform of an RF oscillator.
16. If the impedance of the load connected to the output of an oscillator is extremely high,
- (a) the frequency will drift excessively.
  - (b) the power output will be reduced.
  - (c) the oscillator might fail to start.
  - (d) it is no cause for concern; in fact, it is a good thing.
17. Suppose a certain bipolar-transistor PA is 66 percent efficient. The output power is 33 W. The dc collector power input is
- (a) 22 W.
  - (b) 50 W.
  - (c) 2.2 W.
  - (d) impossible to determine without more information.
18. The arrangement in the block diagram of Fig. 24-20 represents
- (a) a waveform analyzer.
  - (b) an audio oscillator.
  - (c) an RF oscillator.
  - (d) a sine wave generator.
19. A tuned RF PA must always be
- (a) set to work over a wide range of frequencies.
  - (b) adjusted for maximum power output.
  - (c) operated at an even harmonic of the input frequency.
  - (d) operated in class C.



**24-20** Illustration for Quiz Question 18.

20. Class B amplification can be used to obtain low distortion for audio applications
- (a) by connecting two amplifiers in cascade, thereby maximizing the gain and generating a pure sine wave output.
  - (b) by biasing the bipolar transistor or FET beyond cutoff or pinchoff, thereby ensuring that the output is in phase with the input.
  - (c) by connecting two identical bipolar transistors or FETs, biased exactly at cutoff or pinchoff, in a push-pull configuration.
  - (d) by biasing the bipolar transistor or FET in the middle of the straight-line portion of the characteristic curve.

## 25 CHAPTER

# Wireless Transmitters and Receivers

IN RADIO OR WIRELESS COMMUNICATIONS, A *TRANSMITTER* CONVERTS DATA INTO *ELECTROMAGNETIC* (EM) *waves* intended for recovery by one or more *receivers*. In this chapter, we'll look at how data is converted to EM waves and transmitted, and then examine how the resulting *EM fields* can be intercepted and received.

## Oscillation and Amplification

A radio transmitter employs one or more oscillators to generate an RF signal, and amplifiers to generate the required power output. You just learned how these circuits work. Most transmitters have *mixers* in addition to the oscillating and amplifying stages. Signal mixing is commonly done with diodes, and was discussed in Chap. 20.

## Modulation

*Modulation* is the process of writing data onto an electric current or EM wave. The process can be done by varying the amplitude, the frequency, or the phase of the current or wave. Another method is to transmit a series of pulses, whose duration, amplitude, or spacing is made to vary.

### The Carrier

The heart of a wireless signal is a sine wave known as the *carrier*. The lowest carrier frequency used for radio communications is a few kilohertz (kHz). The highest frequency is in the hundreds of gigahertz (GHz). For efficient data transfer, the carrier frequency must be at least 10 times the highest frequency of the modulating signal.

### On/Off Keying

The simplest form of modulation is *on/off keying*. This can be done in the oscillator of a radio transmitter to send Morse code, which is a binary digital mode. The duration of a Morse-code *dot* is one *bit* (binary digit). A *dash* is 3 bits long. The space between dots and dashes within a *charac-*

*ter* is 1 bit. The space between characters in a *word* is 3 bits. The space between words is 7 bits. The key-down (full-carrier) condition is called *mark*, and the key-up (no-signal) condition is called *space*.

Morse code is slow. Human operators use speeds ranging from about 5 words per minute (wpm) to 40 or 50 wpm. A few human operators can work at 60 to 70 wpm. These people usually copy the signals in their heads.

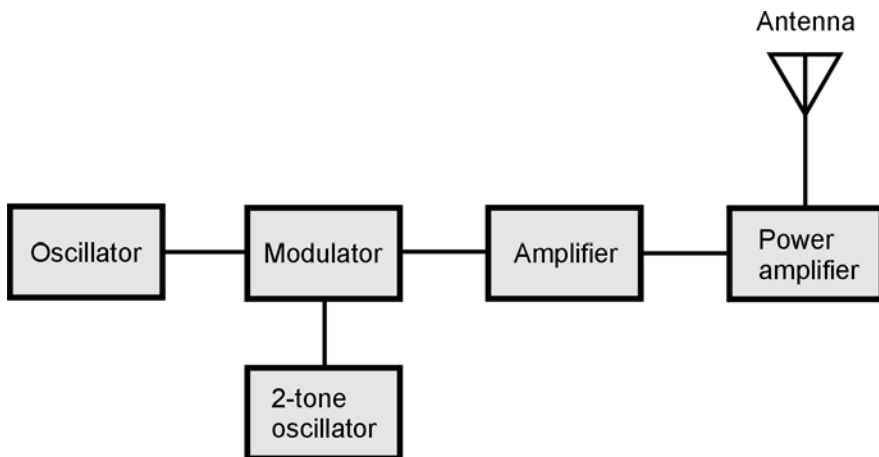
## Frequency-Shift Keying

Digital data can be sent over wireless links by means of *frequency-shift keying* (FSK). In some FSK systems, the carrier frequency is shifted between mark and space conditions, usually by a few hundred hertz or less. In other systems, a two-tone audio-frequency (AF) sine wave modulates the carrier. This is known as *audio-frequency-shift keying* (AFSK).

The two most common codes used with FSK and AFSK are *Baudot* (pronounced “baw-DOE”) and *ASCII* (pronounced “ASK-ee”). The acronym ASCII stands for *American Standard Code for Information Interchange*.

In *radioteletype* (RTTY) FSK and AFSK systems, a *terminal unit* (TU) converts the digital signals into electrical impulses that operate a teleprinter or display characters on a computer screen. The TU also generates the signals necessary to send RTTY as an operator types on a keyboard. A device that sends and receives AFSK is sometimes called a *modem*. This acronym stands for *modulator/demodulator*. A modem is basically the same as a TU. Figure 25-1 is a block diagram of an AFSK transmitter.

The main advantage of FSK or AFSK over on/off keying is the fact that there are fewer errors or misprints, because the space part of the signal is identified as such, rather than existing as a gap or pause in the data. A sudden noise burst in an on/off keyed signal can confuse a receiver into reading the space as a mark signal, but when the space is positively represented by its own signal, this is less likely to happen.



25-1 Simplified block diagram of an AFSK transmitter.



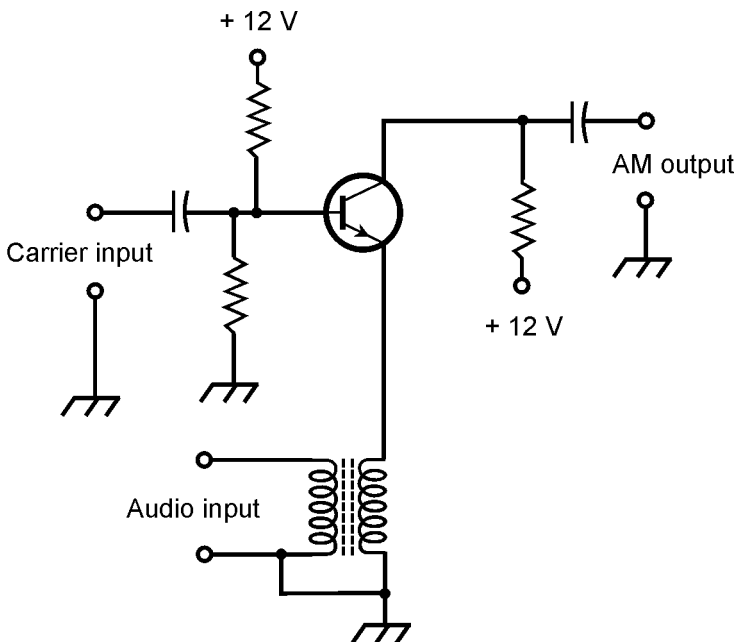
## Amplitude Modulation

An AF voice signal has frequencies mostly in the range between 300 Hz and 3 kHz. Some characteristic of an RF carrier can be varied, or *modulated*, by these waveforms, thereby transmitting voice information. Figure 25-2 shows a simple circuit for obtaining *amplitude modulation* (AM). This circuit can be imagined as an RF amplifier for the carrier, with the instantaneous gain dependent on the instantaneous audio input amplitude. Another way to think of this circuit is as a mixer that combines the RF carrier and audio signals to produce sum and difference signals at frequencies just above and below that of the carrier.

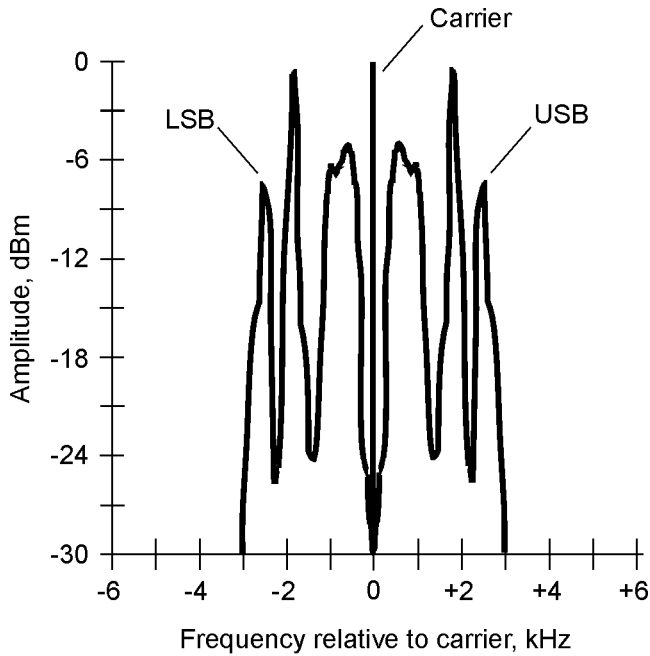
The circuit shown in Fig. 25-2 works well, provided the AF input amplitude is not too great. If the AF input is excessive, then distortion occurs, intelligibility is degraded, system efficiency is reduced, and the bandwidth of the signal is increased unnecessarily.

The extent of AM is expressed as a percentage, from 0 percent (an unmodulated carrier) to 100 percent (full modulation). Increasing the modulation past 100 percent causes the same problems as excessive AF input. In an AM signal modulated 100 percent,  $\frac{1}{3}$  of the power is used to convey the data, and the other  $\frac{2}{3}$  is consumed by the carrier wave.

Figure 25-3 shows a *spectral display* of an AM voice radio signal. The horizontal scale is calibrated in increments of 1 kHz per division. Each vertical division represents 3 dB of change in signal strength. The maximum (reference) amplitude is 0 dB relative to 1 mW (abbreviated as 0 dBm). The data exists in *sidebands* above and below the carrier frequency. These sidebands resemble the sum and difference signals produced by a mixer. In this case the mixing occurs between the AF input signal and the RF carrier. The RF between  $-3$  kHz and the carrier frequency constitutes the *lower*



25-2 An amplitude modulator using an NPN bipolar transistor.



**25-3** Spectral display of a typical AM voice communications signal.

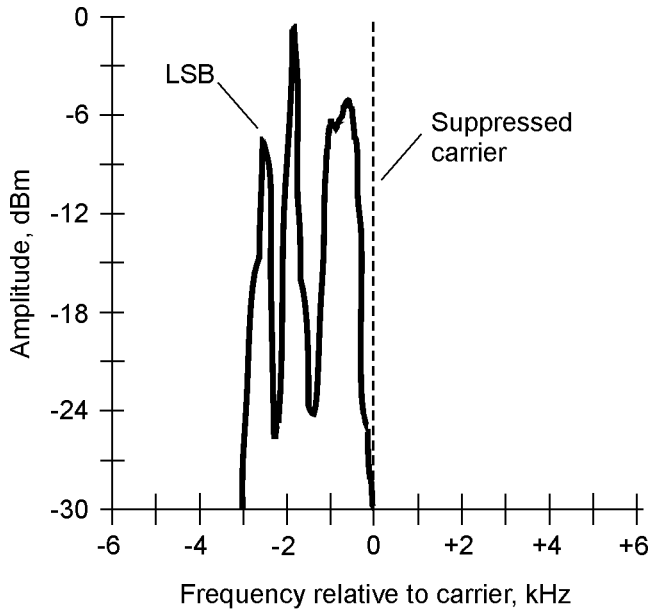
*sideband* (LSB); the RF from the carrier frequency to +3 kHz represents the *upper sideband* (USB). The *bandwidth* is the difference between the maximum and minimum sideband frequencies, in this case 6 kHz.

In an AM signal, the bandwidth is twice the highest audio modulating frequency. In the example of Fig. 25-3, all the voice energy is at or below 3 kHz, so the signal bandwidth is 6 kHz. This is typical of a communications signal. In standard AM broadcasting, the AF energy is spread over a wider bandwidth, nominally 10 kHz.

## Single Sideband

In AM, most of the RF signal power is consumed by the carrier alone; the two sidebands are mirror-image duplicates. This is inefficient, and is also unnecessarily redundant! If the carrier and one of the sidebands is eliminated, these shortcomings can be overcome. That makes the signal stronger for a given amount of RF power, or allows the use of lower RF power in a given communications scenario. Another bonus is the fact that the bandwidth is reduced to less than half that of an AM signal modulated with the same data, so more than twice as many signals can fit into a specific range, or *band*, of frequencies.

When the carrier is removed from an AM signal along with one of the sidebands, the remaining RF energy has a spectral display resembling Fig. 25-4. This is *single sideband* (SSB) transmission. Either the LSB or the USB alone can be used, with equally good results.



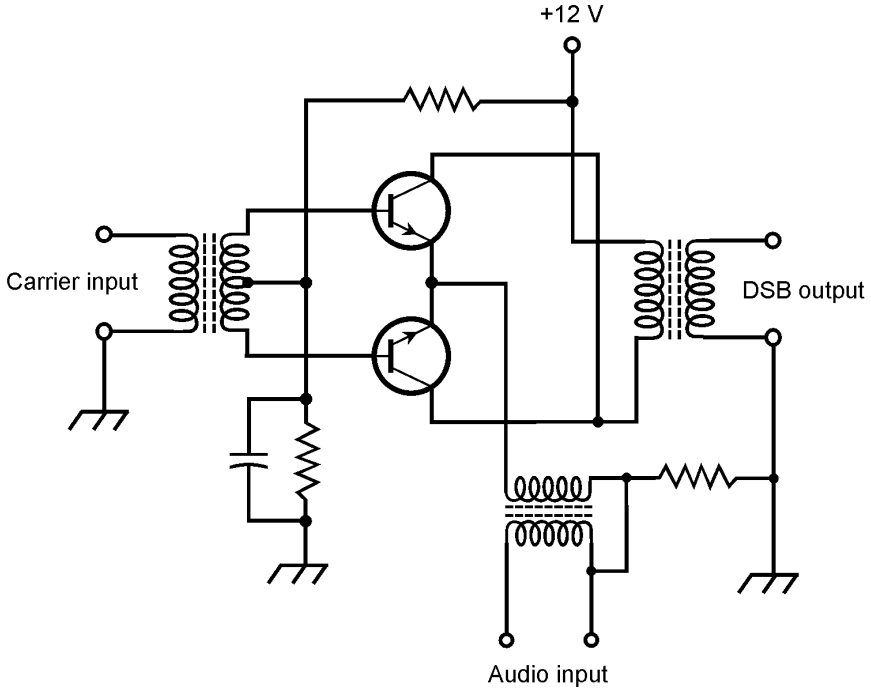
**25-4** Spectral display of a typical SSB voice communications signal. In this example, the carrier and the USB energy are eliminated, leaving only the LSB energy.

An SSB signal can be obtained with a *balanced modulator*, which is an amplitude modulator/amplifier using two transistors with the inputs in push-pull and the outputs in parallel (Fig. 25-5). This cancels the carrier wave in the output, leaving only LSB and USB energy. The result is a *double sideband suppressed carrier* (DSBSC) signal, often called simply *double sideband* (DSB). At some stage following the balanced modulator, one of the sidebands is removed from the DSB signal by a *bandpass filter* to obtain an SSB signal.

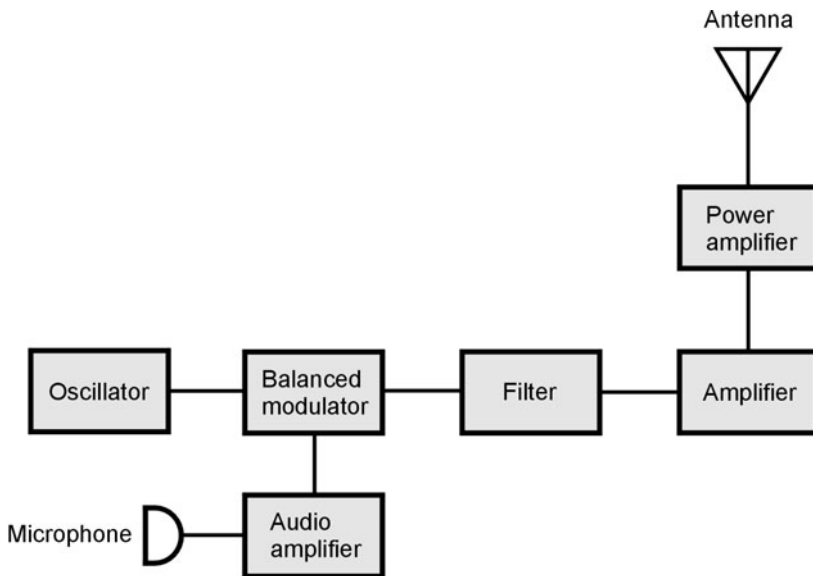
Figure 25-6 is a block diagram of a simple SSB transmitter. The balanced modulator is placed in a low-power section of the transmitter. The RF amplifiers that follow any type of amplitude modulator, including a balanced modulator, must all be linear to avoid distortion and unnecessary spreading of signal bandwidth (“splatter”). They generally work in class A, except for the PA, which works in class AB or class B.

## Frequency and Phase Modulation

In *frequency modulation* (FM), the instantaneous amplitude of a signal remains constant, and the instantaneous frequency is varied instead. A nonlinear PA such as a class C amplifier can be used in an FM transmitter without causing signal distortion, because the amplitude does not fluctuate.



**25-5** A balanced modulator using two NPN bipolar transistors. The inputs are in push-pull, but the outputs are in parallel.



**25-6** Block diagram of a basic SSB transmitter.

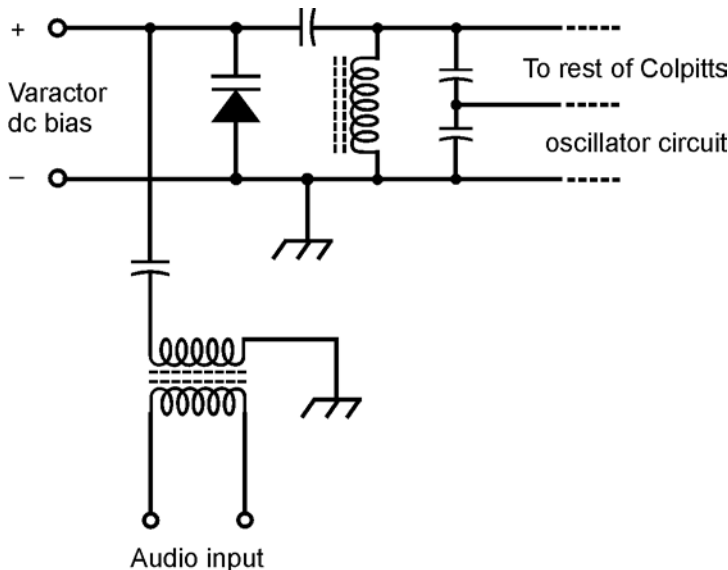
Frequency modulation can be obtained by applying the audio signal to a *varactor* in a tuned oscillator. An example of this scheme, known as *reactance modulation*, is shown in Fig. 25-7. The varying voltage across the varactor causes its capacitance to change in accordance with the audio waveform. The changing capacitance results in variation of the resonant frequency of the inductance-capacitance (LC) tuned circuit, causing small fluctuations in the frequency generated by the oscillator.

Another way to get FM is to modulate the phase of the oscillator signal. This causes small variations in the frequency, because any instantaneous phase change shows up as an instantaneous frequency change (and vice versa). When *phase modulation* is used, the audio signal must be processed, adjusting the frequency response of the audio amplifiers. Otherwise the signal will sound unnatural when it is received.

*Deviation* is the maximum extent to which the instantaneous-carrier frequency differs from the unmodulated-carrier frequency. For most FM voice communications transmitters, the deviation is standardized at  $\pm 5$  kHz. This is known as *narrowband FM* (NBFM). The bandwidth of an NBFM signal is comparable to that of an AM signal containing the same modulating information. In FM hi-fi music broadcasting, and in some other applications, the deviation is much greater than  $\pm 5$  kHz. This is called *wideband FM* (WBFM).

The deviation obtainable by means of direct FM is greater, for a given oscillator frequency, than the deviation that can be obtained by means of phase modulation. The deviation of a signal can be increased by a *frequency multiplier*. When an FM signal is passed through a frequency multiplier, the deviation is multiplied along with the carrier frequency.

The deviation in an FM signal should be equal to the highest modulating audio frequency if optimum fidelity is to be obtained. Thus,  $\pm 5$  kHz is more than enough for voice. For music, a deviation of  $\pm 15$  kHz or even  $\pm 20$  kHz is required for good reproduction when the signal is received.

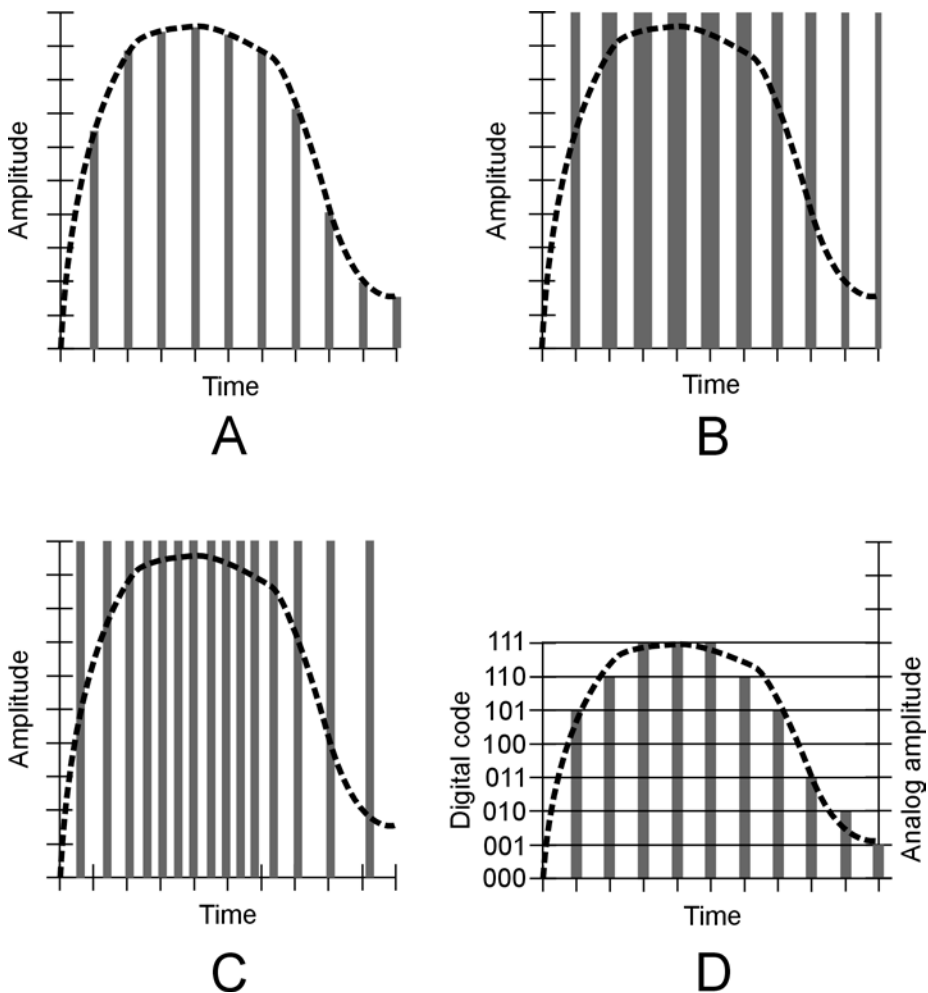


25-7 Generation of FM by reactance modulation of a Colpitts oscillator. Other oscillator types can be similarly modified.

In any FM signal, the ratio of the frequency deviation to the highest modulating audio frequency is called the *modulation index*. Ideally, this figure is between 1:1 and 2:1. If it is less than 1:1, the signal sounds muffled or distorted, and efficiency is sacrificed. Increasing it beyond 2:1 broadens the bandwidth without providing significant improvement in intelligibility or fidelity.

## Pulse Modulation

Another method of modulation works by varying some aspect of a constant stream of signal pulses. Several types of *pulse modulation* (PM) are briefly described in the following sections. They are diagrammed in Fig. 25-8 as amplitude-versus-time graphs. The modulating waveform in each case is shown as a dashed curve, and the pulses are shown as vertical gray bars.



**25-8** Time-domain graphs of various modes of pulse modulation. At A, pulse amplitude modulation; at B, pulse width modulation; at C, pulse interval modulation; at D, pulse code modulation.

## Pulse Amplitude Modulation

In *pulse amplitude modulation* (PAM), the strength of each individual pulse varies according to the modulating waveform. In this respect, PAM resembles AM. An amplitude-versus-time graph of a hypothetical PAM signal is shown in Fig. 25-8A. Normally, the pulse amplitude increases as the instantaneous modulating-signal level increases (*positive PAM*). But this can be reversed, so higher audio levels cause the pulse amplitude to go down (*negative PAM*). Then the signal pulses are at their strongest when there is no modulation. The transmitter works a little harder if negative PAM is used.

## Pulse Width Modulation

Another way to change the transmitter output is to vary the width (duration) of the pulses. This is called *pulse width modulation* (PWM) or *pulse duration modulation* (PDM), and is shown in Fig. 25-8B. Normally, the pulse width increases as the instantaneous modulating-signal level increases (*positive PWM*). But this can be reversed (*negative PWM*). The transmitter must work harder to accomplish negative PWM. Either way, the peak pulse amplitude remains constant.

## Pulse Interval Modulation

Even if all the pulses have the same amplitude and the same duration, modulation can still be accomplished by varying how often they occur. In PAM and PWM, the pulses are always sent at the same time interval, known as the *sampling interval*. But in *pulse interval modulation* (PIM), also called *pulse frequency modulation* (PFM), pulses can occur more or less frequently than they do when there is no modulation. A hypothetical PIM signal is shown in Fig. 25-8C. Every pulse has the same amplitude and the same duration, but the time interval between them changes. When there is no modulation, the pulses are evenly spaced with respect to time. An increase in the instantaneous data amplitude might cause pulses to be sent more often, as is the case in Fig. 25-8C (*positive PIM*). Or, an increase in instantaneous data level might slow down the rate at which the pulses are sent (*negative PIM*).

## Pulse Code Modulation

In recent years, the transmission of data has been done more and more by *digital* means. In *digital communications*, the modulating data attains only certain defined states, rather than continuously varying. Digital transmission offers better efficiency than analog transmission. With digital modes, the *signal-to-noise* (S/N) *ratio* is better, the bandwidth is narrower, and there are fewer errors. In *pulse-code modulation* (PCM), any of the above aspects—amplitude, duration, or interval—of a pulse sequence (or *pulse train*) can be varied. But rather than having infinitely many possible states, there are finitely many. The number of states is a power of 2, such as 4, 8, or 16. The greater the number of states, the better the fidelity. An example of 8-level PCM is shown in Fig. 25-8D.

# Analog-to-Digital Conversion

Pulse code modulation, such as is shown at Fig. 25-8D, is one form of *analog-to-digital* (A/D) *conversion*. A voice signal, or any continuously variable signal, can be *digitized*, or converted into a train of pulses whose amplitudes can achieve only certain defined levels.

## Resolution

In A/D conversion, the number of states is always a power of 2, so that it can be represented as a binary-number code. Fidelity improves as the exponent increases. The number of states is called the

*sampling resolution*, or simply the *resolution*. A resolution of  $2^3 = 8$  (as shown in Fig. 25-8D) is good enough for voice transmission, and is the standard for commercial digital voice circuits. A resolution of  $2^4 = 16$  is adequate for high-fidelity (hi-fi) music reproduction.

## Sampling Rate

The efficiency with which a signal can be digitized depends on the frequency at which sampling is done. In general, the *sampling rate* must be at least twice the highest data frequency. For an audio signal with components as high as 3 kHz, the minimum sampling rate for effective digitization is 6 kHz; the commercial voice standard is 8 kHz. For hi-fi digital transmission, the standard sampling rate is 44.1 kHz, a little more than twice the frequency of the highest audible sound (approximately 20 kHz).

## Image Transmission

Nonmoving images can be sent within the same bandwidth as voice signals. For high-resolution, moving images, the necessary bandwidth is greater.

### Facsimile

Nonmoving images (also called *still images*) are commonly transmitted by *facsimile*, also called *fax*. If data is sent slowly enough, any amount of detail can be transmitted within a 3-kHz-wide band, the standard for voice communications. This is why detailed fax images can be sent over a *plain old telephone service* (POTS) line.

In an electromechanical fax machine, a paper document or photo is wrapped around a *drum*. The drum is rotated at a slow, controlled rate. A spot of light scans from left to right; the drum moves the document so a single line is scanned with each pass of the light spot. This continues, line by line, until the complete *frame* (image) has been scanned. The reflected light is picked up by a *photodetector*. Dark parts of the image reflect less light than bright parts, so the current through the photodetector varies. This current modulates a carrier in one of the modes described earlier, such as AM, FM, or SSB. Typically, black is sent as a 1.5-kHz audio sine wave, and white as a 2.3-kHz audio sine wave. Gray shades produce audio sine waves having frequencies between these extremes.

At the receiver, the scanning rate and pattern can be duplicated, and a *cathode-ray tube* (CRT), *liquid crystal display* (LCD), or printer can be used to reproduce the image in *grayscale* (shades of gray ranging from black to white, without color).

### Slow-Scan Television

One way to think of *slow-scan television* (SSTV) is to imagine “fast fax.” An SSTV signal, like a fax signal, is sent within a band of frequencies as narrow as that of a human voice. And, like fax, SSTV transmission is of still pictures, not moving ones. The big difference between SSTV and fax is that SSTV images are sent in much less time. The time required to send a complete frame (image or scene) is 8 seconds, rather than several minutes. This speed bonus comes with a tradeoff: lower *resolution*, meaning less image detail.

Some SSTV signals are received on CRT displays. A computer can be programmed so that its monitor will act as an SSTV receiver. Converters are also available that allow SSTV signals to be viewed on a consumer-type TV set.



An SSTV frame has 120 lines. The black and white frequencies are the same as for fax transmission; the darkest parts of the picture are sent at 1.5 kHz and the brightest at 2.3 kHz. *Synchronization (sync) pulses*, that keep the receiving apparatus in step with the transmitter, are sent at 1.2 kHz. A *vertical sync pulse* tells the receiver that it's time to begin a new frame; it lasts for 30 milliseconds (ms). A *horizontal sync pulse* tells the receiver that it's time to start a new line in a frame; its duration is 5 ms. These pulses prevent *rolling* (haphazard vertical image motion) or *tearing* (lack of horizontal synchronization).

## Fast-Scan Television

Conventional television is also known as *fast-scan TV* (FSTV). The frames are transmitted at the rate of 30 per second. There are 525 lines per frame. The quick frame time, and the increased resolution, of FSTV make it necessary to use a much wider frequency band than is the case with fax or SSTV. A typical video FSTV signal takes up 6 MHz of spectrum space, or 2000 times the bandwidth of a fax or SSTV signal.

Fast-scan TV is almost always sent using conventional AM. Wideband FM can also be used. With AM, one of the sidebands can be filtered out, leaving just the carrier and the other sideband. This mode is called *vestigial sideband* (VSB) transmission. It cuts the bandwidth of an FSTV signal down to about 3 MHz.

Because of the large amount of spectrum space needed to send FSTV, this mode isn't practical at frequencies below about 30 MHz. All commercial FSTV transmission is done above 50 MHz, with the great majority of channels having frequencies far higher than this. Channels 2 through 13 on your TV receiver are sometimes called the *very high frequency* (VHF) *channels*; the higher channels are called the *ultrahigh frequency* (UHF) *channels*.

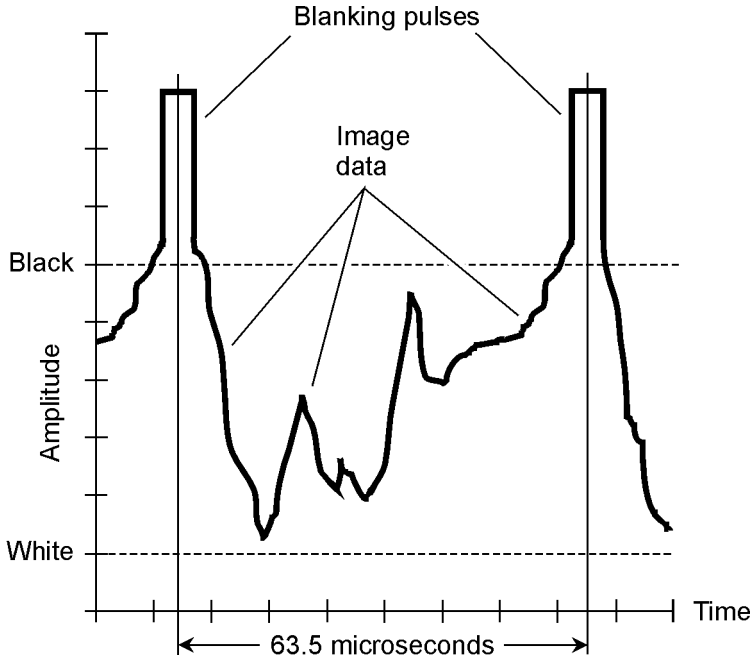
Figure 25-9 is a time-domain graph of the waveform of a single line in an FSTV video signal. This represents  $\frac{1}{525}$  of a complete frame. The highest instantaneous signal amplitude corresponds to the blackest shade, and the lowest amplitude to the lightest shade. Thus, the FSTV signal is sent negatively. The reason that FSTV signals are sent this way is that *retracing* (moving from the end of one line to the beginning of the next) must be synchronized between the transmitter and receiver. This is guaranteed by a defined, strong *blanking pulse*. This pulse tells the receiver when to retrace; it also shuts off the beam while the receiver display is retracing. Have you noticed that weak TV signals have poor *contrast*? (You have, if you're old enough to remember "rabbit ears"!) Weakened blanking pulses result in incomplete retrace blanking. But this is better than having the TV receiver completely lose track of when it should retrace.

Color FSTV works by sending three separate monochromatic signals, corresponding to the primary colors red, blue, and green. The signals are literally black-and-red, black-and-blue, and black-and-green. These are recombined at the receiver and displayed on the screen as a fine, interwoven matrix of red, blue, and green dots. When viewed from a distance, the dots are too small to be individually discernible. Various combinations of red, blue, and green intensities result in reproduction of all possible hues and saturations of color.

## High-Definition Television

The term *high-definition television* (HDTV) refers to any of several similar methods for getting more detail into a TV picture, and for obtaining better audio quality, compared with standard FSTV.

A standard FSTV picture has 525 lines per frame, but HDTV systems have between 787 and 1125 lines per frame. The image is scanned about 60 times per second. High-definition TV is often



25-9 Time-domain graph of a single line in an FSTV video frame.

sent in a digital mode; this offers another advantage over conventional FSTV. Digital signals propagate better, are easier to deal with when they are weak, and can be processed in ways that analog signals cannot.

Some HDTV systems use *interlacing* in which two *rasters* are meshed together. This effectively doubles the image resolution without doubling the cost of the hardware. But it can cause annoying *jitter* in fast-moving or fast-changing images.

## Digital Satellite TV

Until the early 1990s, a satellite television installation required a dish antenna several feet in diameter. A few such systems are still in use. The antennas are expensive, they attract attention (sometimes unwanted), and they are subject to damage from ice storms, heavy snows, and high winds. Digitization has changed this situation. In any communications system, digitization allows the use of smaller receiving antennas, smaller transmitting antennas, and/or lower transmitter power levels. Engineers have managed to get the diameter of the receiving dish down to about 2 ft.

A pioneer in digital TV was RCA (Radio Corporation of America), which developed the *Digital Satellite System* (DSS). The analog signal is changed into digital pulses at the transmitting station via A/D conversion. The digital signal is amplified and sent up to a geostationary satellite. The satellite has a *transponder* that receives the signal, converts it to a different frequency, and retransmits it back toward the earth. The return signal is picked up by a portable dish. A *tuner* selects the channel. *Digital signal processing* (DSP) can be used to improve the quality of reception under marginal conditions. The digital signal is changed back into analog form, suitable for viewing on a conventional FSTV set, by means of *digital-to-analog* (D/A) conversion.

# The Electromagnetic Field

In a radio or television transmitting antenna, electrons are moving back and forth at an extreme speed. Their velocity is constantly changing as they speed up in one direction, slow down, reverse direction, speed up again, and so on. Any change of velocity (that is, of speed and/or direction) constitutes *acceleration*.

## How It Happens

When electrons move, a magnetic (M) field is produced. When electrons accelerate, a changing magnetic field is produced. An alternating M field gives rise to an alternating electric (E) field, and this generates another alternating M field. This process repeats over and over, endlessly, and the effect *propagates* (travels) through space at the speed of light. The E and M fields expand alternately outward from the source in spherical wavefronts. At any given point in space, the E flux is perpendicular to the M flux. The direction of wave travel is perpendicular to both the E and M flux lines. This is an *electromagnetic* (EM) *field*.

An EM field can have any conceivable frequency, ranging from many years per cycle to quadrillions of cycles per second. The sun has a magnetic field that oscillates with a 22-year cycle. Radio waves oscillate at thousands, millions, or billions of cycles per second. Infrared (IR), visible light, ultraviolet (UV), X rays, and gamma rays are EM fields that alternate at many trillions (million millions) of cycles per second.

## Frequency versus Wavelength

All EM fields have two important properties: the *frequency* and the *wavelength*. These are inversely related. You've already learned about frequency. Wavelength, for an EM field, is a rather sophisticated concept. It is measured between any two adjacent points on the wave at which the E and M fields have exactly the same amplitudes, and occur in exactly the same relative directions. The following equations relate the frequency and the wavelength of an EM field in free space (the air or a vacuum).

Let  $f_{\text{MHz}}$  be the frequency of an EM wave in megahertz, and  $L_{\text{ft}}$  be the wavelength in feet. Then the two are related as follows:

$$L_{\text{ft}} = 984 / f_{\text{MHz}}$$

If the wavelength is given as  $L_{\text{m}}$  in meters, then

$$L_{\text{m}} = 300 / f_{\text{MHz}}$$

The inverses of these formulas, for finding the frequency if the wavelength is known, are

$$\begin{aligned} f_{\text{MHz}} &= 984 / L_{\text{ft}} \\ f_{\text{MHz}} &= 300 / L_{\text{m}} \end{aligned}$$

## Velocity Factor

In media other than free space, the speed at which EM fields propagate is slower than the speed of light. As a result, wavelength is shortened by a factor known as the *velocity factor*, symbolized  $v$ . The value of  $v$  can be anything between 0 (representing zero speed of propagation) and 1 (representing the speed of propagation in free space, which is approximately 186,000 mi/s or 300,000 km/s). The velocity factor can also be expressed as a percentage  $v_{\%}$ . In that case, the smallest possible value is 0 percent, and the largest is 100 percent. The velocity factor in practical situations is rarely less than about 0.60, or 60 percent.

Velocity factor is important in the design of RF transmission lines and antenna systems, when sections of cable, wire, or metal tubing must be cut to specific lengths measured in wavelengths or fractions of a wavelength. Taking the velocity factor  $v$ , expressed as a ratio, into account, the preceding four formulas become:

$$L_{ft} = 984v / f_{MHz}$$

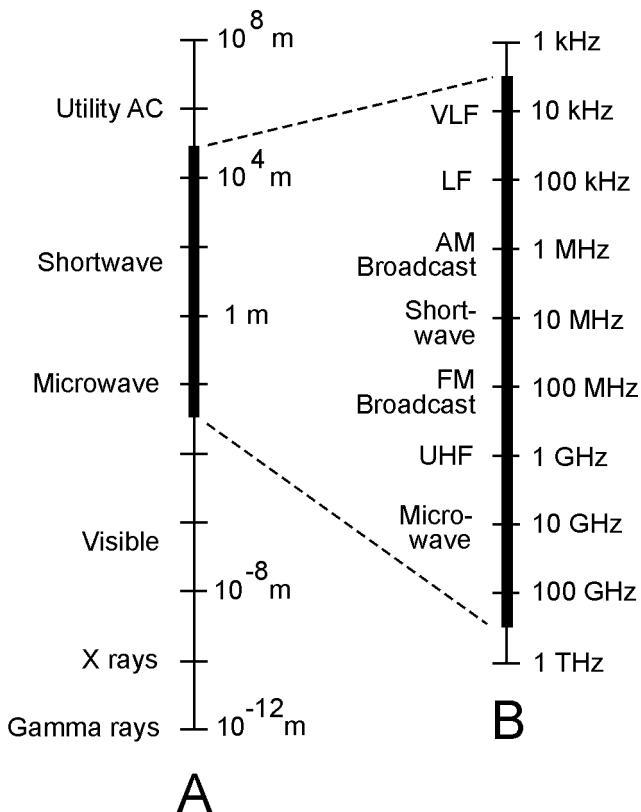
$$L_m = 300v / f_{MHz}$$

$$f_{MHz} = 984v / L_{ft}$$

$$f_{MHz} = 300v / L_m$$

### The Electromagnetic Spectrum

The entire range of EM wavelengths is called the *electromagnetic (EM) spectrum*. Scientists use logarithmic scales to depict the EM spectrum, as shown in Fig. 25-10. The *RF spectrum*, which includes radio, television, and microwaves, is blown up in this illustration, and is labeled for frequency.



**25-10** At A, the EM spectrum from  $10^8$  m to  $10^{-12}$  m, with each vertical division representing two orders of magnitude (an increase or decrease of the wavelength by a factor of 100). At B, the RF spectrum, with each vertical division representing one order of magnitude (an increase or decrease of the frequency by a factor of 10).

# Wave Propagation

Radio-wave propagation has been a fascinating science ever since Marconi and Tesla discovered, around the year 1900, that EM fields can travel over long distances without any supporting infrastructure whatsoever. Let's look at a few of the factors that affect wireless communications at radio frequencies.

## Polarization

The orientation of the E field *lines of flux* is defined as the *polarization* of an EM wave. If the E field flux lines are parallel to the earth's surface, you have *horizontal polarization*. If the E field flux lines are perpendicular to the surface, you have *vertical polarization*. Polarization can also be slanted, of course.

In some situations, the E flux lines rotate as the wave travels through space. This is *circular polarization* if the E field intensity remains constant. If the E field intensity is more intense in some planes than in others, the polarization is *elliptical*. Rotating polarization can be *clockwise* or *counterclockwise*, viewed as the wavefronts approach. The rotational direction is called the *sense* of polarization.

## The Line-of-Sight Wave

Electromagnetic waves follow straight lines unless something makes them bend. *Line-of-sight* propagation can take place even when the receiving antenna cannot be seen from the transmitting antenna. To some extent, radio waves penetrate nonconducting objects such as trees and frame houses. The line-of-sight wave consists of two components: the *direct wave* and the *reflected wave*.

*The direct wave:* The longest wavelengths are least affected by obstructions. At very low, low, and medium frequencies, direct waves can *diffract* around things. As the frequency rises, especially above about 3 MHz, obstructions have a greater and greater blocking effect.

*The reflected wave:* Electromagnetic waves reflect from the earth's surface and from conducting objects like wires and steel beams. The reflected wave always travels farther than the direct wave. The two waves are usually not in phase at the receiving antenna. If they're equally strong but 180° out of phase, a *dead spot* occurs. This phenomenon is most noticeable at the highest frequencies. At VHF and UHF, an improvement in reception can sometimes result from moving the transmitting or receiving antenna just a few inches. In mobile operation, when the transmitter and/or receiver are moving, dead spots produce rapid, repeated interruptions in the received signal. This is called *picket fencing*.

## The Surface Wave

At frequencies below about 10 MHz, the earth's surface conducts ac quite well. Because of this, vertically polarized EM waves follow the surface for hundreds or even thousands of miles, with the earth actually helping to transmit the signals. The lower the frequency, the lower the *ground loss*, and the farther the waves can travel by *surface-wave propagation*. Horizontally polarized waves do not travel well in this mode, because horizontal E field flux is shorted out by the earth. Above about 10 MHz, the earth becomes lossy, and surface-wave propagation is not useful for more than a few miles.

## Sky-Wave EM Propagation

*Ionization* in the upper atmosphere, caused by solar radiation, can return EM waves to the earth at certain frequencies. The ionization takes place at three or four distinct layers.

The lowest ionized region is called the *D layer*. It exists at an altitude of about 30 mi (50 km), and is ordinarily present only on the daylight side of the planet. This layer absorbs radio waves at some frequencies, impeding long-distance ionospheric propagation.

The *E layer*, about 50 mi (80 km) above the surface, also exists mainly during the day, although nighttime ionization is sometimes observed. The E layer can provide medium-range radio communication at certain frequencies.

The uppermost layers are called the *F<sub>1</sub> layer* and the *F<sub>2</sub> layer*. The *F<sub>1</sub>* layer, normally present only on the daylight side of the earth, forms at about 125 mi (200 km) altitude; the *F<sub>2</sub>* layer exists at about 180 mi (300 km) over most, or all, of the earth. Sometimes the distinction between the *F<sub>1</sub>* and *F<sub>2</sub>* layers is ignored, and they are spoken of together as the *F layer*. Communication by means of F-layer propagation can usually be accomplished between any two points on the earth at some frequencies between 5 MHz and 30 MHz.

## **Tropospheric Propagation**

At frequencies above about 30 MHz, the lower atmosphere bends radio waves toward the surface. *Tropospheric banding* occurs because the *index of refraction* of air, with respect to EM waves, decreases with altitude. The effect is similar to the way sound waves sometimes travel long distances over the surface of a calm lake in the early morning or early evening, letting you hear a conversation more than a mile away. Tropospheric propagation makes it possible to communicate for hundreds of miles when the ionosphere will not return waves to the earth.

Another type of tropospheric propagation is called *ducting*. It takes place when EM waves are trapped in a layer of cool, dense air sandwiched between two layers of warmer air. Like bending, ducting occurs almost entirely at frequencies above 30 MHz.

Still another tropospheric-propagation mode is known as *troposcatter*. This takes place because air molecules, dust grains, and water droplets scatter some of the EM field. This effect is commonly seen at VHF and UHF.

Tropospheric propagation in general, without mention of the specific mode, is sometimes called *tropo*.

## **Auroral Propagation**

In the presence of unusual solar activity, the *aurora* (northern lights or southern lights) can return radio waves to the earth. This is known as *auroral propagation*. The aurora occur at altitudes of about 40 to 250 mi (65 to 400 km). Theoretically, auroral propagation is possible, when the aurora are active, between any two points on the surface from which the same part of the aurora lie on a line of sight. Auroral propagation seldom occurs when one end of the circuit is at a latitude less than 35° north or south of the equator.

Auroral propagation is characterized by rapid and deep fading. This almost always renders analog voice and video signals unintelligible. Digital modes are most effective for communication via auroral propagation, but the carrier is often spread out over several hundred hertz as a result of phase modulation induced by auroral motion. This severely limits the maximum data transfer rate. Auroral propagation is often accompanied by deterioration in ionospheric propagation.

## **Meteor Scatter**

Meteors produce ionized trails that persist for approximately 0.5 s up to several seconds, depending on the size of a particular meteor, its speed, and the angle at which it enters the atmosphere. This is not enough time for the transmission of much data, but during a *meteor shower*, multiple trails can result in almost continuous ionization for a period of hours. Such ionized regions reflect radio waves at certain frequencies. This is *meteor scatter propagation*. It can take place at frequencies considerably

above 30 MHz, and occurs over distances ranging from just beyond the horizon up to about 1500 mi (2400 km), depending on the altitude of the ionized trail and the relative positions of the trail, the transmitting station, and the receiving station.

### Moonbounce

The moon, like the earth, reflects EM fields. This makes it possible to communicate by means of *earth-moon-earth* (EME), also called *moonbounce*. High-powered transmitters, sophisticated antenna systems, and sensitive receivers are needed for EME. Some moonbounce communication is done by radio amateurs at frequencies from 50 MHz to over 2 GHz.

## Transmission Media

Data can be transmitted over various *media*. The most common are *cable*, *radio* (also called *wireless*), *satellite links* (a specialized form of wireless), and *fiber optics*. Cable, radio/TV, and satellite communications use the RF spectrum. Fiber optics uses IR or visible light energy.

### Cable

The earliest cables were simple wires that carried dc. Nowadays, data transmission cables more often carry ac at radio frequencies. One advantage of using RF is the fact that the signals can be amplified at intervals on a long span. This greatly increases the distances over which data can be sent by cable. Another advantage of using RF is the fact that numerous signals can be carried over a single cable, with each signal on a different frequency.

Cables can consist of pairs of wires, somewhat akin to lamp cords. But more often coaxial cable, of the type described and illustrated at the end of Chap. 10, is used. This has a center conductor surrounded by a cylindrical shield. The shield is grounded, and the center conductor carries the signals. The shield keeps signals confined to the cable, and also keeps external EM fields from interfering with the signals.

### Radio

All radio and TV signals are electromagnetic waves. The radio or TV transmitter output is coupled into an *antenna system* located at some distance from the transmitter. The energy follows a *transmission line*, also called a *feed line*, from the transmitter output to the antenna itself.

Most radio antenna transmission lines are coaxial cables. There are other types, used in special applications. At microwaves, hollow tubes called *waveguides* are used to transfer the energy from a transmitter to the antenna. A waveguide is more efficient than coaxial cable at the shortest radio wavelengths. Radio amateurs sometimes use *parallel-wire* transmission lines, resembling the ribbon cable popular for use with consumer TV receiving antennas. In a parallel-wire line, the RF currents in the two conductors are always 180° out of phase, so that their EM fields cancel each other. This keeps the transmission line from radiating, guiding the EM field along toward the antenna. The energy is radiated when it reaches the antenna.

The RF bands are generally categorized from *very low frequency* (VLF) through *extremely high frequency* (EHF), according to the breakdown in Table 25-1. As noted in Chap. 24, the exact lower limit of the VLF range is a matter of disagreement in the literature. In Table 25-1, it is defined as 3 kHz, which is consistent with defining the frequency boundaries between RF bands by order of magnitude.

**Table 25-1. Bands in the RF spectrum.**

Frequency designation	Frequency range	Wavelength range
Very Low (VLF)	3 kHz–30 kHz	100 km–10 km
Low (LF)	30 kHz–300 kHz	10 km–1 km
Medium (MF)	300 kHz–3 MHz	1 km–100 m
High (HF)	3 MHz–30 MHz	100 m–10 m
Very High (VHF)	30 MHz–300 MHz	10 m–1 m
Ultra High (UHF)	300 MHz–3 GHz	1 m–100 mm
Super High (SHF)	3 GHz–30 GHz	100 mm–10 mm
Extremely High (EHF)	30 GHz–300 GHz	10 mm–1 mm

## Satellite Systems

At very high frequencies (VHF) and above, many communications circuits use satellites in *geostationary orbits* around the earth. If a satellite is directly over the equator at an altitude of 22,300 mi (36,000 km) and orbits from west to east, it follows the earth's rotation, thereby staying in the same spot in the sky as seen from the surface, and is thus a *geostationary satellite*.

A single geostationary satellite is on a line of sight with about 40 percent of the earth's surface. Three such satellites, placed at 120° ( $\frac{1}{3}$  circle) intervals around the planet, allow coverage of all populated regions. A *dish antenna* can be aimed at a geostationary satellite, and once the antenna is in place, it need not be turned or adjusted.

Another form of satellite system uses multiple satellites in low orbits that take them over the earth's poles. These satellites are in continuous, rapid motion with respect to the surface. But if there are enough of them, they can act like repeaters in a cell phone network, and maintain reliable communications between any two points on the surface at all times. Directional antennas are not necessary in these systems, which are called *low earth orbit* (LEO) satellite networks.

## Fiber Optics

Beams of IR or visible light can be modulated, just as can RF carriers. The frequency of an IR or visible light beam is higher than the frequency of any RF signal, allowing modulation by data at rates higher than anything possible with radio.

Fiber-optic technology offers several advantages over wire cables (which are sometimes called *copper* because the conductors are usually made of that metallic element). A fiber-optic cable is cheap. It is light in weight. It is immune to interference from outside EM fields. A fiber-optic cable does not corrode as metallic wires do. Fiber-optic cables are inexpensive to maintain and easy to repair. An optical fiber can carry far more signals than a cable, because the frequency bands are far wider in terms of megahertz or gigahertz.

The whole RF spectrum, from VLF through EHF, can (at least in theory) be imprinted on a single beam of light and sent through an optical fiber no thicker than a strand of hair!

## Two Basic Receiver Designs

A wireless or *radio receiver* converts EM waves into the original messages sent by a distant transmitter. Let's begin our study of receivers by defining a few of the most important criteria for operation, and then we'll look at two common designs.



## Specifications

The *specifications* of a receiver indicate how well it can do the functions it is designed to perform.

*Sensitivity:* The most common way to express receiver sensitivity is to state the number of microvolts that must exist at the antenna terminals to produce a certain *signal-to-noise (S/N) ratio* or *signal-plus-noise-to-noise (S+N/N) ratio* in decibels (dB). Sensitivity is related to the gain of the *front end* (the amplifier or amplifiers connected to the antenna), but the amount of noise this stage generates is more significant, because subsequent stages amplify the front-end noise output as well as the signal output.

*Selectivity:* The *passband*, or bandwidth that the receiver can hear, is established by a wideband *preselector* in the early RF amplification stages, and is honed to precision by narrowband filters in later amplifier stages. The preselector makes the receiver optimally sensitive within a range of approximately plus-or-minus 10 percent ( $\pm 10\%$ ) of the desired signal frequency. The narrowband filter responds only to the frequency or channel of a specific signal to be received; signals in nearby channels are rejected.

*Dynamic range:* The signals at a receiver input vary over several orders of magnitude (multiples or powers of 10) in terms of absolute voltage. Dynamic range is the ability of a receiver to maintain a fairly constant output, and yet to maintain its rated sensitivity, in the presence of signals ranging from very weak to very strong. The dynamic range in a good receiver is in excess of 100 dB.

*Noise figure:* The less internal noise a receiver produces, in general, the better is the S/N ratio. Excellent S/N ratio in the presence of weak signals is only possible when the *noise figure*, a measure of internally generated receiver noise, is low. This is paramount at VHF, UHF, and microwave frequencies. Gallium-arsenide field effect transistors (GaAsFETs) are well known for the low levels of noise they generate, even at quite high frequencies. Other types of FETs can be used at lower frequencies. *Bipolar transistors* tend to be rather noisy.

## Direct-Conversion Receiver

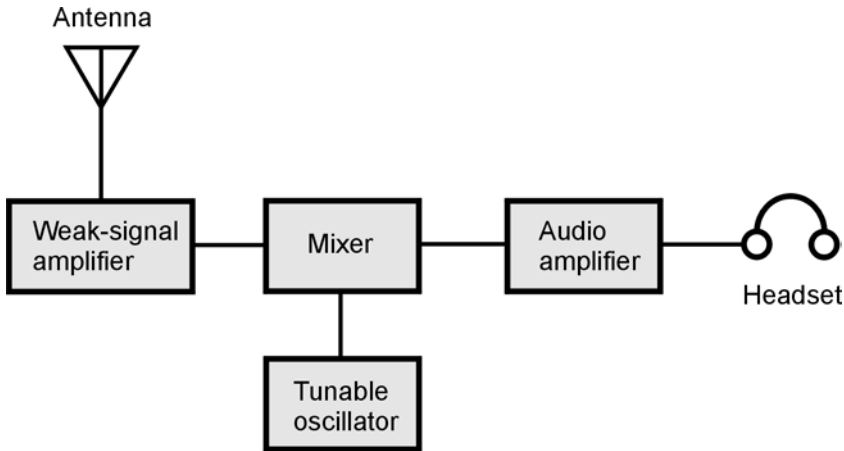
A *direct-conversion receiver* derives its output by mixing incoming signals with the output of a tunable (that is, variable frequency) *local oscillator* (LO). The received signal is fed into a mixer, along with the output of the LO. Figure 25-11 is a block diagram of a direct-conversion receiver.

For the reception of on/off keyed Morse code, also called *radiotelegraphy* or *continuous wave* (CW), the LO, also called a *beat-frequency oscillator* (BFO), is set a few hundred hertz above or below the signal frequency. This can also be done in order to receive FSK signals. The audio output has a frequency equal to the difference between the LO and incoming carrier frequencies. For reception of AM or SSB signals, the LO is set to precisely the same frequency as that of the signal carrier. This condition is known as *zero beat* because the *beat frequency*, or difference frequency, between the LO and the signal carrier is equal to zero.

A direct-conversion receiver provides rather poor *selectivity*. That means it can't separate incoming signals very well when they are close together in frequency. This is because signals on either side of the LO frequency can be heard at the same time. A *selective filter* can theoretically eliminate this. Such a filter must be designed for a fixed frequency if it is to work well. But in a direct-conversion receiver, the RF amplifier works over a wide range of frequencies.

## Superheterodyne Receiver

A *superheterodyne receiver*, also called a *superhet*, uses one or more local oscillators and mixers to obtain a constant-frequency signal. A fixed-frequency signal is more easily processed than a signal that



25-11 Block diagram of a direct-conversion receiver.

changes in frequency. The incoming signal is first passed through a tunable, sensitive front end. The output of the front end is mixed with the signal from a tunable, unmodulated LO. Either the sum or the difference signal is amplified. This is the *first intermediate frequency* (IF), which can be filtered to obtain a high degree of selectivity.

If the first IF signal is detected, the radio is a *single-conversion receiver*. Some receivers use a second mixer and second LO, converting the first IF to a lower-frequency *second IF*. This is a *double-conversion receiver*. The IF bandpass filter can be constructed for use on a fixed frequency, allowing superior selectivity and facilitating adjustable bandwidth. The sensitivity is enhanced because fixed IF amplifiers are easy to keep in tune.

A superheterodyne receiver can intercept or generate unwanted signals. False signals external to the receiver are called *images*; internally generated signals are called *birdies*. If the LO frequencies are carefully chosen, images and birdies do not cause problems during ordinary operation of the receiver.

Figure 25-12 is a block diagram of a generic single-conversion superheterodyne receiver. (Individual receiver designs vary somewhat.) Here's what each stage does.

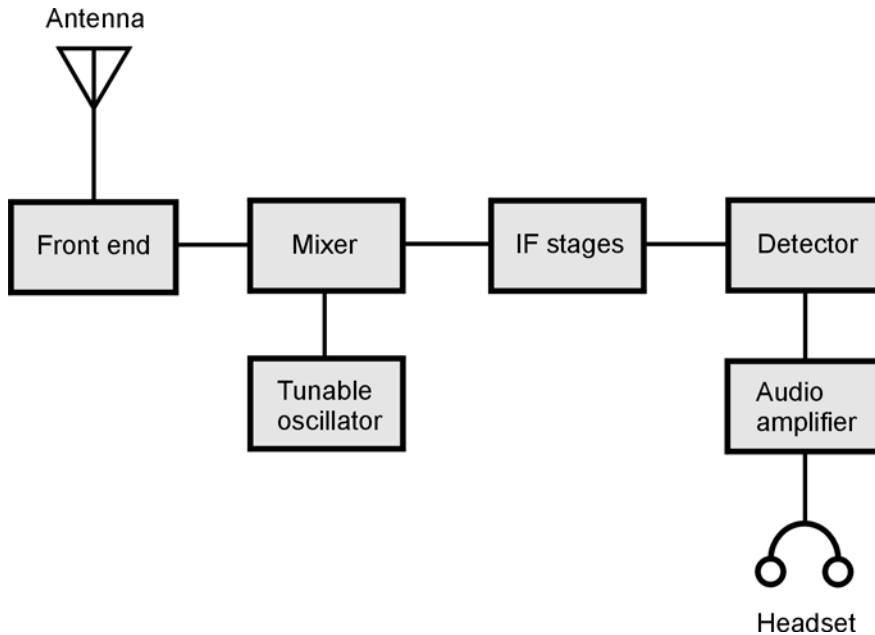
*Front end:* The front end consists of the first RF amplifier, and often includes *LC* bandpass filters between the amplifier and the antenna. The dynamic range and sensitivity of a receiver are determined by the performance of the front end.

*Mixer:* A mixer stage converts the variable signal frequency to a constant IF. The output is either the sum or the difference of the signal frequency and the tunable LO frequency.

*IF stages:* The IF stages are where most of the gain takes place. These stages are also where optimum selectivity is obtained.

*Detector:* The detector extracts the information from the signal. Common circuits are the *envelope detector* for AM, the *product detector* for SSB, FSK, and CW, and the *ratio detector* for FM.

*Audio amplifier:* Following the detector, one or two stages of audio amplification are employed to boost the signal to a level suitable for a speaker or headset. Alternatively, the signal can be fed to a printer, facsimile machine, or computer.



25-12 Block diagram of a single-conversion superheterodyne receiver.

## Predetector Stages

In a superhet, the stages preceding the first mixer must be designed so they provide reasonable gain, but produce as little noise as possible. They must also be capable of handling strong signals without *desensitization* (losing gain), also known as *overloading*.

### Preamplifier

All preamplifiers operate in class A, and most employ FETs. An FET has a high input impedance that is ideally suited to weak-signal work. Figure 25-13 shows a simple RF preamplifier circuit. Input tuning reduces noise and provides some selectivity. This circuit produces 5 dB to 10 dB gain, depending on the frequency and the choice of FET.

It is important that the preamplifier be linear, and that it remain linear in the presence of strong input signals. Nonlinearity results in unwanted mixing in RF amplifiers. The *mixing products* produce *intermodulation distortion* (IMD), or *intermod*. That can wreak havoc in a receiver, producing numerous false signals. It also degrades the S/N ratio by generating *hash*, the result of complex mixing of many false signals over a wide range of frequencies.

### The Front End

At low and medium frequencies, there is considerable atmospheric noise, and the design of a front-end circuit is simple. Above 30 MHz, atmospheric noise diminishes, and the main factor that limits the sensitivity is noise generated within the receiver. For this reason, front-end design becomes increasingly critical as the frequency rises through the VHF, UHF, and microwave spectra.

The front end, like a preamplifier, must be as linear as possible; the greater the degree of nonlinearity, the more susceptible the circuit is to the generation of mixing products. The front end should also have the greatest possible dynamic range.

### Preselector

The preselector provides a bandpass response that improves the S/N ratio, and reduces the likelihood of receiver overloading by a strong signal far removed from the operating frequency. The preselector provides *image rejection* in a superheterodyne circuit. Most preselectors have a 3-dB bandwidth that is a few percent of the received frequency.

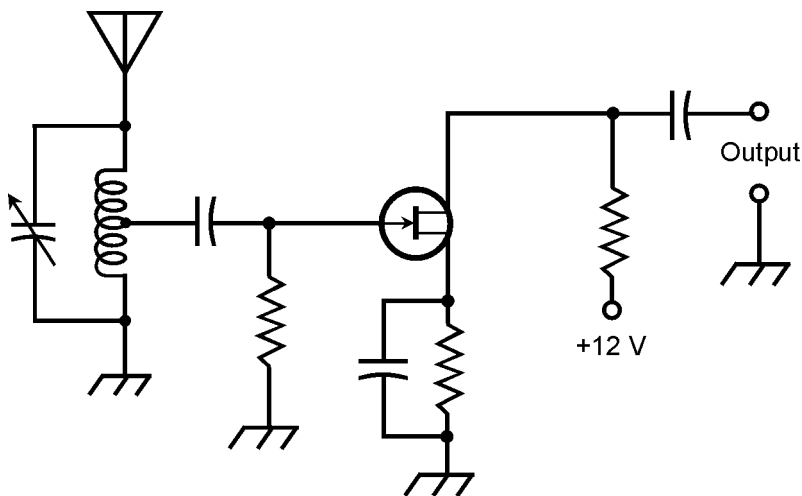
A preselector can be tuned by means of *tracking* with the tuning dial, but this requires careful design and alignment. Some receivers incorporate preselectors that must be adjusted independently of the receiver tuning.

### IF Chain

A high IF (several megahertz) is preferable to a low IF (less than 1 MHz) for image rejection. But a low IF is better for obtaining good selectivity. Double-conversion receivers have a comparatively high first IF and a low second IF to get the best of both worlds.

Intermediate-frequency amplifiers can be cascaded with tuned-transformer coupling. The amplifiers follow the mixer and precede the detector. Double-conversion receivers have two chains of IF amplifiers. The first IF chain follows the first mixer and precedes the second mixer, and the second IF chain follows the second mixer and precedes the detector.

The selectivity of the IF chain in a superheterodyne receiver can be expressed mathematically. The bandwidths are compared for two power-attenuation values, usually 3 dB and 30 dB. This gives an indication of the shape of the bandpass response. The ratio of the 30-dB selectivity to the 3-dB selectivity is called the *shape factor*. A *rectangular response* is desirable in most applications. The smaller the shape factor, the more rectangular the response.



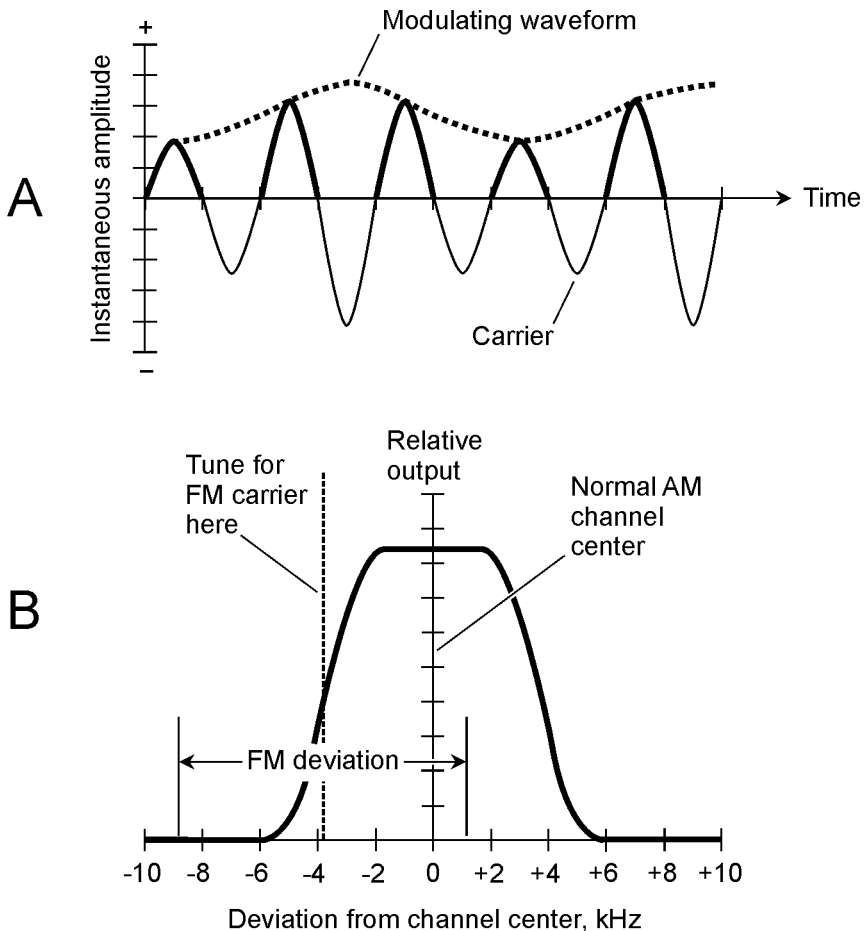
25-13 A tunable FET preamplifier for use in a radio receiver.

# Detectors

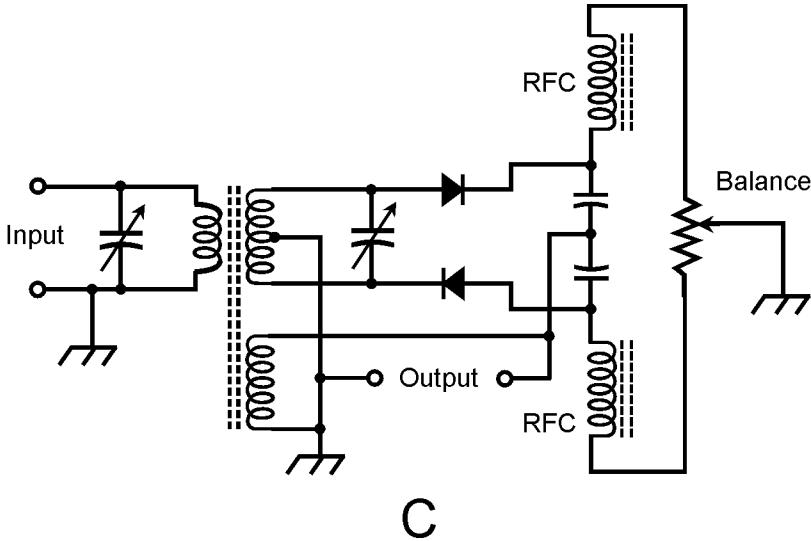
*Detection*, also called *demodulation*, is the recovery of information such as audio, images, or printed data from a signal.

## Detection of AM

The modulating waveform can be extracted from an AM signal by rectifying the carrier wave. A simplified time-domain view of this is shown in Fig. 25-14A. The rapid pulsations occur at the car-



**25-14** At A, envelope detection of AM, shown in the time domain. At B, slope detection of FM, shown in the frequency domain.



25-14 At C, a ratio detector circuit for demodulating FM signals.

rier frequency; the slower fluctuation is a duplication of the modulating data. The carrier pulsations are smoothed out by passing the output through a capacitor large enough to hold the charge for one carrier current cycle, but not so large that it smoothes out the cycles of the modulating signal. This scheme is known as *envelope detection*.

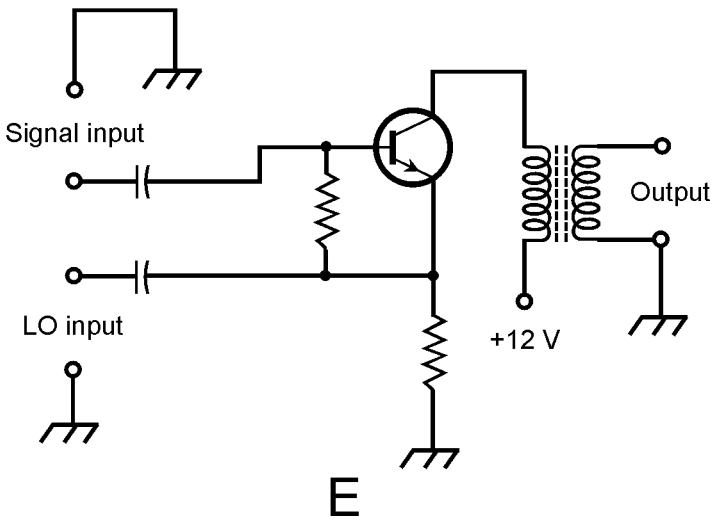
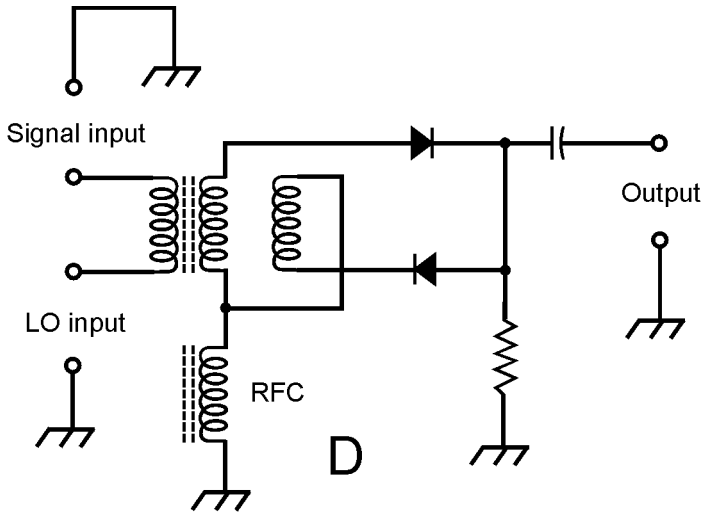
### Detection of CW

For detection of CW signals, it is necessary to inject a signal into the receiver a few hundred hertz from the carrier. The injected signal is produced by a tunable *beat-frequency oscillator* (BFO). The BFO signal and the desired CW signal are mixed, or *heterodyned*, to produce audio output at the difference frequency. The BFO is tuned to a frequency that results in a comfortable listening pitch, usually 500 to 1000 Hz. This is called *heterodyne detection*.

### Detection of FSK

The detection of FSK signals can be done using the same method as CW detection. The carrier beats against the BFO in the mixer, producing an audio tone that alternates between two different pitches.

With FSK, the BFO frequency is set a few hundred hertz above or below both of the carrier frequencies—that is, of both the *mark frequency* and the *space frequency*. The *frequency offset*, or difference between the BFO and signal frequencies, determines the audio output frequencies, and must



25-14 At D, a product detector using diodes. At E, a product detector using an NPN bipolar transistor biased as a class B amplifier. Both of these circuits can also be used as signal mixers.

be set so certain standard tone pitches result (such as 2125 Hz and 2295 Hz in the case of 170-Hz shift). Unlike the situation with CW reception, there is little tolerance for BFO adjustment variation or error.

## Detection of FM

Frequency-modulated (FM) signals can be detected in various ways. These methods also work for phase modulation.

*Slope detection:* An AM receiver can detect FM in a crude manner by setting the receiver frequency near, but not on, the FM unmodulated-carrier frequency. An AM receiver has a filter with a passband of a few kilohertz, having a selectivity curve such as that shown in Fig. 25-14B. If the FM unmodulated-carrier frequency is near either edge, or *skirt*, of the filter response, frequency variations in the incoming signal cause it to swing in and out of the receiver passband. This causes the instantaneous receiver output to vary. The relationship between the instantaneous FM deviation and the instantaneous output amplitude is not linear, however, because the skirt of the passband is not a straight line, as is apparent in the figure. The result is an unnatural-sounding received signal.

*PLL:* If an FM signal is injected into a PLL, the loop produces an error voltage that is a duplicate of the modulating waveform. A *limiter*, which keeps the signal amplitude from varying, can be placed ahead of the PLL so the receiver does not respond to AM. Weak signals tend to abruptly appear and disappear, rather than fading, in an FM receiver that employs limiting.

*Discriminator:* This type of FM detector produces an output voltage that depends on the instantaneous signal frequency. When the signal is at the center of the passband, the output voltage is zero. If the frequency falls below center, the output voltage becomes positive. If the frequency rises above center, the output becomes negative. The relationship between the instantaneous FM deviation and the instantaneous output amplitude is linear, so the output is a faithful reproduction of the incoming signal data. A discriminator is sensitive to amplitude variations, but this can be overcome by a limiter.

*Ratio detector:* This type of FM detector is a discriminator with a built-in limiter. The original design was developed by RCA (Radio Corporation of America), and is used in high-fidelity receivers and in the audio portions of TV receivers. A simple ratio detector circuit is shown in Fig. 25-14C. The balance potentiometer should be adjusted for the best received signal quality.

## Detection of SSB

For reception of SSB signals, a *product detector* is preferred, although a direct-conversion receiver can also do the job. A product detector also works well for the reception of CW and FSK. The incoming signal combines with the output of an unmodulated LO, producing audio or video. Product detection is done at a single frequency, rather than at a variable frequency as in direct-conversion reception. The single, constant frequency is obtained by mixing the incoming signal with the output of the LO.

Two product-detector circuits, which are also representative of the mixers used in superhet receivers, are shown in Fig. 25-14D and E. At D, diodes are used; there is no amplification. At E, a bipolar transistor is employed; this circuit provides some gain. The essential characteristic of either circuit is the nonlinearity of the semiconductor devices. This generates the sum and difference frequency signals that result in audio or video output.



## Audio Stages

Enhanced selectivity can be obtained by tailoring the frequency response in the AF amplifier stages following the detector, in addition to the RF selectivity provided in the IF amplifier stages preceding the detector.

### Filtering

A voice signal requires a band of frequencies ranging from about 300 Hz to 3000 Hz. An *audio bandpass filter*, with a passband of 300 Hz to 3000 Hz, can improve the quality of reception with some voice receivers. An ideal voice audio filter has little or no attenuation within the passband range, and high attenuation outside the range, with a near-rectangular response.

A CW or FSK signal requires only a few hundred hertz of bandwidth to be clearly read. Audio CW filters can narrow the response bandwidth to 100 Hz or less, but passbands narrower than about 100 Hz produce *ringing*, degrading the quality of reception. With FSK, the bandwidth of the filter must be at least as large as the difference (shift) between mark and space, but it need not, and should not, be much greater.

An *audio notch filter* is a *band-rejection filter* with a sharp, narrow response. An interfering carrier that produces a tone of constant frequency in the receiver output can be greatly attenuated with this type of filter. Audio notch filters are tunable from at least 300 Hz to 3000 Hz. Some sophisticated AF notch filters can tune themselves automatically. When an interfering carrier appears and remains for a few tenths of a second, the notch is activated and centers itself on the audio frequency of the detected offending signal.

### Squelching

A *squelch* silences a receiver when no signal is present, allowing reception of signals when they appear. Most FM communications receivers use squelching systems. The squelch is normally closed, allowing no audio output, when no signal is present. The squelch opens, allowing everything to be heard, if the signal amplitude exceeds the *squelch threshold*, which can be adjusted by the operator.

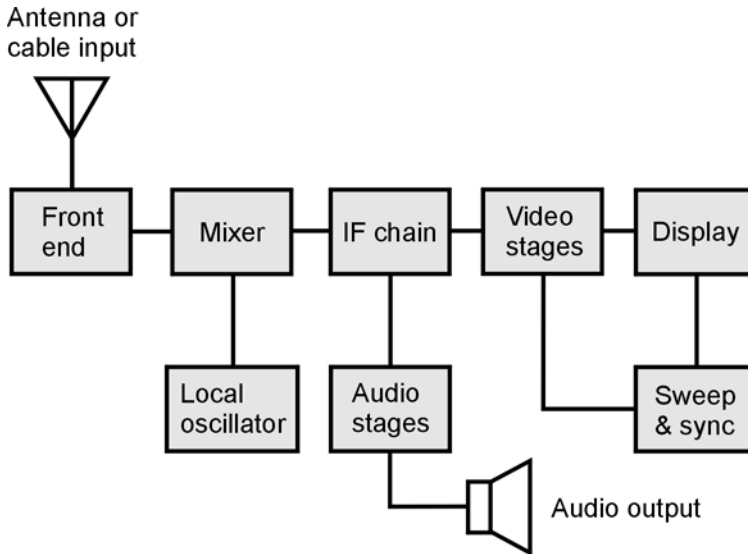
In some systems, the squelch does not open unless the signal has certain characteristics. This is known as *selective squelching*. The most common way to achieve this is the use of subaudible (below 300 Hz) tone generators, or AF tone-burst generators. The squelch opens only when an incoming signal is modulated by a tone, or sequence of tones, having the proper characteristics. This can prevent unauthorized transmissions from accessing repeaters or being picked up by receivers.

## Television Reception

A television (TV) receiver has a tunable front end, an oscillator and mixer, a set of IF amplifiers, a video demodulator, an audio demodulator and amplifier chain, a picture CRT or display with associated peripheral circuitry, and a loudspeaker.

### Fast-Scan TV

Figure 25-15 is a block diagram of a receiver for conventional analog FSTV. In the United States, conventional FSTV broadcasts are made on 68 different channels numbered from 2 through 69. Each channel is 6 MHz wide, including video and audio information. Channels 2 through 13 comprise the *VHF TV broadcast channels*. Channels 14 through 69 are the *UHF TV broadcast channels*.



25-15 Block diagram of a conventional FSTV receiver.

In digital cable television, there are more channels. The number of possible channels is virtually unlimited, because signals are not transmitted over the air and do not consume EM spectrum space.

### Slow-Scan TV

A *slow-scan television* (SSTV) communications station needs a transceiver with SSB capability, a standard TV set or personal computer, a video camera, and a *scan converter* that translates between the SSTV signal and either FSTV imagery or computer video data. The scan converter consists of two data converters (one for receiving and the other for transmitting), some digital *memory*, a *tone generator*, and a TV detector. Scan converters are commercially available. Computers can be programmed to perform this function. Some amateur radio operators build their own scan converters.

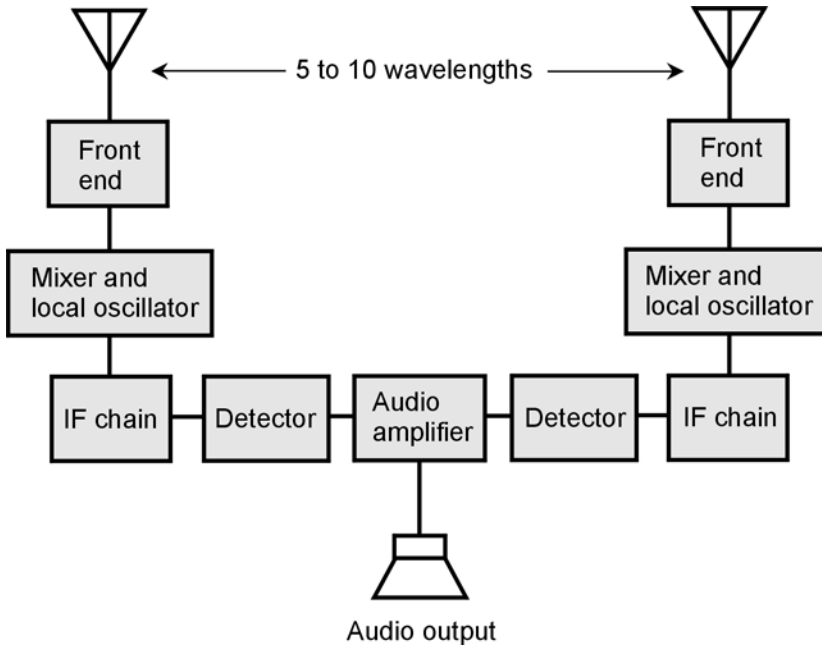
## Specialized Wireless Modes

Some less common wireless communications techniques are effective under certain circumstances.

### Dual-Diversity Reception

A *dual-diversity receiver* can reduce the fading that occurs in radio reception at high frequencies (approximately 3 to 30 MHz) when signals are propagated by means of the ionosphere. Two receivers are used. Both are tuned to the same signal, but they employ separate antennas, spaced several wavelengths apart. The outputs of the receiver detectors are fed into a common audio amplifier, as shown in Fig. 25-16.

Dual-diversity tuning is critical, and the equipment is expensive. In some installations, three or more antennas and receivers are employed. This provides superior immunity to fading, but it compounds the tuning difficulty and further increases the expense.



25-16 Block diagram of a dual-diversity radio receiver system.

## Synchronized Communications

Digital signals require less bandwidth than analog signals to convey a given amount of information per unit time. *Synchronized communications* refers to a specialized digital mode, in which the transmitter and receiver operate from a common time standard to optimize the amount of data that can be sent in a communications channel or band.

In synchronized digital communications, also called *coherent communications*, the receiver and transmitter operate in lock-step. The receiver evaluates each transmitted binary digit, or *bit*, for a block of time lasting for the specified duration of a single bit. This makes it possible to use a receiving filter having extremely narrow bandwidth. The synchronization requires the use of an external frequency/time standard. The broadcasts of standard time-and-frequency radio stations such as WWV or WWVH can be used for this purpose. Frequency dividers are employed to obtain the necessary synchronizing frequencies. A tone or pulse is generated in the receiver output for a particular bit if, but only if, the average signal voltage exceeds a certain value over the duration of that bit. False signals, such as can be caused by filter ringing, sferics, or ignition noise, are generally ignored, because they rarely produce sufficient average bit voltage.

Experiments with synchronized communications have shown that the improvement in S/N ratio, compared with nonsynchronized systems, is several decibels at low to moderate data speeds. Further improvement can be obtained by the use of DSP.

## How DSP Can Improve Reception

In DSP with analog modes such as SSB or SSTV, the signals are first changed into digital form by A/D conversion. Then the digital data is cleaned up so the pulse timing and amplitude adhere

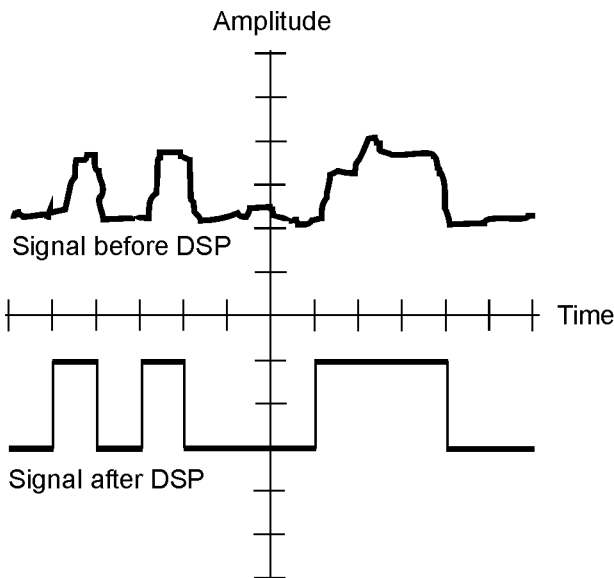
strictly to the protocol (standards) for the type of digital data being used. Finally, the digital signal is changed back to the original voice or video by D/A conversion.

Digital signal processing can extend the range of a wireless communications circuit, because it allows reception under worse conditions than would be possible without it. Digital signal processing also improves the quality of marginal signals, so that the receiving equipment or operator makes fewer errors. In circuits that use only digital modes, A/D and D/A conversion are irrelevant, but DSP can still be used to clean up the signal. This improves the accuracy of the system, and also makes it possible to copy data over and over many times (that is, to produce multigeneration duplicates).

The DSP circuit minimizes noise and interference in a received digital signal as shown in Fig. 25-17. A hypothetical signal before DSP is shown at the top; the signal after processing is shown at the bottom. If the incoming signal is above a certain level for an interval of time, the DSP output is *high* (also called *logic 1*). If the level is below the critical point for a time interval, then the output is *low* (also called *logic 0*).

## Multiplexing

Signals in a communications channel or band can be intertwined, or *multiplexed*, in various ways. The most common methods are *frequency division multiplexing* (FDM) and *time division multiplexing* (TDM). In FDM, the channel is broken down into subchannels. The carrier frequencies of the signals are spaced so they do not overlap. Each signal is independent of the others. In TDM, signals are broken into segments by time, and then the segments are transferred in a rotating sequence. The receiver must be synchronized with the transmitter by means of a time standard such as WWV. Multiplexing requires an *encoder* that combines or intertwines the signals in the transmitter, and a *decoder* that separates or untangles the signals in the receiver.



25-17 Digital signal processing can clean up a signal, improving reception.

## Spread Spectrum

In *spread-spectrum communications*, the main carrier frequency is rapidly varied independently of signal modulation, and the receiver is programmed to follow. As a result, the probability of *catastrophic interference*, in which one strong interfering signal can obliterate the desired signal, is near zero. It is difficult for unauthorized people to eavesdrop on a spread-spectrum communications link unless they gain access to the *sequencing code*, also known as the *frequency spreading function*. Such a function can be complex, and can be kept secret. If the transmitting and receiving operator do not divulge the function to anyone, and if they do not tell anyone about the existence of their contact, then no one else on the band will know the contact is taking place.

During a spread-spectrum contact between a given transmitter and receiver, the operating frequency can fluctuate over a range of kilohertz, megahertz, or tens of megahertz. As a band becomes occupied with an increasing number of spread-spectrum signals, the overall noise level in the band appears to increase. Therefore, there is a practical limit to the number of spread-spectrum contacts that a band can handle. This limit is roughly the same as it would be if all the signals were constant in frequency, and had their own discrete channels.

A common method of generating spread spectrum is *frequency hopping*. The transmitter has a list of channels that it follows in a certain order. The receiver must be programmed with this same list, in the same order, and must be synchronized with the transmitter. The *dwelt time* is the interval at which the frequency changes occur, which is the same as the length of time that the signal remains on any given frequency. The dwell time should be short enough so that a signal will not be noticed, and not cause interference, on any frequency. There are numerous *dwell frequencies*, so the signal energy is diluted to the extent that, if someone tunes to any particular frequency in the sequence, the signal is not noticeable.

Another way to get spread spectrum, called *frequency sweeping*, is to frequency-modulate the main transmitted carrier with a waveform that guides it up and down over the assigned band. This FM is independent of signal intelligence. A receiver can intercept the signal if, but only if, its tuning varies according to the same waveform, over the same band, at the same frequency, and in the same phase as that of the transmitter.

## Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. A radio wave has a frequency of 1.55 MHz. The highest modulating frequency that can be used effectively is about
  - (a) 1.55 kHz.
  - (b) 15.5 kHz.
  - (c) 155 kHz.
  - (d) 1.55 MHz.
2. The reflected wave
  - (a) arrives in phase with the direct wave.
  - (b) arrives out of phase with the direct wave.
  - (c) arrives in a variable phase compared with the direct wave.
  - (d) is always horizontally polarized.

3. An advantage of FSK over on/off keying is the fact that FSK
  - (a) offers better frequency stability.
  - (b) can provide faster data rates.
  - (c) reduces the number of receiving errors.
  - (d) Forget it! On-off keying is just as good as FSK.
  
4. The highest layer of the ionosphere is
  - (a) the D layer.
  - (b) the E layer.
  - (c) the F layer.
  - (d) none of the above.
  
5. If an AM signal is modulated with audio having frequencies up to 5 kHz, then what is the complete signal bandwidth?
  - (a) 10 kHz
  - (b) 20 kHz
  - (c) 50 kHz
  - (d) 500 kHz
  
6. An LSB, suppressed-carrier signal can be demodulated by
  - (a) an envelope detector.
  - (b) a diode.
  - (c) a ratio detector.
  - (d) a product detector.
  
7. Which of the following modes is used to send image data over telephone lines?
  - (a) On/off keying
  - (b) Fax
  - (c) AM
  - (d) Product detection
  
8. The S/N ratio is a measure of
  - (a) the sensitivity of a receiver.
  - (b) the selectivity of a receiver.
  - (c) the dynamic range of a receiver.
  - (d) the efficiency of a transmitter.
  
9. Suppose an SSB suppressed carrier is at 14.335 MHz, and audio data is contained in a band from 14.335 to 14.338 MHz. What is this mode?
  - (a) DSB
  - (b) LSB
  - (c) USB
  - (d) AFSK

10. A receiver that responds to a desired signal, but not to another signal very close by in frequency, has good
- (a) sensitivity.
  - (b) noise figure.
  - (c) dynamic range.
  - (d) selectivity.
11. Fill in the blank in the following sentence to make it true: "The deviation of a narrowband voice FM signal normally extends up to \_\_\_\_ either side of the unmodulated-carrier frequency."
- (a) 3 kHz
  - (b) 5 kHz
  - (c) 10 kHz
  - (d) 3 MHz
12. An FM detector with built-in limiting is
- (a) a ratio detector.
  - (b) a discriminator.
  - (c) an envelope detector.
  - (d) a product detector.
13. In which mode of PM does the peak power level of the pulses vary?
- (a) PAM
  - (b) PDM
  - (c) PIM
  - (d) PFM
14. A continuously variable signal (such as music audio) can be recovered from a signal having only a few discrete levels or states by means of
- (a) a ratio detector.
  - (b) a D/A converter.
  - (c) a product detector.
  - (d) an envelope detector.
15. In which mode are signals intertwined in the time domain at the transmitter, and then separated again at the receiver?
- (a) FDM
  - (b) AFSK
  - (c) PCM
  - (d) None of the above
16. Which of the following modes can be demodulated with an envelope detector?
- (a) AM
  - (b) CW

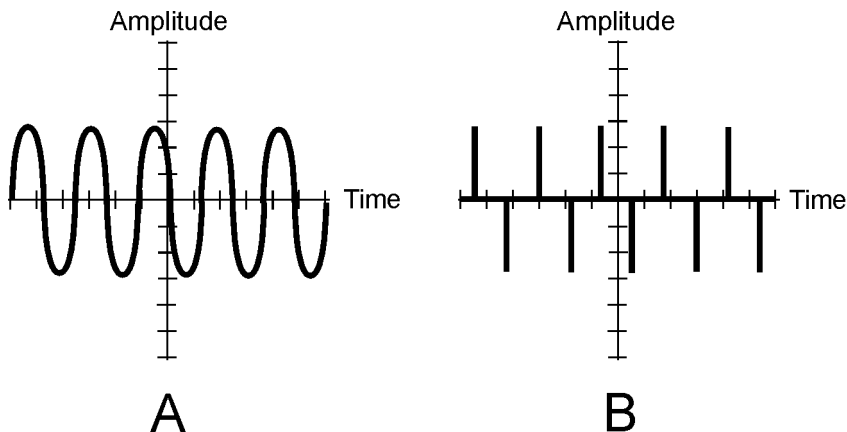
- (c) FSK
  - (d) USB
17. The bandwidth of a fax signal is kept narrow by
- (a) sending the data at a slow rate of speed.
  - (b) maximizing the number of digital states.
  - (c) optimizing the range of colors sent.
  - (d) using pulse modulation.
18. The dynamic range in a superhet is largely influenced by the performance of the
- (a) local oscillator.
  - (b) product detector.
  - (c) front end.
  - (d) selectivity in the IF chain.
19. Frequency sweeping can be used to get a transmitter to produce
- (a) spread-spectrum signals.
  - (b) time division multiplexed signals.
  - (c) narrowband AM signals.
  - (d) double sideband, suppressed carrier signals.
20. Fill in the blank to make the following sentence true: "The reception of \_\_\_\_ can be improved by the use of DSP."
- (a) SSB signals
  - (b) SSTV signals
  - (c) synchronized communications signals
  - (d) any of the above



## 26 CHAPTER

# Digital Basics

A SIGNAL IS *DIGITAL* WHEN IT CAN ATTAIN ONLY SPECIFIC LEVELS. THIS IS IN CONTRAST TO *ANALOG* signals or quantities that vary over a continuous range. A simple analog waveform is shown at Fig. 26-1A; note that the amplitude varies smoothly over time. Figure 26-1B is an example of a digital approximation of the same signal.



26-1 An analog wave (A), and a digital rendition of this wave (B).

## Numbering Systems

People are used to dealing with the *decimal number system*, which has 10 different digits. But machines use schemes that have some power of 2 digits, such as 2 ( $2^1$ ), 4 ( $2^2$ ), 8 ( $2^3$ ), 16 ( $2^4$ ), 32 ( $2^5$ ), 64 ( $2^6$ ), and so on.

## Decimal

The *decimal number system* is also called *base 10* or *radix 10*. The set of possible digits is {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}. The first digit to the left of the radix or “decimal” point is multiplied by  $10^0$ , or 1. The next digit to the left is multiplied by  $10^1$ , or 10. The power of 10 increases as you move farther to the left. The first digit to the right of the decimal point is multiplied by a factor of  $10^{-1}$ , or  $\frac{1}{10}$ . The next digit to the right is multiplied by  $10^{-2}$ , or  $\frac{1}{100}$ . This continues as you go farther to the right. Once the process of multiplying each digit is completed, the resulting values are added up. For example:

$$\begin{aligned}
 & 2 \times 10^3 \\
 & + 7 \times 10^2 \\
 & + 0 \times 10^1 \\
 & + 4 \times 10^0 \\
 & + 5 \times 10^{-1} \\
 & + 3 \times 10^{-2} \\
 & + 8 \times 10^{-3} \\
 & + 1 \times 10^{-4} \\
 & + 6 \times 10^{-5} \\
 & = 2704.53816
 \end{aligned}$$

## Binary

The *binary number system* is a method of expressing numbers using only the digits 0 and 1. It is sometimes called *base 2* or *radix 2*. The digit immediately to the left of the radix point is the “ones” digit. The next digit to the left is a “twos” digit; after that comes the “fours” digit. Moving farther to the left, the digits represent 8, 16, 32, 64, and so on, doubling every time. To the right of the radix point, the value of each digit is cut in half again and again, that is,  $\frac{1}{2}$ ,  $\frac{1}{4}$ ,  $\frac{1}{8}$ ,  $\frac{1}{16}$ ,  $\frac{1}{32}$ ,  $\frac{1}{64}$ , and so on.

Consider the decimal number 94. In the binary number system, this number is written as 1011110. It breaks down as follows:

$$\begin{aligned}
 & 0 \times 2^0 \\
 & + 1 \times 2^1 \\
 & + 1 \times 2^2 \\
 & + 1 \times 2^3 \\
 & + 1 \times 2^4 \\
 & + 0 \times 2^5 \\
 & + 1 \times 2^6 \\
 & = 1011110
 \end{aligned}$$

When you work with a computer or calculator, you give it a decimal number that is converted into binary form. The computer or calculator does its operations entirely using the digits 0 and 1. When the process is complete, the machine converts the result back into decimal form for display.

In a communications system, binary numbers can represent alphanumeric characters, shades of color, frequencies of sound, and other variable quantities.

## Octal

Another scheme, sometimes used in computer programming, is the *octal number system*, so named because it has eight symbols (according to our way of thinking), or  $2^3$ . Every digit is an element of the set {0, 1, 2, 3, 4, 5, 6, 7}. This system is also known as *base 8* or *radix 8*.

## Hexadecimal

Another system used in computer work is the *hexadecimal number system*. It has 16 ( $2^4$ ) symbols. These digits are the usual 0 through 9 plus six more, represented by A through F, the first six letters of the alphabet. The digit set is {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F}. This system is sometimes called *base 16* or *radix 16*.

# Logic

*Logic* refers to the reasoning used by electronic machines. The term is also used in reference to the circuits that make up digital devices and systems.

## Boolean Algebra

*Boolean algebra* is a system of mathematical logic using the numbers 0 and 1 with the operations AND (multiplication), OR (addition), and NOT (negation). Combinations of these operations are NAND (NOT AND) and NOR (NOT OR). This peculiar form of mathematical logic, which gets its name from the nineteenth-century British mathematician George Boole, is used in the design of digital logic circuits.

## Symbology

In Boolean algebra, the AND operation, also called *logical conjunction*, is written using an asterisk (\*), a multiplication symbol ( $\times$ ), or by running two characters together, for example,  $X*Y$ .

The NOT operation, also called *logical inversion*, is denoted by placing a tilde ( $\sim$ ) over the quantity, as a minus sign ( $-$ ) or dash ( $-$ ) followed by the quantity, as a “lazy inverted L” ( $\neg$ ) followed by the quantity, or as the quantity followed by an accent or “prime sign” ( $'$ ). An example is  $\sim X$ .

The OR operation, also called *logical disjunction*, is written using a plus sign (+), for example,  $X + Y$ .

The foregoing are the symbols used by engineers. Table 26-1A shows the values of these functions, where 0 indicates “falsity” and 1 indicates “truth.”

In mathematics and philosophy courses involving logic, you may see other symbols used for conjunction and disjunction. The AND operation in some texts is denoted by a detached arrow-

**Table 26-1A. Boolean operations.**

X	Y	$\sim X$	$X*Y$	$X+Y$
0	0	1	0	0
0	1	1	0	1
1	0	0	0	1
1	1	0	1	1

**Table 26-1B. Common theorems in Boolean algebra.**

Theorem (logic equation)	What it's called
$X+0 = X$	OR identity
$X*1 = X$	AND identity
$X+1 = 1$	
$X*0 = 0$	
$X+X = X$	
$X*X = X$	
$-(-X) = X$	Double negation
$X+(-X) = 1$	
$X*(-X) = 0$	Contradiction
$X+Y = Y+X$	Commutativity of OR
$X*Y = Y*X$	Commutativity of AND
$X+(X*Y) = X$	
$X*(-Y)+Y = X+Y$	
$X+Y+Z = (X+Y)+Z = X+(Y+Z)$	Associativity of OR
$X*Y*Z = (X*Y)*Z = X*(Y*Z)$	Associativity of AND
$X*(Y+Z) = (X*Y)+(X*Z)$	Distributivity
$-(X+Y) = (-X)*(-Y)$	DeMorgan's Theorem
$-(X*Y) = (-X)+(-Y)$	DeMorgan's Theorem

head pointing up ( $\wedge$ ) or by an ampersand (&), and the OR operation is denoted by a detached arrowhead pointing down ( $\vee$ ).

## Theorems

Table 26-1B shows several logic equations. Such facts are called *theorems*. Statements on either side of the equals sign in each case are *logically equivalent*.

When two statements are logically equivalent, it means that one is true *if and only if (iff)* the other is true. For example, the statement  $X = Y$  means that X implies Y, and also that Y implies X. Logical equivalence is sometimes symbolized by a double arrow with one or two shafts ( $\leftrightarrow$  or  $\Leftrightarrow$ ).

Boolean theorems are used to analyze and simplify complicated logic functions. This makes it possible to build a circuit to perform a specific digital function, using the smallest possible number of logic switches.

## Digital Circuits

All binary digital devices and systems employ high-speed electronic switches that perform Boolean operations. These switches are called *logic gates*. By combining logic gates, sophisticated digital systems can be built up. Even the most advanced computers are, at the basic level, comprised of logic gates.

### Positive and Negative Logic

Usually, the binary digit 1 stands for “true” and is represented by a voltage of about +5 V. The binary digit 0 stands for “false” and is represented by about 0 V. This is *positive logic*. There are

other logic forms, the most common of which is *negative logic*. In a common form of negative logic, the digit 1 (the logic high state) is represented by about 0 V, and the digit 0 (the logic low state) is represented by about +5 V. The remainder of this chapter deals with positive logic.

## Basic Gates

An *inverter* or *NOT gate* has one input and one output. It reverses the state of the input. An *OR gate* can have two or more inputs. If both, or all, of the inputs are 0 (low), then the output is 0. If any of the inputs is 1 (high), then the output is 1. An *AND gate* can have two or more inputs. If both, or all, of the inputs are 1, then the output is 1. Otherwise the output is 0.

## Other Gates

Sometimes an inverter and an OR gate are combined. This produces a *NOR gate*. If an inverter and an AND gate are combined, the result is a *NAND gate*.

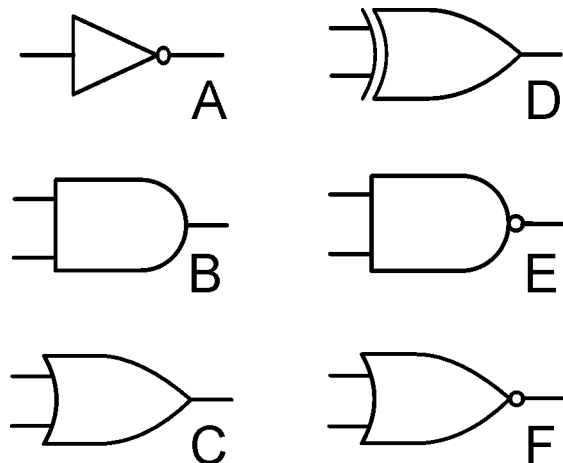
An *exclusive OR gate*, also called an *XOR gate*, has two inputs and one output. If the two inputs are the same (either both 1 or both 0), then the output is 0. If the two inputs are different, then the output is 1.

The functions of logic gates are summarized in Table 26-2. Their schematic symbols are shown in Fig. 26-2.

## Black Boxes

Logic gates can be combined to form circuits with many inputs and outputs. When two or more logic gates are combined, the outputs are always specific *logical functions* of the inputs. A complex combination of logic gates is sometimes called a *black box*.

The functions of a black box can always be determined using Boolean algebra, if the gates inside, and the way they are interconnected, is known. Conversely, if a certain complex logical function is needed for an application, a black box can be designed to perform that function by using Boolean algebra to break the function down into components of NOT, OR, AND, NOR, NAND, and XOR.



**26-2** An inverter or NOT gate (A), an AND gate (B), an OR gate (C), an XOR gate (D), a NAND gate (E), and a NOR gate (F).

**Table 26-2. Logic gates and their characteristics.**

Gate type	Number of inputs	Remarks
NOT	1	Changes state of input.
OR	2 or more	Output high if any inputs are high. Output low if all inputs are low.
AND	2 or more	Output low if any inputs are low. Output high if all inputs are high.
NOR	2 or more	Output low if any inputs are high. Output high if all inputs are low.
NAND	2 or more	Output high if any inputs are low. Output low if all inputs are high.
XOR	2	Output high if inputs differ. Output low if inputs are the same.

## Forms of Binary Data

In communications, binary (two-level) data is less susceptible to noise and other interference than analog or multilevel digital data. There are several forms.

- *Morse code* is the oldest binary means of sending and receiving messages. It is a binary code because it has only two possible states: on (key-down) and off (key-up). It is used mainly by amateur radio operators in their hobby activities. A “human ear/brain machine,” scrutinizing a Morse code signal, is an amazingly effective digital communications receiver.
- *Baudot*, also called the *Murray code*, is a five-unit digital code not widely used by today’s digital equipment, except in some radioteletype communications.
- *ASCII* (American National Standard Code for Information Interchange) is a seven-unit code for the transmission of text and some programs. Letters, numerals, symbols, and control operations are represented. ASCII is designed for computers. There are  $2^7$ , or 128, possible representations. Both upper- and lowercase letters can be represented, along with numerals and certain symbols.

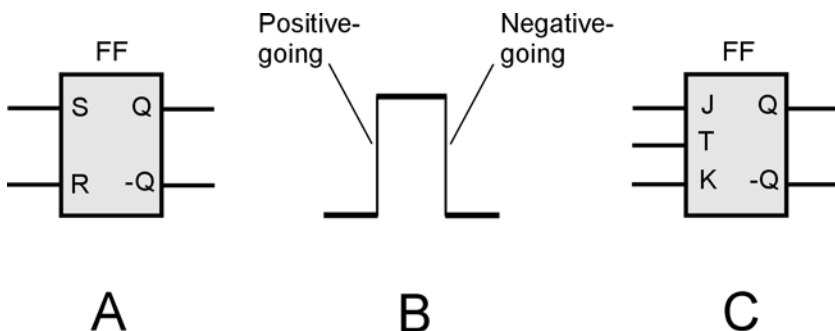
## Flip-flops

A *flip-flop* is also known as a *sequential logic gate*. In a sequential gate, the output state depends on both the inputs and the outputs. A flip-flop has two states, called *set* and *reset*. Usually, the set state is logic 1 (high), and the reset state is logic 0 (low). Here are some common types.

- *R-S flip-flop* inputs are labeled R (reset) and S (set). The outputs are Q and  $\bar{Q}$ . (Often, rather than  $\bar{Q}$ , you will see  $Q'$ , or perhaps Q with a line over it.) The outputs are always in logically opposite states. The symbol for an R-S flip-flop, also known as an *asynchronous flip-flop*, is shown in Fig. 26-3A. The *truth table* (a specialized form of table denoting logic functions) for an R-S flip-flop is at Table 26-3A.
- *Synchronous flip-flop* states change when triggered by the signal from a *clock*. In *static triggering*, the outputs change state only when the clock signal is either high or low. This type of circuit is sometimes called a *gated flip-flop*. In *positive-edge triggering*, the outputs change state at the instant the clock pulse is positive-going. The term *edge triggering* derives from the fact that the abrupt rise or fall of a pulse looks like the edge of a cliff (Fig. 26-3B). In *negative-edge triggering*, the outputs change state at the instant the clock pulse is negative-going.
- *Master/slave (M/S) flip-flop* inputs are stored before the outputs are allowed to change state. This device essentially consists of two R-S flip-flops in series. The first flip-flop is called the *master*, and the second is called the *slave*. The master flip-flop functions when the clock output is high, and the slave acts during the next low portion of the clock output. This time delay prevents confusion between the input and output.
- *J-K flip-flop* operation is similar to that of an R-S flip-flop, except that the J-K has a predictable output when the inputs are both 1. Table 26-3B shows the input and output states for this type of flip-flop. The output changes only when a triggering pulse is received. The symbol for a J-K flip-flop is shown in Fig. 26-3C.
- *R-S-T flip-flop* operation is similar to that of an R-S flip-flop, except that a high pulse at the T input causes the circuit to change state.
- The *T flip-flop* has only one input. Each time a high pulse appears at the T input, the output state is reversed.

## Clocks

In electronics, the term *clock* refers to a circuit that generates pulses at high speed and at precise intervals. It sets the tempo for the operation of digital devices. In a computer, the clock acts like a metronome for the *microprocessor*. Clock speeds are measured and expressed in hertz (Hz), kilohertz (kHz), megahertz (MHz), or gigahertz (GHz).



**26-3** At A, the symbol for an R-S flip-flop. At B, pulse edges are either negative-going or positive-going. At C, the symbol for a J-K flip-flop.

**Table 26-3. Flip-flop states.**

<b>A:</b>			
<b>R-S Flip-flop</b>			
<b>R</b>	<b>S</b>	<b>Q</b>	<b>-Q</b>
0	0	Q	-Q
0	1	1	0
1	0	0	1
1	1	?	?
<b>B:</b>			
<b>J-K Flip-flop</b>			
<b>J</b>	<b>K</b>	<b>Q</b>	<b>-Q</b>
0	0	Q	-Q
0	1	1	0
1	0	0	1
1	1	-Q	Q

## Counters

A *counter* consists of a set of flip-flops or equivalent circuits. Each time a pulse is received, the binary number stored by the counter increases by 1. A *frequency counter* measures the frequency of a wave by tallying the cycles in a given interval of time. The circuit consists of a *gate*, which begins and ends each counting cycle at defined intervals. The accuracy is a function of the length of the *gate time*; the longer the time base, the better the accuracy. The readout is in base-10 digital numerals.

## Binary Digital Communications

The use of binary data provides excellent communications accuracy and efficiency. If multilevel signaling is required, then all the levels can be represented by groups of binary digits. A group of 3 binary digits, for example, can represent  $2^3$ , or 8, levels. A group of 4 binary digits can represent  $2^4$ , or 16, levels. The term *binary digit* is commonly contracted to *bit*.

### Bits, Bytes, and Baud

A bit is almost always represented by either 0 or 1. A group of 8 bits is called an *octet*, and in many systems this also corresponds to a unit called a *byte*. Large quantities of data can be expressed either according to powers of 2, or according to powers of 10. This can cause some confusion and gives rise to endless debates over semantics.

One *kilobit* (kb) is equal to 1000 bits. A *megabit* (Mb) is 1000 kilobits, or 1,000,000 bits. A *gigabit* (Gb) is 1000 megabits, or 1,000,000,000 bits. When data is expressed in bits, powers of 10 are used to define large quantities. If you hear about a modem that operates at 56 kbps, it means 56,000 *bits per second* (bps). Bits, kilobits, megabits, and gigabits per second (bps, kbps, Mbps, and Gbps) are commonly used to express data in communications.



Data quantity in storage and memory is usually specified in *kilobytes* (units of  $2^{10} = 1,024$  bytes), *megabytes* (units of  $2^{20} = 1,048,576$  bytes), and *gigabytes* (units of  $2^{30} = 1,073,741,824$  bytes). The abbreviations for these units are KB, MB, and GB, respectively. Note that the uppercase K represents  $2^{10}$  or 1024, while the lowercase k represents  $10^3$  or 1000. But M and G are always uppercase, no matter whether powers of 2 or 10 are used. (Is all of this confusing to you? Don't be discouraged. It confuses almost everybody.)

Larger data units are being used as memory and storage media continue to grow. The *terabyte* (TB) is  $2^{40}$  bytes, or 1024 GB. The *petabyte* (PB) is  $2^{50}$  bytes, or 1024 TB. The *exabyte* (EB) is  $2^{60}$  bytes, or 1024 PB.

The term *baud* refers to the number of times per second that a signal changes state. The units of bps and baud are not equivalent, even though people often speak of them as if they are. These days, baud (or "baud rate") is seldom used to express data speed.

When computers are linked in a *network*, each computer has a *modem* (modulator/demodulator) connecting it to the communications medium. The slowest modem determines the speed at which the machines communicate. Table 26-4 shows common data speeds and the approximate time periods required to send 1, 10, and 100 pages of double-spaced, typewritten text at each speed.

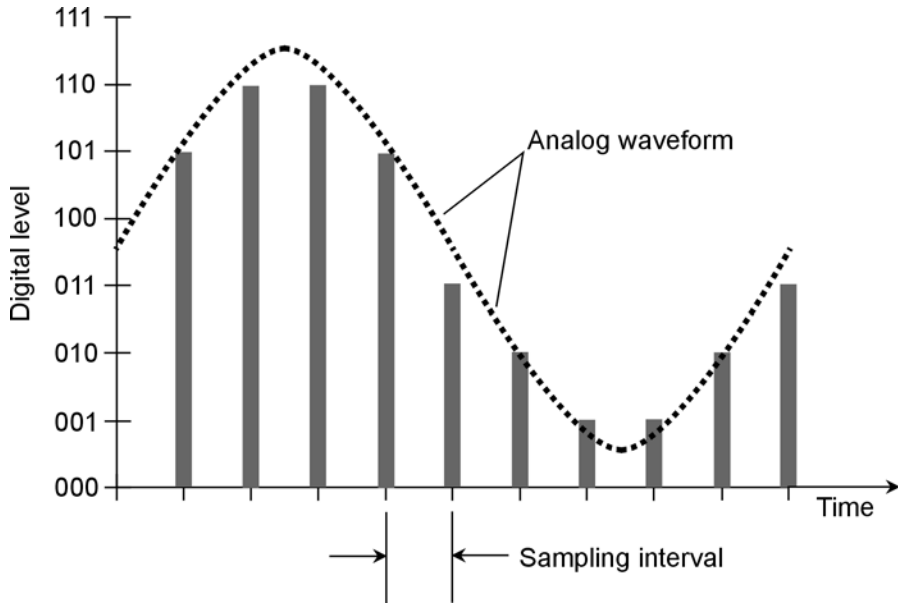
## Forms of Conversion

Any analog (continuously variable) signal can be converted into a string of pulses whose amplitudes have a finite number of states, usually some power of 2. This is analog-to-digital (A/D) conversion, and its reverse is digital-to-analog (D/A) conversion, as you've already learned. The difference between analog and digital signals can be intuitively seen by examining Fig. 26-4. This is essentially a rendition of 8-level pulse code modulation (PCM), as described in Chap. 25 and illustrated in Fig. 25-8D. Imagine the curve being sampled as a train of pulses (A/D), or the pulses being smoothed into the curve (D/A).

**Table 26-4. Time needed to send data at various speeds.**

Abbreviations: s = second, ms = millisecond (0.001 s),  $\mu$ s = microsecond (0.000001 s).

<b>A:</b>			
<b>Speed, kbps</b>	<b>Time for one page</b>	<b>Time for 10 pages</b>	<b>Time for 100 pages</b>
28.8	0.38 s	3.8 s	38 s
38.4	280 ms	2.8 s	28 s
57.6	190 ms	1.9 s	19 s
100	110 ms	1.1 ms	11 s
250	44 ms	440 ms	4.4 s
500	22 ms	220 ms	2.2 s
<b>B:</b>			
<b>Speed, Mbps</b>	<b>Time for one page</b>	<b>Time for 10 pages</b>	<b>Time for 100 pages</b>
1.00	11 ms	110 ms	1.1 s
2.50	4.4 ms	44 ms	440 ms
10.0	1.1 ms	11 ms	110 ms
100	110 $\mu$ s	1.1 ms	11 ms



**26-4** An analog waveform (dashed curve) and an 8-level digital representation (vertical bars).

Binary digital data can be sent and received one bit at a time along a single line or channel. This is *serial data transmission*. Higher data speeds can be obtained by using multiple lines or a wideband channel, sending independent sequences of bits along each line or subchannel. This is *parallel data transmission*.

*Parallel-to-serial (P/S) conversion* receives bits from multiple lines or channels, and transmits them one at a time along a single line or channel. A *buffer* stores the bits from the parallel lines or channels while they are awaiting transmission along the serial line or channel. *Serial-to-parallel (S/P) conversion* receives bits from a serial line or channel, and sends them in batches along several lines or channels. The output of an S/P converter cannot go any faster than the input, but the circuit is useful when it is necessary to interface between a serial-data device and a parallel-data device.

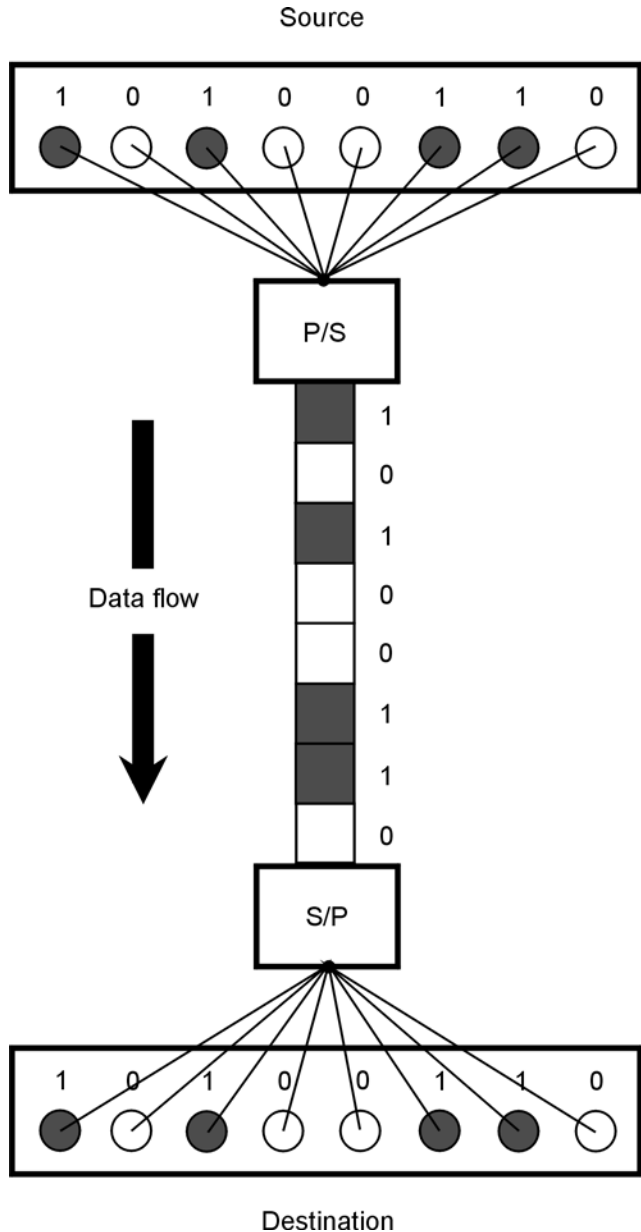
Figure 26-5 illustrates a circuit in which a P/S converter is used at the source (transmitting station), and an S/P converter is used at the destination (receiving station). In this example, the words are 8-bit bytes. However, the words could have 16, 32, 64, or even 128 bits, depending on the communications scheme.

## Data Compression

*Data compression* is a way of maximizing the amount of digital information that can be stored in a given space, or sent in a certain period of time.

Text files can be compressed by replacing often-used words and phrases with symbols such as =, #, &, \$, and @, as long as none of these symbols occurs in the uncompressed file. When the data is received, it is uncompressed by substituting the original words and phrases for the symbols.

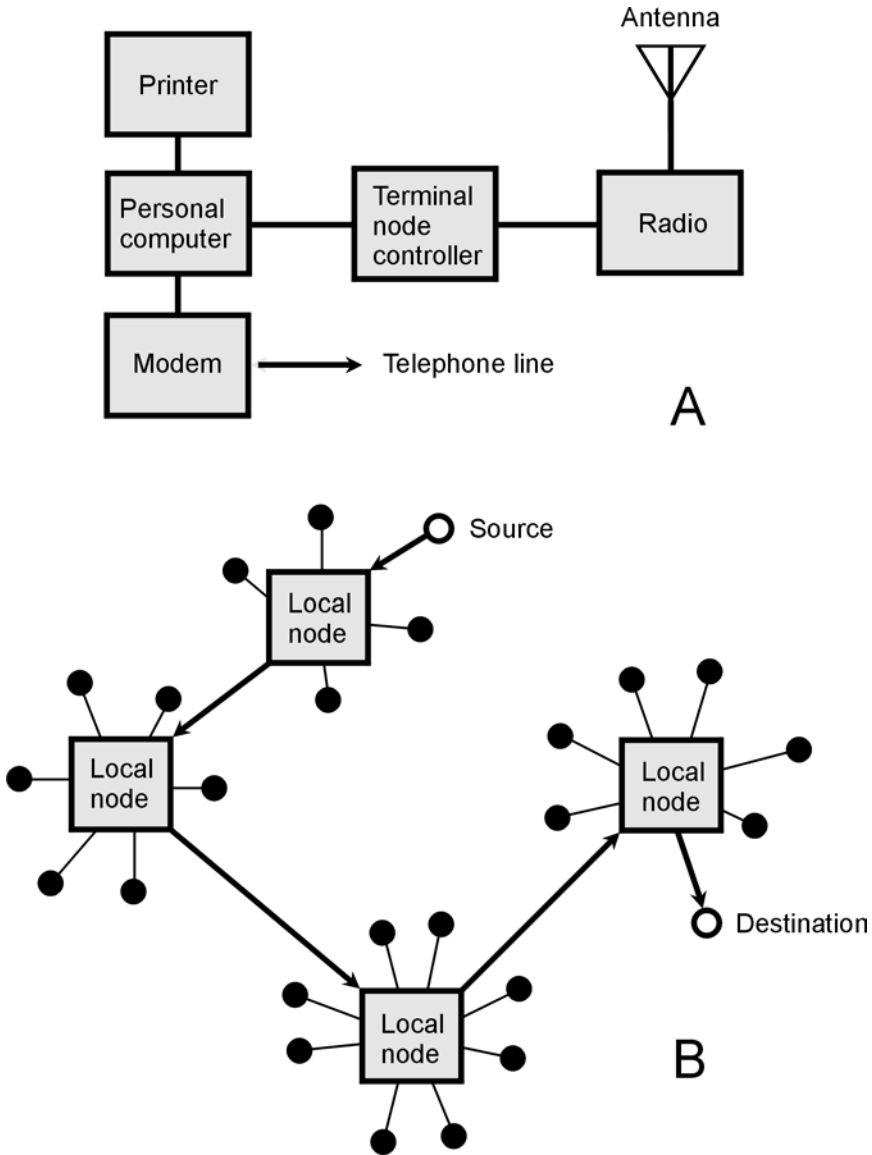
**26-5** A communications circuit employing parallel-to-serial (P/S) conversion at the source, and serial-to-parallel (S/P) conversion at the destination.



Digital images can be compressed in either of two ways. In *lossless image compression*, detail is not sacrificed; only the redundant bits are eliminated. In *lossy image compression*, some detail is lost, although the loss is usually not significant.

### Packet Wireless

In *packet wireless*, a computer is connected to a radio transceiver using a *terminal node controller* (TNC), which is similar to a modem. An example is shown in Fig. 26-6A. The computer has a tele-



**26-6** At A, a packet-wireless station. At B, passage of a packet through nodes in a wireless communications circuit.

phone modem as well as a TNC, so messages can be sent and received using conventional online services as well as radio.

Figure 26-6B shows how a packet wireless message is routed. Black dots represent *subscribers*. Rectangles represent local nodes, each of which serves subscribers by means of short-range links at very high, ultra high, or microwave radio frequencies. The nodes are interconnected by terrestrial radio links if they are relatively near each other. If the nodes are widely separated, satellite links are used.

# The RGB Color Model

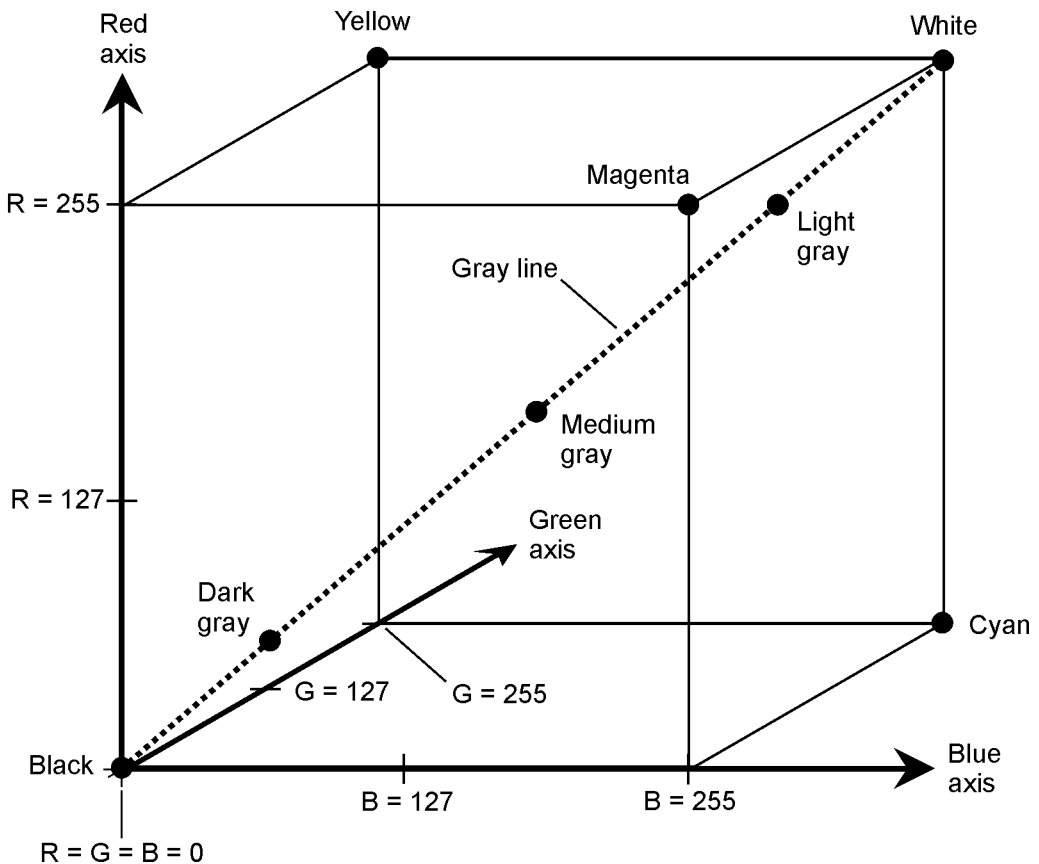
All visible colors can be obtained by combining red, green, and blue light. The *red/green/blue* (RGB) *color model* takes advantage of this fact.

## Hue, Saturation, and Brightness

Color, as perceived by the human eye, is a function of the wavelength of optical EM energy. When most of the energy is concentrated near a single wavelength, you see an intense *hue*. The vividness of a hue is called the *saturation*. The *brightness* of a color is a function of how much total energy the light contains. In most video displays, there is a control for adjusting the brightness, also called *brilliance*.

## 3D Color

A color *palette* is obtained by combining pure red, green, and blue in various ratios. Assign each primary color an axis in Cartesian three-dimensional (3D) space as shown in Fig. 26-7. Call the axes R (for red), G (for green), and B (for blue). Color brightness can range from 0 to 255, or binary



26-7 The RGB color model, depicted as a cube in Cartesian three-dimensional (3D) space.

00000000 to 11111111. The result is 16,777,216 ( $256^3$ ) possible colors. Any point within the cube represents a unique color.

Colors in the RGB model can also be represented by six-digit hexadecimal numbers, such as 005CFF. In this scheme, the first two digits represent the red (R) intensity in 256 levels ranging from 00 to FF. The middle two digits represent the green (G) intensity, and the last two digits represent the blue (B) intensity.

Some RGB systems use only 16 levels for each primary color (binary 0000 through 1111). This results in 4096 possible colors.

## Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. Suppose inverters are placed in series with both inputs of an AND gate. Under what conditions is the output of the resulting black box high?
  - (a) If and only if both inverter inputs are high
  - (b) If and only if both inverter inputs are low
  - (c) If and only if one inverter input is high and the other is low
  - (d) Under no conditions (the output is always low)
2. Suppose an AND gate is followed by an inverter. Under what conditions is the output of the resulting black box low?
  - (a) If and only if both inputs are high
  - (b) If and only if both inputs are low
  - (c) If and only if one input is high and the other is low
  - (d) Under no conditions (the output is always high)
3. What is the binary equivalent of decimal 29?
  - (a) 10101
  - (b) 11101
  - (c) 10111
  - (d) 11011
4. In Boolean algebra, addition represents
  - (a) the logical NOT operation.
  - (b) the logical AND operation.
  - (c) the logical OR operation.
  - (d) the logical NAND operation.
5. If the output to a logical inverter is low, it means that
  - (a) both of the two inputs are high.
  - (b) both of the two inputs are low.

- (c) the single input is high.
  - (d) the single input is low.
6. What is the decimal equivalent of binary 1111?
- (a) 15
  - (b) 31
  - (c) 63
  - (d) 127
7. The output of an AND gate is high if and only if
- (a) one input is high and the others are low.
  - (b) one input is low and the others are high.
  - (c) any of the inputs is low.
  - (d) all inputs are high.
8. DeMorgan's Theorem states that, for all logical statements X and Y,
- (a)  $\neg(X*Y)$  is equivalent to  $X+Y$ .
  - (b)  $X*Y$  is equivalent to  $\neg(X+Y)$ .
  - (c)  $\neg X+\neg Y$  is equivalent to  $X*Y$ .
  - (d)  $\neg X+\neg Y$  is equivalent to  $\neg(X*Y)$ .
9. If you see a number represented by FF in the documentation for an electronic circuit or system, you can be certain that the number is
- (a) radix 16.
  - (b) radix 10.
  - (c) radix 8.
  - (d) radix 2.
10. Which of the following codes is digital?
- (a) Morse
  - (b) Baudot
  - (c) ASCII
  - (d) All of the above
11. Which of the following voltages could normally represent a 1 in positive logic?
- (a) 0.0 V
  - (b) +0.5 V
  - (c) +5 V
  - (d) +50 V
12. Which of the following voltages might normally represent a 1 in negative logic?
- (a) 0.0 V
  - (b) +12 V

- (c) +5 V
  - (d) +50 V
13. Fill in the blank to make the following sentence true: “ \_\_\_\_ data transmission refers to digital bits sent along a single line, one bit after another.”
- (a) Serial
  - (b) Synchronous
  - (c) Parallel
  - (d) Analog
14. The inputs of an R-S flip-flop are known as
- (a) low and high.
  - (b) octal and binary.
  - (c) synchronous and asynchronous.
  - (d) reset and set.
15. When both inputs of an R-S flip-flop are 0,
- (a) the outputs stay as they are.
  - (b)  $Q = 0$  and  $\bar{Q} = 1$ .
  - (c)  $Q = 1$  and  $\bar{Q} = 0$ .
  - (d) the resulting outputs can be absurd.
16. The fifth digit from the right in a binary number carries which decimal value?
- (a) 64
  - (b) 32
  - (c) 24
  - (d) 16
17. What is the octal equivalent of binary 1010?
- (a) 4
  - (b) 10
  - (c) 12
  - (d) There is no way to tell without more information.
18. What is the largest possible radix 10 number that can be represented as a six-digit binary number?
- (a) Decimal 256
  - (b) Decimal 128
  - (c) Decimal 64
  - (d) Decimal 63
19. If one of the inputs to a two-input NOR gate is high, what is the output state?
- (a) There is not enough information to tell.
  - (b) Low



- (c) High
- (d) It depends on the state of the other input.

20. Suppose a logic circuit has four inputs W, X, Y, and Z. How many possible input combinations are there?

- (a) 4
- (b) 8
- (c) 16
- (d) 32

# Test: Part 3

DO NOT REFER TO THE TEXT WHEN TAKING THIS TEST. A GOOD SCORE IS AT LEAST 37 CORRECT. Answers are in the back of the book. It's best to have a friend check your score the first time, so you won't memorize the answers if you want to take the test again.

1. In a JFET, the control electrode is usually the
  - (a) source.
  - (b) emitter.
  - (c) drain.
  - (d) base.
  - (e) gate.
2. A diode can be used as a frequency multiplier because of its
  - (a) junction capacitance.
  - (b) nonlinearity.
  - (c) avalanche voltage.
  - (d) forward breakover.
  - (e) charge carrier concentration.
3. Which of the following is not a common form of data transmission?
  - (a) Polarization modulation
  - (b) Frequency modulation
  - (c) Amplitude modulation
  - (d) Phase modulation
  - (e) Pulse modulation
4. A very brief, high-voltage spike on an ac power line is called
  - (a) a bleeder.
  - (b) an arc.

- (c) a transient.
  - (d) an avalanche.
  - (e) a clipped peak.
5. Which of the following is not characteristic of an oscillator?
- (a) Negative feedback
  - (b) Good output-to-input coupling
  - (c) Reasonably high transistor gain
  - (d) Alternating current signal output
  - (e) Usefulness as a signal generator
6. Which layer of the ionosphere exists at the lowest altitude?
- (a) The F layer
  - (b) The E layer
  - (c) The D layer
  - (d) The C layer
  - (e) The B layer
7. The beta of a bipolar transistor is another name for its
- (a) current amplification factor.
  - (b) voltage amplification factor.
  - (c) power amplification factor.
  - (d) maximum amplification frequency.
  - (e) optimum amplification frequency.
8. In a schematic diagram, the symbol for a PNP bipolar transistor can be recognized by
- (a) a broken vertical line inside the symbol circle.
  - (b) an arrow inside the symbol circle, pointing from the base toward the emitter.
  - (c) an arrow inside the symbol circle, pointing outward from the collector.
  - (d) a collector biased at a positive voltage with respect to the emitter.
  - (e) an arrow inside the symbol circle, pointing from the emitter toward the base.
9. Fill in the blank in the following sentence to make it true: "In an oscillator, \_\_\_\_\_ is an expression of the extent to which the circuit maintains constant signal frequency output under variable operating conditions, and is also an expression of its overall reliability."
- (a) sensitivity
  - (b) drift ratio
  - (c) gain
  - (d) selectivity
  - (e) stability
10. A Zener diode would most likely be found in
- (a) the mixer in a superheterodyne receiver.
  - (b) the PLL in a circuit for detecting FM.

- (c) the product detector in a receiver for SSB.
  - (d) the voltage regulator in a power supply.
  - (e) the AF oscillator in an AFSK transmitter.
11. When the bias in an FET stops the flow of current, the condition is called
- (a) forward breakover.
  - (b) cutoff.
  - (c) reverse bias.
  - (d) pinchoff.
  - (e) avalanche.
12. The VA rating of a transformer is an expression of
- (a) the maximum frequency at which it can function.
  - (b) the type of core material it has.
  - (c) the voltage step-up or step-down ratio.
  - (d) the impedance transfer ratio.
  - (e) none of the above.
13. In an N-type semiconductor, the minority carriers are
- (a) electrons.
  - (b) protons.
  - (c) holes.
  - (d) neutrons.
  - (e) positrons.
14. A disadvantage of a half-wave rectifier is the fact that
- (a) the output voltage is excessive compared to that of a full-wave rectifier.
  - (b) the output current is excessive compared to that of a full-wave rectifier.
  - (c) the output waveform is harder to filter than is the case with a full-wave rectifier.
  - (d) it requires several expensive diodes, whereas a full-wave rectifier requires only a single cheap diode.
  - (e) it requires an expensive center-tapped transformer, whereas a full-wave rectifier does not need a transformer at all.
15. A power gain of 30 dB is equivalent to which amplification factor?
- (a) 0.001
  - (b)  $\frac{1}{30}$
  - (c) 30
  - (d) 1000
  - (e) None of the above

16. Suppose an RF power amplifier has a dc collector input of 60 W, and is 75 percent efficient. What is the RF signal output power?

- (a) 80 W
- (b) 60 W
- (c) 45 W
- (d) 40 W
- (e) Impossible to determine from this data

17. If a listener is not expecting any change in the intensity of a signal or sound, the smallest change that can be noticed is approximately which of the following?

- (a)  $\pm 0.3$  dB
- (b)  $\pm 1$  dB
- (c)  $\pm 3$  dB
- (d)  $\pm 10$  dB
- (e)  $\pm 30$  dB

18. A common base circuit is commonly employed as

- (a) a microwave oscillator.
- (b) a low-pass filter.
- (c) a noise generator.
- (d) a phase-locked loop.
- (e) an RF power amplifier.

19. A semiconductor material especially noted for its photoconductivity, and which is often used in making photocells, is

- (a) selenium.
- (b) lithium salt.
- (c) copper chloride.
- (d) aluminum oxide.
- (e) polyethylene.

20. Which type of PA circuit provides the highest efficiency when used with the appropriate amount of driving power?

- (a) Class A
- (b) Class AB
- (c) Class B
- (d) Class C
- (e) All of the above circuits are equally efficient.

21. Baudot is a form of

- (a) video modulation.
- (b) diode.

- (c) digital code.
  - (d) voice modulation.
  - (e) AM detector.
22. The most stable type of oscillator circuit uses
- (a) a tapped coil.
  - (b) a split capacitor.
  - (c) negative feedback.
  - (d) a common base arrangement.
  - (e) a quartz crystal.
23. If the source-gate junction in an FET conducts,
- (a) it is a sign of improper bias.
  - (b) the device will work in class C.
  - (c) the device will oscillate.
  - (d) the device will work in class A.
  - (e) the circuit will have good stability.
24. What is the radix of the octal number system?
- (a) 2
  - (b)  $2^2$
  - (c)  $2^3$
  - (d)  $2^4$
  - (e)  $2^5$
25. Signal-to-noise ratio (S/N) is often specified when stating
- (a) the selectivity of a receiver.
  - (b) the stability of an oscillator.
  - (c) the modulation coefficient of a transmitter.
  - (d) the sensitivity of a receiver.
  - (e) the polarization of an EM wave.
26. In a reverse-biased semiconductor diode, the junction capacitance depends on
- (a) the width of the depletion region.
  - (b) the reverse current.
  - (c) the P:N ratio.
  - (d) the gate bias.
  - (e) the avalanche voltage.
27. Suppose the bandwidth of the channel in a receiver's IF filter is 4.5 kHz at 30-dB power attenuation, and 3.0 kHz at 3-dB power attenuation. What is the shape factor?
- (a) 10:1
  - (b) 1.5:1

- (c) 1:1
  - (d) 1:10
  - (e) More information is necessary to determine this.
28. A simple power supply filter can be built with
- (a) a capacitor in series with the dc output.
  - (b) an inductor in parallel with the dc output.
  - (c) a rectifier in parallel with the dc output.
  - (d) a resistor in series and an inductor in parallel with the dc output.
  - (e) a capacitor in parallel with the dc output.
29. Which of the following bipolar transistor circuits can, in theory, provide the most amplification?
- (a) Common emitter
  - (b) Common base
  - (c) Common collector
  - (d) Common gate
  - (e) Common drain
30. The ratio of the difference in  $I_C$  to the difference in  $I_B$  that occurs when a small signal is applied to the base of a bipolar transistor is called the
- (a) static current amplification.
  - (b) dynamic current amplification.
  - (c) power factor.
  - (d) efficiency.
  - (e) current attenuation factor.
31. An example of a device that is commonly used as an oscillator at microwave frequencies is
- (a) a rectifier diode.
  - (b) a weak-signal diode.
  - (c) a Gunn diode.
  - (d) a Zener diode.
  - (e) an avalanche diode.
32. In the operation of a PNP bipolar transistor, which of the following is normal concerning the emitter and collector voltages?
- (a) The collector is positive relative to the emitter.
  - (b) The collector is at the same voltage as the emitter.
  - (c) The collector is negative relative to the emitter.
  - (d) The collector can be either positive or negative relative to the emitter.
  - (e) The collector must be at ground potential.

33. What can be done to minimize the capacitance of the P-N junction in a semiconductor diode, thereby making the component effective as a high-speed RF switch?
- (a) The surface area of the P-N junction can be maximized.
  - (b) A layer of intrinsic semiconductor can be placed between the P- and N-type materials.
  - (c) The frequency of the applied signal can be made as high as possible.
  - (d) The diode can be forward-biased with a high voltage.
  - (e) A small-value resistor can be connected in series with each of several diodes, and the series diode-resistor combinations can all be connected in parallel.
34. Which type of modulation consists of one voice sideband, with a suppressed carrier?
- (a) AM
  - (b) LSB
  - (c) FM
  - (d) RTTY
  - (e) PCM
35. Suppose a series-parallel network of resistors is assembled into an *RF attenuator pad*, which is a circuit deliberately intended to produce a constant signal loss over a wide band of frequencies. If the input RF voltage to this attenuator pad is  $500\ \mu\text{V}$  and the output RF voltage is  $100\ \mu\text{V}$ , what is the loss in decibels?
- (a) 14.0 dB
  - (b) 6.99 dB
  - (c) 5.00 dB
  - (d) 1.40 dB
  - (e) 0.699 dB
36. In an AND gate, the output is high
- (a) if any input is high.
  - (b) only when all inputs are low.
  - (c) if any input is low.
  - (d) only when all inputs are high.
  - (e) only when all inputs have identical logic states.
37. A voltage-controlled oscillator makes use of
- (a) a varactor diode.
  - (b) a Zener diode.
  - (c) negative feedback.
  - (d) a split capacitance.
  - (e) adjustable gate or base bias.
38. The best electrical ground for a power supply is obtained by
- (a) connecting a small-value capacitor between the hot wire and a good earth ground, such as an 8-ft copper rod driven into the earth.



- (b) ensuring that the ground wire from the power supply is connected to the third wire hole in one of the utility outlets of a properly installed three-wire electrical circuit.
  - (c) using a center-tapped transformer and a full-wave rectifier circuit, in conjunction with good filtering, voltage regulation, and transient suppression.
  - (d) connecting a large-value inductor in series with the hot wire, thereby ensuring that the voltage on that wire cannot be shunted to ground.
  - (e) making sure that all equipment connected to the supply is provided with fuses and/or circuit breakers of the proper ratings.
39. Suppose the output signal voltage from an amplifier is 35 times the input signal voltage. If the input and output impedances are both pure resistances of the same value, what is the gain?
- (a) 15 dB
  - (b) 31 dB
  - (c) 35 dB
  - (d) 350 dB
  - (e) 700 dB
40. In an exclusive OR gate, the output is high
- (a) if either input is high.
  - (b) only when both inputs are low.
  - (c) if either input is low.
  - (d) only when both inputs are high.
  - (e) only when the inputs have opposite logic states.
41. A ratio detector is a circuit for demodulating which of the following?
- (a) AM
  - (b) PM
  - (c) FM
  - (d) SSB
  - (e) AFSK
42. Suppose there is a binary digital black box with two inputs, called X and Y, and an output called Z. If  $X = 0$  and  $Y = 1$ , then  $Z = 0$ . In all other instances,  $Z = 1$ . Which of the following logical expressions represents the contents of the black box?
- (a)  $X + (-Y)$
  - (b)  $X * (-Y)$
  - (c)  $X - (*Y)$
  - (d)  $X = (-Y)$
  - (e)  $X - (+Y)$
43. A method of modulation in which the strength of pulses varies is called
- (a) pulse ratio modulation.
  - (b) pulse position modulation.

- (c) pulse frequency modulation.
  - (d) pulse amplitude modulation.
  - (e) pulse width modulation.
44. Boolean algebra is
- (a) useful for calculating amplifier gain in decibels.
  - (b) a useful tool in binary digital circuit design.
  - (c) used to calculate the value of an unknown.
  - (d) used with negative pulse modulation schemes.
  - (e) used with positive pulse modulation schemes.
45. A voltage-doubler power supply is best for use in
- (a) circuits that need high voltage but do not draw much current.
  - (b) low-voltage devices.
  - (c) high-current appliances.
  - (d) all kinds of electronic equipment.
  - (e) broadcast transmitter power amplifiers.
46. An optoisolator consists of
- (a) two Zener diodes back-to-back.
  - (b) an LED and a photodiode.
  - (c) two NPN transistors in series.
  - (d) an NPN transistor followed by a PNP transistor.
  - (e) a PNP transistor followed by an NPN transistor.
47. When a semiconductor is reverse-biased with a large enough voltage, it will conduct. This phenomenon is known as
- (a) bias effect.
  - (b) avalanche effect.
  - (c) forward breakover.
  - (d) saturation.
  - (e) conduction effect.
48. Synchronizing pulses in a conventional analog TV signal
- (a) keep the brightness constant.
  - (b) keep the contrast constant.
  - (c) keep the image from tearing or rolling.
  - (d) ensure that the colors are right.
  - (e) keep the image in good focus.
49. In an enhancement-mode MOSFET,
- (a) the channel conducts fully with zero gate bias.
  - (b) the channel conducts partially with zero gate bias.

- (c) the channel conducts ac but not dc.
  - (d) the channel conducts dc but not ac.
  - (e) the channel does not conduct with zero gate bias.
50. In a step-up power transformer,
- (a) the primary voltage is more than the secondary voltage.
  - (b) the secondary voltage is more than the primary voltage.
  - (c) the primary and secondary voltages are the same.
  - (d) the secondary must be center-tapped.
  - (e) the primary must be center-tapped.

*This page intentionally left blank*

# 4 PART

# Specialized Devices and Systems

*This page intentionally left blank*

## 27 CHAPTER

# Antennas

ANTENNAS CAN BE CATEGORIZED INTO TWO MAJOR CLASSES: RECEIVING TYPES AND TRANSMITTING types. Most transmitting antennas can also function effectively for reception. Some receiving antennas can efficiently transmit EM signals; others cannot.

## Radiation Resistance

When RF current flows in an electrical conductor, some EM energy is radiated into space. If a resistor is substituted for the antenna, in combination with a capacitor or inductor to mimic any inherent reactance in the antenna, the transmitter behaves in the same manner as when connected to the actual antenna. For any antenna operating at a specific frequency, there exists a specific resistance, in ohms, for which this can be done. This is known as the *radiation resistance* ( $R_R$ ) of the antenna at the frequency in question. Radiation resistance is specified in ohms. This phenomenon was introduced in Chap. 17. Now you'll see why it's important!

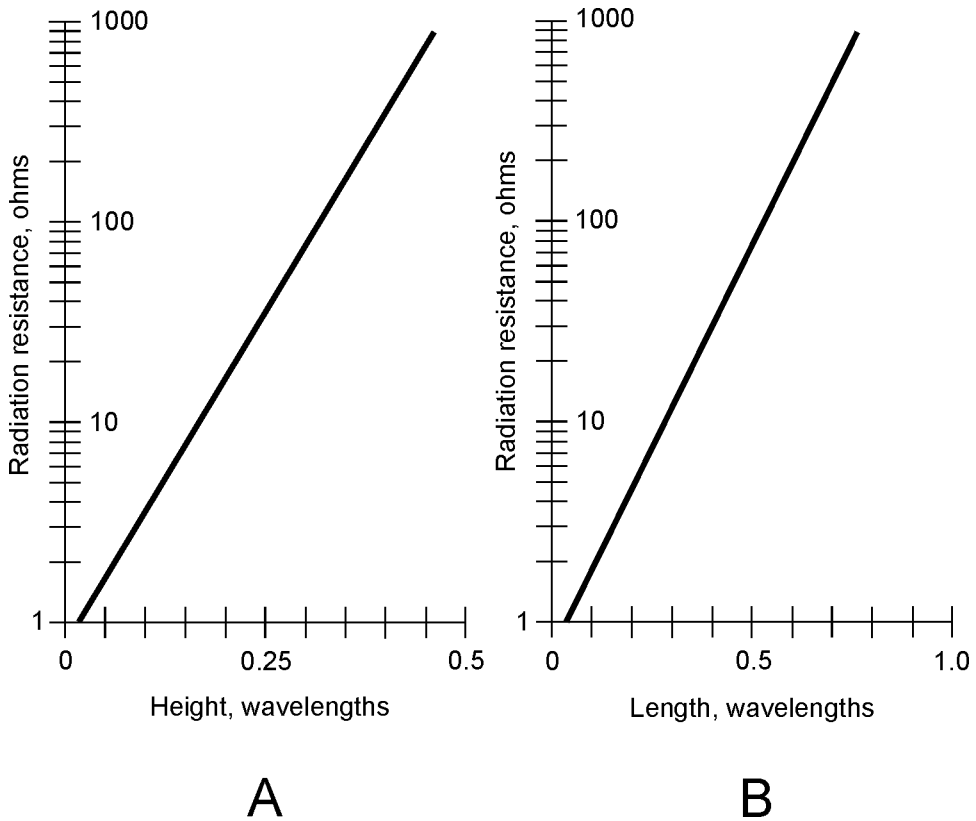
### Determining Factors

Suppose a thin, straight, lossless vertical antenna is placed over perfectly conducting ground. Then  $R_R$  is a function of the vertical-antenna conductor height in wavelengths (Fig. 27-1A). Suppose a thin, straight, lossless wire is placed in free space and fed at the center. Then  $R_R$  is a function of the overall conductor length in wavelengths (Fig. 27-1B).

### Antenna Efficiency

Efficiency is rarely crucial to the performance of a receiving antenna system, but it is important in any transmitting antenna system. Radiation resistance always appears in series with *loss resistance* ( $R_L$ ). The antenna efficiency,  $Eff$ , is given by:

$$Eff = R_R / (R_R + R_L)$$



27-1 Approximate values of radiation resistance for vertical antennas over perfectly conducting ground (A) and for center-fed antennas in free space (B).

which is the ratio of the radiation resistance to the total antenna system resistance. As a percentage,

$$Eff_{\%} = 100 R_R / (R_R + R_L)$$

High efficiency in a transmitting antenna is obtained when the radiation resistance is much larger than the loss resistance. Then most of the power goes into useful EM radiation, and relatively little is dissipated as heat in the earth and in objects surrounding the antenna. When the opposite situation exists—radiation resistance comparable to, or smaller than, the loss resistance, a transmitting antenna is inefficient. This is often the case for extremely short radiators, because they always have low radiation resistance. To some extent this can be overcome by careful antenna design and location to minimize loss resistance, but even the most concerted efforts rarely reduce the loss resistance to less than a few ohms.

If an antenna system has a high loss resistance, it can nevertheless work efficiently if its radiation resistance is high enough. When an antenna radiator is just the right length at a given frequency, and if it is constructed of wire or thin tubing, its radiation resistance can be in excess of 1000  $\Omega$ . This makes it easy to construct an efficient antenna even when the loss resistance is fairly high.



# Half-Wave Antennas

A half wavelength in free space is given by the equation:

$$L_{ft} = 492/f_o$$

where  $L_{ft}$  is the linear distance in feet, and  $f_o$  is the fundamental frequency, in megahertz, at which the antenna exhibits resonance. A half wavelength in meters,  $L_m$ , is given by:

$$L_m = 150/f_o$$

For ordinary wire, the results as obtained above should be multiplied by a velocity factor,  $v$ , of 0.95 (95 percent). For tubing or large-diameter wire,  $v$  can range down to about 0.90 (90 percent).

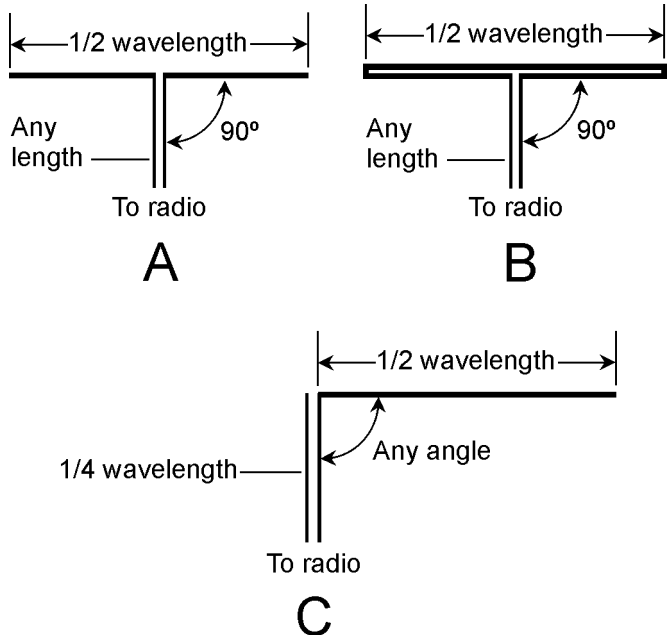
## Open Dipole

An *open dipole* or *doublet* is a half-wavelength ( $\lambda/2$ ) radiator fed at the center (Fig. 27-2A). Each leg of the antenna is a quarter wavelength ( $\lambda/4$ ) long. For a straight wire radiator, the length  $L_{ft}$ , in feet, at a design frequency  $f_o$ , in megahertz, for a center-fed,  $\lambda/2$  dipole is approximately:

$$L_{ft} = 468/f_o$$

The length in meters is approximately:

$$L_m = 143/f_o$$



27-2 Basic half-wave antennas. At A, dipole antenna. At B, folded-dipole antenna. At C, zep antenna.

These values assume  $v = 0.95$ . In free space, the impedance at the feed point is a pure resistance of approximately  $73 \Omega$ . This represents the radiation resistance alone, in the absence of reactance at the resonant frequency.

### Folded Dipole

A *folded dipole antenna* is a  $\lambda/2$ , center-fed antenna constructed of two parallel wires with their ends connected together (Fig. 27-2B). The feed-point impedance of the folded dipole is a pure resistance of approximately  $290 \Omega$ . This makes the folded dipole ideal for use with high-impedance, parallel-wire transmission lines in applications where gain and directivity are not especially important.

### Half-Wave Vertical

A radiator measuring  $\lambda/2$  can be stood up, fed at the base (the bottom end) against an earth ground with a transmatch or antenna tuner designed for high impedances, and connected to a radio by a coaxial cable feed line. This type of antenna is an efficient radiator even in the presence of considerable loss resistance, because the radiation resistance is extremely high.

### Zepp

A *zeppelin antenna*, also called a *zepp*, is a  $\lambda/2$  radiator, fed at one end with a  $\lambda/4$  section of open-wire line (Fig. 27-2C). The impedance at the feed point is an extremely high, pure resistance. At the transmitter end of the line, the impedance is a low, pure resistance. A zeppelin antenna can operate well at all harmonics of the design frequency. If an *antenna tuner*, also called a *transmatch*, is available to tune out reactance, the feed line can be of any length. Feed-line radiation can be minimized by carefully cutting the radiator to  $\lambda/2$  at the fundamental frequency, and by using the antenna only at this frequency or one of its harmonics. Zepp antennas are rarely used at frequencies above 30 MHz, except when modified to form a *J pole*.

### J Pole

A zepp can be oriented vertically, and the feed line placed so it lies in the same line as the radiating element. This antenna, called a J pole, radiates equally well in all horizontal directions. The J pole is used as a low-budget antenna at frequencies from approximately 10 MHz up through 300 MHz. It is, in effect, a  $\lambda/2$  vertical fed with an *impedance matching section* consisting of a length of transmission line measuring  $\lambda/4$ . It does not require any radials, and this makes it convenient in locations where space is at a premium.

Some radio amateurs hang long J poles, cut for 3.5 MHz or 1.8 MHz, from kites or helium-filled balloons. Such antennas work well, but they are dangerous if they are not properly tethered to prevent them from breaking off and flying away, or if they are placed where they might fall on power lines. They can develop considerable electrostatic charge, even in clear weather. They are deadly if flown in or near thunderstorms.

## Quarter-Wave Verticals

A *quarter wavelength* is related to frequency by the equation:

$$L_{ft} = 246v/f_0$$

where  $L_{ft}$  represents  $\lambda/4$  in feet,  $f_o$  represents the frequency in megahertz, and  $v$  represents the velocity factor. If the length is expressed in meters as  $L_m$ , then the formula is:

$$L_m = 75v/f_o$$

For a typical wire conductor,  $v = 0.95$  (95 percent); for metal tubing,  $v$  can range down to approximately 0.90 (90 percent).

A  $\lambda/4$  antenna must be operated against a low-loss RF ground. The feed-point impedance over perfectly conducting ground is approximately  $37 \Omega$ . This represents radiation resistance in the absence of reactance, and provides a reasonable impedance match to most coaxial transmission lines.

### Ground-Mounted Vertical

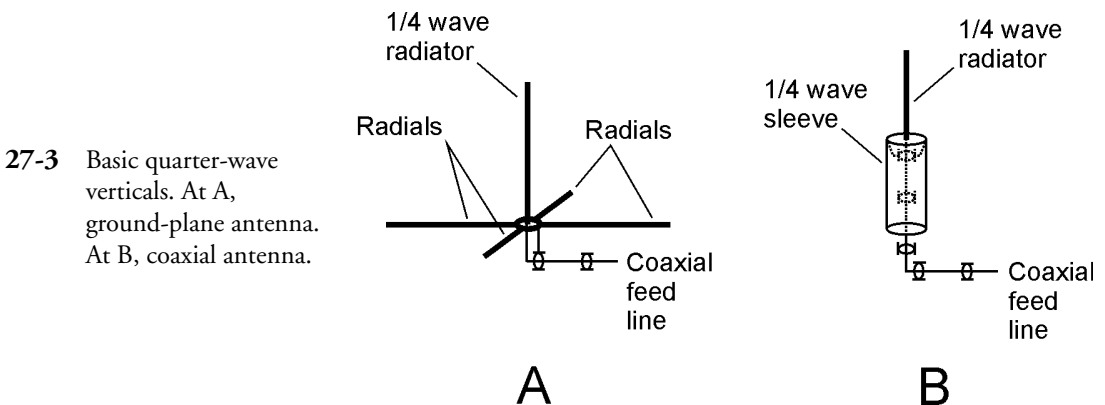
The simplest vertical antenna is a  $\lambda/4$  radiator mounted at ground level. The radiator is fed with a coaxial cable. The center conductor is connected to the base of the radiator, and the shield is connected to a ground system.

Unless an extensive *ground radial* system is installed, a ground-mounted vertical is likely to be inefficient unless the surface happens to be an excellent conductor (saltwater or a salt marsh, for example). Another problem is the fact that vertically polarized antennas receive more human-made noise than horizontal antennas. In addition, the EM fields from ground-mounted transmitting antennas are more likely to cause interference to nearby electronic devices than are the EM fields from antennas installed at a height.

### Ground Plane

A *ground-plane antenna* is a vertical radiator, usually measuring  $\lambda/4$ , operated against a system of  $\lambda/4$  conductors called *radials*. The *feed point*, where the feed line joins the antenna radiator and the hub of the radial system, is elevated. When the feed point is at least  $\lambda/4$  above the ground, only three or four radials are necessary to obtain a low-loss RF ground system. The radials extend outward from the base of the antenna at an angle between  $0^\circ$  and  $45^\circ$  with respect to the horizon. Figure 27-3A illustrates a typical ground-plane antenna with four horizontal radials.

A ground-plane antenna should be fed with coaxial cable. The feed-point impedance of a ground-plane antenna having a  $\lambda/2$  radiator is about  $37 \Omega$  if the radials are horizontal; the impedance increases as the radials droop, reaching about  $50 \Omega$  at a *droop angle* of  $45^\circ$ .



## Coaxial Antenna

The radials in a ground-plane antenna can extend straight downward, and then can be merged to form a  $\lambda/2$  cylinder or sleeve that is concentric with the feed line. In this case, the feed-point impedance is approximately  $73 \Omega$ . This is known as a *coaxial antenna* (Fig. 27-3B).

## Loops

Any receiving or transmitting antenna, consisting of one or more turns of wire forming a dc short circuit, comprises a *loop antenna*.

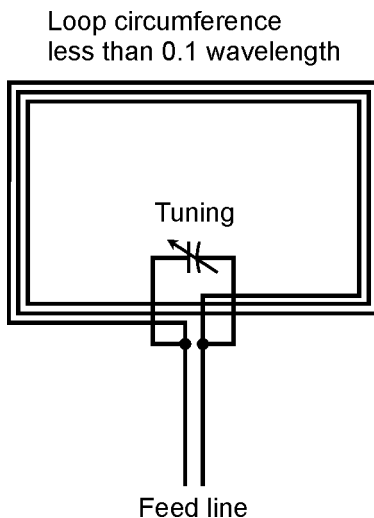
### Small Loop

A *small loop antenna* has a circumference of less than  $0.1\lambda$  (for each turn) and is suitable for receiving, but generally not for transmitting. It is the least responsive to signals coming from along its axis, and most responsive in the plane perpendicular to its axis. A capacitor can be connected in series or parallel with the loop to provide a resonant response. An example of such an arrangement is shown in Fig. 27-4.

Small loops are useful for *radio direction finding* (RDF), and also for reducing interference caused by human-made noise or strong local signals. The *null* along the axis is sharp and deep, and can be pointed in the direction of an offending signal or noise source.

### Loopstick

For receiving applications at frequencies up to approximately 20 MHz, a *loopstick antenna* is sometimes used. This antenna, a variant of the small loop, consists of a coil wound on a solenoidal (rod-shaped), powdered-iron core. A series or parallel capacitor, in conjunction with the coil, forms a tuned circuit. A loopstick displays directional characteristics similar to those of the small loop antenna shown in Fig. 27-4. The sensitivity is maximum off the sides of the coil, and a sharp null occurs off the ends.



27-4 A small loop antenna with a capacitor for adjusting the resonant frequency.

## Large Loop

A *large loop antenna* usually has a circumference of either  $\lambda/2$  or  $\lambda$  (a full wavelength), is circular or square in shape, and lies entirely in a single plane. It can work well for transmitting or receiving. The  $\lambda/2$  loop presents a high impedance at the feed point, and maximum radiation/response occurs in the plane of the loop. The  $\lambda$  loop presents an impedance of about  $100\ \Omega$  at the feed point, and the maximum radiation/response occurs along the axis (that is, perpendicular to the plane containing the loop).

The  $\lambda/2$  loop exhibits a slight power loss relative to a  $\lambda/2$  dipole in its favored directions. The  $\lambda$  loop shows a slight gain over a  $\lambda/2$  dipole in its favored directions. These properties hold for loops up to several percent larger or smaller than exact  $\lambda/2$  or  $\lambda$  circumferences (as determined for the wavelength in free space). Resonance can be obtained by means of an antenna tuner if the loop is fed with open-wire transmission line.

Sometimes, loop antennas measuring several wavelengths in circumference are strung up horizontally among multiple supports. These are technically large loops, but their gain and directional characteristics are hard to predict. If fed with open-wire line and an antenna tuner, and if placed at least  $\lambda/4$  above the surface, such an antenna can be exceptionally effective.

## Ground Systems

End-fed  $\lambda/4$  antennas, such as the ground plane, require low-loss RF ground systems in order to perform efficiently. Center-fed  $\lambda/2$  antennas, such as the dipole, do not. However, good grounding is advisable for any antenna system in order to minimize interference and electrical hazards.

### Electrical versus RF Ground

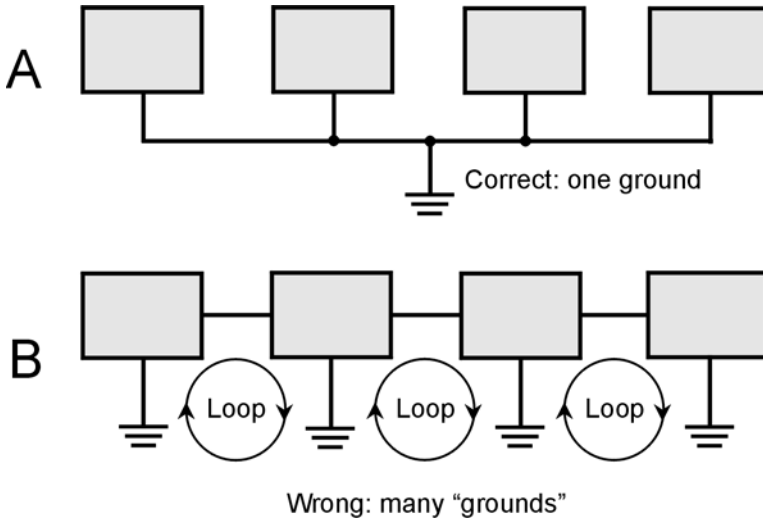
Electrical grounding is important for personal safety. It can help protect equipment from damage if lightning strikes in the vicinity. It also minimizes the risk of *electromagnetic interference* (EMI) to and from radio equipment. In a three-wire electrical system, the ground prong on the plug should never be defeated, because such modification can result in dangerous voltages appearing on exposed metal surfaces.

A good RF ground system can help minimize EMI, even if it is not necessary for efficient antenna operation. Figure 27-5 shows a proper RF ground scheme (A) and an improper one (B). In a good RF ground system, each device is connected to a common *ground bus*, which in turn runs to the earth ground through a single conductor. This conductor should be as short as possible. A poor ground system contains *ground loops* that can act like loop antennas and increase the risk of EMI.

### Radials and the Counterpoise

With a surface-mounted vertical antenna, there should be as many radials as possible, and they should be as long as possible. They can lie on the surface or be buried a few inches underground. The greater the number of radials of a given length, the better the antenna will work. Also, the longer the radials for a given number, the better. The radials should all converge toward, and be connected to, a ground rod at the feed point.

A *counterpoise* is a means of obtaining an RF ground or ground plane without a direct earth-ground connection. A grid of wires, a screen, or a metal sheet is placed above the surface and oriented horizontally, to provide *capacitive coupling* to the earth. This minimizes RF ground loss. Ideally, the radius of a counterpoise should be at least  $\lambda/4$  at the lowest operating frequency.



27-5 At A, the correct method for grounding multiple units. At B, an incorrect method creates RF ground loops.

## Gain and Directivity

The *power gain* of a transmitting antenna is the ratio of the maximum *effective radiated power* (ERP) to the actual RF power applied at the feed point. Power gain is expressed in decibels (dB).

Suppose the ERP, in watts, for a given antenna is  $P_{\text{ERP}}$ , and the applied power, also in watts, is  $P$ . Then the following equation holds:

$$\text{Power gain (dB)} = 10 \log_{10} (P_{\text{ERP}}/P)$$

Power gain is always measured in the favored direction or directions of an antenna. These are the directions in which the antenna performs the best.

For power gain to be defined, a *reference antenna* must be chosen with a gain that is defined as 0 dB. This reference antenna is usually a  $\lambda/2$  dipole in free space. Power-gain figures taken with respect to a dipole (in its favored directions) are expressed in units called dBd. The reference antenna for power-gain measurements can also be an *isotropic antenna*, which theoretically radiates and receives equally well in all directions in three dimensions. In this case, units of power gain are called dBi. For any given antenna, the power gains in dBd and dBi are different by approximately 2.15 dB:

$$\text{Power gain (dBi)} = 2.15 + \text{Power gain (dBd)}$$

### Directivity Plots

Antenna radiation and response patterns are represented by plots such as those shown in Fig. 27-6. The location of the antenna is assumed to be at the center (or *origin*) of a *polar coordinate system*. The

greater the radiation or reception capability of the antenna in a certain direction, the farther from the center the points on the chart are plotted.

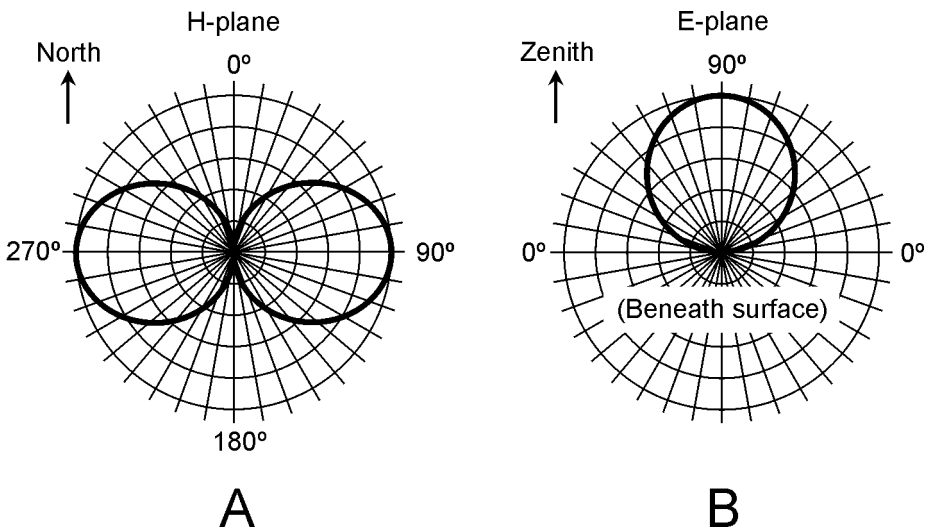
A dipole antenna, oriented horizontally so that its conductor runs in a north-south direction, has a *horizontal plane* (or *H plane*) pattern similar to that in Fig. 27-6A. The *elevation plane* (or *E plane*) pattern depends on the height of the antenna above effective ground at the viewing angle. With the dipole oriented so that its conductor runs perpendicular to the page, and the antenna  $\frac{1}{4}$  wavelength above effective ground, the E plane antenna pattern for a half-wave dipole resembles the graph shown at B.

## Forward Gain

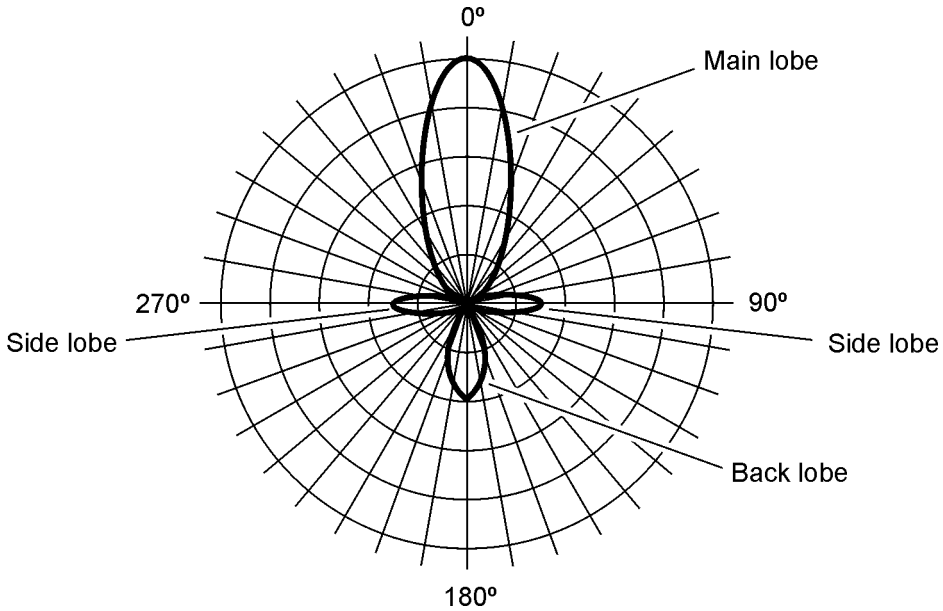
*Forward gain* is expressed in terms of the ERP in the *main lobe* (favored direction) of a *unidirectional* (one-directional) antenna compared with the ERP from a reference antenna, usually a half-wave dipole, in its favored directions. This gain is calculated and defined in dBd at microwave frequencies; large dish antennas can have forward gain upward of 35 dBd. In general, as the wavelength decreases (the frequency gets higher), it becomes easier to obtain high forward gain figures.

## Front-to-Back Ratio

The *front-to-back* (f/b) *ratio* of a unidirectional antenna is an expression of the concentration of radiation/response in the main lobe, relative to the direction opposite the center of the main lobe. Figure 27-7 shows a hypothetical directivity plot for a unidirectional antenna pointed north. The outer circle depicts the RF *field strength* in the direction of the center of the main lobe, and represents 0 dB. The next smaller circle represents a field strength 5 dB down with respect to the main lobe. Continuing inward, circles represent 10 dB down, 15 dB down, and 20 dB down. The origin represents 25 dB down, and also shows the location of the antenna. The f/b ratio is found, in this case, by comparing the signal levels between north (azimuth  $0^\circ$ ) and south (azimuth  $180^\circ$ ).



27-6 Directivity plots for a dipole antenna. At A, the H plane (horizontal plane) plot; at B, the E plane (elevation plane) plot.



27-7 Directivity plot for a hypothetical antenna. Front-to-back and front-to-side ratios can be determined from such a graph. This plot is in the H plane.

### Front-to-Side Ratio

The *front-to-side (f/s) ratio* is another expression of the directivity of an antenna system. The term applies to unidirectional antennas, and also to *bidirectional* antennas. The *f/s* ratio is expressed in decibels (dBd), just as is the *f/b* ratio. The EM field strength in the favored direction is compared with the field strength at right angles to the favored direction. An example is shown in Fig. 27-7. The *f/s* ratios are found, in this case, by comparing the signal levels between north and east (right-hand *f/s*), or between north and west (left-hand *f/s*). The right-hand and left-hand *f/s* ratios are usually the same in theory, although they can differ slightly in practice.

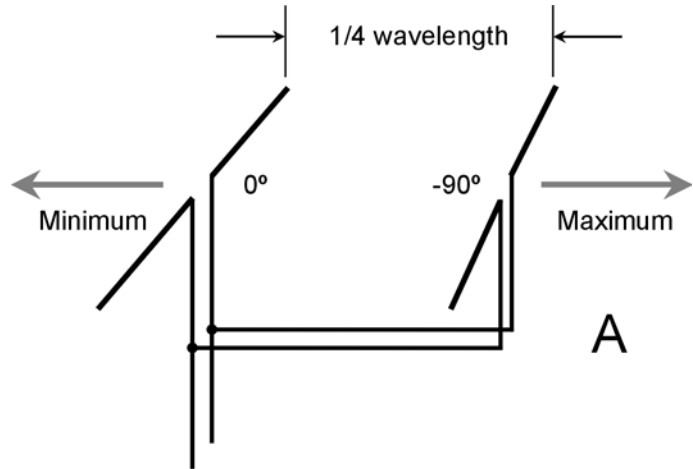
## Phased Arrays

A *phased array* uses two or more *driven elements* (radiators connected directly to the feed line) to produce gain in some directions at the expense of other directions.

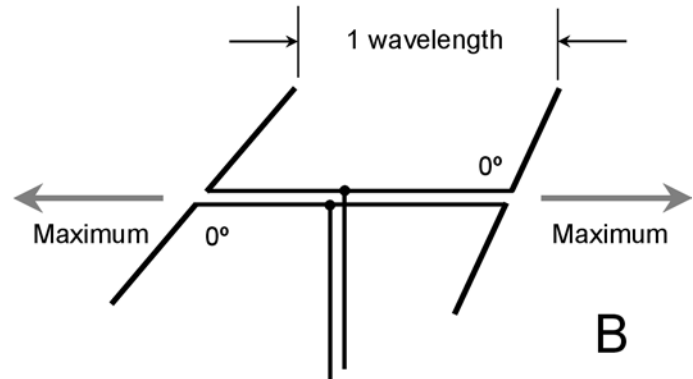
### End-Fire Array

A typical *end-fire array* consists of two parallel  $\lambda/2$  dipoles fed  $90^\circ$  out of phase and spaced  $\lambda/4$  apart (Fig. 27-8A). This produces a unidirectional radiation pattern. Or, the two elements can be driven in phase and spaced at a separation of  $\lambda$  (Fig. 27-8B). This results in a bidirectional radiation pattern. In the phasing system, the branches of the transmission line must be cut to precisely the correct lengths, and the velocity factor of the line must be known and be taken into account.





27-8 At A, a unidirectional phased system. At B, a bidirectional phased system.

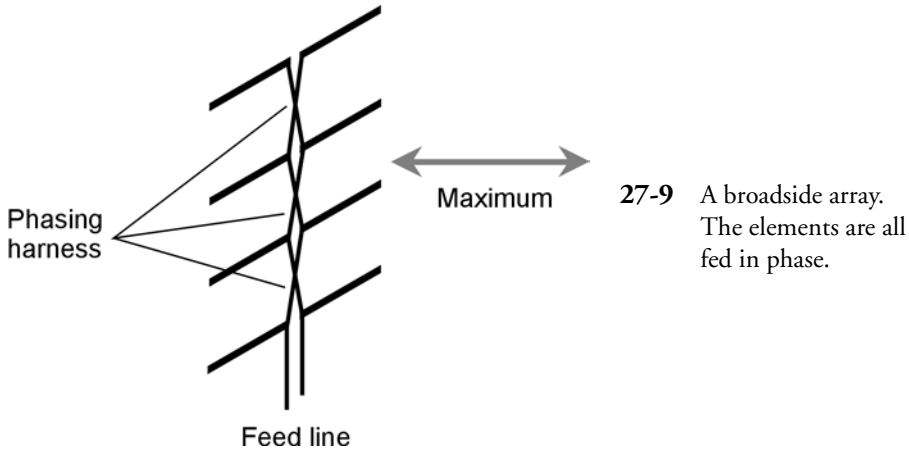


## Longwire

A wire antenna measuring  $\lambda$  or more, and fed at a high-current point or at one end, is a *longwire antenna*. A longwire antenna offers gain over a half-wave dipole. As the wire is made longer, the main lobes get more nearly in line with the antenna, and their amplitudes increase. The gain is a function of the length of the antenna; the longer the wire, the greater the gain. A longwire antenna must be as straight as possible for proper operation.

## Broadside Array

Figure 27-9 shows the geometric arrangement of a *broadside array*. The driven elements can each consist of a single radiator, as shown in the figure, or they can consist of more complex antennas with directive properties. In any case, all the elements are identical. If a reflecting screen is placed behind the array of dipoles in Fig. 27-9, the system is known as a *billboard antenna*. The directional properties depend on the number of elements, whether or not the elements have gain themselves, and on the spacing among the elements. In general, the larger the number of elements, the greater the forward gain, and the greater the  $f/b$  and  $f/s$  ratios.



## Parasitic Arrays

*Parasitic arrays* are used at frequencies ranging from approximately 5 MHz into the microwave range for obtaining directivity and forward gain. Examples include the *Yagi antenna* and the *quad antenna*.

### Concept

A *parasitic element* is a conductor that forms an important part of an antenna system, but is not directly connected to the feed line. Parasitic elements operate by means of *EM coupling* to the driven element. When gain is produced in the direction of the parasitic element, the element is a *director*. When gain is produced in the direction opposite the parasitic element, the element is a *reflector*. Directors are a few percent shorter than the driven element; reflectors are a few percent longer.

### Yagi

The *Yagi antenna*, sometimes called a “beam,” is an array of parallel, straight elements. A two-element Yagi is formed by placing a director or a reflector parallel to, and a specific distance away from, a single  $\lambda/2$  driven element. The optimum spacing for a driven-element/director Yagi is  $0.1\lambda$  to  $0.2\lambda$ , with the director tuned 5 percent to 10 percent higher than the resonant frequency of the driven element. The optimum spacing for a driven-element/reflector Yagi is  $0.15\lambda$  to  $0.2\lambda$ , with the reflector tuned 5 percent to 10 percent lower than the resonant frequency of the driven element. The gain of a well-designed *two-element Yagi* is approximately 5 dBd.

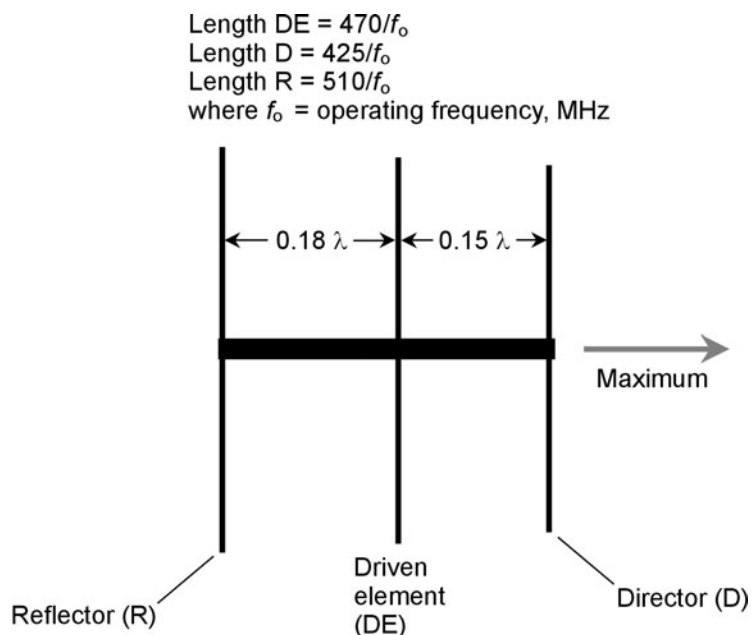
A Yagi with one director and one reflector, along with the driven element, increases the gain and *f/b* ratio compared with a two-element Yagi. An optimally designed *three-element Yagi* has approximately 7 dBd gain. An example is shown in Fig. 27-10. (This drawing should not be used as an engineering blueprint.)

The gain, *f/b* ratio, and *f/s* ratio of a properly designed Yagi antenna all increase as elements are added. This is usually done by placing extra directors in front of a three-element Yagi. Each director is slightly shorter than its predecessor.

### Quad

A *quad antenna* operates according to the same principles as the Yagi, except full-wavelength loops are used instead of straight half-wavelength elements.

**27-10** A three-element Yagi antenna. Dimensions are discussed in the text.



A *two-element quad* can consist of a driven element and a reflector, or it can have a driven element and a director. A *three-element quad* has one driven element, one director, and one reflector. The director has a perimeter of  $0.95\lambda$  to  $0.97\lambda$ , the driven element has a perimeter of exactly  $\lambda$ , and the reflector has a perimeter of  $1.03\lambda$  to  $1.05\lambda$ . These dimensions, as all antenna element dimensions, are electrical dimensions (taking the velocity factor of wire or tubing into account), and not free-space dimensions.

Additional directors can be added to the basic three-element quad design to form quads having any desired numbers of elements. The gain increases as the number of elements increases. Each succeeding director is slightly shorter than its predecessor. Long quad antennas are practical at frequencies above 100 MHz. At lower frequencies, their construction tends to be unwieldy.

## Antennas for Ultrahigh and Microwave Frequencies

At ultrahigh frequencies (UHF) and microwave frequencies, high-gain antennas are reasonable in size because the wavelengths are short.

### Waveguides

In Chap. 25, waveguides were briefly mentioned as an option for feeding antenna systems at UHF and microwave frequencies. Here's a little more about them.

A waveguide is a hollow metal pipe having a rectangular or circular cross section. The EM field travels down the pipe, provided that the wavelength is short enough (or the cross-sectional dimensions of the pipe are large enough). In order to efficiently propagate an EM field, a *rectangular waveguide* must have height and width that both measure at least  $0.5\lambda$ , and preferably more than  $0.7\lambda$ . A *circular waveguide* should be at least  $0.6\lambda$  in diameter, and preferably  $0.7\lambda$  or more.

The characteristic impedance ( $Z_0$ ) of a waveguide varies with frequency. In this sense, it differs from coaxial or parallel-wire lines, whose  $Z_0$  values are generally independent of the frequency.

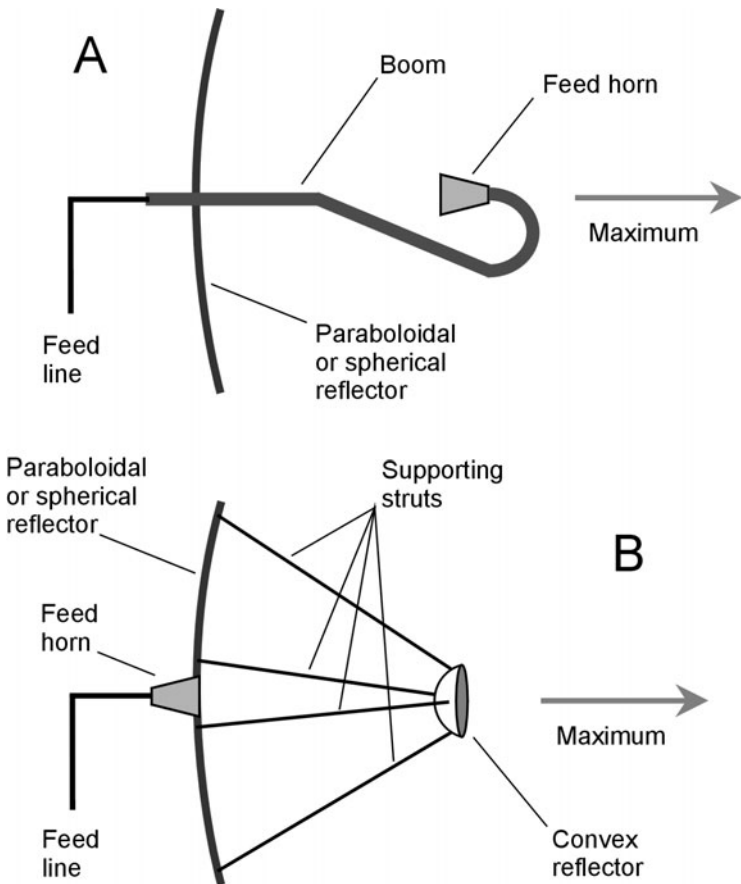
A properly installed and well-maintained waveguide is an exceptional transmission line because dry air has practically no loss, even at UHF and microwave frequencies. But it is important that the interior of a waveguide be kept free from dirt, dust, insects, spiderwebs, and condensation. Even a small obstruction can seriously degrade the performance and cause the waveguide to become *lossy*.

The main limitation of a waveguide, from a practical standpoint, is inflexibility. You can't run a waveguide from one point to another in a haphazard fashion, as you can do with coaxial cable. Bends or turns in a waveguide present a particular problem, because they must be made gradually. Installing a waveguide for a UHF or microwave antenna is a little like putting in the conduit for a new electrical circuit in a home or business. It's a significant construction project!

Waveguides are impractical for use at frequencies below approximately 300 MHz, because the required cross-sectional dimensions become prohibitively large.

**Dish**

A *dish antenna* must be correctly shaped and precisely aligned. The most efficient shape, especially at the shortest wavelengths, is a *paraboloidal reflector*. However, a *spherical reflector*, having



27-11 Dish antennas with conventional feed (A) and Cassegrain feed (B).

the shape of a section of a sphere, can also work well. The feed system consists of a coaxial line or waveguide from the receiver and/or transmitter, and a horn or helical driven element at the focal point of the reflector. *Conventional dish feed* is shown in Fig. 27-11A. *Cassegrain dish feed* is shown in Fig. 27-11B.

The larger the diameter of the reflector in wavelengths, the greater the gain, the  $f/b$  ratio, and the  $f/s$  ratio, and the narrower the main lobe. A dish antenna must be at least several wavelengths in diameter for proper operation. The reflecting element can be sheet metal, a screen, or a wire mesh. If a screen or mesh is used, the spacing between the wires must be a small fraction of a wavelength.

## Helical

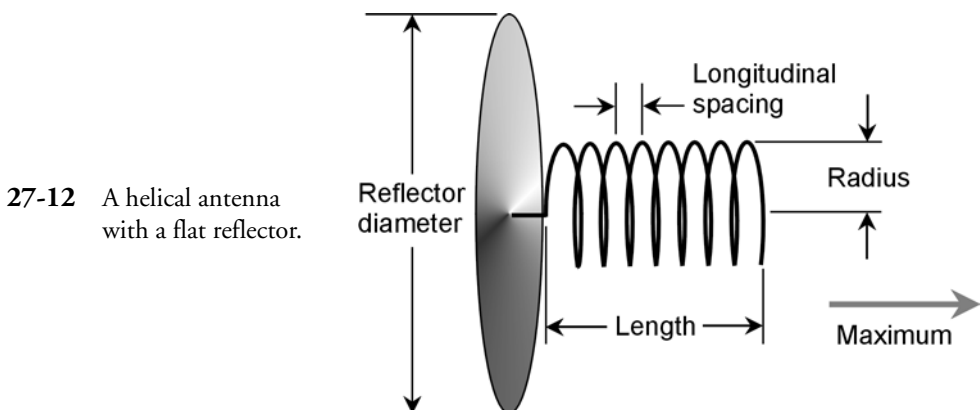
A *helical antenna* is a circularly polarized, high-gain, unidirectional antenna. A typical helical antenna is shown in Fig. 27-12. The reflector diameter should be at least  $0.8\lambda$  at the lowest operating frequency. The radius of the helix should be approximately  $0.17\lambda$  at the center of the intended operating frequency range. The longitudinal spacing between helix turns should be approximately  $0.25\lambda$  in the center of the operating frequency range. The overall length of the helix should be at least  $\lambda$  at the lowest operating frequency. A helical antenna can provide about 15 dBd forward gain. Helical antennas are sometimes used in space communications systems.

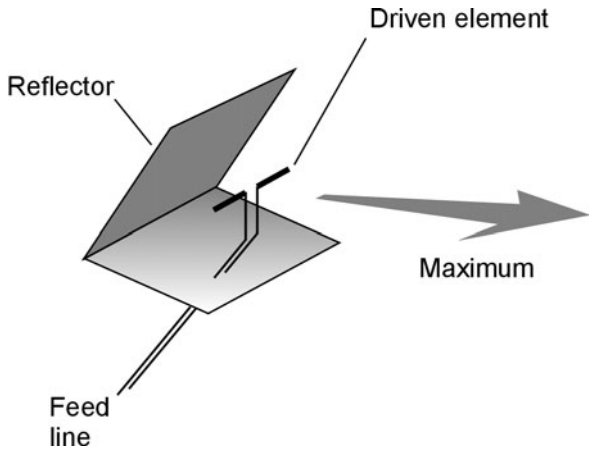
## Corner Reflector

A *corner reflector*, employed with a  $\lambda/2$  dipole driven element, is illustrated in Fig. 27-13. This provides some gain over a  $\lambda/2$  dipole by itself. The reflector is made of wire mesh, screen, or sheet metal. The *flare angle* of the reflecting element is approximately  $90^\circ$ . Corner reflectors are widely used in terrestrial communications at UHF and microwave frequencies. Several  $\lambda/2$  dipoles can be fed in phase and placed along a common axis with a single, elongated reflector, forming a *collinear corner reflector array*.

## Horn

The *horn antenna* is shaped like a squared-off trumpet or trombone horn. It provides a unidirectional radiation and response pattern, with the favored direction coincident with the opening of the horn. The feed line is a waveguide that joins the antenna at the narrowest point (throat) of the horn.





**27-13** A corner reflector with a dipole as the driven element.

Horns are sometimes used all by themselves, but they are also used to feed large dish antennas. This optimizes the  $f/s$  ratio by minimizing extraneous radiation and response that occurs if a dipole is used as the driven element for the dish.

## Safety

Antennas should never be placed in such a way that they can fall or blow down on power lines. Also, it should not be possible for power lines to fall or blow down on an antenna.

Wireless equipment having outdoor antennas should not be used during thundershowers, or when lightning is anywhere in the vicinity. Antenna construction and maintenance should never be undertaken when lightning is visible or thunder can be heard, even if a storm appears to be many miles away. Ideally, antennas should be disconnected from electronic equipment, and connected to a substantial earth ground, at all times when the equipment is not in use.

Tower and antenna climbing is a job for professionals. Under no circumstances should an inexperienced person attempt to climb such a structure.

Indoor transmitting antennas can expose operating personnel to EM field energy. The extent of the hazard, if any, posed by such exposure has not been firmly established. However, there is sufficient concern to warrant checking the latest publications on the topic.

For detailed information concerning antenna safety, consult a professional antenna engineer and/or a comprehensive text on antenna design and construction.

## Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. Imagine a center-fed, straight wire antenna in free space, whose length can be varied at will. At which of the following lengths is the radiation resistance at the feed point the highest?

- (a)  $\lambda/4$
- (b)  $\lambda$

- (c)  $\lambda/3$   
(d)  $\lambda/2$
2. What is the approximate length, in meters (m), of a center-fed,  $\lambda/2$  wire antenna at a frequency of 18.1 MHz? Assume the velocity factor is typical for wire.
- (a) 7.90 m  
(b) 8.29 m  
(c) 13.6 m  
(d) 25.9 m
3. What is the approximate height, in feet (ft), of a ground-mounted,  $\lambda/4$  vertical antenna, made of aluminum tubing, at a frequency of 7.025 MHz? Assume the velocity factor is 92 percent.
- (a) 64.43 ft  
(b) 32.22 ft  
(c) 19.64 ft  
(d) 9.822 ft
4. In order to obtain reasonable efficiency with a  $\lambda/4$  vertical antenna mounted over poorly conducting soil at a frequency of 10 MHz, it is necessary to
- (a) make the antenna out of large-diameter aluminum or copper tubing.  
(b) suspend the antenna from a kite or balloon.  
(c) feed the antenna with a waveguide.  
(d) minimize the loss resistance by installing numerous ground radials.
5. A driven element in a parasitic array is
- (a) grounded for dc.  
(b) a loop with a circumference of  $\lambda/2$ .  
(c) connected to the feed line.  
(d) a straight length of wire.
6. In a zepp antenna, feed-line radiation can be minimized by
- (a) using the antenna only at frequencies where its length is a whole-number multiple of  $\lambda/2$ .  
(b) using the antenna only at frequencies where its length is a whole-number multiple of  $\lambda/4$ .  
(c) using the antenna only at frequencies where its length is an even-number multiple of  $\lambda/10$ .  
(d) using the antenna only at frequencies where its length is an odd-number multiple of  $\lambda/4$ .
7. Consider a single-turn, closed loop having a circumference of 95 ft. Suppose this loop is circular, lies entirely in a single plane, is fed with open-wire transmission line and an antenna tuner, and is operated at 10 MHz. The maximum radiation and response for this loop can be expected to occur
- (a) in the plane containing the loop.  
(b) along a line perpendicular to the plane containing the loop.

- (c) along a line coinciding with the path of the feed line.
  - (d) equally in all directions in three-dimensional space.
8. Suppose an antenna system has a radiation resistance of  $40\ \Omega$  at 8.5 MHz. What is the efficiency of this system?
- (a) 25 percent
  - (b) 40 percent
  - (c) 80 percent
  - (d) Impossible to determine without more information
9. Suppose an antenna system has a radiation resistance of  $40\ \Omega$  and a loss resistance of  $10\ \Omega$ . What is the efficiency of this system?
- (a) 25 percent
  - (b) 40 percent
  - (c) 80 percent
  - (d) Impossible to determine without more information
10. The null in a loopstick antenna exists
- (a) in the plane perpendicular to the coil axis.
  - (b) in the line that coincides with the coil axis.
  - (c) at a  $45^\circ$  angle with respect to the coil axis.
  - (d) nowhere, because a loopstick is an isotropic antenna.
11. As elements are added to a properly designed Yagi antenna,
- (a) the forward gain decreases.
  - (b) the  $f/b$  ratio decreases.
  - (c) the  $f/s$  ratio increases.
  - (d) None of the above apply.
12. When the radiation resistance in an antenna system represents most of the total system resistance,
- (a) the system can be expected to have high loss.
  - (b) the system can be expected to have high efficiency.
  - (c) the system can be expected to have a high  $f/s$  ratio.
  - (d) the system can be expected to have a high  $f/b$  ratio.
13. What state of affairs is optimal in a waveguide?
- (a) A colony of spiders inside
  - (b) A cross-sectional measurement of less than  $\lambda/4$
  - (c) Operation at frequencies below 3 MHz
  - (d) None of the above



14. Consider a single-turn, closed loop having a circumference of 65 ft. Suppose this loop is circular, lies entirely in a single plane, is fed with open-wire transmission line and an antenna tuner, and is operated at 7 MHz. The maximum radiation and response for this loop can be expected to occur
- (a) in the plane containing the loop.
  - (b) along a line perpendicular to the plane containing the loop.
  - (c) along a line coinciding with the path of the feed line.
  - (d) equally in all directions in three-dimensional space.
15. Suppose the total resistance in an antenna system is  $80\ \Omega$ , and the loss resistance is  $20\ \Omega$ . The efficiency of the system is
- (a) 80 percent.
  - (b) 75 percent.
  - (c) 25 percent.
  - (d) impossible to determine without more information.
16. In a parasitic array, a director is usually
- (a) tuned to a slightly higher frequency than the driven element.
  - (b) connected to a system of ground radials.
  - (c) a closed loop that is grounded for RF.
  - (d) a straight length of wire or tubing.
17. A J pole is a modified form of
- (a) ground-plane antenna.
  - (b) folded dipole antenna.
  - (c) quad antenna.
  - (d) zepp antenna.
18. Which of the following antenna types is well suited to use with parallel-wire transmission line?
- (a) The ground plane
  - (b) The ground-mounted vertical
  - (c) The folded dipole
  - (d) None of the above
19. Which of the following antenna types is not designed for transmitting at 10 GHz?
- (a) A horn antenna
  - (b) A dish antenna
  - (c) A zepp antenna
  - (d) A helical antenna

20. A vertical antenna measuring  $\lambda/2$  in height does not require a ground system with low resistance because

- (a) its loss resistance is high.
- (b) its radiation resistance is low.
- (c) its main lobe is sharp.
- (d) None of the above apply.

## 28 CHAPTER

# Integrated Circuits

*INTEGRATED CIRCUITS* (ICs), ALSO CALLED *CHIPS*, HAVE STIMULATED AS MUCH EVOLUTION AS ANY single development in the history of electronic technology. Most ICs look like gray or black plastic boxes with protruding pins. The schematic symbols are triangles or rectangles with lines emerging from them, labeled according to their functions.

## Advantages of IC Technology

Integrated circuits have advantages over *discrete components* (individual transistors, diodes, capacitors, and resistors). Here are some of the most important considerations.

### Compactness

An obvious asset of IC design is economy of space. They are far more compact than equivalent circuits made from discrete components. A corollary to this is the fact that more complex circuits can be built, and kept down to a reasonable size, using ICs as compared with discrete components.

### High Speed

The interconnections among the components within an IC are physically tiny, making high switching speeds possible. Electric currents travel fast, but not instantaneously. The less time charge carriers need to get from component *X* to component *Y*, in general, the more computations are possible within a given span of time, and the less time is needed for complex operations.

### Low Power Consumption

Another advantage of ICs is the fact that they use less power than equivalent discrete-component circuits. This is crucial if batteries are to be used for operation. Because ICs use so little current, they produce less heat than their discrete-component equivalents. This translates into better efficiency. It also minimizes the problems that bedevil equipment that gets hot with use, such as frequency drift and generation of internal noise.

## Reliability

Integrated circuits fail less often, per component-hour of use, than systems that use discrete components. This is because all the component interconnections are sealed within the IC case, preventing corrosion or the intrusion of dust. The reduced failure rate translates into less *downtime*, or time during which the equipment is out of service for repairs.

## Ease of Maintenance

Integrated-circuit technology lowers maintenance costs. Repair procedures are simplified when failures occur. Many appliances use sockets for ICs, and replacement is simply a matter of finding the faulty IC, unplugging it, and plugging in a new one. Special desoldering equipment is used with appliances having ICs soldered directly to the circuit boards.

## Modular Construction

Modern IC appliances use *modular construction*. In this scheme, individual ICs perform defined functions within a circuit board; the circuit board or *card*, in turn, fits into a socket and has a specific purpose. Computers, programmed with customized software, are used by repair technicians to locate the faulty card in an appliance. The whole card can be pulled and replaced, getting the appliance back to the consumer in the shortest possible time. Then a computer can be used to troubleshoot the faulty card, getting the card ready for use in the next appliance that happens to come along with a failure in the same card.

# Limitations of IC Technology

No technological advancement ever comes without a downside. Integrated circuits have some limitations that must be considered when designing an electronic system.

## Inductors Impractical

While some components are easy to fabricate onto chips, other components defy the IC manufacturing process. Inductances, except for extremely low values, are an example. Devices using ICs must generally be designed to work with discrete inductors, external to the ICs themselves. However, resistance-capacitance (RC) circuits are capable of doing most things that inductance-capacitance (LC) circuits can do, and these can exist inside an IC.

## High Power Impossible

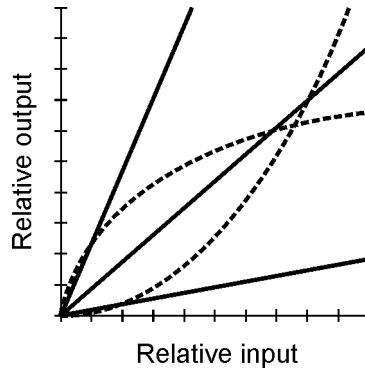
The small size and low current consumption of ICs comes with an inherent limitation: high-power amplifiers cannot, in general, be fabricated onto chips. High power necessitates a certain amount of physical mass and volume. High-power devices and systems invariably generate large amounts of heat, which must be efficiently radiated or conducted away.

## Linear ICs

A *linear IC* is used to process *analog signals* such as voices, music, and radio transmissions.

The term *linear* arises from the fact that, in general, the amplification factor is constant as the input amplitude varies. That is, the output signal strength is a *linear function* of the input signal strength (Fig. 28-1).

**28-1** In a linear IC, the relative output is a linear (straight-line) function of the relative input. The solid lines show examples of linear IC characteristics. The dashed curves show functions not characteristic of properly operating linear ICs.



### Operational Amplifier

An *operational amplifier*, or *op amp*, is a specialized linear IC that consists of several bipolar transistors, resistors, diodes, and capacitors, interconnected so that high gain is possible over a wide range of frequencies. An op amp might comprise an entire IC. Or, an IC might consist of two or more op amps. Thus, you'll sometimes hear of *dual op amps* or *quad op amps*. Some ICs have op amps in addition to other circuits.

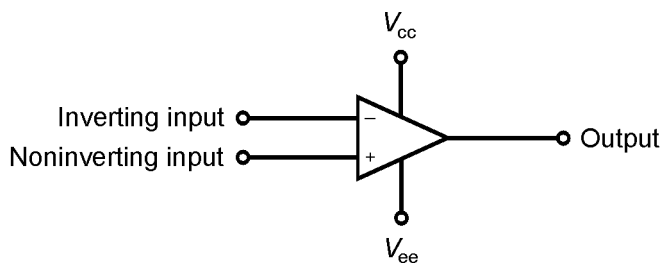
An op amp has two inputs, one *noninverting* and one *inverting*, and one output. When a signal goes into the noninverting input, the output is in phase with the input; when a signal goes into the inverting input, the output is 180° out of phase with the input. An op amp has two power supply connections, one for the emitters of the transistors ( $V_{ee}$ ) and one for the collectors ( $V_{cc}$ ). The usual schematic symbol for an op amp is a triangle. The inputs, output, and power-supply connections are drawn as lines emerging from it (Fig. 28-2).

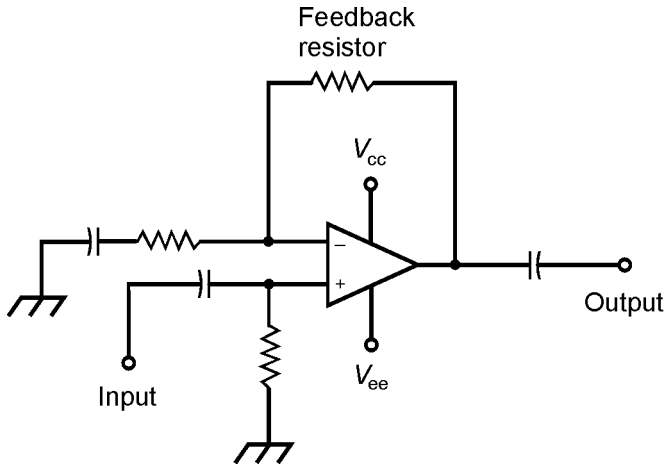
The gain of an op amp is determined by one or more external resistors. Normally, a resistor is connected between the output and the inverting input. This is called the *closed-loop configuration*. The feedback is negative (out of phase), causing the gain of the op amp to be less than it would be if there were no feedback (the *open-loop configuration*). Figure 28-3 is a schematic diagram of a non-inverting closed-loop amplifier.

The reason for providing negative feedback in an op amp circuit is the fact that without it, the gain may be too great. Excessive amplifier gain can cause problems. Open-loop op amps are prone to instability, especially at low frequencies. They also generate a lot of internal noise.

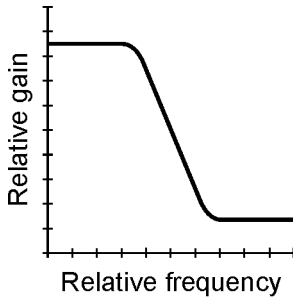
When an RC combination is used in the feedback loop of an op amp, the gain depends on the input-signal frequency. It is possible to get a *lowpass response*, a *highpass response*, a *resonant peak*, or a *resonant notch* using an op amp and various RC feedback arrangements. These four types of responses are shown in Fig. 28-4.

**28-2** Schematic symbol for an op amp. Connections are discussed in the text.

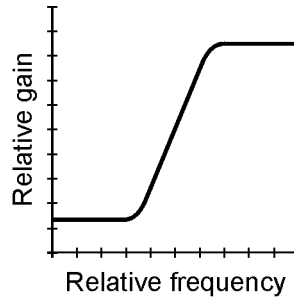




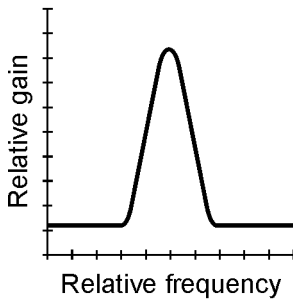
**28-3** A closed-loop op amp circuit with negative feedback. If the feedback resistor is removed, it becomes an open-loop circuit.



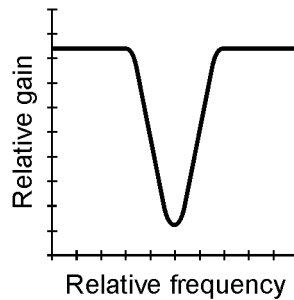
**A**



**B**

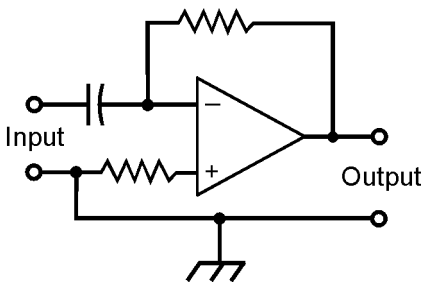


**C**



**D**

**28-4** Gain-versus-frequency response curves. At A, lowpass; at B, highpass; at C, resonant peak; at D, resonant notch.



**28-5** A differentiator circuit that uses an op amp.

## Op Amp Differentiator

A *differentiator* is an electronic circuit whose instantaneous output amplitude is proportional to the rate at which the input amplitude changes. The circuit mathematically *differentiates* the input signal. Op amps can be used as differentiator circuits. An example is shown in Fig. 28-5.

When the input to a differentiator is a constant dc voltage, the output is zero (no signal). When the input amplitude is increasing, the output is a positive dc voltage. When the input decreases, the output is a negative dc voltage. If the input waveform fluctuates periodically (the usual case), the output varies according to the instantaneous rate of change of the input amplitude. This results in an output signal with the same frequency as that of the input signal, although the waveform is often quite different. A pure sine wave input produces a pure sine wave at the output, but the phase is shifted  $90^\circ$  to the left (that is,  $\frac{1}{4}$  cycle earlier in time). Complex input waveforms can produce a wide variety of outputs from a differentiator.

## Op Amp Integrator

An *integrator* is an electronic circuit whose instantaneous output amplitude is proportional to the accumulated input signal amplitude as a function of time. The circuit mathematically *integrates* the input signal. The function of an integrator is basically the inverse, or opposite, of the function of a differentiator circuit. Figure 28-6 shows how an op amp can be connected to obtain an integrator.

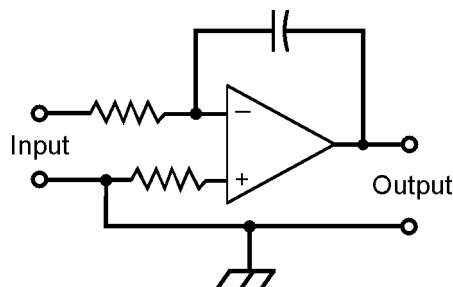
If an integrator circuit is supplied with an input signal waveform that fluctuates periodically (the usual case), the output voltage varies according to the *integral*, or *antiderivative*, of the input voltage. This results in an output signal with the same frequency as that of the input signal, although the waveform is likely to be different. A pure sine wave input produces a pure sine wave output, but the phase is shifted  $90^\circ$  to the right (that is,  $\frac{1}{4}$  cycle later in time). Complex input waveforms can produce many types of output waveforms.

An indefinite rise, either negatively or positively, in output voltage cannot occur in a practical integrator. If the mathematical integral (in pure theory) of an input function is an endlessly increasing output function, the actual output voltage rises to a certain maximum, either positive or negative, and stays there. This maximum is less than or equal to the power supply or battery voltage.

## Voltage Regulator

A *voltage regulator IC* acts to control the output voltage of a power supply. This is important with precision electronic equipment. These ICs are available with various voltage and current ratings. Typical voltage regulator ICs have three terminals, and because of this, they are sometimes mistaken for power transistors.

28-6 An integrator circuit that uses an op amp.



## Timer

A *timer IC* is a specialized oscillator that produces a delayed output. The delay is adjustable to suit the needs of a particular device. The delay is generated by counting the number of oscillator pulses; the length of the delay can be adjusted by means of external resistors and capacitors. Timers are commonly used in circuits such as digital frequency counters, where a precise time interval or window must be provided.

## Multiplexer

A *multiplexer IC* allows several different signals to be combined in a single channel by means of a process called *multiplexing*. An analog multiplexer can also be used in reverse; then it works as a *demultiplexer*. Thus, you'll sometimes hear or read about a *multiplexer/demultiplexer IC*.

## Comparator

A *comparator IC* has two inputs. It compares the voltages at the two inputs, which are called A and B. If the voltage at input A is significantly higher than the voltage at input B, the output is about +5 V. This is logic 1, or high. If the voltage at input A is lower than or equal to the voltage at input B, the output voltage is about +2 V. This is designated as logic 0, or low.

Voltage comparators are available for a variety of applications. Some can switch between low and high states at a rapid rate of speed, while others are slow. Some have low input impedance, and others have high impedance. Some are intended for AF or low-frequency RF use; others are fabricated for video or high-frequency RF applications. Voltage comparators can be used to actuate, or *trigger*, other devices such as relays and electronic switching circuits.

# Digital ICs

A *digital IC*, also sometimes called a *digital-logic IC*, operates using two discrete states: high (logic 1) and low (logic 0). Digital logic is discussed in Chap. 26. Digital ICs contain massive arrays of logic gates that perform Boolean operations at high speed.

## Transistor-Transistor Logic

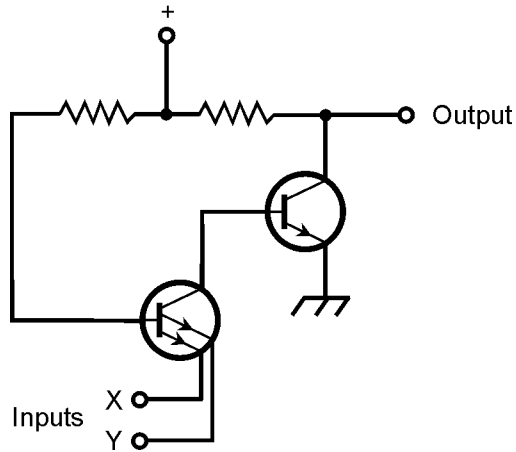
In *transistor-transistor logic* (TTL), arrays of bipolar transistors, some with multiple emitters, operate on dc pulses. This technology has several variants, some of which date back to around 1970. A basic TTL gate is illustrated in Fig. 28-7. This gate uses two NPN bipolar transistors, one of which is a *dual-emitter* device. The transistors are always either completely cut off, or else completely saturated. Because of this, TTL is relatively immune to external noise.

## Emitter-Coupled Logic

Another bipolar-transistor logic form is known as *emitter-coupled logic* (ECL). In an ECL device, the transistors are not operated at saturation, as they are with TTL. This increases the speed. But noise pulses have a greater effect in ECL, because unsaturated transistors are sensitive to, and can actually amplify, external signals and noise. The schematic of Fig. 28-8 shows a basic ECL gate using four NPN bipolar transistors.



- 28-7** A transistor-transistor logic (TTL) gate. In this gate, two NPN bipolar transistors are used. Note that one of the transistors has two emitters.

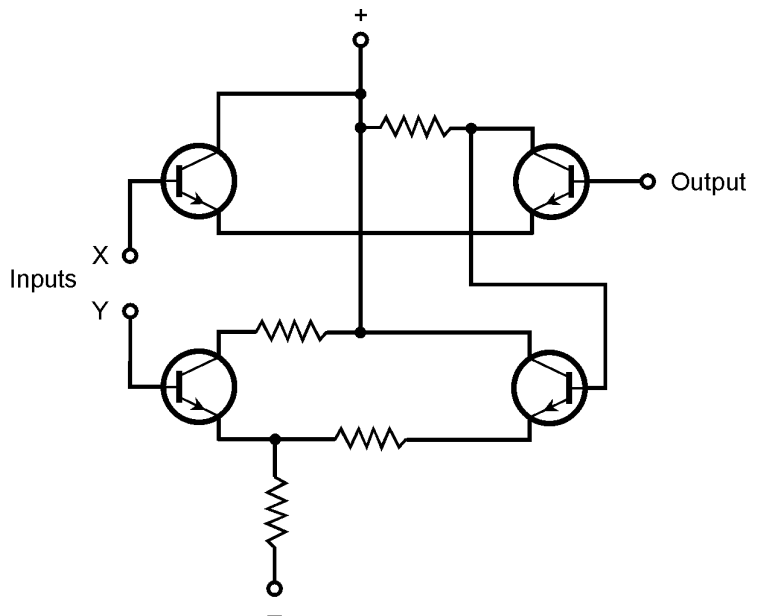


### Metal-Oxide-Semiconductor Logic

Digital ICs can also be constructed using metal-oxide-semiconductor (MOS) devices. *N-channel MOS (NMOS) logic*, pronounced “EN-moss logic,” offers simplicity of design, along with high operating speed. *P-channel MOS logic*, pronounced “PEA-moss logic,” is similar to NMOS logic, but the speed is slower. An NMOS or PMOS digital IC is like a circuit that uses only N-channel MOSFETs, or only P-channel MOSFETs, respectively.

*Complementary-metal-oxide-semiconductor (CMOS) logic*, pronounced “SEA-moss logic,” employs both N-type and P-type silicon on a single chip. This is analogous to using both N-channel

- 28-8** An emitter-coupled logic (ECL) gate using four NPN bipolar transistors.



and P-channel MOSFETs in a circuit. The advantages of CMOS technology include extremely low current drain, high operating speed, and immunity to noise.

All forms of MOS logic ICs require care in handling to prevent destruction by electrostatic discharges. The precautions are the same as those that are required when handling MOSFETs. All technical personnel who work with the devices should be grounded. This can be achieved by having technicians wear metal wrist straps connected to a good electrical ground, and by ensuring that the relative humidity in the lab is not allowed to get too low. When MOS ICs are stored, the pins should be pushed into special conductive foam that is manufactured for that purpose.

## Component Density

The number of elements per chip in an IC is called the *component density*. There has been a steady increase in the number of components that can be fabricated on a single chip. There is an absolute limit on component density, imposed by the atomic structure of the semiconductor material—literally, the size of the atoms! Technology has begun to approach that barrier.

### Small Scale

In *small scale integration* (SSI), there are fewer than 10 transistors on a chip. These types of ICs can carry the largest currents, and can be useful in voltage regulation and other moderate-power applications.

### Medium Scale

In *medium scale integration* (MSI), there are 10 to 100 transistors per chip. This allows for considerable miniaturization, but it is not a high level of component density, relatively speaking. An advantage of MSI (in a few applications) is that fairly large currents can be carried by the individual gates. Both bipolar and MOS technologies can be adapted to MSI.

### Large Scale

In *large scale integration* (LSI), there are 100 to 1000 transistors per semiconductor chip. This is an *order of magnitude* (a factor of 10 times) more dense than MSI. Electronic wristwatches, single-chip calculators, and simple microcomputers are examples of devices using LSI ICs.

### Very-Large-Scale Devices

*Very-large-scale integration* (VLSI) devices have from 1000 to 1,000,000 transistors per chip. This can be up to three orders of magnitude more dense than LSI. Microcomputers and *memory* ICs are made using VLSI.

### Ultra-Large-Scale Devices

You might sometimes hear of *ultra-large-scale integration* (ULSI). Devices of this kind have more than 1,000,000 transistors per chip. The principal uses for this technology are in the fields of high-level computing, supercomputing, robotics, and artificial intelligence (AI).

# IC Memory

Binary digital data, in the form of high and low levels (logic ones and zeros), can be stored in *memory ICs*. These devices can take various physical forms.

## Random Access

A *random access memory* (RAM) chip stores binary data in *arrays*. The data can be *addressed* (selected) from anywhere in the matrix. Data is easily changed and stored back in RAM, in whole or in part. A RAM chip is sometimes called a *read/write memory*.

An example of the use of RAM is a word-processing computer file that you are actively working on. This paragraph, this chapter, and in fact the whole text of this book was written in semiconductor RAM in small sections before being incrementally stored on the computer hard drive, and ultimately on external media.

There are two major categories of RAM: *dynamic RAM* (DRAM) and *static RAM* (SRAM). A DRAM chip contains transistors and capacitors, and data is stored as charges on the capacitors. The charge must be replenished frequently, or it will be lost through discharge. Replenishing is done automatically several hundred times per second. An SRAM chip uses a flip-flop to store the data. This gets rid of the need for constant replenishing of charge, but the tradeoff is that SRAM ICs require more elements than DRAM chips to store a given amount of data.

With any RAM chip, the data in it will vanish when power is removed, unless some provision is made for *memory backup*. The most common means of memory backup is the use of a small cell or battery with a long shelf life. Modern IC memories need so little current to store their data that a *backup battery* lasts as long in the circuit as it would on the shelf. Memory that disappears when power is removed is called *volatile memory*. If memory is retained when power is removed, it is *nonvolatile*.

## Read Only

By contrast to RAM chips, the data in a *read-only memory* (ROM) chip can be easily accessed, in whole or in part, but not easily written over. A standard ROM chip is *programmed* at the factory. This permanent programming is known as *firmware*. But there are also ROM chips that you can program and reprogram yourself. The data contents of ROM chips are generally nonvolatile. That means a power failure is no cause for concern.

An *erasable programmable ROM* (EPROM) chip is an IC whose memory is of the read-only type, but that can be reprogrammed by a certain procedure. It is more difficult to rewrite data in an EPROM than in a RAM; the usual process for erasure involves exposure to ultraviolet (UV). An EPROM IC can be recognized by the presence of a transparent window with a removable cover, through which the UV is focused to erase the data. The IC must be taken from the circuit in which it is used, exposed to UV for several minutes, and then reprogrammed.

The data in some EPROM chips can be erased by electrical means. Such an IC is called an *electrically erasable programmable read-only memory* (EEPROM) chip. These do not have to be removed from the circuit for reprogramming.

# Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. Because of the small size of ICs compared with equivalent circuits made from discrete components,
  - (a) more heat is generated.
  - (b) higher power output is possible.
  - (c) higher switching speeds are attainable.
  - (d) fewer calculations need to be done per unit time.
2. Which of the following is not an advantage of ICs over discrete components?
  - (a) Higher component density
  - (b) Ease of maintenance
  - (c) Lower gain
  - (d) Lower current consumption
3. In which of the following devices would you never find an IC as the main component?
  - (a) The final amplifier in a large radio broadcast transmitter
  - (b) The microprocessor in a personal computer
  - (c) A battery-powered calculator
  - (d) A low-power audio amplifier
4. Which, if any, of the following component types is not practical for direct fabrication on a chip?
  - (a) Resistors
  - (b) Capacitors
  - (c) Diodes
  - (d) All three of the above component types are practical for direct fabrication on a chip.
5. An op amp circuit usually employs negative feedback to
  - (a) maximize the gain.
  - (b) control the gain.
  - (c) allow oscillation over a wide band of frequencies.
  - (d) Forget it! Op amp circuits never have negative feedback.
6. Suppose a communications channel carries several signals at once. Which type of IC might be used to single one of the signals out for reception?
  - (a) An op amp
  - (b) A timer
  - (c) A comparator
  - (d) A multiplexer/demultiplexer
7. Which type of IC is commonly used to determine whether two dc voltages are the same or different?
  - (a) An op amp
  - (b) A timer

- (c) A comparator
  - (d) A multiplexer/demultiplexer
8. Which type of digital IC is least susceptible to noise?
- (a) Transistor-transistor logic
  - (b) Base-coupled logic
  - (c) Emitter-coupled logic
  - (d) N-channel-coupled logic
9. Which of the following is *not* a characteristic of CMOS chips?
- (a) Sensitivity to damage by electrostatic discharge
  - (b) Low current demand
  - (c) Ability to work at high speed
  - (d) Ability to handle extremely high power
10. An absolute limit to the component density that can be achieved in IC technology is determined by
- (a) the maximum current levels needed.
  - (b) the maximum attainable impedance.
  - (c) the size of the semiconductor atoms.
  - (d) Forget it! There is no limit to the component density that can be achieved in IC technology.
11. With a ROM chip,
- (a) it is easier to read data from the device than to write data into it.
  - (b) it is more difficult to read data from the device than to write data into it.
  - (c) it is easy to read data from the device, and just as easy to write data into it.
  - (d) it is difficult to read data from the device, and equally difficult to write data into it.
12. With a RAM chip,
- (a) it is easier to read data from the device than to write data into it.
  - (b) it is more difficult to read data from the device than to write data into it.
  - (c) it is easy to read data from the device, and just as easy to write data into it.
  - (d) it is difficult to read data from the device, and equally difficult to write data into it.
13. Which type of IC must be physically removed from the circuit to have its memory contents changed?
- (a) An EEPROM chip
  - (b) An EPROM chip
  - (c) An SRAM chip
  - (d) A DRAM chip

14. With respect to memory, which of the following terms means that the data contents vanish if power is removed from a chip?

- (a) Volatility
- (b) Component density
- (c) Multiplexing
- (d) Corruptibility

15. In terms of the maximum number of transistors on a single chip, how many orders of magnitude larger is VLSI than MSI?

- (a) Two
- (b) Three
- (c) Four
- (d) Forget it! The maximum number of transistors on a VLSI chip is not larger, but smaller than the maximum number of transistors on an MSI chip.

16. Which of the following illustrations in Fig. 28-9 shows the input and output of a differentiator circuit?

- (a) A
- (b) B
- (c) C
- (d) None of them

17. Which of the following illustrations in Fig. 28-9 shows the input and output of an integrator circuit?

- (a) A
- (b) B
- (c) C
- (d) None of them

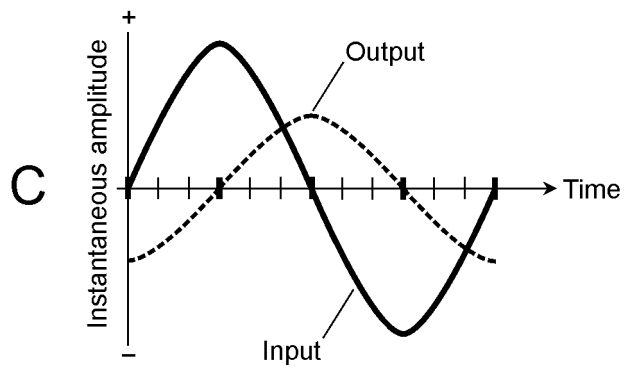
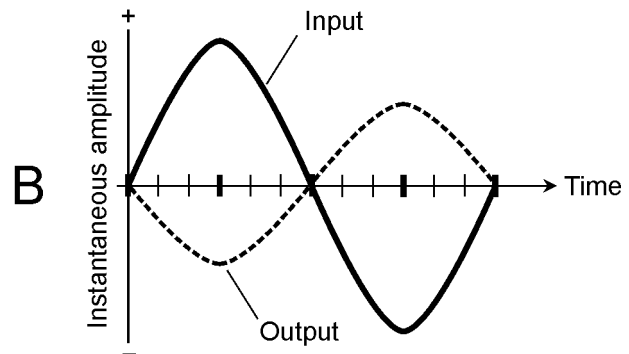
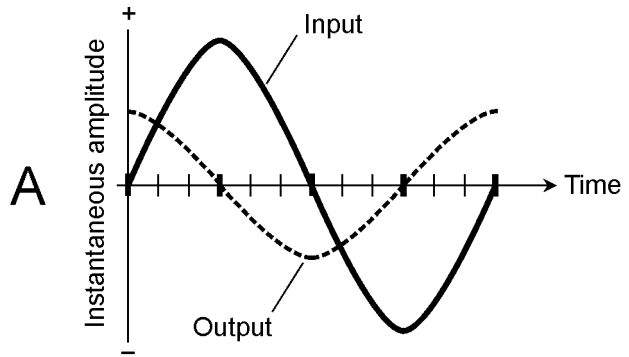
18. Flip-flops are used to store data in

- (a) timers.
- (b) op amps.
- (c) multiplexers.
- (d) none of the above.

19. Capacitors are used to store data in

- (a) timers.
- (b) op amps.
- (c) multiplexers.
- (d) none of the above.

28-9 Illustration for Quiz Questions 16 and 17.



20. The phase of a sine wave signal can be shifted  $180^\circ$  by
- applying the signal to the input of a timer.
  - applying the signal to the inverting input of an op amp.
  - applying the signal to the input of a flip-flop.
  - applying the signal to the input of an SRAM chip.

## 29 CHAPTER

# Electron Tubes

*ELECTRON TUBES*, ALSO CALLED *TUBES* OR *VALVES* (IN ENGLAND), ARE USED IN SOME ELECTRONIC equipment. In a tube, the *charge carriers* are free electrons that travel through space between electrodes inside the device. This makes tubes fundamentally different from semiconductor devices, in which charge carriers move among atoms in a solid medium.

## Tube Forms

There are two basic types of electron tube: the *vacuum tube* and the *gas-filled tube*. As their names imply, vacuum tubes have virtually all the gases removed from their envelopes. Gas-filled tubes contain elemental vapor at low pressure.

### Vacuum Tube

Vacuum tubes accelerate electrons to high speeds, resulting in large currents. This current can be focused into a beam and guided in a particular direction. The intensity and/or beam direction can be changed with extreme rapidity, producing effects such as rectification, detection, oscillation, amplification, signal mixing, waveform displays, spectral displays, and video imaging.

### Gas-Filled Tube

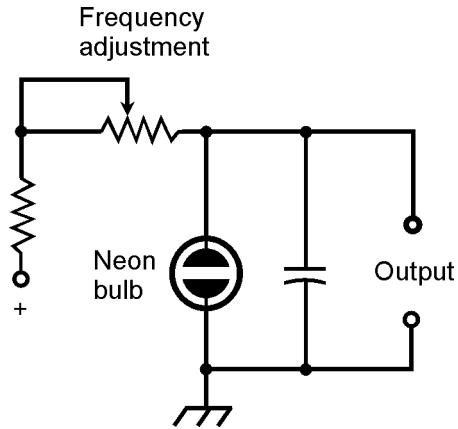
Gas-filled tubes have a constant voltage drop, no matter what the current. This makes them useful as voltage regulators for high-voltage, high-current power supplies. Gas-filled tubes can withstand conditions that would destroy semiconductor regulating devices. Gas-filled tubes emit infrared (IR), visible light, and/or ultraviolet (UV). This property can be put to use for decorative lighting. A small *neon bulb* can be employed to construct an AF *relaxation oscillator* (Fig. 29-1).

### Diode Tube

Even before the year 1900, scientists knew that electrons could carry electric current through a vacuum. They also knew that hot electrodes emit electrons more easily than cool ones. These phenom-



29-1 A neon bulb oscillator, also known as a relaxation oscillator.



ena were put to use in the first electron tubes, known as *diode tubes*, for the purpose of rectification. Diode tubes are rarely used nowadays, although they can still be found in some power supplies that are required to deliver several thousand volts for long periods at a 100 percent *duty cycle* (that is, continuous operation).

## Electrodes in a Tube

In a tube, the electron-emitting electrode is the *cathode*. The cathode is usually heated by means of a wire *filament*, similar to the glowing element in an incandescent bulb. The heat drives electrons from the cathode. The cathode of a tube is analogous to the source of an FET, or to the emitter of a bipolar transistor. The electron-collecting electrode is the *anode*, also called the *plate*. The plate is the tube counterpart of the drain of an FET or the collector of a bipolar transistor. In most tubes, intervening *grids* control the flow of electrons from the cathode to the plate. The grids are the counterparts of the gate of an FET or the base of a bipolar transistor.

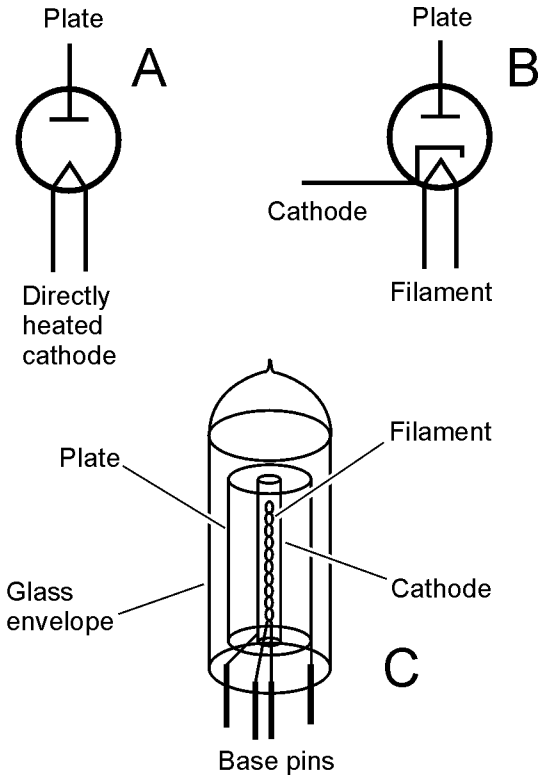
### Directly Heated Cathode

In some tubes, the filament also serves as the cathode. This type of electrode is called a *directly heated cathode*. The negative supply voltage is applied directly to the filament. The filament voltage for most tubes is 6 V or 12 V dc. It is important that dc be used to heat the filament in this type of tube, because ac will tend to modulate the output. The schematic symbol for a diode tube with a directly heated cathode is shown in Fig. 29-2A.

### Indirectly Heated Cathode

In many types of tubes, the filament is enclosed within a cylindrical cathode, and the cathode gets hot from the IR radiated by the filament. This is an *indirectly heated cathode*. The filament normally receives 6 V or 12 V ac or dc. In an indirectly heated cathode arrangement, ac does not cause modulation problems, as it can with a directly heated cathode tube. The schematic symbol for a diode tube with an indirectly heated cathode is shown in Fig. 29-2B.

Because the electron emission in a tube depends on the filament or *heater*, tubes need a certain amount of time to warm up before they can operate properly. This time can vary from a few seconds



**29-2** At A, schematic symbol for diode tube with directly heated cathode. At B, symbol for diode tube with indirectly heated cathode. At C, simplified rendition of the construction of a diode tube.

(for a small tube with a directly heated cathode) to a couple of minutes (for massive power-amplifier tubes with indirectly heated cathodes). The warm-up time for a large tube is about the same as the boot-up time for a personal computer.

### Cold Cathode

In a gas-filled tube, the cathode does not have a filament to heat it. Such an electrode is called a *cold cathode*. Various chemical elements are used in gas-filled tubes. In fluorescent devices, neon, argon, and xenon are common. In gas-filled voltage-regulator (VR) tubes, mercury vapor is used. In a mercury-vapor VR tube, the warm-up period is the time needed for the elemental mercury, which is a liquid at room temperature, to vaporize (approximately 2 minutes).

### Plate

The plate, or anode, of a tube is a cylinder concentric with the cathode and filament, as shown in Fig. 29-2C. The plate is connected to the positive dc supply voltage. Tubes operate at plate voltages ranging from 50 V to more than 3 kV. These voltages are potentially lethal. Technicians unfamiliar with vacuum tubes should not attempt to service equipment that contains them. The output of a tube-type amplifier circuit is almost always taken from the plate circuit. The plate exhibits high impedance for signal output, similar to that of an FET.

## Control Grid

The flow of current can be controlled by means of an electrode between the cathode and the plate. This electrode, the *control grid* (or simply the *grid*), is a wire mesh or screen that lets electrons pass through. The grid impedes the flow of electrons if it is provided with a negative voltage relative to the cathode. The greater the negative *grid bias*, the more the grid obstructs the flow of electrons through the tube.

## Triode Tube

A tube with one grid is a *triode*. The schematic symbol is shown at A in Fig. 29-3. In this case the cathode is indirectly heated; the filament is not shown. This omission is standard in schematics showing tubes with indirectly heated cathodes. When the cathode is directly heated, the filament symbol serves as the cathode symbol. The control grid is usually biased with a negative dc voltage ranging from near 0 to as much as half the positive dc plate voltage.

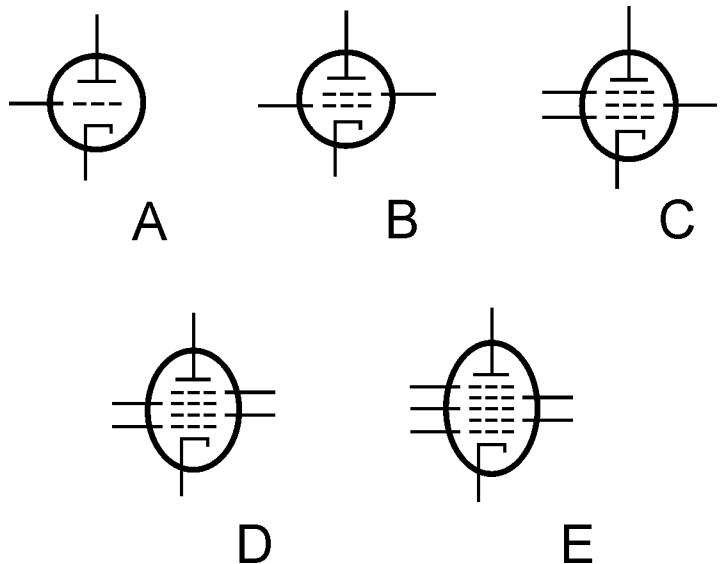
## Tetrode Tube

A second grid can be added between the control grid and the plate. This is a spiral of wire or a coarse screen, and is called the *screen grid* or *screen*. This grid normally carries a positive dc voltage at 25 to 35 percent of the plate voltage.

The screen grid reduces the capacitance between the control grid and plate, minimizing the tendency of a tube amplifier to oscillate. The screen grid can also serve as a second control grid, allowing two signals to be injected into a tube. This tube has four elements, and is known as a *tetrode*. Its schematic symbol is shown at B in Fig. 29-3.

## Pentode Tube

The electrons in a tetrode can bombard the plate with such force that some of them bounce back, or knock other electrons from the plate. This so-called secondary emission can hinder tube perfor-



**29-3** Schematic symbols for vacuum tubes with grids: triode (A), tetrode (B), pentode (C), hexode (D), and heptode (E).

mance and, at high power levels, cause screen current so high that the electrode is destroyed. This problem can be eliminated by placing another grid, called the *suppressor grid* or *suppressor*, between the screen and the plate. The suppressor repels *secondary electrons* emanating from the plate, preventing most of them from reaching the screen. The suppressor also reduces the capacitance between the control grid and the plate more than a screen grid by itself.

Greater gain and stability are possible with a *pentode*, or tube with five elements, than with a tetrode or triode. The schematic symbol for a pentode is shown at C in Fig. 29-3. The suppressor is often connected directly to the cathode.

### Hexode and Heptode Tubes

In some older radio and TV receivers, tubes with four or five grids were sometimes used. These tubes had six and seven elements, respectively, and were called *hexode* and *heptode*. The usual function of such tubes was signal mixing. The schematic symbol for a hexode is shown at D in Fig. 29-3; the symbol for a heptode is at E. You will not encounter hexodes and heptodes in modern electronics, because solid-state components are used for signal mixing.

### Interelectrode Capacitance

In a vacuum tube, the cathode, grid(s), and plate exhibit *interelectrode capacitance* that is the primary limiting factor on the frequency range in which the device can produce gain. The interelectrode capacitance in a typical tube is a few picofarads. This is negligible at low frequencies, but at frequencies above approximately 30 MHz, it becomes a significant consideration. Vacuum tubes intended for use as RF amplifiers are designed to minimize this capacitance.

## Circuit Configurations

The most common application of vacuum tubes is in amplifiers, especially in radio and television transmitters at power levels of more than 1 kW. Some high-fidelity audio systems also employ vacuum tubes. In recent years, tubes have gained favor with some popular music bands. Some musicians insist that “tube amps” provide richer sound than amplifiers using power transistors. There are two basic vacuum-tube amplifier circuit arrangements: the *grounded-cathode* configuration and the *grounded-grid* configuration.

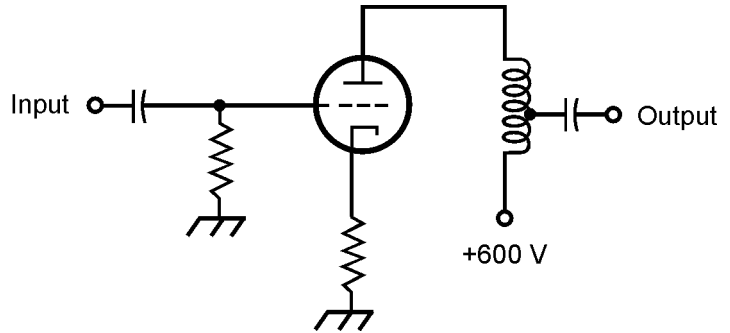
### Grounded Cathode

Figure 29-4 is a simplified schematic diagram of a grounded-cathode circuit using a triode tube. This circuit is the basis for many tube-type RF power amplifiers and audio amplifiers. The input impedance is moderate, and the output impedance is high. Impedance matching between the amplifier and the load can be obtained by tapping a coil in the output circuit, or by using a transformer.

### Grounded Grid

Figure 29-5 shows a basic grounded-grid RF amplifier circuit. The input impedance is low, and the output impedance is high. The output impedance is matched by the same means as with the grounded-cathode arrangement. The grounded-grid configuration requires more driving (input) power than the grounded-cathode scheme. A grounded-cathode amplifier might produce 1 kW of

- 29-4 Simplified schematic diagram of a grounded-cathode RF amplifier circuit using a triode tube.



RF output for 10-W input, but a grounded-grid amplifier needs 50 W to 100 W of drive to produce 1 kW of RF output. A grounded-grid amplifier has a significant advantage, however: it is less likely to break into unwanted oscillation than a grounded-cathode circuit.

### Plate Voltage

The plate voltages (+600 V dc) in the circuits of Figs. 29-4 and 29-5 are given as examples. The amplifiers shown could produce 75- to 150-W signal output provided they receive sufficient drive and are properly biased. An amplifier rated at 1-kW output would require a plate voltage of +2 kV dc to +5 kV dc. In high-power radio and TV broadcast transmitters producing in excess of 50-kW RF output, even higher dc plate voltages are used.

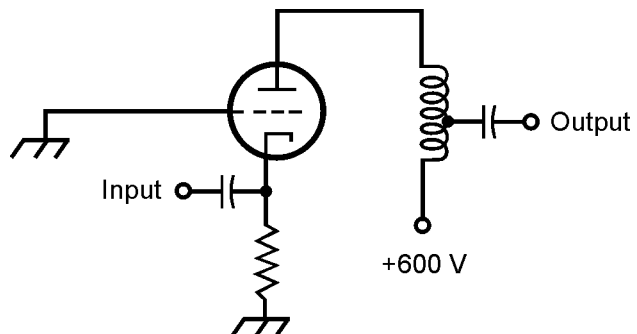
## Cathode-Ray Tubes

Many TV receivers, and some desktop computer monitors, use *cathode-ray tubes* (CRTs). So do older oscilloscopes, spectrum analyzers, and radar sets.

### Electron Beam

In a CRT, a specialized cathode called an *electron gun* emits an electron beam that is focused and accelerated as it passes through positively charged *anodes*. The beam then strikes a glass screen whose inner surface is coated with *phosphor*. The phosphor glows visibly, as seen from the face of the CRT, because of the effect of the high-speed electrons striking it.

- 29-5 Simplified schematic diagram of a grounded-grid RF amplifier circuit using a triode tube.

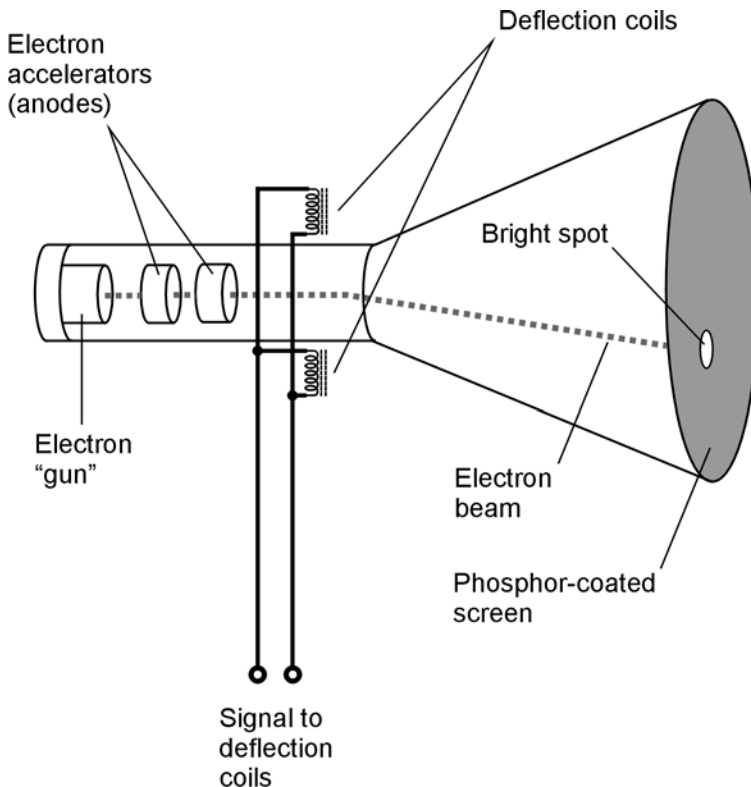


The beam *scanning* pattern is controlled by magnetic or electrostatic fields. One field causes the beam to scan rapidly across the screen in a horizontal direction. Another field moves the beam vertically. When complex waveforms are applied to the electrodes that produce the deflection of the electron beam, a display pattern results. This pattern can be the graph of a signal wave, a fixed image, an animated image, a computer text display, or any other type of visible image.

### Electromagnetic CRT

A simplified cross-sectional drawing of an *electromagnetic CRT* is shown in Fig. 29-6. There are two sets of *deflecting coils*, one for the horizontal plane and the other for the vertical plane. (To keep the illustration reasonably clear, only one set of deflecting coils is shown.) The greater the current in the coils, the greater the intensity of the magnetic field, and the more the electron beam is deflected. The electron beam is bent at right angles to the magnetic lines of flux.

In an *oscilloscope*, the horizontal deflecting coils receive a sawtooth waveform. This causes the beam to scan, or *sweep*, at a precise, adjustable speed across the screen from left to right as viewed from in front. After each timed left-to-right sweep, the beam returns, almost instantly, to the left side of the screen for the next sweep. The vertical deflecting coils receive the waveform to be analyzed. This waveform makes the electron beam move up and down. The combination of vertical and horizontal beam motion produces a display of the input waveform as a function of time.



29-6 Simplified cross-sectional rendition of an electromagnetic CRT.

## Electrostatic CRT

In an *electrostatic CRT*, charged metal plates, rather than current-carrying coils, are used to deflect the electron beam. When voltages appear on these *deflecting plates*, the beam is bent in the direction of the electric lines of flux. The greater the voltage applied to a deflecting plate, the stronger the electric field, and the greater the extent to which the beam is deflected.

The principal advantage of an electrostatic CRT is the fact that it generates a far less intense magnetic field than an electromagnetic CRT. This so-called extremely low frequency (ELF) energy is a cause for concern, because it might have adverse effects on people who use CRT-equipped devices, such as desktop computers, for extended periods of time. In recent years, with the evolution of *liquid crystal displays* (LCDs) and *plasma displays* as alternatives to the CRT type of display, ELF has become a much less significant concern.

## Camera Tubes

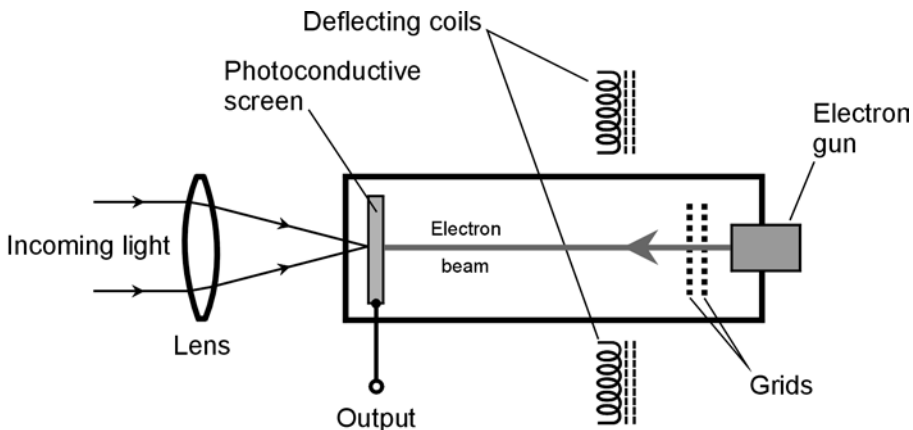
Some video cameras use a form of electron tube that converts visible light into varying electric currents. The two most common types of *camera tube* are the *vidicon* and the *image orthicon*.

### Vidicon

In the vidicon, a lens focuses the incoming image onto a photoconductive screen. An electron gun generates a beam that sweeps across the screen as a result of the effects of deflecting coils, in a manner similar to the operation of an electromagnetic CRT. The sweep in the vidicon is synchronized with any CRT that displays the image.

As the electron beam scans the photoconductive surface, the screen becomes charged. The rate of discharge in a certain region on the screen depends on the intensity of the visible light striking that region. A simplified cutaway view of a vidicon tube is shown in Fig. 29-7.

The main advantage of the vidicon is its small physical size and mass. A vidicon is sensitive, but its response can be sluggish when the level of illumination is low. This causes images to persist for a short while, resulting in poor portrayal of fast-motion scenes.



29-7 Functional diagram of a vidicon.

## Image Orthicon

Another type of camera tube, also quite sensitive but having a quicker response to image changes, is the image orthicon. It is constructed much like the vidicon, except that there is a target electrode behind the *photocathode* (Fig. 29-8). When a single electron from the photocathode strikes the target electrode, multiple secondary electrons are emitted as a result. The image orthicon thus acts as a video signal amplifier, as well as a camera.

A fine beam of electrons, emitted from the electron gun, scans the target electrode. The secondary electrons cause some of this beam to be reflected back toward the electron gun. Areas of the target electrode with the most secondary electron emission produce the greatest return beam intensity, and regions with the least emission produce the lowest return beam intensity. The greatest return beam intensity corresponds to the brightest parts of the video image. The return beam is modulated as it scans the target electrode and is picked up by a receptor electrode.

One significant disadvantage of the image orthicon is that it produces considerable noise in addition to the signal output. But when a fast response is needed and the illumination ranges from dim to very bright, the image orthicon is the camera tube of choice.

## Photomultiplier

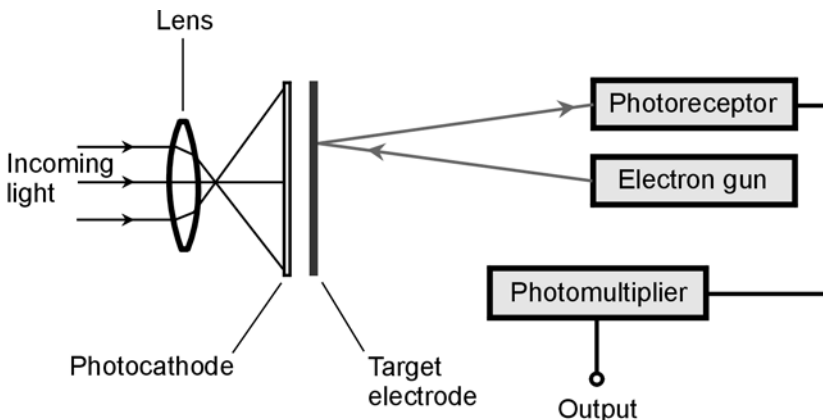
A *photomultiplier* is a vacuum tube that generates a variable current depending on the intensity of the light that strikes it. It multiplies its own output, thereby obtaining high sensitivity. Photomultipliers can be used to measure light intensity at low levels.

The photomultiplier consists of a photocathode, which emits electrons in proportion to the intensity of the light striking it. These electrons are focused into a beam, and this beam strikes an electrode called a *dynode*. The dynode emits several secondary electrons for each electron that strikes it. The resulting beam is collected by the anode.

A photomultiplier can have several dynodes, resulting in high gain. The extent to which the sensitivity can be increased by cascading dynodes is limited by the amount of background electron emission or *dark noise* from the photocathode.

## Dissector

A *dissector*, also known as an *image dissector*, is a form of photomultiplier in which the light is focused by a lens onto a translucent photocathode. This surface emits electrons in proportion to the



29-8 Functional diagram of an image orthicon.



light intensity. The electrons from the photocathode are directed to a barrier containing a small aperture. The vertical and horizontal deflection plates, supplied with synchronized scanning voltages, move the beam from the photocathode across the aperture. The electron stream passing through the aperture is modulated depending on the light and dark nature of the image.

The *image resolution* of the dissector tube depends on the size of the aperture. The smaller the aperture, the sharper the image, down to a certain limiting point. However, there is a limit to how small the aperture can be, while still allowing enough electrons to pass, and avoiding the generation of *interference patterns*. The image dissector tube produces very little dark noise, and this allows for excellent sensitivity.

## Tubes for Use above 300 MHz

Specialized vacuum tubes are required for RF operation at frequencies above 300 MHz. These bands are known as *ultrahigh frequency (UHF) band*, which ranges from 300 MHz to 3 GHz, and the *microwave band*, which ranges from 3 GHz up. The *magnetron* and the *Klystron* are examples of tubes that are used to generate and amplify signals at these frequencies.

### Magnetron

A magnetron contains a cathode and a surrounding anode. The anode is divided into sections, or *cavities*, by radial barriers. The output is taken from an opening in the anode, and passes into a waveguide that serves as a transmission line for the RF output energy.

The cathode is connected to the negative terminal of a high-voltage source, and the anode is connected to the positive terminal. Therefore, electrons flow radially outward. A magnetic field is applied lengthwise through the cavities. As a result, the electron paths are bent into spirals. The electric field produced by the high voltage, interacting with the longitudinal magnetic field and the effects of the cavities, causes the electrons to bunch up into *clouds*. The swirling movement of the electron clouds causes a fluctuating current in the anode. The frequency depends on the shapes and sizes of the cavities. Small cavities result in the highest oscillation frequencies; larger cavities produce oscillation at relatively lower frequencies.

A magnetron can generate more than 1 kW of RF power at a frequency of 1 GHz. As the frequency increases, the realizable power output decreases. At 10 GHz, a typical magnetron generates about 20 W of RF power output.

### Klystron

A Klystron has an electron gun, one or more cavities, and a device that modulates the electron beam. There are several different types. The most common are the *multicavity Klystron* and the *reflex Klystron*.

In a multicavity Klystron, the electron beam is *velocity-modulated* in the first cavity. This causes the density of electrons (the number of particles per unit volume) in the beam to change as the beam moves through subsequent cavities. The electrons tend to bunch up in some regions and spread out in other regions. The intermediate cavities increase the magnitude of the electron beam modulation, resulting in amplification. Output is taken from the last cavity. Peak power levels in some multicavity Klystrons can exceed 1 MW ( $10^6$  W), although the average power is much less.

A reflex Klystron has a single cavity. A *retarding field* causes the electron beam to periodically reverse direction. This produces a phase reversal that allows large amounts of energy to be drawn from

the electrons. A typical reflex Klystron can produce signals on the order of a few watts at frequencies of 300 MHz and above.

## Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. A major difference between a triode tube and an N-channel FET is the fact that
  - (a) triodes work with lower signal voltages.
  - (b) triodes are more compact.
  - (c) triodes need higher power-supply voltages.
  - (d) triodes don't need filaments.
2. The control grid of a vacuum tube is the electrical counterpart of the
  - (a) source of a MOSFET.
  - (b) collector of a bipolar transistor.
  - (c) anode of a diode.
  - (d) gate of an FET.
3. In a tetrode tube, the charge carriers are
  - (a) free electrons, which pass among the electrodes.
  - (b) holes, which are conducted within the electrodes.
  - (c) holes or electrons, depending on whether the device is P type or N type.
  - (d) nuclei of whatever elemental gas happens to exist in the tube.
4. Which factor is most significant in limiting the maximum frequency at which a tube can operate?
  - (a) The power-supply voltage
  - (b) The capacitance among the electrodes
  - (c) The physical size of the tube
  - (d) The current passing through the filament
5. In a tube with a directly heated cathode,
  - (a) the filament is separate from the cathode.
  - (b) the grid is connected to the filament.
  - (c) the filament serves as the cathode.
  - (d) there is no filament.
6. In a tube with a cold cathode,
  - (a) the filament is separate from the cathode.
  - (b) the grid is connected to the filament.
  - (c) the filament serves as the cathode.
  - (d) there is no filament.

7. A screen grid enhances tube operation by
  - (a) increasing the gain, helping the circuit to oscillate more easily.
  - (b) decreasing the plate voltage required to produce oscillation.
  - (c) minimizing the risk that a tube amplifier will break into oscillation.
  - (d) pulling excess electrons from the plate.
8. A tube with three grids is called a
  - (a) triode.
  - (b) tetrode.
  - (c) pentode.
  - (d) hexode.
9. A tube with four grids is called a
  - (a) triode.
  - (b) tetrode.
  - (c) pentode.
  - (d) hexode.
10. An advantage of a grounded-grid RF power amplifier over a grounded-cathode RF power amplifier is the fact that the grounded-grid circuit
  - (a) has excellent sensitivity.
  - (b) exhibits high input impedance.
  - (c) produces little or no noise in the input.
  - (d) is more stable.
11. A heptode tube has
  - (a) one plate.
  - (b) two plates.
  - (c) three plates.
  - (d) four plates.
12. The electron gun in a CRT is another name for its
  - (a) cathode.
  - (b) anode.
  - (c) control grid.
  - (d) screen grid.
13. The electron beam in an electrostatic CRT is bent by
  - (a) magnetic fields produced by current-carrying coils.
  - (b) electric fields produced by charged electrodes.
  - (c) a variable voltage on the screen grid.
  - (d) visible light striking the electrodes.

14. A grounded-grid RF power amplifier
  - (a) requires more driving power, for a given RF power output, than a grounded-cathode RF power amplifier.
  - (b) requires less driving power, for a given RF power output, than a grounded-cathode RF power amplifier.
  - (c) oscillates at a more stable frequency than a grounded-cathode RF power amplifier.
  - (d) oscillates at a less stable frequency than a grounded-cathode RF power amplifier.
15. In a Klystron, the electron-beam density varies as a result of
  - (a) amplitude modulation.
  - (b) pulse modulation.
  - (c) velocity modulation.
  - (d) frequency modulation.
16. A vidicon camera tube is noted for its
  - (a) poor signal-to-noise ratio.
  - (b) large size and heavy weight.
  - (c) slow response to image movement in dim light.
  - (d) excellent selectivity and electrical ruggedness.
17. The plate in a tetrode tube is normally connected to
  - (a) a positive dc power-supply voltage.
  - (b) a negative dc power-supply voltage.
  - (c) electrical ground.
  - (d) RF ground.
18. The screen grid in a tetrode tube is normally connected to
  - (a) a positive dc power-supply voltage.
  - (b) a negative dc power-supply voltage.
  - (c) electrical ground.
  - (d) RF ground.
19. Which of the following is most suitable for measuring the intensity of dim light?
  - (a) A triode gas-filled tube
  - (b) A photomultiplier tube
  - (c) An electrostatic CRT
  - (d) An electromagnetic CRT
20. In a dissector tube, the aperture size is directly related to the
  - (a) operating voltage.
  - (b) signal-to-noise ratio.
  - (c) response speed.
  - (d) image resolution.

## 30 CHAPTER

# Transducers, Sensors, Location, and Navigation

IN THIS CHAPTER, YOU'LL LEARN ABOUT ELECTRONIC DEVICES THAT CONVERT ENERGY FROM ONE form to another, devices that can detect phenomena and measure their intensity, systems that can help you find out where you are (or where some other object is), and devices that facilitate navigation for vessels such as ships, aircraft, and robots.

## Wave Transducers

In electronics, *wave transducers* convert ac or dc into acoustic or electromagnetic (EM) waves. They can also convert these waves into ac or dc signals.

### Dynamic Transducer for Sound

A *dynamic transducer* is a coil-and-magnet device that translates mechanical vibration into varying electrical current, and can also do the reverse. The most common examples are the *dynamic microphone* and the *dynamic speaker*.

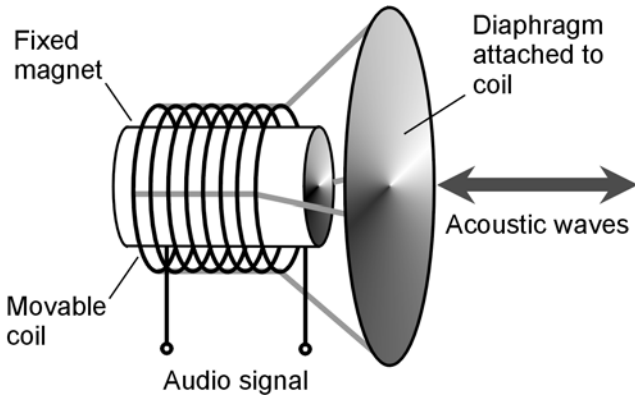
Figure 30-1 shows a functional diagram of a dynamic transducer. A *diaphragm* is attached to a coil that is mounted so it can move back and forth rapidly along its axis. A permanent magnet is placed inside the coil. Sound waves cause the diaphragm to move; this moves the coil, which causes fluctuations in the magnetic field within the coil. The result is ac output from the coil, having the same waveform as the sound waves that strike the diaphragm.

If an audio signal is applied to the coil, it generates a magnetic field that produces forces on the coil. This causes the coil to move, pushing the diaphragm back and forth, creating acoustic waves in the surrounding medium.

### Electrostatic Transducer for Sound

An *electrostatic transducer* takes advantage of the forces produced by electric fields. Two metal plates, one flexible and the other rigid, are placed parallel to each other and close together (Fig. 30-2).

In an *electrostatic pickup*, incoming sound waves vibrate the flexible plate. This produces small, rapid changes in the spacing, and therefore the capacitance, between the two plates. A dc voltage is



30-1 Functional diagram of a dynamic sound transducer.

applied between the plates. As the interplate capacitance varies, the electric field intensity between them fluctuates. This produces variations in the current through the primary winding of the transformer. Audio signals appear across the secondary.

In an *electrostatic emitter*, fluctuating currents in the transformer produce changes in the voltage between the plates. This results in electrostatic field variations, pulling and pushing the flexible plate in and out. The motion of the flexible plate produces sound waves.

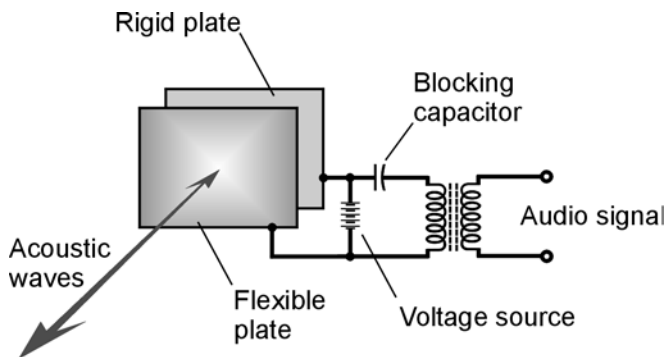
Electrostatic transducers can be used in most applications where dynamic transducers are employed. Advantages of electrostatic transducers include light weight and good sensitivity. The relative absence of magnetic fields can also be an asset in certain situations.

### Piezoelectric Transducer for Sound and Ultrasound

Figure 30-3 shows a *piezoelectric transducer*. This device consists of a *crystal* of quartz or ceramic material, sandwiched between two metal plates. When sound waves strike one or both of the plates, the metal vibrates. This vibration is transferred to the crystal. The crystal generates weak electric currents when subjected to this mechanical stress. Therefore, an ac voltage develops between the two metal plates, with a waveform similar to that of the sound.

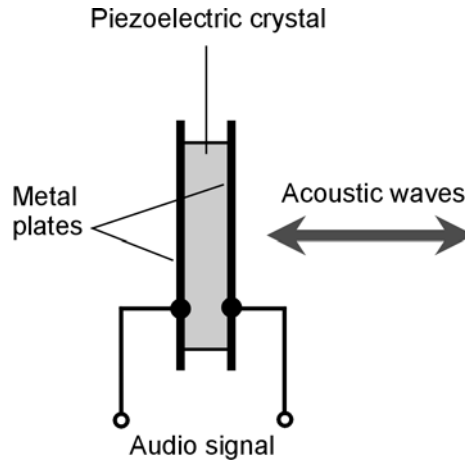
If an ac signal is applied to the plates, it causes the crystal to vibrate in sync with the current. The metal plates vibrate also, producing an acoustic disturbance.

Piezoelectric transducers can function at higher frequencies than can dynamic or electrostatic transducers. For this reason, they are favored in ultrasonic applications, such as intrusion detectors and alarms.



30-2 Functional diagram of an electrostatic sound transducer.

**30-3** Functional diagram of a piezoelectric transducer for sound and ultrasound.



### Transducers for RF Energy

The term *radio-frequency (RF) transducer* is a fancy name for an *antenna*. There are two basic types: the *receiving antenna* and the *transmitting antenna*. You learned about antennas in Chap. 27.

### Transducers for IR and Visible Light

Many wireless devices transmit and receive energy at IR wavelengths. Infrared energy has a frequency higher than that of radio waves, but lower than that of visible light. Some wireless devices transmit and receive their signals in the visible range, although these are encountered much less often than IR devices.

The most common IR transmitting transducer is the infrared-emitting diode (IRED). Fluctuating dc is applied to the device, causing it to emit IR rays. The fluctuations in the current constitute modulation, and this produces rapid variations in the intensity of the rays emitted by the semiconductor P-N junction. The modulation contains information, such as which channel your television set should seek, or whether the volume is to be raised or lowered. Infrared energy can be focused by optical lenses and reflected by optical mirrors. This makes it possible to *collimate* IR rays (make them parallel) so they can be transmitted for distances up to several hundred meters.

Infrared receiving transducers resemble photodiodes or photovoltaic cells. The fluctuating IR energy from the transmitter strikes the P-N junction of the receiving diode. If the receiving device is a photodiode, a current is applied to it, and this current varies rapidly in accordance with the signal waveform on the IR beam from the transmitter. If the receiving device is a photovoltaic cell, it produces the fluctuating current all by itself, without the need for an external power supply. In either case, the current fluctuations are weak, and must be amplified before they are delivered to whatever equipment (television set, garage door, oven, security system, etc.) is controlled by the wireless system. Infrared wireless devices work best on a line of sight.

## Displacement Transducers

A *displacement transducer* measures a distance or angle traversed, or the distance or angle separating two points. Conversely, it can convert a signal into movement over a certain distance or angle. A device that measures or produces movement in a straight line is a *linear displacement transducer*. If it measures or produces movement through an angle, it is an *angular displacement transducer*.

## Pointing and Control Devices

A *joystick* is a control device capable of producing movement, or controlling variable quantities, in two dimensions. The device consists of a movable lever and a ball bearing within a control box. The lever can be moved by hand up and down, and to the right and left. Joysticks are used in computer games, for entering coordinates into a computer, and for the remote control of robots. In some joysticks, the lever can be rotated, allowing control in a third dimension.

A *mouse* is a peripheral commonly used with personal computers. By sliding the mouse around on a flat surface, a cursor or arrow is positioned on the display. Pushbutton switches on the top of the unit actuate the computer to perform whatever function the cursor or arrow shows. These actions are called *clicks*.

A *trackball* resembles an inverted mouse, or a two-dimensional joystick without the lever. Instead of the device being pushed around on a surface, the user moves a ball bearing, causing the display cursor to move vertically and horizontally. Pushbutton switches on a computer keyboard, or on the trackball box itself, actuate the functions.

An *eraser-head pointer* is a rubber button approximately 5 mm in diameter, usually placed in the center of a computer keyboard. The user moves the cursor on the display by pushing against the button. Clicking and double clicking are done with button switches on the keyboard.

A *touch pad* is a sensitive plate that is about the size of a business card. The user places an index finger on the plate and moves the finger around. This results in intuitive movement of the display cursor. Clicking and double clicking are done in the same way as with the trackball and eraser-head pointer.

## Electric Motor

An *electric motor* converts electrical energy into angular (and in some cases linear) mechanical energy. Motors can operate from ac or dc, and range in size from tiny devices used in microscopic robots to huge machines that pull passenger trains.

The basics of dc motors were discussed in Chap. 8. In a motor designed to work with ac, there is no commutator. The alternations in the current keep the polarity correct at all times, so the shaft does not lock up. The rotational speed of an ac motor depends on the frequency of the applied ac. With 60-Hz ac, for example, the rotational speed is 60 revolutions per second (60 rps) or 3600 revolutions per minute (3600 rpm).

When a motor is connected to a load, the rotational force required to turn the shaft increases. The greater the required force becomes, the more power is drawn from the source.

## Stepper Motor

A *stepper motor* turns in small increments, rather than continuously. The *step angle*, or extent of each turn, varies depending on the particular motor. It can range from less than 1° of arc to a quarter of a circle (90°). A stepper motor turns through its step angle and then stops, even if the current is maintained. When a stepper motor is stopped with a current going through its coils, the shaft resists external rotational force.

Conventional motors run at hundreds, or even thousands, of revolutions per minute (rpm). But a stepper motor usually runs at much lower speeds, almost always less than 180 rpm. A stepper motor has the most turning power when it is running at its slowest speeds, and the least turning power when it runs at its highest speeds.

When a pulsed current is supplied to a stepper motor, the shaft rotates in increments, one step for each pulse. In this way, a precise speed can be maintained. Because of the braking effect, this



speed is constant for a wide range of mechanical turning resistances. Stepper motors can be controlled using microcomputers. This type of motor is especially well suited for point-to-point motion. Complicated, intricate tasks can be done by computer-controlled robots using stepper motors.

## Selsyn

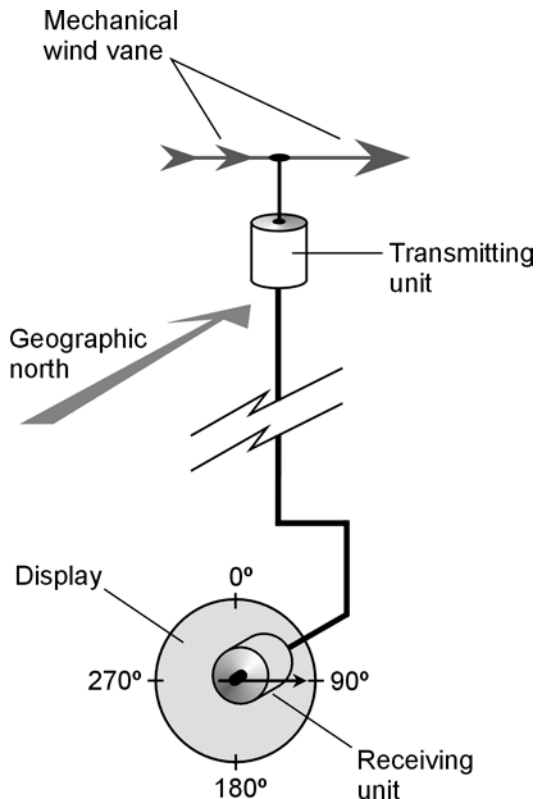
A *selsyn* is an indicating device that shows the direction in which an object is oriented. It consists of a transmitting unit and a receiving (or indicator) unit. As the shaft of the transmitting unit rotates, the shaft of the receiving unit, which is a stepper motor, follows along exactly. A common application is as a direction indicator for a wind vane (Fig. 30-4). When the wind vane rotates, the indicator unit shaft moves through the same number of angular degrees as the transmitting unit shaft. A selsyn for azimuth (compass) bearings has a range of  $0^\circ$  to  $360^\circ$ . A selsyn for elevation bearings has a range of  $0^\circ$  to  $90^\circ$ .

A *synchro* is a selsyn used for the control of mechanical devices. Synchros are well suited for robotic *teleoperation*, or remote control. Some synchros are programmable. The operator inputs a number into the generator, and the receiver changes position accordingly. Synchros are commonly used as rotators and direction indicators for directional communications antennas such as the Yagi or dish.

## Electric Generator

An *electric generator* is constructed in much the same way as an ac motor, although it functions in the opposite sense. Some generators can also be used as motors; devices of this type are called *motor/generators*.

**30-4** A selsyn can be used to remotely indicate the direction in which a mechanical device is pointed.



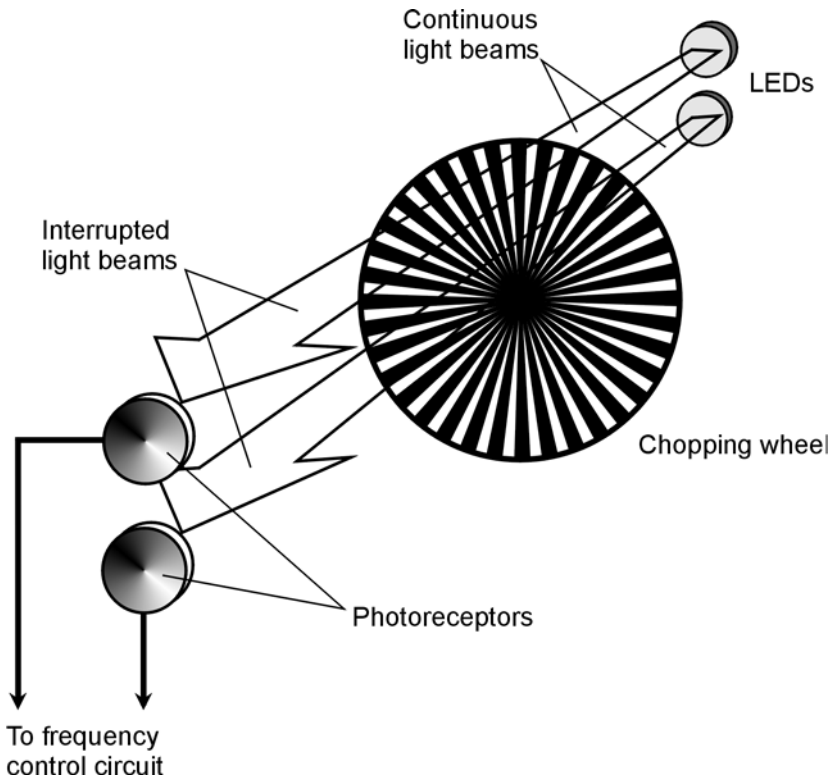
A typical generator produces ac from the mechanical rotation of a coil in a strong magnetic field. Alternatively, a permanent magnet can be rotated within a coil of wire. The rotating shaft can be driven by a gasoline engine, a steam turbine, a water turbine, a windmill, or even by human power. A commutator can be used with a generator to produce pulsating dc output, which can be filtered to obtain pure dc for use with electronic equipment.

Small portable gasoline-powered generators, capable of delivering a few kilowatts, can be purchased in department stores or home-and-garden stores. Larger generators, which usually burn propane or methane (“natural gas”), allow homes or buildings to keep their electrical power in the event of an interruption in the utility. The largest generators are found in power plants, and can produce many kilowatts.

Small generators can be used in synchro systems. These specialized generators allow remote control of robotic devices. A generator can be used to measure the speed at which a vehicle or rolling robot moves. The shaft of the generator is connected to one of the wheels, and the generator output voltage and frequency vary directly with the angular speed of the wheel. This is an example of a *tachometer*, a device familiar to people with automotive experience.

### Optical Encoder

In digital radios, frequency adjustment is done in discrete steps. A typical increment is 10 Hz for shortwave radios and 200 kHz for FM broadcast radios. An alternative to mechanical switches or



**30-5** An optical encoder uses LEDs and photodetectors to sense the direction and extent of shaft rotation.

gear-driven devices, which wear out with time, is the *optical encoder*, also called the *optical shaft encoder*.

An optical encoder consists of two LEDs, two photodetectors, and a device called a *chopping wheel*. The LEDs shine on the photodetectors through the wheel. The wheel has radial bands, alternately transparent and opaque (Fig. 30-5). The wheel is attached to the tuning shaft, which is attached to a large knob. As the tuning knob is rotated, the light beams are interrupted. Each interruption causes the frequency to change by a specified increment. The difference between “frequency up” and “frequency down” (clockwise and counterclockwise shaft rotation, respectively) is determined according to which photodetector senses each sequential beam interruption first.

## Detection and Measurement

A *sensor* employs one or more transducers to detect and/or measure phenomena such as temperature, humidity, barometric pressure, texture, proximity, and the presence of certain substances.

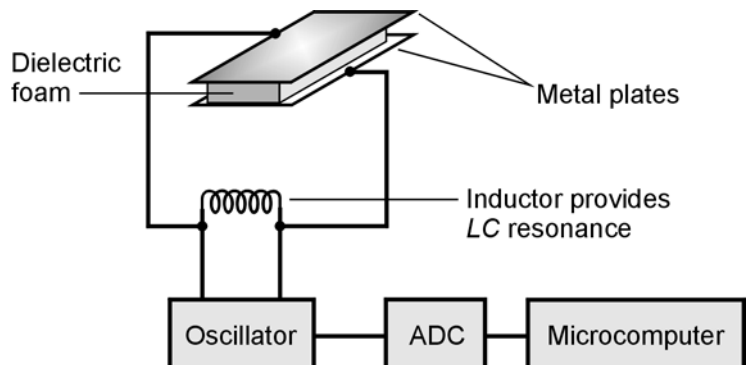
### Capacitive Pressure Sensor

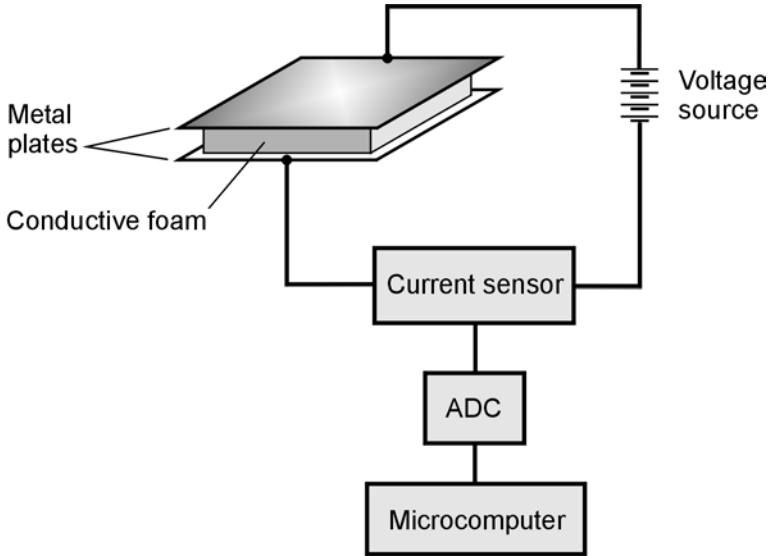
A *capacitive pressure sensor* is shown in Fig. 30-6. Two metal plates are separated by a layer of dielectric (electrically insulating) foam, forming a capacitor. This component is connected in parallel with an inductor. The resulting inductance/capacitance (*LC*) circuit determines the frequency of an oscillator. If an object strikes the sensor, the plate spacing momentarily decreases. This increases the capacitance, causing a drop in the oscillator frequency. When the object moves away from the transducer, the foam springs back, the plates return to their original spacing, and the oscillator frequency returns to normal.

The output of a capacitive pressure sensor can be converted to digital data using an *analog-to-digital converter* (ADC). This signal can be sent to a microcomputer such as a robot controller. Pressure sensors can be mounted in various places on a mobile robot, such as the front, back, and sides. Then, for example, physical pressure on the sensor in the front of the robot can send a signal to the controller, which tells the machine to move backward.

A capacitive pressure sensor can be fooled by massive conducting or semiconducting objects in its vicinity. If such a mass comes near the transducer, the capacitance may change even if direct contact is not made. This phenomenon is known as *body capacitance*. When the effect must be avoided, an *elastomer* device can be used for pressure sensing.

**30-6** A capacitive pressure sensor. When force is applied, the spacing between the plates decreases, causing the capacitance to increase and the oscillator frequency to go down.





**30-7** An elastomer pressure sensor detects applied force without unwanted capacitive effects.

## Elastomer

An *elastomer* is a flexible substance resembling rubber or plastic that can be used to detect the presence or absence of mechanical pressure. Figure 30-7 illustrates how an elastomer can be used to detect, and locate, a pressure point. The elastomer conducts electricity fairly well, but not perfectly. It has a foam-like consistency, so that it can be compressed. Conductive plates are attached to the pad.

When pressure appears at some point in the elastomer pad, the material is compressed, and this lowers its electrical resistance. This is detected as an increase in the current between the plates. The greater the pressure becomes, the more the elastomer is compressed, and the greater is the increase in the current. The current-change data can be sent to a microcomputer such as a robot controller.

## Back-Pressure Sensor

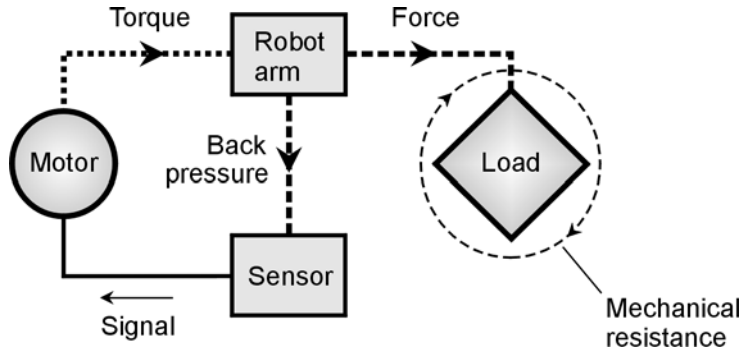
A motor produces a measurable pressure that depends on the torque being applied. A back-pressure sensor detects and measures the torque that the motor is applying at any given time. The sensor produces a signal, usually a variable voltage, that increases as the torque increases. Figure 30-8 is a functional block diagram of a back-pressure sensor.

Back-pressure sensors are used to limit the force applied by robot grippers, arms, drills, hammers, or other *end effectors*. The *back voltage*, or signal produced by the sensor, reduces the torque applied by the motor. This can prevent damage to objects being handled by the robot. It also helps to ensure the safety of people working around the robot.

## Capacitive Proximity Sensor

A *capacitive proximity sensor* uses an RF oscillator, a frequency detector, and a metal plate connected into the oscillator circuit (Fig. 30-9). The oscillator is designed so that a change in the capacitance

**30-8** A back-pressure sensor governs the force applied by a robot arm or other mechanical device.



of the plate, with respect to the environment, causes the oscillator frequency to change. This change is sensed by the frequency detector, which sends a signal to a microcomputer or robot controller.

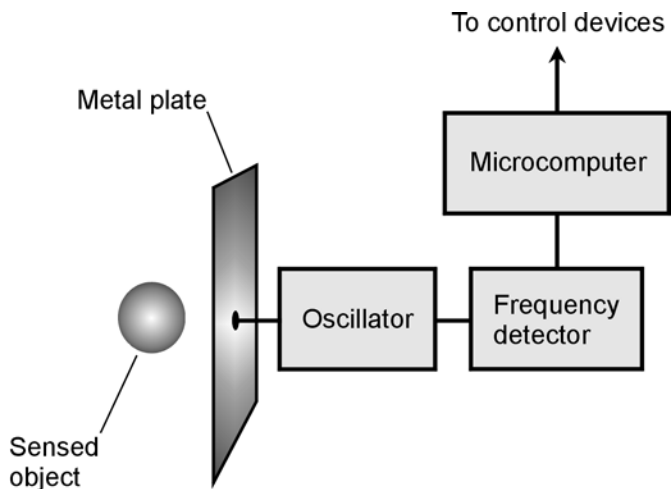
Substances that conduct electricity to some extent, such as metal, saltwater, and living tissue, are sensed more easily by capacitive transducers than are materials that do not conduct, such as dry wood, plastic, glass, or dry fabric.

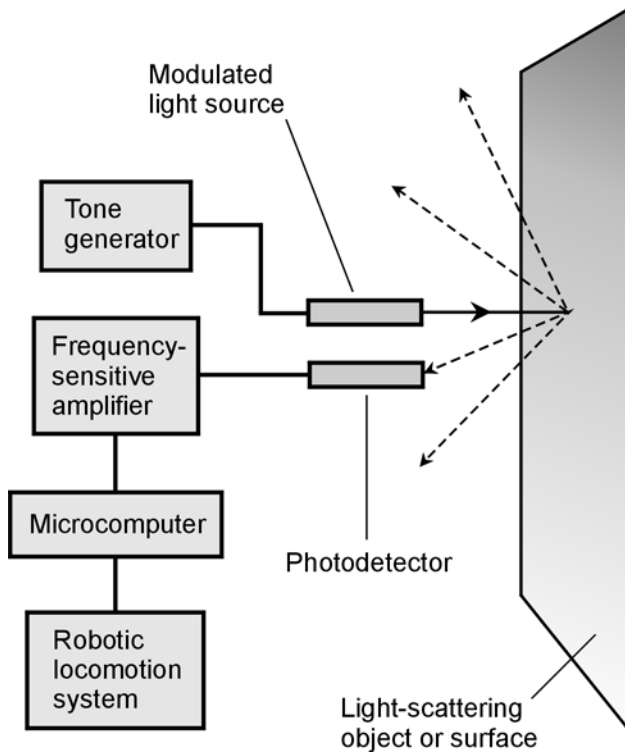
### Photoelectric Proximity Sensor

Reflected light can provide a way for a mobile robot to tell if it is approaching something. A *photoelectric proximity sensor* uses a light-beam generator, a photodetector, a frequency-sensitive amplifier, and a microcomputer (Fig. 30-10).

The light beam reflects from the object and is picked up by the photodetector. The light beam is modulated at a certain frequency, say 1000 Hz. The amplifier responds only to light modulated at that frequency. This prevents false imaging that can otherwise be caused by lamps or sunlight. (Such light sources are unmodulated, and will not actuate a sensor designed to respond only to modulated light.) If the robot is approaching an object, its controller senses that the reflected, modulated beam is getting stronger. The robot can then steer clear of the object.

**30-9** A capacitive proximity sensor can detect nearby conducting or semiconducting objects.





**30-10** A photoelectric proximity sensor. Modulation of the light beam allows the device to distinguish between sensor-generated light and background illumination.

This method of proximity sensing does not work for objects that do not reflect light, or for windows or mirrors approached at a sharp angle. In these scenarios, the light beam is not reflected back toward the photodetector, so the object is invisible.

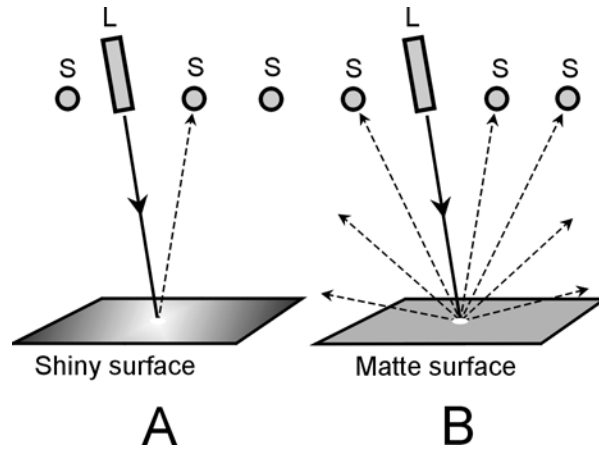
## Texture Sensor

*Texture sensing* is the ability of a machine to determine whether a surface is shiny or rough (matte). Basic texture sensing involves the use of a laser and several light-sensitive sensors.

Figure 30-11 shows how a laser (L) and sensors (S) can be used to tell the difference between a shiny surface (at A) and a rough or matte surface (at B). The shiny surface, such as the polished hood of a car, reflects light at the incidence angle only. But the matte surface, such as a sheet of paper, scatters light in all directions. The shiny surface reflects the beam back entirely to the sensor in the path of the beam whose reflection angle equals its incidence angle. The matte surface reflects the beam back to all the sensors. A microcomputer can be programmed to tell the difference.

Certain types of surfaces can confuse the texture sensor shown in Fig. 30-11. For example, a pile of glass marbles can be defined as shiny on a small scale but irregular on a large scale. Depending on the diameter of the laser beam, the texture sensor might interpret such a surface as either shiny or matte. The determination can also be affected by motion of the sensor relative to the surface. A surface that is interpreted as shiny when standing still relative to the sensor might be interpreted as matte when moving relative to the sensor.

- 30-11** In texture sensing, lasers (L) and sensors (S) analyze a shiny surface (at A) and a matte surface (at B). Solid lines represent incident light; dashed lines represent reflected light.



## Location Systems

Transducers and sensors can operate over long distances. The above-described devices are intended mainly for short-range applications (with the exception of RF antennas). In this section, a few medium-range and long-range applications of transducers and sensors are described. These applications fall into the broad category of *location systems*.

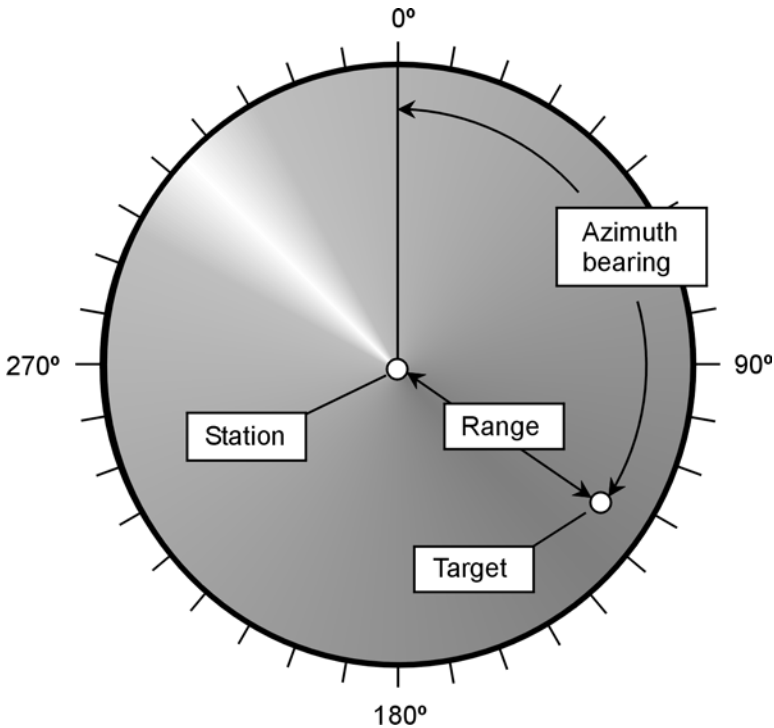
### Radar

The term *radar* is derived from the words *radio detection and ranging*. Electromagnetic (EM) waves having certain frequencies reflect from various objects, especially if those objects contain metals or other good electrical conductors. By ascertaining the direction(s) from which radio signals are returned, and by measuring the time it takes for a pulsed beam of EM energy to travel from the transmitter location to a target and back again, it is possible to pinpoint the geographic positions of distant objects. During the Second World War in the 1940s, this property of radio waves was put to use for the purpose of locating aircraft.

In the years following the war, it was discovered that radar can be useful in a variety of applications, such as measurement of automobile speed (by the police), weather forecasting (rain and snow reflect radar signals), and even the mapping of the moon and the planet Venus. Radar is extensively used in aviation, both commercial and military. In recent years, radar has also found uses in robot guidance systems.

A complete radar set consists of a transmitter, a directional antenna with a narrow main lobe and high gain, a receiver, and an indicator or display. The transmitter produces intense pulses of radio microwaves at short intervals. The pulses are propagated outward in a narrow beam from the antenna, and they strike objects at various distances. The reflected signals, or *echoes*, are picked up by the antenna shortly after the pulse is transmitted. The farther away the reflecting object, or target, the longer the time before the echo is received. The transmitting antenna is rotated so that all azimuth bearings (compass directions) can be observed.

A typical circular radar display consists of a CRT or LCD. Figure 30-12 shows the basic display configuration. The observing station is at the center of the display. Azimuth bearings are indicated



**30-12** A radar display. The light radial band shows the azimuth direction in which the microwave beam is currently transmitted and received. (Not all radar displays show this band.)

in degrees clockwise from true north, and are marked around the perimeter of the screen. The distance, or *range*, is indicated by the radial displacement of the echo; the farther away the target, the farther from the display center the echo or blip. The radar display is, therefore, a set of polar coordinates. In the drawing, a target is shown at an azimuth of about  $125^\circ$  (east-southeast). Its range is near the maximum for the display.

The maximum range of a radar system depends on the height of the antenna above the ground, the nature of the terrain in the area, the transmitter output power and antenna gain, the receiver sensitivity, and the weather conditions in the vicinity. Airborne long-range radar can detect echoes from several hundred kilometers (km) away under ideal conditions. A low-power radar system, with the antenna at a low height above the ground, might receive echoes from only 50 to 70 km.

The fact that precipitation reflects radar echoes is a nuisance to aviation personnel, but it is invaluable for weather forecasting and observation. Radar has made it possible to detect and track severe thunderstorms and hurricanes. A *mesocyclone*, which is a severe thunderstorm likely to produce tornadoes, causes a hook-shaped echo on radar. The eye of a hurricane, and the eyewall and rainbands surrounding it, all show up clearly on a radar display.

Some radar sets can detect changes in the frequency of the returned pulse, thereby allowing measurement of wind speeds in hurricanes and tornadoes. This is called *Doppler radar*. This type of radar is also employed to measure the speeds of approaching or receding targets.



## Sonar

*Sonar* is a medium-range method of proximity sensing. The acronym derives from the words *sonic navigation and ranging*. The basic principle is simple: Bounce acoustic waves off of objects, and measure the time it takes for the echoes to return.

An elementary sonar system consists of an ac pulse generator, an acoustic emitter, an acoustic pickup, a receiver, a delay timer, and an indicating device such as a numeric display, CRT, LCD, or pen recorder. The transmitter sends out acoustic waves through the medium, usually water or air. These waves are reflected by objects, and the echoes are picked up by the receiver. The distance to an object is determined on the basis of the echo delay, provided the speed of the acoustic waves in the medium is known.

A simple sonar system is diagrammed in Fig. 30-13A. A *computer map* can be generated on the basis of sounds returned from various directions in two or three dimensions. This can help a mobile robot or vessel navigate in its environment. However, the system can be fooled if the echo delay is equal to or longer than the time interval between pulses, as shown at B. To overcome this, a computer can instruct the generator to send pulses of various frequencies in a defined sequence. The computer keeps track of which echo corresponds to which pulse.

Acoustic waves travel faster in water than in air. The amount of salt in water makes a difference in the propagation speed when sonar is used on boats, for example, in depth finding. The density of water can vary because of temperature differences as well. If the true speed of the acoustic waves is not accurately known, false readings will result. In freshwater, the speed of sound is about 1400 meters per second (m/s), or 4600 feet per second (ft/s). In saltwater, it is about 1500 m/s (4900 ft/s). In air, sound travels at approximately 335 m/s (1100 ft/s).

In the atmosphere, sonar can make use of audible sound waves, but ultrasound is often used instead. Ultrasound has a frequency too high to hear, ranging from about 20 kHz to more than 100 kHz. One advantage of ultrasound is that the signals will not be heard by people working around machines equipped with sonar. Another advantage is the fact that it is less likely to be fooled by people talking, machinery operating, and other noises. At frequencies higher than the range of human hearing, acoustical disturbances do not normally occur as often, or with as much intensity, as they do within the hearing range.

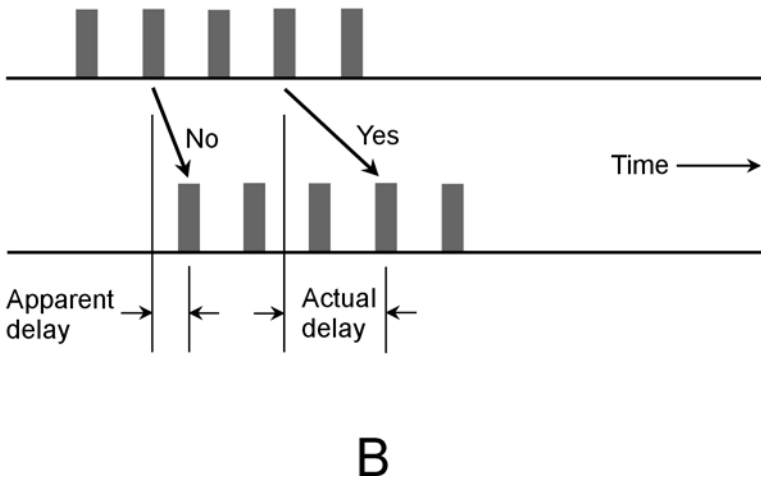
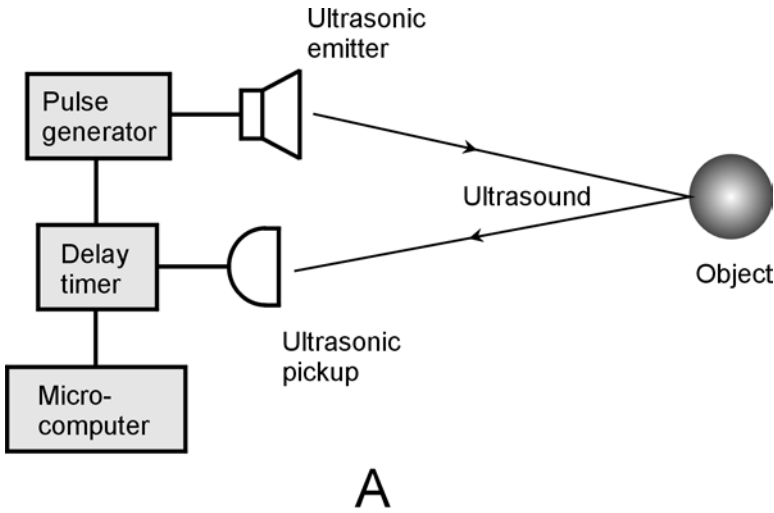
In its most advanced forms, sonar can rival *vision systems* (also called *machine vision*) as a means of mapping the environment for a mobile robot or vessel. Sonar has one significant limitation: all acoustic waves, including sound and ultrasound, require a gaseous or liquid medium in order to propagate. Therefore, sonar is useless in outer space, which is practically a vacuum.

## Signal Comparison

A machine or vessel can find its geographical position by comparing the signals from two fixed stations whose positions are known, as shown in Fig. 30-14A. By adding 180° to the bearings of the sources X and Y, the machine or vessel (small square block) obtains its bearings as seen from the sources (round dots). The machine or vessel can determine its direction and speed by taking two readings separated by a certain amount of time.

In the old days, diagrams such as the one in Fig. 30-14A were actually plotted on maps by the captains of oceangoing vessels and aircraft. Nowadays, computers do that work, with more accurate results.

Figure 30-14B is a block diagram of an *acoustic direction finder* such as can be used by a mobile robot. The receiver has a signal-strength indicator and a servo that turns a directional ultrasonic

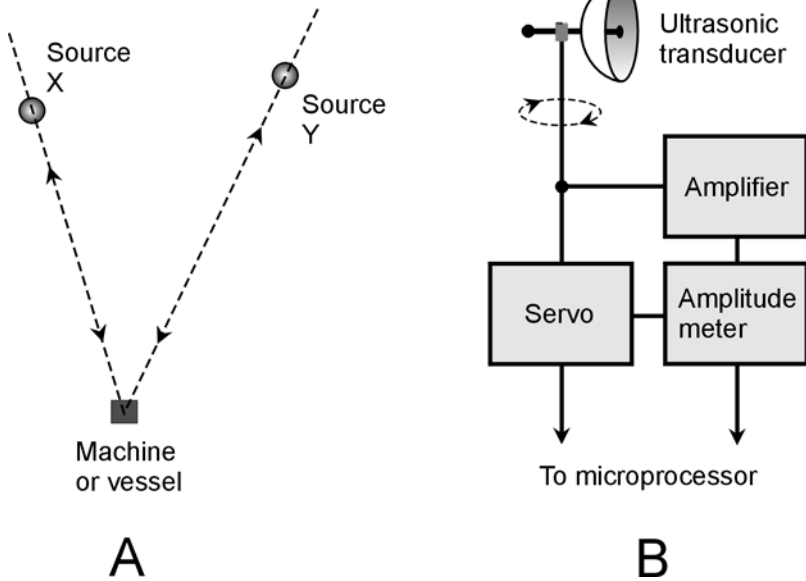


**30-13** At A, a block diagram of medium-range sonar system. At B, a sonar system can be confused by long delays.

transducer. There are two signal sources at different frequencies. When the transducer is rotated so the signal from one source is maximum, a bearing is obtained by comparing the orientation of the transducer with some known standard such as a magnetic compass. The same is done for the other source. A computer determines the precise location of the robot, based on this data.

### Radio Direction Finding (RDF)

A radio receiver, equipped with a signal-strength indicator and connected to a rotatable, directional antenna, can be used to determine the direction from which signals are coming. *Radio direction find-*



**30-14** A simple direction-finding scheme (A) and an ultrasonic direction finder (B).

ing (RDF) equipment aboard a mobile vehicle facilitates determining the location of a transmitter. An RDF receiver can also be used to find one’s own position with respect to two or more transmitters operating on different frequencies.

In an RDF receiver for use at frequencies below about 300 MHz, a small loop or loopstick antenna is used. It is shielded against the electric component of radio waves, so it picks up only the magnetic part of the EM field. The loop is rotated until a sharp dip, or null, occurs in the received signal strength, indicating that the axis of the loop lies along a line toward the transmitter. When readings are taken from two or more locations separated by a sufficient distance, the transmitter can be pinpointed by finding the intersection point of the azimuth bearing lines on a map.

At frequencies above approximately 300 MHz, a directional transmitting and receiving antenna, such as a Yagi, quad, dish, or helical type, gives better results than a small loop. When such an antenna is employed for RDF, the azimuth bearing is indicated by a signal peak rather than by a null.

## Navigational Methods

Navigation involves the use of location devices over a period of time, thereby deriving a function of position versus time. This technique can be used to determine whether or not a vessel is on course. It can also be used to track the paths of military targets, severe thunderstorms, and hurricanes.

### Fluxgate Magnetometer

When conventional position sensors do not function in a particular environment for a mobile robot, a *fluxgate magnetometer* can be used. This system employs sensitive magnetic receptors and a

microcomputer to sense the presence of, and detect changes in, an artificially generated magnetic field. Navigation within a room can be done by checking the orientation of magnetic lines of flux generated by electromagnets in the walls, floor, and ceiling of the room. For each point in the room, the magnetic flux lines have a unique direction and intensity. There is a one-to-one correspondence between the magnetic flux intensity/direction and the points within the robot's operating environment. This correspondence can be represented as a two-variable mathematical function of every location in the room. The robot controller is programmed to know this function. This makes it possible for the machine to pinpoint its position with extreme accuracy, in some cases to within a few millimeters.

### **Epipolar Navigation**

*Epipolar navigation* works by evaluating the way an image changes as viewed from a moving perspective. Suppose that you are piloting an aircraft over the ocean. The only land in sight is a small island. The on-board computer sees an image of the island that constantly changes shape. Figure 30-15 shows three sample sighting positions (A, B, C) and the size/shape of the island as seen by a machine vision system in each case. The computer has the map data, so it knows the true size, shape, and location of the island. The computer compares the shape and size of the image it sees at each point in time, from the vantage point of the aircraft, with the actual shape and size of the island from the map data. From this, the computer can ascertain the altitude of the aircraft, its speed and direction of movement relative to the surface, its latitude, and its longitude.

### **Log Polar Navigation**

In *log polar navigation*, a computer converts an image in *polar coordinates* to an image in *rectangular coordinates*. The polar radius is mapped onto the vertical rectangular axis, and the polar angle is mapped onto the horizontal rectangular axis.

Radial coordinates are unevenly spaced in the polar map, but are uniform in the rectangular map. During the transformation, the logarithm of the radius is taken to maximize the geographical area that the system can observe. As a result of this logarithmic transformation (that's where the "log" comes from in the term), long-distance resolution is sacrificed, but the close-in resolution is enhanced.

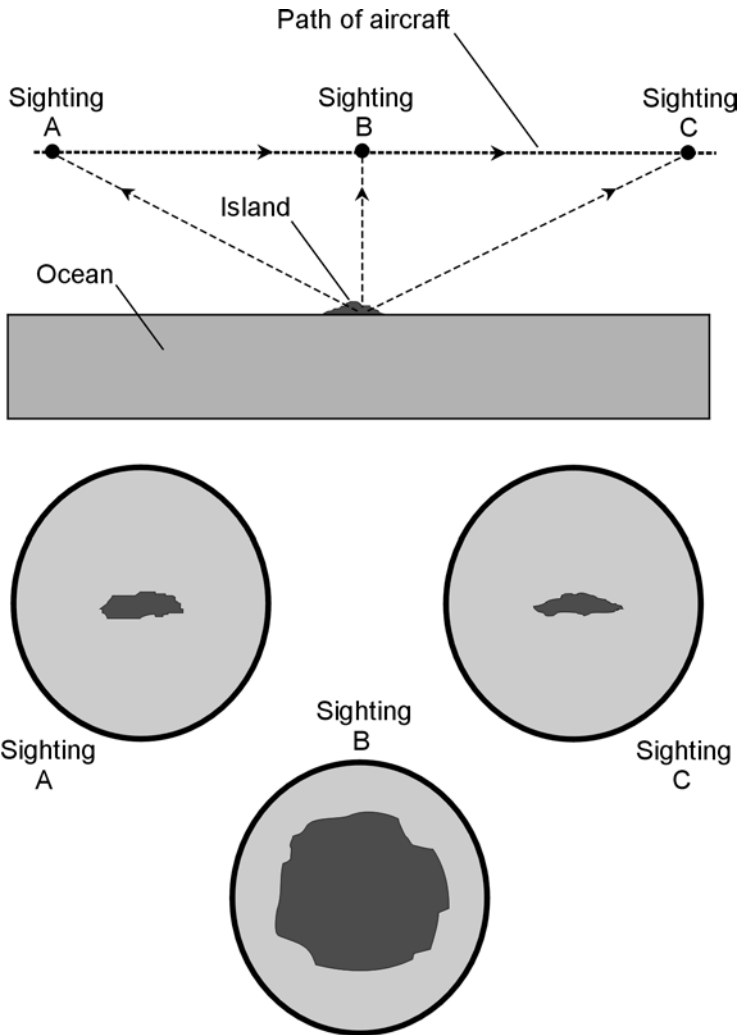
The so-called log polar transform distorts the way a scene appears to human observers, but translates images and motions into data that can be efficiently dealt with by a computerized scanning system.

### **Loran**

The term *loran* is an acronym derived from the words *long range navigation*. Loran is one of the oldest electronic navigation schemes, and is still used by some ships and aircraft. The system employs RF pulse transmission at low frequencies from multiple transmitters at specific geographic locations.

In loran, a computer on board a vessel can determine the location of the ship by comparing the time difference in the arrival of the signals from two different transmitters at known locations. Based on the fact that radio waves propagate at the speed of light in free space (approximately 299,792 km/s or 186,282 mi/s), it is possible to determine the distance to each transmitter, and from this, the location of the ship relative to the transmitters.

In recent years, loran has been largely supplanted by the *Global Positioning System* (GPS).



**30-15** Epipolar navigation is the optoelectronic counterpart of human spatial perception.

### **Global Positioning System (GPS)**

The Global Positioning System (GPS) is a network of radiolocation/radionavigation apparatus that operates on a worldwide basis. The system employs several satellites, and allows determination of latitude, longitude, and altitude.

All GPS satellites transmit signals in the microwave part of the radio spectrum. The signals are modulated with codes that contain timing information used by the receiving apparatus to make measurements. A GPS receiver determines its location by measuring the distances to several different satellites. This is done by precisely timing the signals as they travel between the satellites and the receiver. The process is similar to triangulation, except that it is done in three dimensions (in space) rather than in two dimensions (on the surface of the earth).

There is some reduction in the propagation speed of EM microwaves in the ionosphere, as compared with the propagation speed in free space. The extent of this reduction depends on the signal frequency. The GPS employs dual-frequency transmission to compensate for this effect. A GPS receiver uses a computer to process the information received from the satellites. From this information, it can give the user an indication of position down to a few meters.

Increasingly, automobiles, trucks, and pleasure boats come with GPS receivers installed as standard equipment. If you are driving in a remote area and you get lost, you can use GPS to locate your position. Using a cell phone, citizens band (CB) radio transceiver, or amateur (ham) radio transceiver, you can call for help and inform authorities of your exact position, which is displayed on a detailed map of the area in which you are located.

## Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. Which of the following devices would most likely be used for adjusting the frequency setting of a digital radio transmitter or receiver?
  - (a) A piezoelectric transducer
  - (b) A dynamic transducer
  - (c) An optical shaft encoder
  - (d) A capacitive proximity sensor
2. Collimation of IR rays can be done by means of
  - (a) an ordinary lens.
  - (b) a selsyn.
  - (c) a log polar transform system.
  - (d) any of the above.
3. A computer map
  - (a) can help a robot find its way around.
  - (b) requires the use of IR beacons.
  - (c) is generated using optical encoders.
  - (d) requires the use of a low-frequency RDF loop antenna.
4. The distance of a target from a radar station is called
  - (a) the resolution.
  - (b) the azimuth.
  - (c) the range.
  - (d) the transform.
5. Which of the following devices is best for use as an ultrasonic pickup?
  - (a) A back-pressure transducer
  - (b) A piezoelectric transducer

- (c) An elastomer pressure transducer
  - (d) A capacitive pressure transducer
6. What do Loran and the GPS have in common?
- (a) They both operate using IR energy.
  - (b) They both employ loop antennas at the transmitting and receiving stations.
  - (c) They both involve measurement of azimuth angles to determine the location of a transmitter.
  - (d) They both involve measurement of distances to determine the location of a receiver.
7. A bright spot close to, and directly to the left of, the center of a conventional radar display indicates the presence of a target
- (a) right over the observing station.
  - (b) at close range, west of the observing station.
  - (c) at close range, south of the observing station.
  - (d) far from the observing station, azimuth unknown.
8. Which of the following systems enables a vessel to navigate entirely from its own frame of reference, without the need for any electronic devices external to itself?
- (a) An epipolar navigation system
  - (b) The GPS
  - (c) Loran
  - (d) Forget it! There is no such thing.
9. Ultrasonic waves travel through a vacuum
- (a) at the same speed as they travel in air.
  - (b) at a slightly higher speed than they travel in air.
  - (c) at a slightly lower speed than they travel in air.
  - (d) Forget it! No form of acoustic wave can travel through a vacuum.
10. Which of the following devices or techniques makes use of artificially generated magnetic fields to facilitate location within a work environment?
- (a) A dynamic transducer
  - (b) Epipolar navigation
  - (c) Capacitive proximity sensing
  - (d) None of the above
11. Which of the following is not an example of a transducer?
- (a) A microphone
  - (b) A radio antenna
  - (c) A headset
  - (d) All of the above are examples of transducers.

12. A foam-like material, having a resistance that varies depending on how much it is compressed, is known as
- (a) an elastomer.
  - (b) a fluxgate magnetometer.
  - (c) a piezoelectric substance.
  - (d) none of the above.
13. A stepper motor
- (a) has torque that increases as its speed increases.
  - (b) resists applied torque when stopped with current going through its coils.
  - (c) has turning power that does not depend on the speed.
  - (d) is easily fooled by body capacitance.
14. A permanent magnet is a key component of
- (a) an elastomer pressure sensor.
  - (b) a capacitive pressure sensor.
  - (c) a dynamic microphone.
  - (d) a piezoelectric microphone.
15. An electric motor is an example of
- (a) an electromechanical transducer.
  - (b) a pressure sensor.
  - (c) a proximity sensor.
  - (d) a texture sensor.
16. A capacitive proximity sensor should not be expected to detect the presence of
- (a) a wooden chair.
  - (b) a metal desk.
  - (c) a large dog.
  - (d) a steel post.
17. A well-designed photoelectric proximity sensor
- (a) is easily fooled by loud noises.
  - (b) cannot detect the presence of objects that reflect light.
  - (c) is sensitive to magnetic fields.
  - (d) is not easily fooled by stray sources of light.
18. An acoustic transducer can translate an ac electrical signal into
- (a) visible light.
  - (b) sound or ultrasound.
  - (c) a dc electrical signal.
  - (d) mechanical torque.



19. Body capacitance can be a problem with
- (a) electrostatic transducers.
  - (b) elastomer pressure sensors.
  - (c) capacitive pressure sensors.
  - (d) all of the above.
20. Which of the following devices can be used to limit the torque or force applied by an electromechanical device such as a robot?
- (a) A back-pressure sensor
  - (b) A capacitive proximity sensor
  - (c) A dynamic transducer
  - (d) An optical shaft encoder

## 31 CHAPTER

# Acoustics, Audio, and High Fidelity

IN SOUND RECORDING AND REPRODUCTION, ESPECIALLY WITH MUSIC, FIDELITY SUPERSEDES ALL other considerations. In these applications, low distortion and “sound esthetics” are more important than amplifier efficiency or gain. In this chapter, we’ll look at basic acoustics principles, and outline some *high-fidelity* (hi-fi) components and design concepts.

## Acoustics

*Acoustics* is the science of sound waves. Sound consists of molecular vibrations at audio frequency (AF), ranging from about 20 Hz to 20 kHz. Young people can hear the full range of AF; older people lose hearing sensitivity at the upper and lower extremes.

## Audio Frequencies

In music, the AF range is divided into three broad, vaguely defined parts, called *bass* (pronounced “base”), *midrange*, and *treble*. The bass frequencies start at 20 Hz and extend to 150 or 200 Hz. Midrange begins at this point, and extends up to 2 or 3 kHz. Treble consists of the audio frequencies higher than midrange. As the frequency increases, the wavelength becomes shorter.

In air, sound travels at about 1100 feet per second (ft/s), or 335 meters per second (m/s). The relationship between the frequency  $f$  of a sound wave in hertz, and the wavelength  $\lambda_{ft}$  in feet, is as follows:

$$\lambda_{ft} = 1100/f$$

The relationship between  $f$  in hertz and  $\lambda_m$  in meters is given by:

$$\lambda_m = 335/f$$

This formula is also valid for frequencies in kilohertz and wavelengths in millimeters.

A sound disturbance in air at 20 Hz has a wavelength of 55 ft (17 m). A sound of 1.0 kHz produces a wave measuring 1.1 ft (34 cm). At 20 kHz, a sound wave in air is only 0.055 ft (17 mm)

long. In substances other than air at sea level, such as air at extreme altitudes, freshwater, saltwater, or metals, the preceding formulas do not apply.

## Waveforms

The frequency, or *pitch*, of a sound is only one of several variables that acoustic waves can possess. Another important factor is the shape of the wave itself. The simplest acoustic waveform is a sine wave (or *sinusoid*), in which all of the energy is concentrated at a single frequency. Sinusoidal sound waves are rare in nature. A good artificial example is the beat note, or heterodyne, produced by a steady carrier in a communications receiver.

In music, most of the notes are complex waveforms, consisting of energy at a specific fundamental frequency and its harmonics. Examples are sawtooth, square, and triangular waves. The shape of the waveform depends on the distribution of energy among the fundamental and the harmonics. There are infinitely many different shapes that a wave can have at a single frequency such as 1 kHz. As a result, there is infinite variety in the *timbre* that a single musical note can have.

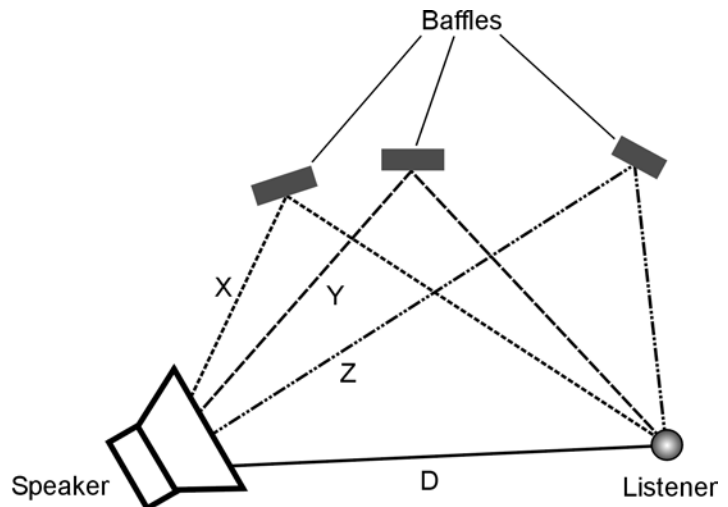
## Path Effects

A flute, a clarinet, a guitar, and a piano can each produce a sound at 1 kHz, but the tone quality is different for each instrument. The waveform affects the way a sound is reflected from objects. Acoustics engineers must consider this when designing sound systems and concert halls. The goal is to make sure that all the instruments sound realistic everywhere in the room.

Suppose you have a sound system set up in your living room, and that, for the particular placement of speakers with respect to your ears, sounds propagate well at 1, 3, and 5 kHz, but poorly at 2, 4, and 6 kHz. This affects the way musical instruments sound. It distorts the sounds from some instruments more than the sounds from others. Unless all sounds, at all frequencies, reach your ears in the same proportions that they come from the speakers, you do not hear the music the way it originally came from the instruments.

Figure 31-1 shows a listener, a speaker, and three sound reflectors, also known as *baffles*. The waves reflected by the baffles (X, Y, and Z), along with the direct-path waves (D), add up to some-

**31-1** Sounds from reflected paths (such as X, Y, and Z) combine with direct-path sound (D) to produce what a listener hears.



thing different, at the listener's ears, for each frequency of sound. This phenomenon is impossible to prevent. That is why it is so difficult to design an acoustical room, such as a concert auditorium, to propagate sound well at all frequencies for every listener.

## Loudness and Phase

You do not perceive the *loudness* (also called *volume*) of sound in direct proportion to the power contained in the disturbance. Your ears and brain sense sound levels according to the logarithm of the actual intensity. Another variable is the *phase* with which waves arrive at your ears. Phase allows you to perceive the direction from which a sound is coming, and it also affects perceived sound volume.

### The Decibel in Acoustics

You have already learned about decibels in terms of signal voltage, current, and power. Decibels are also used in acoustics, and in this application, they are considered in terms of relative power. If you change the volume control on a hi-fi set until you can just barely tell the difference, the increment is one decibel (1 dB). If you use the volume control to halve or double the actual acoustic-wave power coming from a set of speakers, you perceive a change of 3 dB.

For decibels to have meaning in acoustics, there must be a reference level against which everything is measured. Have you read that a vacuum cleaner produces 80 dB of sound? This is determined with respect to the *threshold of hearing*, which is the faintest sound that a person with good hearing can detect in a *quiet room* specially designed to have a minimum of background noise.

### Phase in Acoustics

Even if there is only one sound source, acoustic waves reflect from the walls, ceiling, and floor of a room. In Fig. 31-1, imagine the baffles as two walls and the ceiling in a room. As is the case with baffles, the three sound paths X, Y, and Z are likely to have different lengths, so the sound waves reflected from these surfaces will not arrive in the same phase at the listener's ears. The direct path (D), a straight line from the speaker to the listener, is always the shortest path. In this situation, there are at least four different paths by which sound waves can propagate from the speaker to the listener. In some practical scenarios, there are dozens.

Suppose that, at a certain frequency, the acoustic waves for all four paths happen to arrive in exactly the same phase in the listener's ears. Sounds at that frequency will be exaggerated in volume. The same phase coincidence might also occur at harmonics of this frequency. This is undesirable because it causes acoustic peaks, called *antinodes*, distorting the original sound. At certain other frequencies, the waves might mix in phase opposition. This produces acoustic nulls called *nodes* or *dead zones*. If the listener moves a few feet, the volume at any affected frequency will change. As if this isn't bad enough, a new antinode or node might then present itself at another set of frequencies.

One of the biggest challenges in acoustical design is the avoidance of significant antinodes and nodes. In a home hi-fi system, this can be as simple as minimizing the extent to which sound waves reflect from the ceiling, the walls, the floor, and the furniture. Acoustical tile can be used on the ceiling, the walls can be papered or covered with cork tile, the floor can be carpeted, and the furniture can be upholstered with cloth. In large auditoriums and music halls, the problem becomes more complex because of the larger sound propagation distances involved, and also because of the fact that sound waves reflect from the bodies of the people in the audience!

## Technical Considerations

Regardless of its size, a good hi-fi sound system must have certain characteristics. Here are two of the most important technical considerations.

### Linearity

*Linearity* is the extent to which the output waveform of an amplifier is a faithful reproduction of the input waveform. In hi-fi equipment, all the amplifiers must be as linear as the state of the art allows.

If you connect a dual-trace oscilloscope (one that lets you observe two waveforms at the same time) to the input and output terminals of a hi-fi audio amplifier with good linearity, the output waveform is a vertically magnified duplicate of the input waveform. When the input signal is applied to the horizontal scope input and the output signal is applied to the vertical scope input, the display is a straight line. In an amplifier with poor linearity, the instantaneous output-versus-input function is not a straight line. The output waveform is not a faithful reproduction of the input, and distortion occurs. In some RF amplifiers this is all right. In hi-fi audio systems, it is not.

Hi-fi amplifiers are designed to work with input signals up to a certain peak (maximum instantaneous) amplitude. If the peak input exceeds this level, the amplifier becomes nonlinear, and distortion is inevitable. In a hi-fi system equipped with VU or distortion meters, excessive input causes the needles to kick up into the red range of the scale during peaks.

### Dynamic Range

*Dynamic range* is a prime consideration in hi-fi recording and reproduction. As the dynamic range increases, the sound quality improves for music or programming having a wide range of volume levels. Dynamic range is expressed in decibels (dB).

At low volume levels, the limiting factor in dynamic range is the *background noise* in the system. In an analog system, most of this noise comes from the audio amplification stages. In tape recording, there is also some *tape hiss*. A scheme called *Dolby* (a trademark of Dolby Laboratories) is used in professional recording studios, and also in high-end consumer tape equipment, to minimize this hiss. Digital recording systems produce less internal noise than analog systems.

At high volume levels, the power-handling capability of an audio amplifier limits the dynamic range. If all other factors are equal, a 100-W audio system can be expected to have greater dynamic range than a 50-W system. The speaker size is also important. As speakers get physically larger, their ability to handle high power improves, resulting in increased dynamic range. This is why serious audio enthusiasts sometimes purchase sound systems with amplifiers and speakers that seem unnecessarily large.

## Components

There are myriad ways to set up a hi-fi system. A true *audiophile* (“sound lover”) assembles a complex system over a period of time. Here are some basic considerations that can serve as guidelines when choosing system components.

### Configurations

The simplest type of home stereo system is contained in a single box, with an AM/FM radio receiver and a *compact disk* (CD) player. The speakers are generally external, but the connecting cables are short. The assets of a so-called compact hi-fi system are small size and low cost.

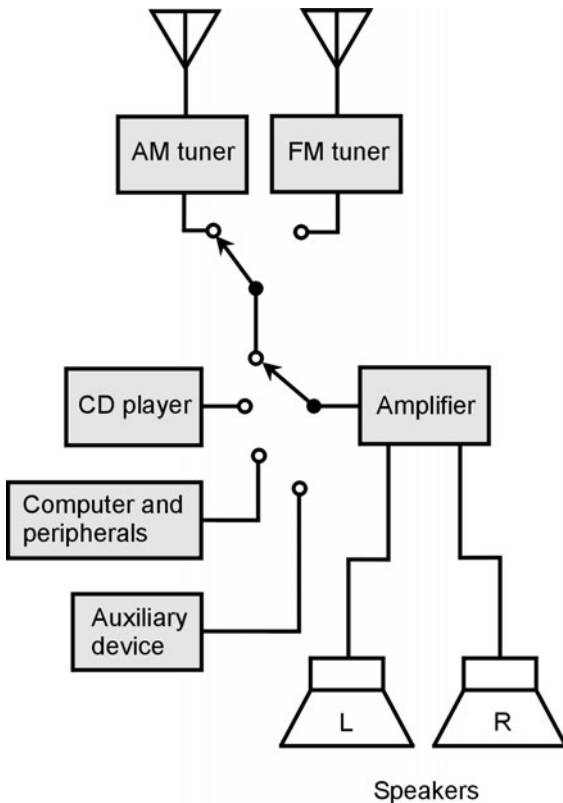
More sophisticated hi-fi systems have separate boxes containing components such as the following:

- An AM tuner
- An FM tuner
- An amplifier or pair of amplifiers
- A CD player
- A computer and its peripherals

The computer is optional, but it facilitates downloading music files or *streaming audio* from the Internet, creating (“burning”) CDs, and composing and editing electronic music. A *satellite radio* receiver, a *tape player*, a *turntable*, or other nonstandard peripheral may also be included. The individual hardware units in this type of system, known as a *component hi-fi system*, are interconnected with shielded cables. A component system costs more than a compact system, but the sound quality is better, you get more audio power, you can do more tasks, and you can tailor the system to your preferences.

Some hi-fi manufacturers build all their equipment cabinets to a standard width so they can be mounted one above the other in a *rack*. A so-called rack-mounted hi-fi system saves floor space and makes the system look professional. The rack can be mounted on wheels so the whole system, except for the external speakers, can be rolled from place to place.

Figure 31-2 is a block diagram of a typical home stereo hi-fi system. The amplifier chassis is grounded to minimize hum and noise, and to minimize susceptibility to interference from external



**31-2** A basic stereo hi-fi system.

sources. The AM antenna is usually a loopstick built into the cabinet or mounted on the rear panel. The FM antenna can be an indoor type, such as television rabbit ears, or a directional outdoor antenna equipped with lightning protection hardware.

## The Tuner

A *tuner* is a radio receiver capable of receiving signals in the standard AM broadcast band (535 to 1605 kHz) and/or the standard FM broadcast band (88 to 108 MHz). Tuners don't have built-in amplifiers. A tuner can provide enough power to drive a headset, but an amplifier is necessary to provide sufficient power to a pair of speakers.

Modern hi-fi tuners employ frequency synthesizers and have digital readouts. Most tuners have several *memory channels*. These are programmable, and allow you to select your favorite stations with a push of a single button, no matter where the stations happen to be in the frequency band. Most tuners also have *seek* and/or *scan* modes that allow the radio to automatically search the band for any station strong enough to be received clearly.

## The Amplifier

In hi-fi, an amplifier delivers medium or high audio power to a set of speakers. There is at least one input, but more often there are three or more: one for a CD player, another for a tuner, and still others for auxiliary devices such as a tape player, turntable, or computer. Input requirements are a few milliwatts; the output can range up to hundreds of watts.

Amplifier prices increase with increasing power output. A simplified hi-fi amplifier forms the basis for a public-address system. Massive amplifiers are used by popular music bands. Some such systems employ vacuum tubes, because tubes offer electrical ruggedness and excellent linearity.

## Speakers and Headsets

No amplifier can deliver sound that is better than the speakers will allow. Speakers are rated according to the audio power they can handle. It's a good idea to purchase speakers that can tolerate at least twice the audio output power that the amplifier can deliver. This will ensure that speaker distortion will not occur during loud, low-frequency sound bursts. It will also prevent physical damage to the speakers that might otherwise result from accidentally overdriving them.

Good speakers contain two or three individual units within a single cabinet. The *woofer* reproduces bass. The *midrange speaker* handles medium and, sometimes, treble (high) audio frequencies. A *tweeter* is designed especially for enhanced treble reproduction.

Headsets are rated according to how well they reproduce sound. This is a subjective consideration. Equally expensive headsets can, and often do, exhibit huge differences in the quality of the sound that they put out. Not only that, but people disagree about what constitutes good sound.

## Balance Control

In hi-fi stereo sound equipment, the *balance control* allows adjustment of the relative volumes of the left and right channels.

In a basic hi-fi system, the balance control consists of a single rotatable knob connected to a pair of potentiometers. When the knob is rotated counterclockwise, the left-channel volume increases and the right-channel volume decreases. When the knob is rotated clockwise, the right-channel volume increases and the left-channel volume decreases. In more sophisticated sound systems, the balance is adjusted by means of two independent volume controls, one for the left channel and the other for the right channel.

Proper balance is important in stereo hi-fi. A balance control can compensate for such factors as variations in speaker placement, relative loudness in the channels, and the acoustical characteristics of the room in which the equipment is installed.

## Tone Control

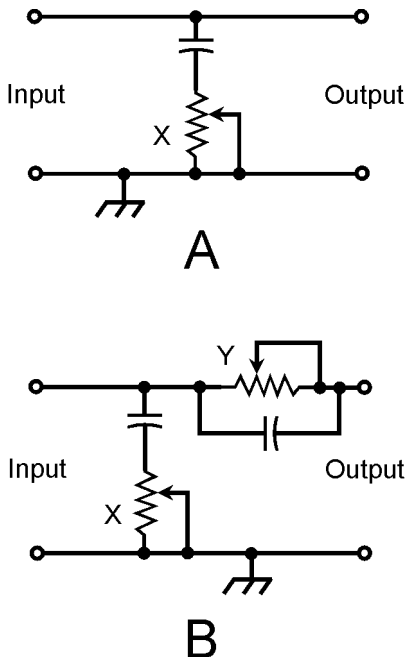
The amplitude versus frequency characteristics of a hi-fi sound system are adjusted by means of a tone control or controls. In its simplest form, a tone control consists of a single knob or slide device. The counterclockwise, lower, or left-hand settings of this control result in strong bass and weak treble audio output. The clockwise, upper, or right-hand settings result in weak bass and strong treble. When the control is set to midposition, the audio response of the amplifier is more or less flat; that is, the bass, midrange, and treble are in roughly the same proportions as in the recorded or received signal.

Figure 31-3A shows how a single-knob tone control can be incorporated into an audio amplifier. The amplifier is designed so that the treble response is exaggerated. The potentiometer attenuates the treble to a variable extent.

A more versatile tone control has two capacitors and two potentiometers, as shown in Fig. 31-3B. One combination is in series, and the other is in parallel. The series-connected resistance-capacitance ( $RC$ ) circuit is connected in parallel with the audio output, and it attenuates the treble to a variable extent. The parallel  $RC$  circuit is in series with the audio path, and it attenuates the bass to a variable extent. The two potentiometers can be adjusted separately, although there is some interaction.

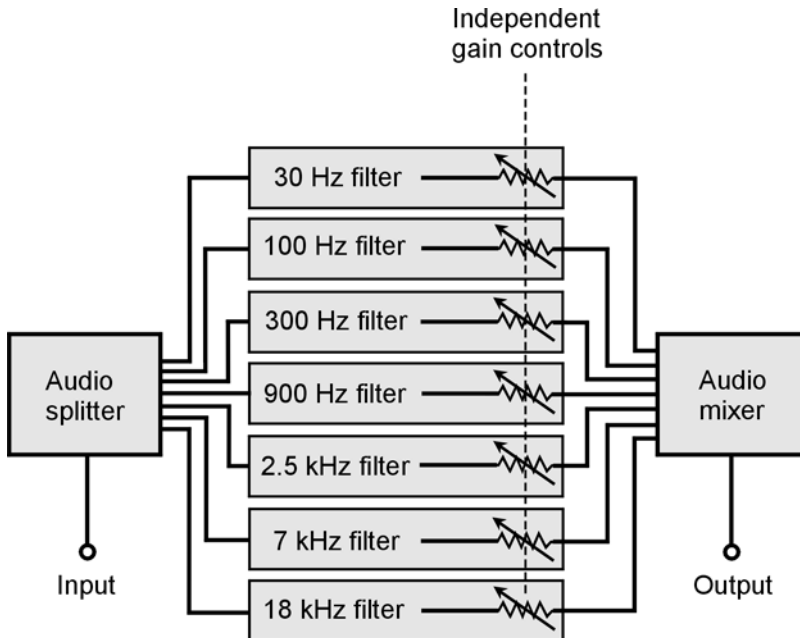
## Audio Mixer

If you simply connect two or more audio sources to the same input of an amplifier, you can't expect good results. Different signal sources (such as a computer, a tuner, and a CD player) are likely to



**31-3** Methods of tone control. At A, a single potentiometer/capacitor combination (X) provides treble attenuation only. At B, one potentiometer/capacitor combination (X) attenuates the treble, and the other (Y) attenuates the bass.





**31-4** A graphic equalizer consists of an audio splitter, several bandstop filters, several gain controls, and an audio mixer.

have different impedances. When connected together, the impedances appear in parallel. This can cause impedance mismatches for most or all of the sources, as well as at the amplifier input. The result will be degradation of system efficiency and poor overall performance.

Another problem arises from the fact that the signal amplitudes from various sources almost always differ. A microphone produces minuscule audio-frequency currents, whereas some tuners produce enough power to drive a pair of small loudspeakers. Connecting both of these together will cause the microphone signal to be obliterated by the signal from the tuner. In addition, the tuner output audio might damage the microphone.

An *audio mixer* eliminates all of the problems involved with the connection of multiple devices to a single channel. First, it isolates the inputs from each other, so there is no impedance mismatch or competition among the sources. Second, the gain at each input can be varied independently. This allows adjustment of amplitudes so the signals blend in the desired ratio.

## Graphic Equalizer

A *graphic equalizer* is a device for adjusting the relative loudness of audio signals at various frequencies. It allows tailoring of the amplitude versus frequency output of hi-fi sound equipment. Equalizers are used in recording studios and by serious hi-fi stereo enthusiasts. There are several independent gain controls, each one affecting a different part of the audible spectrum. The controls are slide potentiometers with calibrated scales. The slides move up and down, or, in some cases, left to right. When the potentiometers are set so that the slides are all at the same level, the audio output or response is flat, meaning that no particular range is amplified or attenuated with respect to the whole AF spectrum. By moving any one of the controls, the user can adjust the gain within a

certain frequency range without affecting the gain outside that range. The positions of the controls on the front panel provide an intuitive graph of the output or response curve.

Figure 31-4 is a block diagram of a hypothetical graphic equalizer with seven gain controls. The input is fed to an *audio splitter* that breaks the signal into seven paths of equal impedance, and prevents interaction among the circuits. The seven signals are fed to *audio attenuators*, also called *band-stop filters*, each filter having its own gain control. (The center frequencies of the attenuators in this example are at 30, 100, 300, and 900 Hz, and 2.5, 7, and 18 kHz. These are not standard frequencies, and are given here only for illustrative purposes.) The slide potentiometers affect the extent to which each filter affects the gain within its frequency range. Finally, the signals pass through an audio mixer, and the composite is sent to the output.

There are several challenges in the design and proper use of graphic equalizers. The filter gain controls must not interact. Judicious choice of filter frequencies and responses is important. The filters must not introduce distortion. The active devices must not generate significant audio noise. Graphic equalizers are not built to handle high power, so they must be placed at low-level points in an audio amplifier chain. In a multichannel circuit such as a stereo sound system, a separate graphic equalizer can be used for each channel.

## Specialized Systems

Mobile and portable hi-fi systems operate at low dc voltages. Typical audio power levels are much lower than in home hi-fi systems. Speakers are much smaller also. In portable systems, headsets are often used in place of speakers.

### Mobile Systems

Mobile hi-fi systems, designed for cars and trucks, usually have four speakers. The left and right channels each supply a pair of speakers. The left stereo channel drives the left front and left rear speakers; the right stereo channel drives the right front and right rear speakers. The balance control adjusts the ratio of sound volume between the left and right channels for both the front and rear speaker sets. Another control adjusts the ratio of sound volume between the front and rear sets.

A mobile hi-fi system typically has an AM/FM receiver and a CD player. Some older vehicles have systems with cassette tape players. Newer, high-end cars and trucks have satellite radio receivers as well as conventional AM/FM receivers and CD players. One note of caution: compact disks and cassettes are heat-sensitive, so they should not be stored in a car or truck that will be left out in the sun on a warm day.

### Portable Systems

Portable hi-fi systems operate from sets of dry cells or rechargeable cells. The most well-known portable hi-fi set is the so-called headphone radio. There are literally hundreds of different designs. Some include only an FM radio; some have AM/FM reception capability. Some have a small box with a cord that runs to the headset; others are entirely contained within the headset. There are portable CD players and even portable satellite radio receivers. The sound quality from these systems can be excellent. The defining factor is the quality of the headset.

Another form of portable hi-fi set, sometimes called a *boom box*, can produce several watts of audio output, and delivers the sound to a pair of speakers built into the box. A typical boom box is

about 8 in high by 18 in wide by 6 in deep. It includes an AM/FM radio and a CD player. The system gets its name from the loud bass acoustic energy peaks its speakers can deliver.

## Quadraphonic Sound

*Quadraphonic sound* refers to four-channel audio recording and reproduction. It is also called *quad stereo* or *four-channel stereo*. Each of the four channels is independent of the other three. In a well-designed quadraphonic sound system, the speakers should be level with the listener, equidistant from the listener, and separated by angles of  $90^\circ$  from the listener's point of view. If the listener is facing north, for example, the left front speaker is to the northwest, the right front speaker is to the northeast, the left rear speaker is to the southwest, and the right rear speaker is to the southeast. This provides optimum balance, and also facilitates the greatest possible left-to-right and front-to-rear contrast in the perceived sounds.

## Recorded Media

Methods of recording sound, particularly music, have evolved dramatically since the ascent of digital technology. Several types of media are available, but the compact disk (CD) is the most common. Other media include *analog audio tape*, *digital audio tape*, and (of historical interest) *vinyl disk*.

### Compact Disk

A compact disk (often spelled “disc” in hi-fi applications), also called a CD, is a plastic disk with a diameter of 4.72 in (12.0 cm), on which data can be recorded in digital form. Any kind of data can be digitized: sound, images, and computer programs and files. This data can also be stored on other digital media of sufficient capacity, such as a computer hard drive or a backup tape drive.

Digital sound, recorded on the surface of a CD, is practically devoid of the hiss and crackle that bedevil recordings on analog media. This is because the information on the disk is binary: a bit (binary digit) is either 1 (high) or 0 (low). The distinction between these two states is more clear-cut than the subtle fluctuations of an analog signal.

When a CD is prepared, the sound is first subjected to *analog-to-digital (A/D) conversion*. This changes the continuously variable audio waves into logic bits. These bits are then burned (literally) into the surface of the disk in the form of microscopic pits. The pits are arranged in a spiral track that would, if unwound, measure several kilometers in length. *Digital signal processing (DSP)* minimizes the noise introduced by environmental factors such as microscopic particles on the disk or random electronic impulses in circuit hardware. A *scrambling* process “smears” recordings throughout the disk, rather than burning the pits in a direct linear sequence. This further improves the signal-to-noise (S/N) ratio.

Compact-disk players recover the sound from a disk without any hardware physically touching the surface. A laser beam scans the disk. The beam is scattered by the pits and is reflected from the unpitted plastic. The result is a digitally modulated beam that is picked up by a sensor and converted into electrical currents. These currents proceed to a *descrambling* circuit, a *digital-to-analog (D/A) converter*, a DSP circuit, and audio amplifiers. Speakers or headphones convert the audio currents into sound waves.

With a CD player, the track location processes are entirely electronic, and they can all be done quickly. Tracks are assigned numbers that you select by pressing buttons. It is impossible to damage

the CD, no matter how much you skip around among the songs. You can move instantly to any desired point within an individual track. You can program the system to play only those tracks you want, ignoring the others.

## Analog Audio Tape

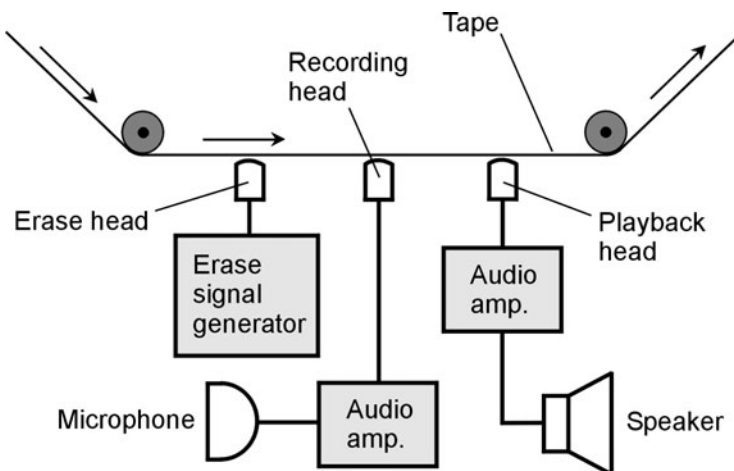
Analog audio tape recorders can be classified as either cassette or reel-to-reel. A typical audio cassette plays for 30 minutes on each side; longer-playing cassettes allow recording for as much as 60 minutes per side. The longer tapes are thinner and more subject to stretching than the shorter tapes.

A reel-to-reel tape feed system resembles that of an old-fashioned movie projector. The tape is wound on two flat spools called the *supply reel* and the *take-up reel*. The reels rotate counterclockwise as the tape passes through the recording/playback mechanism. When the take-up reel is full and the supply reel is empty, both reels can be flipped over and interchanged. This allows recording or playback on the other side of the tape. (Actually, the process takes place on the same side of the tape, but on different tracks.) The speed is usually 1.875, 3.75, or 7.5 inches per second (ips).

Figure 31-5 is a simplified functional diagram of an analog audio tape recording/playback mechanism.

In the *record mode*, the tape moves past the *erase head* before anything is recorded on it. If the tape is not blank (that is, if magnetic impulses already exist on it), the erase head removes these before anything else is recorded. This prevents *doubling*, or the simultaneous presence of two programs on the tape. The erase head can be disabled in some tape recorders when doubling is desired. The *recording head* is a small electromagnet. It generates a fluctuating magnetic field whose instantaneous flux density is proportional to the instantaneous level of the audio input signal. This magnetizes the tape in a pattern that duplicates the waveform of the signal. The *playback head* is usually not activated in the record mode. However, the playback head and circuits can be used while recording to create an echo effect.

In the *playback mode*, the erase head and recording head are not activated. The playback head acts as a sensitive magnetic-field detector. As the tape moves past it, the playback head is exposed to a fluctuating magnetic field whose waveform is identical to that produced by the recording head when the audio was originally recorded on the tape. This magnetic field induces weak alternating



**31-5** Simplified functional diagram of the recording and playback hardware in an analog audio tape recorder/player.

currents in the playback head. These currents are amplified and delivered to a speaker, headset, or other output device.

### Digital Audio Tape

Digital audio tape (DAT) is magnetic recording tape on which binary digital data can be recorded. In digital audio recording, tape noise is practically eliminated because such noise is analog in nature. Some electronic noise is generated in the analog amplification stages following D/A conversion, but this is minimal compared with the noise generated in older, fully analog systems. The reduced noise in DAT equipment provides more true-to-life reproduction than is possible with analog methods.

With DAT, multigeneration copies can be made with practically no degradation in audio fidelity. The reason for this is the same as the reason a computer can repeatedly read and overwrite data on a magnetic disk. On DAT, the bits are represented by distinct magnetized regions on the tape. While analog signals are fuzzy in the sense that they vary continuously, digital signals are crisp. Imperfections in the recording apparatus, the tape itself, and the pickup head affect digital signals less than they affect analog signals. Digital signal processing can eliminate the minute flaws that creep into a digital signal each time it is recorded and played back.

### Vinyl Disk

Vinyl disks were superseded years ago by CDs and Internet downloads, but some audiophiles are still intrigued by vinyl. Some vinyl disks, and the turntables that can play them, have attained considerable value as collectors' items. The main trouble with vinyl is that it can be physically damaged by even the slightest mishandling. Electrostatic effects can produce noise when the humidity is low, as in alpine or far northern regions in the winter.

Vinyl disks require a turntable that spins at speeds of 33 and 45 revolutions per minute (rpm). There are various drive systems. These are called *rim drive*, *belt drive*, and *direct drive*. The best type is a matter of individual taste because there are various factors to consider, such as cost, audio quality, ruggedness, and durability.

## Electromagnetic Interference

As hi-fi equipment becomes more sophisticated and complex, the circuits become more susceptible to interference from outside sources, particularly electromagnetic (EM) fields. This problem is known as *electromagnetic interference* (EMI). Sometimes it is called *radio-frequency interference* (RFI).

If a radio transmitter is operated near a hi-fi stereo system, the radio signals can be intercepted by the hi-fi wiring and peripherals, and delivered to the amplifier. Unshielded interconnecting cables act as radio receiving antennas. This problem is exacerbated if any of the connecting cables happen to resonate at the operating frequency of the radio transmitter. In the hi-fi amplifier, the RF currents are rectified, causing changes in the audio gain. Sometimes the signal data can be heard in the speakers or headset.

In most cases when EMI takes place in a hi-fi setup, the fault exists in the stereo system design, not in the radio transmitter. The transmitter system is merely doing its job: generating and radiating electromagnetic signals!

Several steps can be taken when installing a stereo hi-fi system to minimize the likelihood that EMI will occur:

- Connect the stereo amplifier chassis to a good electrical ground.
- Use shielded interconnecting cables as much as possible.
- Use shielded (coaxial) speaker cables.
- Keep all cables as short as possible.

If you have an amateur or citizens band (CB) radio station in your house and it causes EMI to your hi-fi system, perform these additional steps:

- Locate the radio transmitting antenna as far from the hi-fi equipment as possible.
- Use the lowest possible transmitter output power that will ensure reliable communications.

Unfortunately, EMI problems can sometimes prove nigh impossible to eliminate. This can be especially troublesome for amateur and CB radio operators when it damages relations with neighbors. In these cases, old-fashioned diplomacy may work better than engineering-based attempts at resolution.

## Quiz

Refer to the text in this chapter if necessary. A good score is 18 correct. Answers are in the back of the book.

1. A young person can hear sounds at frequencies as high as approximately which of the following?
  - (a) 20 Hz
  - (b) 20 kHz
  - (c) 20 MHz
  - (d) 20 GHz
2. Electromagnetic interference to a hi-fi amplifier can occur in the presence of
  - (a) a nearby radio broadcast station.
  - (b) improperly designed receiving antennas.
  - (c) excessive utility voltage.
  - (d) improper balance between the left and right channels.
3. The midrange audio frequencies
  - (a) are exactly halfway between the lowest and highest audible frequencies.
  - (b) represent sounds whose volume levels are not too loud or too soft.
  - (c) are above the treble range but below the bass range.
  - (d) None of the above are true.
4. In the acoustical design of a room intended for a home audio system,
  - (a) the use of small speakers can minimize distortion.
  - (b) reflection of sound waves from walls should be minimized.
  - (c) the walls should all intersect at perfect 90° angles.
  - (d) wooden furniture, without upholstery, should be used.

5. A change of +10 dB in an audio signal represents
  - (a) a doubling of acoustic power.
  - (b) a threefold increase in acoustic power.
  - (c) a tenfold increase in acoustic power.
  - (d) no change in acoustic power, but a change in frequency.
6. What is the frequency of an acoustic disturbance whose wavelength is 120 mm in air?
  - (a) 279 Hz
  - (b) 2.79 kHz
  - (c) 35.8 Hz
  - (d) 358 Hz
7. What is the frequency of a sound wave that propagates at a speed of 1100 ft/s?
  - (a) 33.5 Hz
  - (b) 335 Hz
  - (c) 3.35 kHz
  - (d) Forget it! The frequency of a sound wave is independent of the propagation speed.
8. The relative phase of two acoustic waves from the same source at the same time, one wave direct and one wave reflected from a wall, can affect
  - (a) the positions of antinodes and nodes.
  - (b) the perceived frequency.
  - (c) the positions of antinodes and nodes, and the perceived frequency.
  - (d) neither the positions of the antinodes and nodes, nor the perceived frequency.
9. In an acoustic sine wave,
  - (a) the frequency and phase are identical.
  - (b) the sound power is inversely proportional to the frequency.
  - (c) the sound power is directly proportional to the frequency.
  - (d) all of the sound power is concentrated at a single frequency.
10. Vinyl disks are
  - (a) susceptible to physical damage.
  - (b) useful primarily in high-power sound systems.
  - (c) digital media.
  - (d) preferred for off-the-air sound recording.
11. If an amplifier introduces severe distortion in the waveforms of input signals, then that amplifier is
  - (a) not delivering enough power.
  - (b) operating at the wrong frequency.
  - (c) operating in a nonlinear fashion.
  - (d) being underdriven.

12. Suppose a 10-W amplifier is used with speakers designed for a 100-W amplifier. Which of the following statements is true?
- (a) The speakers are capable of handling the amplifier output.
  - (b) The amplifier might be damaged by the speakers.
  - (c) Electromagnetic interference is more likely to occur than would be the case if the speakers were designed for a 10-W amplifier.
  - (d) The speakers are likely to produce excessive distortion.
13. Which of the following frequencies cannot be received by an AM/FM tuner?
- (a) 830 kHz
  - (b) 95.7 kHz
  - (c) 90.1 MHz
  - (d) 107.3 MHz
14. Which of the following statements about woofers is true?
- (a) They are especially useful for reproducing the sounds of barking dogs.
  - (b) They are designed to handle short, intense bursts of sound.
  - (c) They should not be used with graphic equalizers.
  - (d) They are specifically designed to reproduce low-frequency sounds.
15. Suppose you have an amateur radio station and its transmitter causes EMI to your hi-fi system. Which of the following would almost certainly *not* help?
- (a) Build a new transmitter that works on the same frequencies with the same power output as your existing transmitter.
  - (b) Reduce the transmitter output power.
  - (c) Install shielded speaker wires in the hi-fi system, and be sure the system is well grounded.
  - (d) Move the amateur radio transmitting antenna to a location farther away from the hi-fi system.
16. In an analog audio tape recorder/player, the recording head
- (a) converts sound waves to radio signals.
  - (b) converts sound waves to fluctuating electric current.
  - (c) converts AF current to a fluctuating magnetic field.
  - (d) converts dc to AF current.
17. A rack-mounted hi-fi system
- (a) can save floor space.
  - (b) is more susceptible to EMI than a compact system.
  - (c) is cheaper than a compact system.
  - (d) is designed especially for use with vinyl disks and turntables.
18. An audio mixer
- (a) cannot match impedances among interconnected components.
  - (b) cannot increase the audio output of an amplifier.



- (c) eliminates EMI problems in all but the worst cases.
  - (d) allows a microphone to be used as a speaker.
19. Which of the following devices or circuits ordinarily employs D/A conversion?
- (a) A portable CD player
  - (b) A turntable for use with vinyl disks
  - (c) An analog audio tape recording/playback system
  - (d) An *RC* tone control
20. Which of the following is an advantage of digital audio over analog audio?
- (a) Digital media can be used to record and play back audio at higher frequencies than can analog media.
  - (b) Multigeneration copies of digital audio programs can be made without degradation in fidelity, but this is not true of analog audio programs.
  - (c) Digital audio is compatible with vinyl disks and old-fashioned reel-to-reel and cassette tape, but analog audio is not.
  - (d) There are no genuine advantages of digital audio over analog audio. In fact, analog audio is superior to digital audio in every respect.

## 32 CHAPTER

# Personal and Hobby Wireless

THE TERM *WIRELESS* AROSE IN THE EARLY 1900S WHEN INVENTORS AND EXPERIMENTERS BEGAN sending and receiving messages using electromagnetic fields. Gradually, the terminology changed, and wireless became known as *radio*, *television*, and *electromagnetic communications*. In the 1980s and 1990s, the term *wireless* emerged again, this time in a consumer context.

## Cellular Communications

Wireless telephone sets are used in a specialized communications system called *cellular*. Originally, the cellular communications network was used mainly by traveling business people. Nowadays, many people regard *cell phones* as necessities, and some of them come equipped with extra features such as text messaging, Web browsing, and digital cameras.

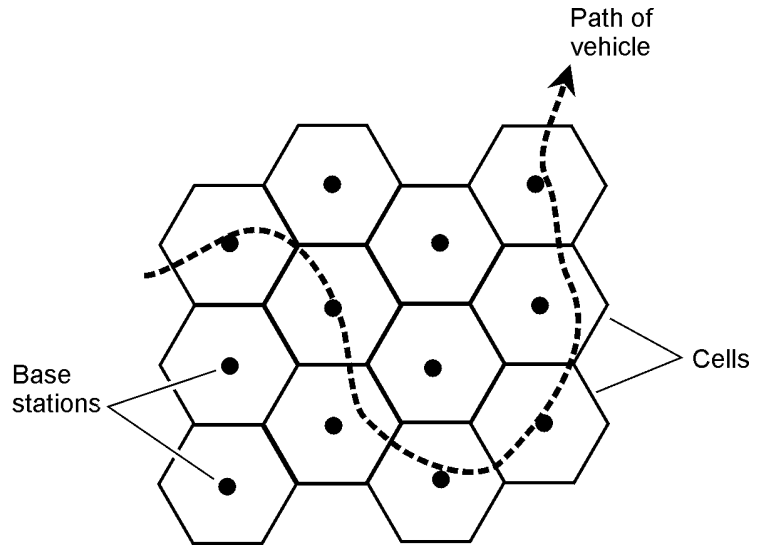
### How Cellular Systems Work

A cell phone looks like a cross between a cordless telephone receiver and a walkie-talkie, but smaller. The unit contains a radio transmitter and receiver combination called a *transceiver*. Transmission and reception take place on different frequencies, so you can talk and listen at the same time, and easily interrupt the other party if necessary. This capability is known as *full duplex*.

In an ideal cellular network, all the transceivers are always within range of at least one *repeater*. The repeaters pick up the transmissions from the portable units and retransmit the signals to the telephone network and to other portable units. The region of coverage for any repeater (also known as a *base station*) is called a *cell*.

When a cell phone is in motion, say in a car or on a boat, the set goes from cell to cell in the network. This situation is shown in Fig. 32-1. The curved, dashed line is the path of the vehicle. Base stations (dots) transfer access to the cell phone. This is called *handoff*. The hexagons show the limits of the transmission/reception range for each base station. All the base stations are connected to the regional telephone system. This makes it possible for the user of the portable unit to place

**32-1** In an ideal cellular system, a moving vehicle is always within range of at least one base station.



calls to, or receive calls from, anyone else in the system, whether those other subscribers have cell phones or regular phones.

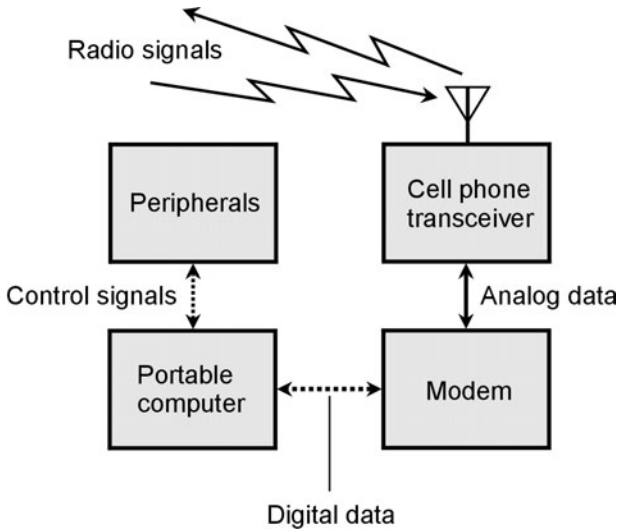
Older cellular systems occasionally suffer from call loss or breakup when signals are handed off from one repeater to another. This problem has been largely overcome by a technology called *code-division multiple access* (CDMA). In CDMA, the repeater coverage zones overlap significantly, but signals do not interfere with each other because every phone set is assigned a unique signal code. Rather than abruptly switching from one base-station zone to the next, the signal goes through a region in which it is sent through more than one base station at a time. This *make-before-break* scheme gets rid of one of the most annoying problems inherent in cellular communications.

In order to use a cellular network, you must purchase or rent a transceiver and pay a monthly fee. The fees vary, depending on the location and the amount of time per month you use the service. When using such a system, it is important to keep in mind that your conversations are not necessarily private. It is easier for unauthorized people to eavesdrop on wireless communications than to intercept wire or cable communications.

## Cell Phones and Computers

You can connect a *personal computer* (PC) to a cell phone with a portable *modem* that converts incoming computer data from analog to digital form, and also converts outgoing data from digital to analog form. In this way, you can access the Internet from anywhere within range of a cellular base station. Figure 32-2 is a block diagram of this scheme.

Most commercial aircraft have telephones at each row of seats, complete with jacks into which you can plug a modem. If you plan to access the Internet from an aircraft, you must use the phones provided by the airline, not your own cell phone, because radio transceivers can cause interference to flight instruments. You must also observe the airline's restrictions concerning the operation of electronic equipment while in flight.



**32-2** A cell phone can be equipped with a modem, allowing portable or mobile access to online computer networks.

## Satellites and Networks

A satellite system is like a huge cellular network with the base stations (repeaters) located in space rather than on the earth's surface. The zones of coverage are large, and they change in size and shape if the satellite moves relative to the earth's surface.

### Geostationary-Orbit Satellites

You learned about geostationary orbits in Chap. 25. Geostationary satellites are used in television (TV) broadcasting, in telephone and data communication, for gathering weather and environmental data, and for radiolocation.

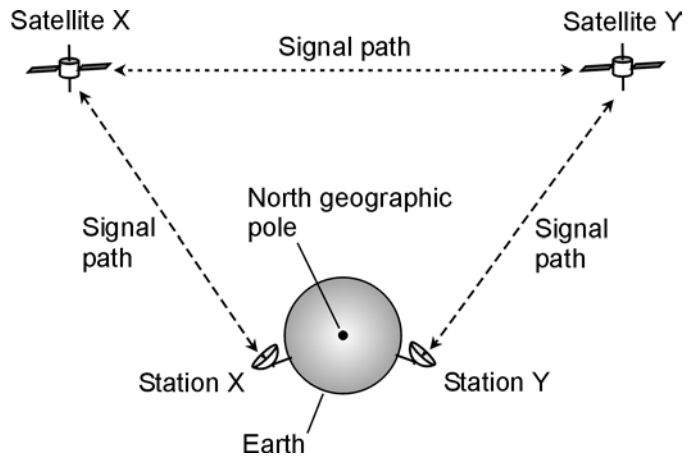
In geostationary satellite networks, earth-based stations can communicate through a single “bird” only when the stations are both on a line of sight with the satellite. If two stations are nearly on opposite sides of the planet, say in Australia and Wisconsin, they must operate through two satellites to obtain a link (Fig. 32-3). In this situation, signals are relayed between the two satellites, as well as between either satellite and its respective earth-based station.

A potential problem with geostationary satellite links is the fact that the signal path is long enough so that perceptible propagation delays occur. This delay, and its observed effect, is known as *latency*. This doesn't cause problems with casual communications or Web browsing, but it slows things down when computers are linked with the intention of combining their processing power.

### Low-Earth-Orbit Satellites

The earliest communications satellites orbited only a few hundred miles above the earth. They were *low-earth-orbit* (LEO) *satellites*. Because of their low orbits, LEO satellites took only about 90 minutes to complete one revolution. Communication was spotty, because a satellite was in range of any given ground station for only a few minutes at a time. This is the main reason why geostationary satellites became predominant once rocket technology progressed to the point where the necessary altitude and orbital precision could be obtained.

**32-3** A communications link that employs two geostationary satellites.



However, geostationary satellites have certain limitations. A geostationary orbit requires constant adjustment, because a tiny change in altitude will cause the satellite to get out of sync with the earth's rotation. Geostationary satellites are expensive to launch and maintain. When communicating through them, there is always a delay because of the path length. It takes high transmitter power, and a sophisticated, precisely aimed antenna, to communicate reliably. These problems with geostationary satellites have brought about a revival of the LEO scheme. Instead of one single satellite, the new concept is to have a large fleet of them.

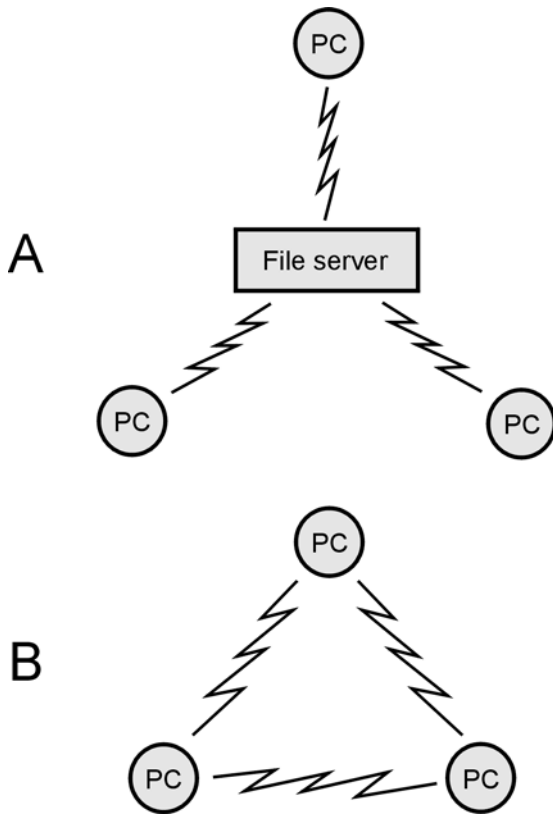
A good LEO satellite system is launched and maintained in such a way that, for any point on the earth, there is always at least one satellite in direct line-of-sight range. The satellites can relay messages throughout the fleet. Thus, any two points on the surface can always make, and maintain, contact through the satellites. The satellites are placed in *polar orbits* (routes that pass over or near the earth's geographic poles) to optimize the geographical coverage. Even if you're at or near the north geographic pole or the south geographic pole, you can use a LEO satellite system. This is not true of geostationary satellite networks, where the regions immediately around the geographic poles are not seen by the satellites.

A LEO satellite wireless communications link is easier to access and use than a geostationary satellite link. A small, simple antenna will suffice, and it doesn't have to be aimed in any particular direction. The transmitter can reach the network using only a few watts of power. The latency is less than 100 milliseconds (ms), compared with as much as 400 ms for geostationary satellite links.

## Medium-Earth-Orbit Satellites

Some satellites revolve in orbits higher than those normally considered low-earth, but at altitudes lower than the geostationary level of 22,300 mi (36,000 km). These intermediate birds are called *medium-earth-orbit* (MEO) *satellites*. A MEO satellite takes several hours to complete each orbit. MEO satellites operate in fleets, in a manner similar to the way LEO satellites are deployed. Because the average MEO altitude is higher than the average LEO altitude, each bird can cover a larger region on the surface at any given time. A fleet of MEO satellites can be smaller than a comparable fleet of LEO satellites, and still provide continuous, worldwide communications.

The orbits of geostationary satellites are essentially perfect circles, and most LEO satellites orbit in near-perfect circles. But MEO satellites often have elongated, or elliptical, orbits. The point of



**32-4** At A, a client-server wireless LAN. At B, a peer-to-peer wireless LAN.

lowest altitude is called *perigee*; the point of greatest altitude is called *apogee*. The apogee can be, and often is, much greater than the perigee. Such a satellite orbits at a speed that depends on its altitude. The lower the altitude, the faster the satellite moves. A satellite with an elliptical orbit crosses the sky rapidly when it is near perigee, and slowly when it is near apogee; it is easiest to use when its apogee is high above the horizon, because then it stays in the visible sky for a long time.

Every time a MEO satellite completes one orbit, the earth rotates beneath it. The rotation of the earth rarely coincides with the orbital period of the satellite. Therefore, successive apogees for a MEO satellite occur over different points on the earth's surface. This makes the tracking of individual satellites a complicated business, requiring computers programmed with accurate orbital data. For a MEO system to be effective in providing worldwide coverage without localized periodic blackouts, the orbits must be diverse, yet coordinated in a precise and predictable way. In addition, there must be enough satellites so that each point on the earth is always on a line of sight with one or more of the satellites, and preferably, there should be at least one bird in sight near apogee at all times.

## Wireless Networks

A *local area network* (LAN) is a group of computers linked together within a building, campus, or other small region. The interconnections in early LANs were made with wire cables, but wireless links are increasingly common. A *wireless LAN* offers flexibility, because the computer users can move around without having to bother with plugging and unplugging cables. This arrangement is

ideal when notebook computers (also known as laptops) are used. The way in which a LAN is arranged is called the *LAN topology*. There are two major wireless LAN topologies: the *client-server wireless LAN* and the *peer-to-peer wireless LAN*.

In the client-server topology (Fig. 32-4A), there is one large, powerful, central computer called a *file server*, to which all the smaller personal computers (labeled PC) are linked. The file server has enormous computing power, high speed, and large storage capacity, and can contain all the data for every user. End users do not communicate directly. All the data must pass through the file server.

In a peer-to-peer LAN (Fig. 32-4B), all of the computers in the network are PCs with more or less equal computing power, speed, and storage capacity. Each user maintains his or her own data. Subscribers can, and almost always do, communicate directly without the data having to pass through any intermediary. This offers greater privacy and individuality than the client-server topology, but it is slower when a large number of users need to share data.

Client-server LANs are favored by large institutions. Small businesses and schools, or departments within a larger corporation or university, prefer to use peer-to-peer LANs, mainly because they are cheaper and easier to maintain. In these illustrations, only three PCs are shown in the networks. But any LAN can have as few as two, or as many as several dozen, computers.

Home Internet users sometimes employ a modified version of the arrangement shown in Fig. 32-4A. In place of the file server, a device called a *wireless router* provides a hub through which the computers can communicate. The router is connected to the Internet by a high-speed interface such as a *cable modem*, allowing several computers in a household to have Internet access at the same time.

## Amateur and Shortwave Radio

In most countries of the world, people can obtain government-issued licenses to send and receive messages by radio for nonprofessional purposes. In America, this hobby is called *amateur radio* or *ham radio*. If you want only to listen to communications and broadcasting, and not to transmit signals, you do not need a license in the United States (although you do need one in some countries).

### Who Uses Amateur Radio?

Anyone can use ham radio, provided they can pass the tests necessary to obtain a license. Amateur radio operators (or hams) can communicate in numerous modes, including voice, Morse code, television, and various forms of text messaging. Text messaging can be done in real time, or by posting messages similar to *electronic mail* (e-mail). Radio amateurs have set up their own radio networks. Some of these networks have Internet gateways. This mode is known as *packet* radio.

Some radio hams chat about anything they can think of (except business matters, which are illegal to discuss via ham radio). Others like to practice emergency communications skills, so they can be of public service during crises such as hurricanes, earthquakes, or floods. Still others like to go out into the wilderness and talk to people far away while sitting out under the stars and using battery power.

Amateur radio operators communicate from cars, boats, aircraft, and bicycles; this is called *mobile operation*. When transceivers are used while walking or hiking, it is known as *portable* or *hand-held operation*.

### Amateur Equipment and Licensing

A simple ham radio station has a transceiver (transmitter/receiver), a microphone, and an antenna. A small station can fit on a desktop, and is about the size of a home computer or hi-fi stereo system.

Accessories can be added until a ham “rig” is a large installation, comparable to a small commercial broadcast station.

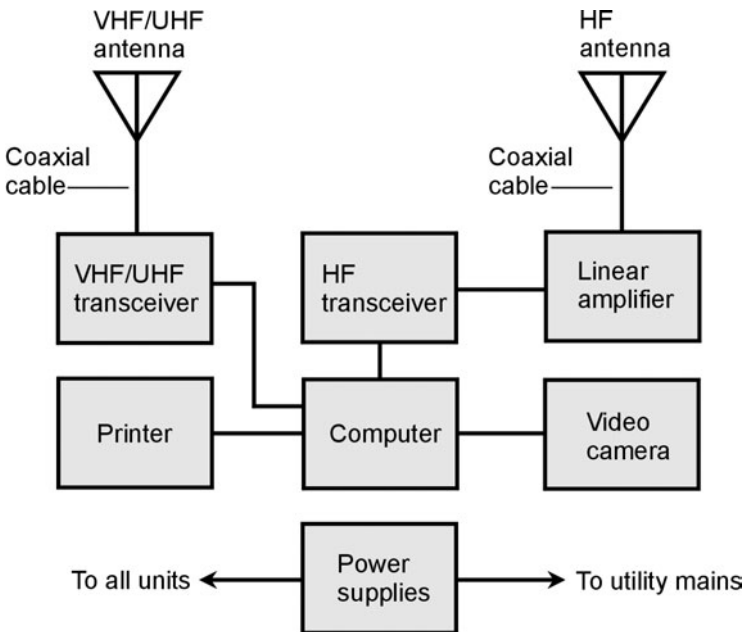
Figure 32-5 shows an example of a fixed, computer-controlled amateur radio station. The computer can be used to control the functions of the transceivers, and also to communicate using digital modes with other hams who own computers. The station can be equipped to interface with the telephone services, also called *landline*. The computer can control the antennas for the station, and can keep a log of all stations that have been contacted.

The best way to learn about ham radio in the United States is to contact the headquarters of the American Radio Relay League (ARRL), 225 Main Street, Newington, CT 06111. They maintain a Web site at [www.arrl.org](http://www.arrl.org). If you live outside the United States, the ARRL can direct you to an organization in your home country that can help you obtain a license and get on the air.

### Shortwave Listening

The high-frequency (HF) portion of the radio spectrum, at frequencies between 3 and 30 MHz, is sometimes called the *shortwave band*. This is a misnomer by contemporary standards; the waves are long compared with ultrahigh frequencies (UHF), microwaves, and IR, which are commonly used in wireless devices. In free space, a frequency of 3 MHz corresponds to a wavelength of 100 m; 30 MHz corresponds to 10 m. In the early years of radio when the shortwave band got its name, the wavelengths between 3 and 30 MHz were short compared with the wavelengths of most broadcast and communications signals, which had wavelengths in the kilometer range.

Anyone can build or obtain a shortwave or general-coverage radio receiver, install a modest outdoor antenna, and listen to signals from all around the world. This hobby is called *shortwave listening* or *SWLing*. In the United States, the proliferation of computers and online communications has, to some extent, overshadowed SWLing, and many young people grow up today ignorant of a



32-5 A computer-controlled amateur radio station.



realm of broadcasting and communications that still predominates in much of the world. But some people are still fascinated by the fact that people can contact each other using wireless devices alone, without the need for any human-made infrastructure other than an antenna at the source and another antenna at the destination. The ionosphere returns shortwave signals to the earth's surface, and allows reliable global broadcasting and communication to take place today, just as it has since the early 1900s.

Various commercially manufactured shortwave receivers are available. Most electronics stores carry one or more models, along with antenna equipment, for a complete installation. Amateur radio conventions, called "hamfests," can be a good source of shortwave receiving equipment at bargain prices. For information about events of this sort in your area, you can contact the American Radio Relay League at [www.arrl.org](http://www.arrl.org), or pay a visit to your local amateur radio club.

## Security and Privacy

People are becoming more and more concerned about the security and privacy of electronic communications, particularly wireless. When a wireless system is compromised, the intrusion is not detected until harm has been done to the system or to its subscribers.

### Wireless versus Wired

*Wireless eavesdropping* differs from conventional *wiretapping* in two fundamental ways. First, eavesdropping is easier to do in wireless systems than in hard-wired systems. Those old-fashioned hard-wired phone sets might not always be convenient, but your privacy is more likely to be maintained than in the case with a system that uses any form of wireless. Second, eavesdropping of a wireless link is nigh impossible to physically detect, but a tap can usually be found in a hard-wired system.

If any portion of a communications link is done by wireless, then an eavesdropping receiver can be positioned within range of the RF transmitting antenna (Fig. 32-6) and the signals intercepted. The existence of a *wireless tap* has no effect on the electronic characteristics of any equipment in the system.

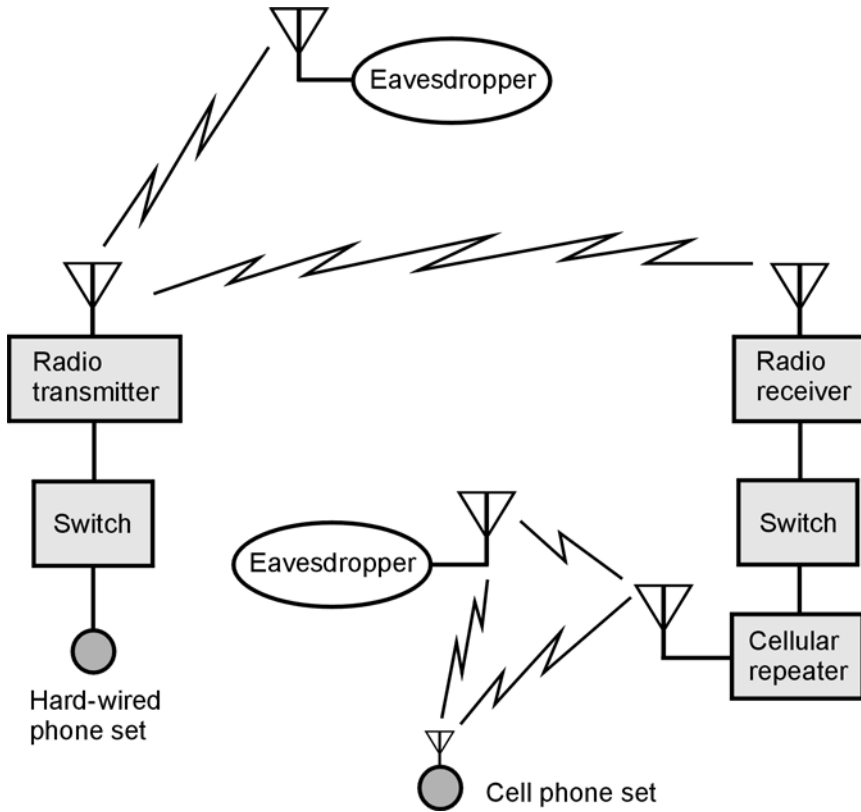
### Levels of Security

There are four levels of telecommunications security, ranging from zero (no security) to the most secure connections technology allows.

*No Security (Level 0)*: In a communications system with *level 0 security*, anyone can eavesdrop on a connection at any time, provided they are willing to spend the money and time to obtain the necessary equipment. Two examples of level 0 links are amateur radio and citizens band (CB) voice communications.

*Wire Equivalent Security (Level 1)*: An end-to-end hard-wired connection requires considerable effort to tap, and sensitive detection apparatus can usually reveal the existence of any wiretap. A communications system with *level 1 security* must have certain characteristics in order to be effective and practical:

- The cost must be affordable.
- The system must be reasonably safe for transactions such as credit-card purchases.
- When network usage is heavy, the degree of privacy afforded to each subscriber should not decrease, relative to the case when network usage is light.



**32-6** Eavesdropping on RF links in a telephone system. Heavy, straight lines represent wires or cables; zigzag lines represent RF signals.

- Ciphers, if used, should be unbreakable for at least 12 months, and preferably for 24 months or more.
- Encryption technology, if used, should be updated at least every 12 months, and preferably every six months.

*Security for Commercial Transactions (Level 2):* Some financial and business data demands protection even beyond the wire equivalent level. Many companies and individuals refuse to transfer money by electronic means because they fear criminals will gain access to an account. In a communications system with *level 2 security*, the encryption used in commercial transactions should be such that it would take a potential intruder (also called a *hacker*) at least 10 years, and preferably 20 years or more, to break the cipher. The technology should be updated at least every 10 years, but preferably every 3 to 5 years.

*Military Level Security (Level 3):* Security to *military specifications* (also called *mil spec*) involves the most sophisticated encryption available. Technologically advanced countries, and entities with economic power, have an advantage here. However, as technology gains ever more (and arguably too much) power over human activities, aggressor nations and terrorists might injure powerful nations by seeking out, and striking at, the weak points in communications infrastructures. In a communi-

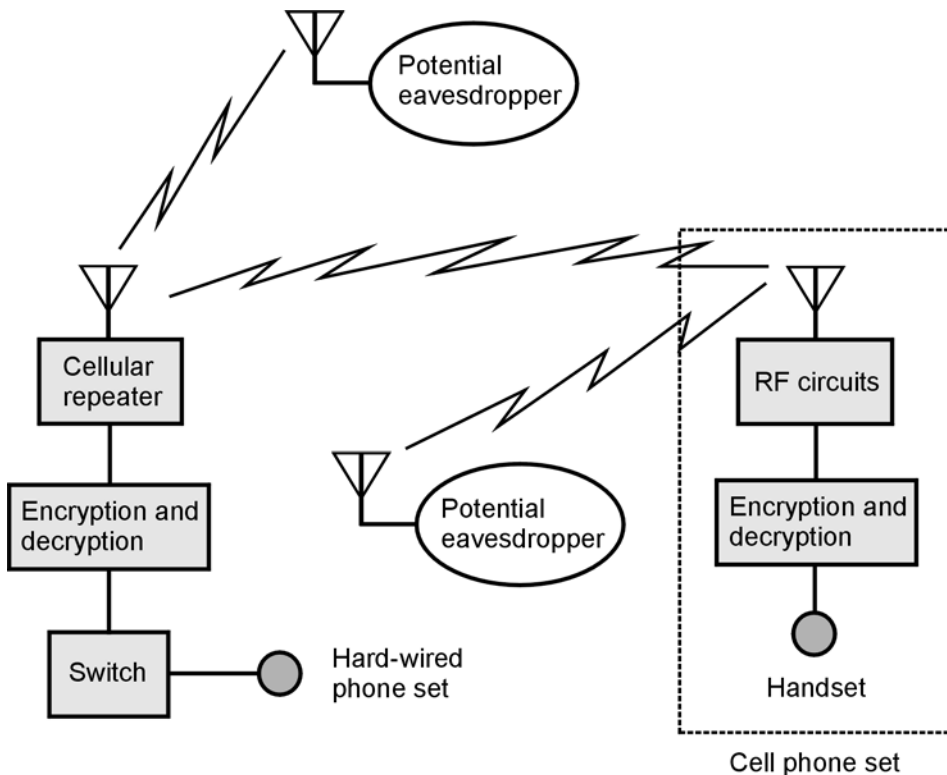
communications system with *level 3 security*, the encryption scheme should be such that engineers believe it would take a hacker at least 20 years, and preferably 40 years or more, to break the cipher. The technology should be updated as often as economics allow.

### Extent of Encryption

Security and privacy in wireless networks and communications systems can be achieved by means of *digital encryption*. The idea is to render signals readable only to receivers with the necessary *decryption key*. This makes it difficult for unauthorized people to gain access to the system.

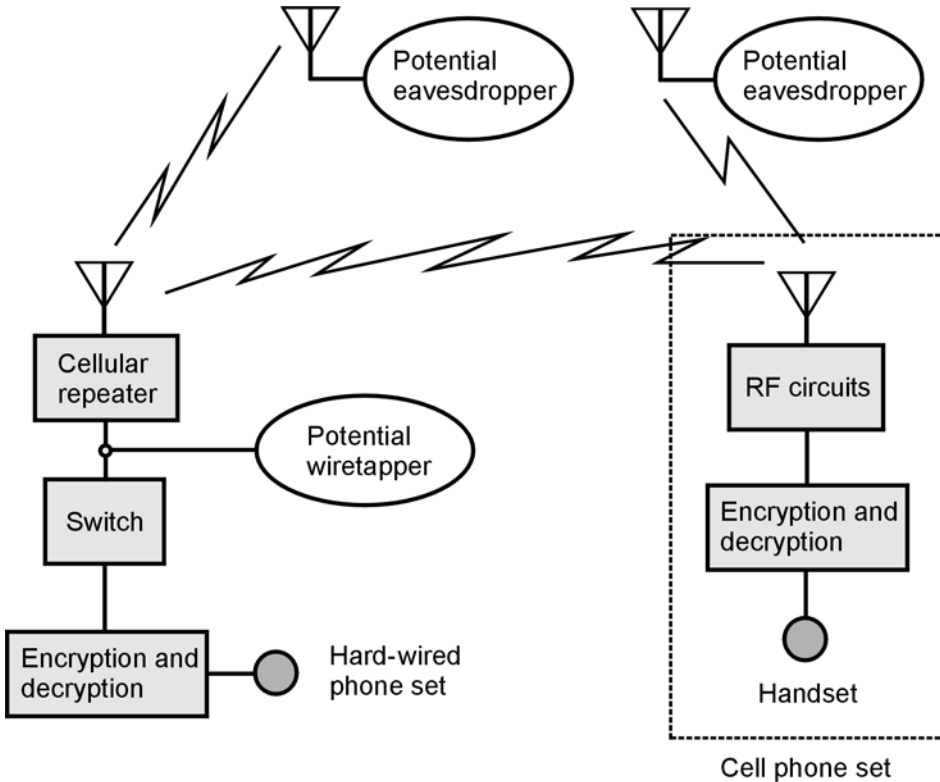
For level 1 security, encryption is required only for the wireless portion(s) of the circuit. The cipher should be changed at regular intervals to keep it fresh. The block diagram of Fig. 32-7A shows *wireless-only encryption* for a hypothetical cellular telephone connection.

For security at levels 2 and 3, *end-to-end encryption* is necessary. The signal is encrypted at all intermediate points, even those for which signals are transmitted by wire or cable. Figure 32-7B shows this scheme in place for the same hypothetical cellular connection as depicted at A.



A

32-7A Wireless-only encryption. Heavy, straight lines represent wires or cables; zigzag lines represent RF signals.



**B**

**32-7B** End-to-end encryption. Heavy, straight lines represent wires or cables; zigzag lines represent RF signals.

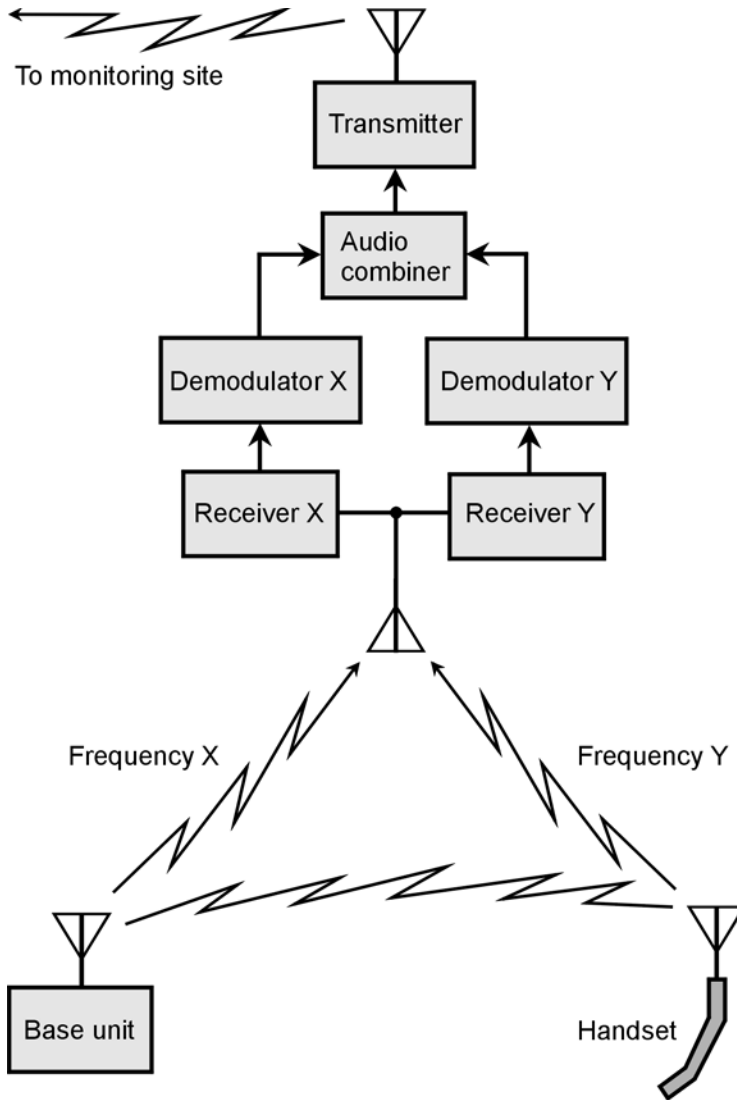
**Security with Cordless Phones**

Most cordless phones are designed to make it difficult for unauthorized people to pirate a telephone line. Prevention of eavesdropping is a lower priority, except in expensive cordless systems. If there is concern about using a cordless phone in a particular situation, a hard-wired phone set should be used.

If someone knows the frequencies at which a cordless handset and base unit operate, and if that person is determined to eavesdrop on conversations that take place using that system, it is possible to place a *wireless tap* on the line. The conversation can be intercepted at a point near the cordless phone set and its base unit, and then transmitted to a remote site (Fig. 32-8) and recorded there.

**Security with Cell Phones**

Cellular telephones are, in effect, long-range cordless phones. The wider coverage of cellular repeaters, as compared with cordless base units, increases the risk of eavesdropping and unauthorized use. Some cell phone vendors advertise their systems as “snoop proof.” Some of these claims have more merit than others. The word “proof” (meaning “immune”) should be regarded with skepticism. Digital encryption is the most effective way to maintain privacy and security of cellular communications. Nothing short of this is really of any use against a determined hacker.



32-8 Wireless tapping of a cordless telephone.

Access and privacy codes, as well as data, must be encrypted if a cell phone system is to be maximally secure. If an unauthorized person knows the code with which a cell phone set accesses the system (the “name” of the set), rogue cell phone sets can be programmed to fool the system into thinking they belong to the user of the authorized set. This is known as *cell phone cloning*.

In addition to digital encryption of data, *user identification* (user ID) must be employed. The simplest is a *personal identification number* (PIN). More sophisticated systems can employ *voice-pattern recognition*, in which the phone set functions only when the designated user’s voice speaks into it. *Hand-print recognition*, *electronic fingerprinting*, or *iris-print recognition* can also be employed. These are examples of *biometric security* measures.

## Quiz

Refer to the text in this chapter if necessary. A good score is 18 correct. Answers are in the back of the book.

1. A network that employs one central computer and multiple personal computers is called
  - (a) a wireless network.
  - (b) a local area network.
  - (c) a client-server network.
  - (d) a peer-to-peer network.
2. In order for a receiver to render a digitally encrypted signal intelligible, that receiver must
  - (a) have a directional antenna.
  - (b) have a biometric security system.
  - (c) have a decryption key.
  - (d) be licensed by the government.
3. Which of the following devices or systems is *not* wireless?
  - (a) An SWL station
  - (b) An amateur radio station
  - (c) A cordless phone set
  - (d) A computer with a telephone modem
4. Eavesdropping of a wireless link
  - (a) requires digital encryption.
  - (b) requires a decryption key.
  - (c) is difficult or impossible to detect.
  - (d) includes all of the above.
5. In the United States, a license is required for
  - (a) using a receiver on amateur radio frequencies.
  - (b) transmitting on amateur radio frequencies.
  - (c) using cordless phone sets.
  - (d) using cell phone sets.
6. In a LEO satellite system, the repeaters are located
  - (a) in space.
  - (b) at strategic locations in most countries of the world.
  - (c) at the north and south geographic poles.
  - (d) over the equator.
7. In a geostationary satellite link, the repeater is located
  - (a) on a buoy in the ocean.
  - (b) at strategic locations in most countries of the world.
  - (c) at the north and south geographic poles.
  - (d) over the equator.

8. The term *shortwave*, in reference to radio, refers to signals having wavelengths of approximately
- (a) 10 mm to 100 mm.
  - (b) 100 mm to 1 m.
  - (c) 1 m to 10 m.
  - (d) 10 m to 100 m.
9. Which of the following is illegal to do in the United States using a ham radio station?
- (a) Advertise used cars for sale
  - (b) Transmit Morse code
  - (c) Receive outside the amateur radio bands
  - (d) Receive signals without a license
10. A file server is
- (a) a sensitive radio receiver equipped with D/A conversion.
  - (b) a radio transmitter equipped with digital signal processing (DSP).
  - (c) a central computer in a local area network.
  - (d) a repeater in a LEO satellite system.
11. A device consisting of a receiver and transmitter in the same case or enclosure, and commonly used for wireless communications, is called
- (a) a file server.
  - (b) a transceiver.
  - (c) a transverter.
  - (d) an encoder/decoder.
12. End-to-end encryption provides
- (a) maximum range in a wireless LAN.
  - (b) a higher degree of security than wireless-only encryption.
  - (c) simultaneous transmission and reception in an amateur radio station.
  - (d) an easy means for hackers to clone cell phone sets.
13. A LAN in which each computer has more or less equal status, and in which each computer stores its own data, is known as a
- (a) wireless router.
  - (b) wide-area LAN.
  - (c) LAN topology.
  - (d) peer-to-peer LAN.
14. With respect to security, the term *mil spec* refers to
- (a) wireless-only encryption.
  - (b) a peer-to-peer network.
  - (c) the use of a wireless router.
  - (d) the highest obtainable level of security.

15. A radio signal is considered to be in the shortwave band if its frequency is which of the following:

- (a) 5 MHz
- (b) 50 MHz
- (c) 500 MHz
- (d) Any of the above

16. In a cellular network, the region covered by a single base station is called a

- (a) zone.
- (b) cell.
- (c) radius.
- (d) link.

17. An advantage of a conventional hard-wired telephone over a cell phone is the fact that a hard-wired phone set

- (a) affords better privacy.
- (b) offers greater portability.
- (c) is more convenient for use around the house.
- (d) employs digital encryption.

18. An advantage of a cell phone over a hard-wired phone is the fact that a cell phone set

- (a) affords better privacy.
- (b) costs less.
- (c) offers better mobility.
- (d) provides higher data speed.

19. In a two-way communications system with full duplex,

- (a) either party can hear the other at all times.
- (b) a cell phone transmits on two frequencies at the same time.
- (c) end-to-end encryption is necessary.
- (d) a file server or router is necessary.

20. An advantage of LEO satellite systems over geostationary satellite systems is the fact that

- (a) the latency is much greater with LEO satellite systems.
- (b) the equatorial regions are covered by LEO satellite systems, but not by geostationary satellite systems.
- (c) the regions near the geographic poles are covered by LEO satellite systems, but not by geostationary satellite systems.
- (d) All of the above are advantages of LEO satellite systems over geostationary satellite systems.

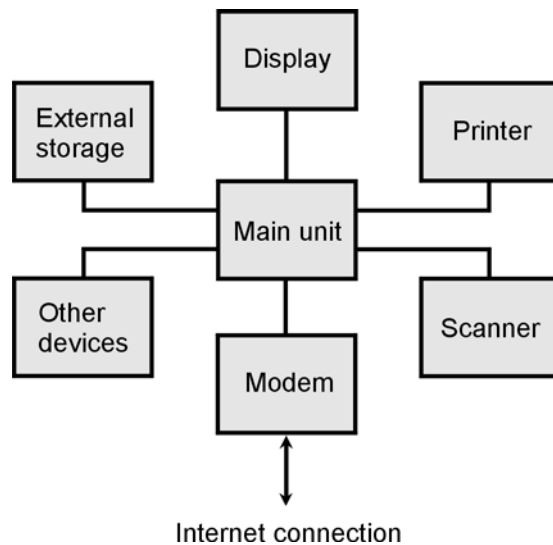


## 33 CHAPTER

# A Computer and Internet Primer

COMPUTERS ARE USED FOR CALCULATIONS, COMMUNICATIONS, WORD PROCESSING, DATA PROCESSING, drawing, photo processing, music composition and editing, location, navigation, information searches, robot control, and many other purposes. Figure 33-1 is a block diagram showing the major parts of a typical computer system used in a home or small business.

**33-1** Components of a basic computer system for personal or small business use.



## The Central Processing Unit

The *microprocessor* is the IC, or chip, that forms the core of your computer's "brain." It coordinates all the action and does all the calculations. It is located on the *motherboard*, or main circuit board. This board is sometimes called the *logic board*. The microprocessor, together with various other circuits, comprises the *central processing unit* (CPU). Auxiliary circuits can be fabricated onto the same

chip as the microprocessor, but they are often separate. The external chips contain *memory*, *storage*, and *programming* instructions.

You might think of the microprocessor as the computer's conscious mind, which directs the behavior of the machine by deliberate control. The CPU, controlled by the microprocessor, represents the PC's entire mind, conscious and subconscious. All the ancillary circuits, in conjunction with the CPU, make up the nervous system. Peripherals such as printers, disk drives, mice, speech recognition/synthesis apparatus, modems, and displays are the hands, ears, eyes, and mouth of the system. In advanced computer systems there can also be robots, vision systems, various home appliances, surveillance apparatus, medical devices, and other exotic equipment under the control of the CPU.

## Units of Digital Data

You learned about bits and bytes in Chap. 26. Let's examine them again in more detail, with emphasis on their relevance to computers. Recall that one bit (1 b) is a single binary digit, and one byte (1 B) is a unit of digital data consisting of a string of eight contiguous bits (8 b) in most systems. One byte constitutes roughly the same amount of data as one character, such as a letter, numeral, punctuation mark, or space.

### Memory and Storage Capacity

Computer memory and storage involves files that are huge in terms of bytes. Therefore, *kilobytes* (units of  $2^{10} = 1024$  bytes), *megabytes* (units of  $2^{20} = 1,048,576$  bytes), and *gigabytes* (units of  $2^{30} = 1,073,741,824$  bytes) are used. The abbreviations for these units are KB, MB, and GB, respectively. Alternatively you might see them abbreviated as K, M, and G.

As computer technology advances, you'll hear more and more about a unit of data called a *terabyte* (TB or T). This is equivalent to  $2^{40}$  bytes, or 1,048,576 MB. Someday we will commonly use the terms *petabyte* (PB or P), which refers to  $2^{50}$  bytes or 1,048,576 GB, and *exabyte* (EB or E), which refers to  $2^{60}$  bytes or 1,048,576 TB.

Here are all these data units listed as a hierarchy:

1 KB = 1024 B  
 1 MB = 1024 KB  
 1 GB = 1024 MB  
 1 TB = 1024 GB  
 1 PB = 1024 TB  
 1 EB = 1024 PB

Computer memory is usually specified in megabytes or gigabytes. The same holds true for removable data storage media. The *hard drive* in a computer generally has capacity measured in gigabytes, although a few get into the terabyte range. Some external storage media, used for data *archiving* (saving it for long-term reference), have capacity measured in terabytes and petabytes.

### Data Speed

The speed at which computers send digital data to and from each other is almost always expressed in *bits per second* (bps) and large multiples thereof. Multiples of bits per second involve the same pre-

fixes as multiples for bytes, but when talking or writing about bits per second, these prefixes refer to powers of 10 rather than powers of 2.

A *kilobit per second* (1 kbps) is equal to  $10^3 = 1000$  bps, a *megabit per second* (1 Mbps) is equal to  $10^6 = 1,000,000$  bps, and a *gigabit per second* (1 Gbps) is equal to  $10^9 = 1,000,000,000$  bps. (You won't often hear of data speed units larger than the gigabit per second.) Hierarchically:

$$\begin{aligned} 1 \text{ kbps} &= 1000 \text{ bps} \\ 1 \text{ Mbps} &= 1000 \text{ kbps} \\ 1 \text{ Gbps} &= 1000 \text{ Mbps} \end{aligned}$$

Note the lowercase “k” in reference to “kilo-” meaning  $10^3$  or 1000, and the uppercase “K” in reference to “kilo-” meaning  $2^{10}$  or 1024. Note also that for the other prefix multipliers, uppercase letters are always used. These are not typos! The “k versus K” distinction is a notational peculiarity, and is often ignored or overlooked. You will sometimes see “Kbps” rather than “kbps,” or “kB” rather than “KB,” in technical papers and other documents. As long as you know whether the author is writing about data speed (bits per second) or memory/storage (bytes), it should not be a problem. But if you want to be rigorous, this peculiarity is worth remembering.

## The Hard Drive

A hard drive, also known as a *hard disk*, is a common form of mass storage for computer data. The drive consists of several disks, called *platters*, arranged in a stack. They are made of rigid, durable material that is coated with a ferromagnetic substance similar to that used in audio or video tape. The platters are spaced a fraction of a centimeter apart. Each has two sides (top and bottom) and two *read/write heads* (one for the top and one for the bottom). The assembly is enclosed in a sealed cabinet. Figure 33-2A is an edgewise, cutaway view of the platters and heads in a typical hard drive.

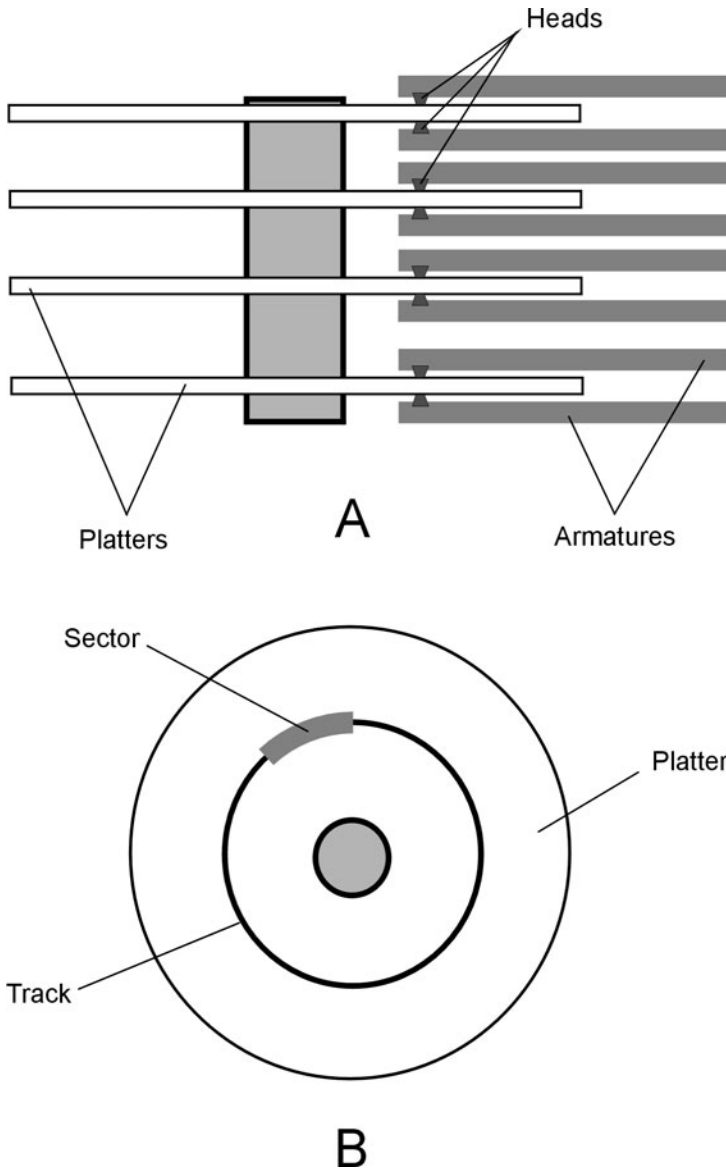
### Drive Action

When the computer is switched off, the hard-drive mechanism locks the heads in a position away from the platters. This prevents damage to the heads and platters if the computer is moved. When the computer is powered up, the platters spin at several thousand revolutions per minute (rpm). The heads hover a few millionths of a centimeter above and below the platter surfaces.

When you type a command or click on an icon telling the computer to read or write data, the hard-drive mechanism goes through a series of rapid, complex, and precise movements. The head positions itself over the particular spot on the platter where the data is located or is to be written; then the head detects the magnetic fields and translates them into tiny electric currents. All this takes place in a small fraction of a second.

### Data Arrangement and Capacity

The data on a hard drive is arranged in concentric, circular *tracks*. There are hundreds or even thousands of tracks per radial centimeter of the platter surface. Each circular track is broken into a number of arcs called *sectors*. A *cylinder* is the set of equal-radius tracks on all the platters in the drive. Tracks and sectors are set up on the hard drive during the initial *formatting* process. There are also data units called *clusters*. These are units consisting of one to several sectors, depending on the arrangement of data on the platters. Figure 33-2B is a face-on view of a single hard-disk platter, showing a track and one of its constituent sectors.



**33-2** At A, an edgewise view of a hard drive, showing four platters. At B, a top-down view of a platter, showing one track and one sector.

When you buy a computer, whether it is a desktop, notebook (also called laptop), or portable (also called handheld) unit, it will have a hard drive built in. The drive comes installed and formatted. Most new computers are sold with several commonly used programs installed on the hard drive. Some computer users prefer to buy new computers with only the operating system, by means of which the programs run, installed; this frees up hard-drive space and gives the user control over which programs to install (or not to install).

# External Storage

There are several types of *external storage* (besides the hard drive) in which data can be kept in large quantities. Computer experts categorize external storage in two ways: *access time* and *cost per megabyte* (or gigabyte, or terabyte).

## Disk Media

Disk media offer the advantage of speed, convenience, and reliability. For personal computers, there are many forms of external disks. Here are three of the most well-known.

An *external hard drive* is exactly what its name suggests. This type of device is exceptionally fast, and has storage capacity similar to the hard drives inside computers. They can be easily connected to any personal computer with a short cable. Most of these devices require a power supply, often found in the form of a “brick” that contains a transformer, rectifier, and filter that converts 120 V ac into the necessary dc for the device.

*Compact disk recordable* (CD-R) and *compact disk rewritable* (CD-RW) are popular for backing up and archiving computer data. You can buy these disks for various other applications, too, such as storage for digital photos and home videos. They are the same size, physically, as conventional CD-ROMs, which are used for commercial software, databases, and digital publications. The main asset of CD-R and CD-RW is moderately large capacity and long shelf life. Most new computers have built-in drives for these disks.

*Diskettes*, also called (imprecisely) “floppies,” are about 9 cm (3.5 in) in diameter and enclosed in a rigid, square case about 4 mm (0.15 in) thick. Their capacity, individually, is limited. They are all but obsolete. Increasingly, new computers are sold without drives for floppies.

## Tape Media

The earliest computers used magnetic tape to store data. This is still done in some systems. You can get a *tape drive* for making an emergency backup of the data on your hard drive, or for archiving data you rarely need to use. Magnetic tape has high storage capacity. There are microcassettes that can hold more than 1 GB of data; standard cassettes can hold many gigabytes. But tapes are extremely slow because, unlike their disk-shaped counterparts, they are a *serial-access* storage medium. This means that the data bits are written in a string, one after another, along the entire length of the tape. The drive might have to mechanically rewind or fast-forward through a football field’s length of tape to get to a particular data bit, whereas on a disk medium, the read/write head never has to travel farther than the diameter of the disk to reach a given data bit.

## Flash Memory

*Flash memory* is an all-electronic form of storage that is useful especially in high-level graphics, big-business applications, and scientific work. The capacity is comparable to that of a small hard drive, but there are no moving parts. Because there are no mechanical components, flash memory is faster than any other mass-storage scheme, provided it does not cause a *software conflict* with other programs in the computer. (Software conflicts can cause a computer to slow down or “freeze up.”)

Flash memory is available in small modules roughly the size of your index finger, and can be plugged directly into one of the *Universal Serial Bus* (USB) ports provided in all new computers. Some flash memory modules come in the form of *PC cards* (also called *PCMCIA cards*), which are credit-card-sized, removable components.

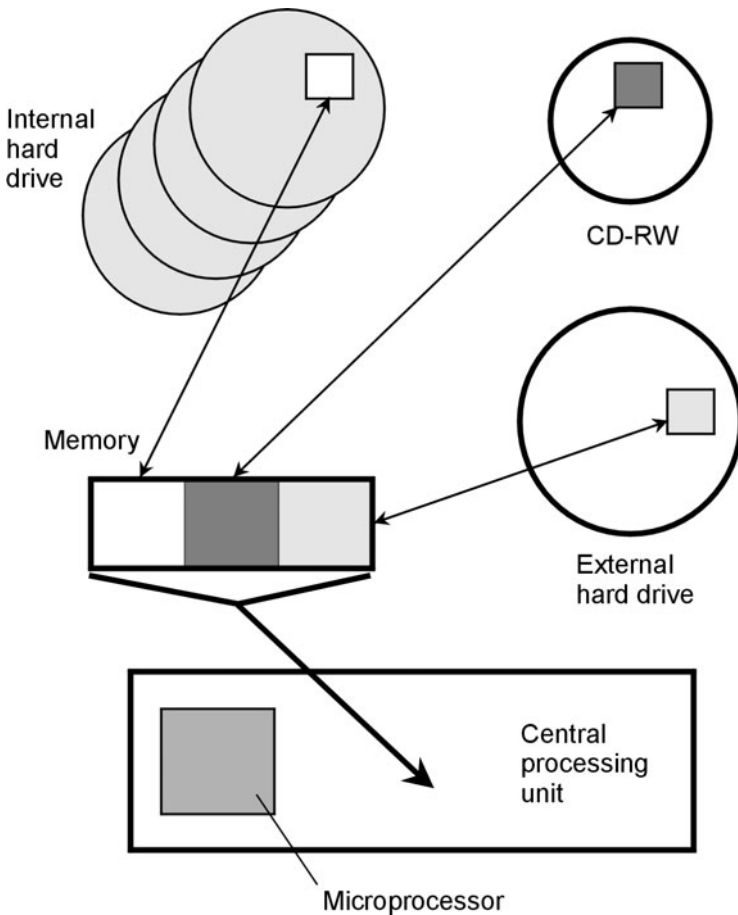
# Memory

In a computer, the term *random access memory* (RAM), often called simply *memory*, refers to ICs that store working data. The amount, and speed, of memory are crucial factors in determining what a computer can and cannot do.

## Data Flow

Figure 33-3 shows how data moves in a computer among an internal hard drive, an external hard drive, and a CD-RW storage medium. This process is controlled by the CPU. When you open a file on any medium, the data goes immediately into the memory. The CPU, under direction of the microprocessor, manipulates the data in the memory as you work on the file. Thus, the data in memory changes from moment to moment.

When you hit a key to add a character, or drag the mouse to draw a line that shows up on your display, that character or line goes into memory at the same time. If you hit the backspace key to delete a character, or drag the mouse to erase a line on the screen, it disappears from the memory.



**33-3** The flow of data between storage media and memory is controlled by the CPU.

During this time the original file on the storage medium stays as it was before you accessed it. No change is made to that data until you specifically instruct the computer to overwrite it. When you're done working on a file, you tell the microprocessor to close it. Then the data leaves memory and goes back to the medium from which it came, or to some other place, as you direct. If you tell the computer to overwrite the file on the medium from which it came, many programs send the new data (containing the changes you have made) to unused space there; the old data (as it was before you opened the file) stays in its old location. This is a safeguard, in case you decide to undo the changes you made.

All the data passing between the storage media and memory, and between the memory and the CPU, is in *machine language*. This consists of binary digits (bits) 0 and 1. But the data passing between you and the CPU is in plain English (or whatever other language you prefer), or in some high-level programming language, having been translated by the machine into a form you can understand.

## Memory Capacity

The maximum number of megabytes or gigabytes of data that can be stored in a computer's memory is known as the *memory capacity*. The main factor that determines memory capacity is the number of transistors that can be fabricated onto a single memory chip. Other factors, such as microprocessor speed, have a practical effect on the usable memory capacity.

A gigantic memory is of little practical value if the microprocessor is slow. Nor is a fast microprocessor worth much if the memory capacity is too small for the applications (programs) you want to run. Most software packages tell you how much memory you need to run the programs they contain. They'll often quote two specifications: a minimum memory requirement and a figure for optimum performance (approximately twice the minimum requirement). If possible, you should equip your computer with enough memory for optimum performance.

When buying a new computer, keep in mind that this year's high-end machine may become next year's ordinary one when it comes to popular software, and in a few years it will no longer be adequate to run many of the programs that will be available. If your computer lacks the memory to run a given application, you can usually add more. But this can only be done up to a certain point. Eventually, your microprocessor will no longer be able to run contemporary software at reasonable speed, no matter how much memory you have.

How much memory do you think your machine will require to run two or three of your favorite applications at the same time? Double or triple it, and you will come close to the amount you are likely to need for the next two or three years, or until you are overcome by the urge to buy a new computer again.

## The Display

The visual interface between you and your computer is known as the *display*. In desktop computers, an external display is often called a *monitor*. There are two popular types in use for personal computing today.

A *cathode-ray tube* (CRT) *monitor* resembles a television set without the tuning or volume controls. This type of display is large and heavy, and although some people consider CRTs obsolete, other computer users prefer them, especially the larger ones, which can have a diagonal screen measure of 53 cm (21 in) or more.

A *liquid crystal display* (LCD) is lightweight and thin. This type of display is used in notebook and portable computers. It has become increasingly popular for desktop computers because the

technology has improved and has become somewhat less expensive than it used to be. Light weight and a shallow desk profile are decided advantages of this type of display. A good one can provide image crispness and color at least as good as the best CRT display. An LCD also consumes less power.

Other types of displays exist. Of special note is the *plasma display*, which obtains its images from the properties of electric charges passing through rarefied gases. These displays are often found in large department stores, where you can see them hanging from the ceilings, displaying advertisements.

## Resolution

The *image resolution*, often called simply the *resolution*, of a computer display is one of its most important specifications. This is the extent to which it can show detail. The better the resolution, the sharper the image.

Resolution can be specified in terms of *dot size* or *dot pitch*. This is the diameter, in millimeters (mm), of the individual elements in the display—the “smallest unbreakable pieces.” A good display has a dot pitch that is a small fraction of a millimeter. A typical CRT has a dot pitch of 0.25 or 0.26 mm. The smaller the number, the higher the resolution and, all other factors being equal, the crisper the image in an absolute sense.

Image resolution can be specified in a general sense as a pair of numbers, representing the number of *pixels* (picture elements) the screen shows horizontally and vertically. For a particular screen size, the greater the number of pixels the unit can display, the crisper the image. In personal computers, typical displays have 800 × 600 resolution (800 pixels wide by 600 pixels high) or 1024 × 768 resolution. A few can work up to 1600 × 1200 or even higher.

## Screen Size

Screen sizes are given in terms of diagonal measure; a popular size is 43 cm (17 in). This will work quite well at 800 × 600 resolution. For higher resolution, a larger screen is preferable, such as 48 cm (19 in) or 53 cm (21 in). A high-end display is crucial for doing graphics work, when doing serious research on the Internet, in remote-control robotics, and in computer *gaming*. Besides these practical advantages, a sharp display is more pleasant to work with than a marginal one.

Along with memory and hard-drive capacity, the display is one of the most important parts of a computer from a user-friendliness standpoint. On the job, long hours at a computer can get tedious even if the machine is perfect. An inadequate display can give rise to eye strain and headaches, and can also degrade the quality and accuracy of work done by people using the computer.

## Interlacing and Refresh Rate

Another important consideration in the choice of a display is whether or not it uses *interlacing*. Interlacing increases the obtainable resolution, but it also results in a lower *refresh rate* (number of times the entire image is renewed). A low refresh rate can cause noticeable flickering in the image, and can be especially disruptive in applications where rapid motion must be displayed, such as high-end computer gaming.

A good refresh rate specification is 70 Hz or more. For applications not involving much motion, 60 or 66 Hz is adequate for some people, but others complain of eye fatigue because they can vaguely sense the flicker. Most people find 56-Hz refresh rates too slow for comfort in any application.



## ELF Fields

*Extremely low frequency* (ELF) fields are electromagnetic (EM) fields that fluctuate or alternate more slowly than conventional radio waves. Such fields are produced by various consumer electronic devices and appliances. An ELF field is nothing like X rays or gamma rays, which cause radiation sickness. Nor is ELF radiation like ultraviolet (UV), which can cause skin cancer over long periods. An ELF field cannot make anything radioactive. In computer systems, the ELF that you've heard about is emitted mainly by electromagnetic CRT monitors. Other parts of a computer system are not responsible for much ELF energy. The LCDs used in many systems today produce essentially none.

As you learned in Chap. 29, the characters and images in a CRT are created as electron beams strike a phosphor coating on the inside of the glass. The electrons constantly change direction as they sweep from left to right, and from top to bottom, on your screen. The sweeping is caused by deflecting coils that steer the beam across the screen. The coils generate magnetic fields that interact with the negatively charged electrons, forcing them to change direction. Because of the positions of the coils, and the shapes of the fields surrounding them, there is more magnetic energy radiated from the sides of an electromagnetic CRT than from the front. If there's any health hazard with ELF, therefore, it is greater for someone sitting off to the side of an electromagnetic CRT monitor, and less for someone watching the screen from directly in front at the same distance.

If you're concerned about the ELF fields produced by CRT monitors, you might consider buying an electrostatic CRT unit that has been designed to minimize ELF fields, or buying a stand-alone LCD display panel. It's a good idea to arrange your workstation so your eyes are at least 0.5 m (about 18 in) away from the screen of a CRT monitor, and if you are working near other computer users, workstations should be at least 1 m (3 ft) apart.

## The Printer

There are several types of *printer* in common use in personal and business computing applications today. The two most common are the *inkjet printer* and the *laser printer*. Less often used are the *thermal printer* and the *dot matrix printer*.

### Inkjet Printers

In an inkjet printer, tiny nozzles spray ink onto the paper. Most of these printers are comparatively slow to produce an image, and the ink needs time to dry even after the image comes out, but the quality can be excellent. Inkjet printers are available in single-color and multicolor designs. The best color machines produce images of photo quality. In fact, some are designed especially to print digital photos.

Inkjet printers require periodic replacement of the ink cartridges. These can be quite expensive, and for this reason, inkjet printers are not well suited for high-volume printing. Inkjet printers require paper with low fiber content. This keeps the ink from being carried along by capillary action before it dries, muddling or blurring the printout. Look specifically for "inkjet paper" in computer supply stores and department stores.

### Laser Printers

A laser printer works like a photocopier machine. The main difference is that, while a photocopier creates a copy of a real image (the paper original), a laser printer makes a copy of a digital computer image.

When data arrives at the printer from the computer, the encoded image is stored in the printer memory. The memory then sends it along to the laser and other devices. The laser blinks rapidly while it scans a cylindrical *drum*. The drum has special properties that cause it to attract the printing chemical, called *toner*, in some places but not others, creating an image pattern that will ultimately appear on the paper. A sheet of paper is pulled past the drum and also past an *electrostatic charger*. Toner from the drum is attracted to the paper. The image thus goes onto the paper, although it has not yet been permanently fused, or bonded, to the paper. The *fuser*, a hot pair of roller/squeezers, does this job, completing the printing process.

The main asset of laser printers is their ability to produce a large number of copies at high speed. They generally have excellent print and graphics quality for grayscale. Color laser printers are available as well, but they can be quite expensive.

The image resolution of a laser printer ranges from about 300 dots per inch (dpi) for older units to 1200 dpi, 2400 dpi, or even higher resolutions in state-of-the-art machines. As far as the untrained eye can tell, 600 dpi is as good as a photograph. Laser printers can handle graphics and text equally well. If an image can be rendered on a photocopy machine, it can be rendered just as well on a laser printer.

## **Thermal Printers**

A thermal printer uses temperature-sensitive dye and/or paper to create hard-copy text and images. Some thermal printers produce only grayscale images, while others can render full color. Thermal printers are often preferred by traveling executives who use portable computers, because these printers are physically small and light.

A simple grayscale thermal printer employs special paper that darkens when it gets hot. A color thermal printer uses thick, heat-sensitive dyes of the primary pigments: magenta (pinkish red), yellow, and cyan (bluish green). Sometimes black dye is also used, although it can be obtained by combining large, equal amounts of the primary pigments. The print head uses heat to liquefy the dye, so it bleeds onto the paper. This is done for each pigment separately.

Some, if not most, thermal printouts fade after awhile. Have you ever pulled out an old store receipt and found that it was washed-out or blank? Thermal printers can be convenient in a pinch, but you should be aware that some of them have this problem. If you're keeping a receipt for tax purposes or for proof-of-purchase and it has been printed on thermal paper, make a photocopy or a digital scan of the receipt right away. You can recognize the output of a thermal printer because the paper curls up when it's fresh out of the machine.

## **Dot Matrix Printers**

The dot matrix printer is the horse and buggy of the printing family. This type of printer is the least expensive, in terms of both the purchase price and the long-term operating cost. Dot matrix printers produce fair text quality for most manuscripts, reports, term papers, and theses. The mechanical parts are rugged, and maintenance requirements are minimal. Dot matrix printers can render some simple graphic images, but the quality is fair at best, and it can take a long time to print a single image. Dot matrix printers cannot reproduce detailed artwork or photographs with acceptable quality.

## **The Scanner**

A *scanner* is an electromechanical device that converts hard-copy text and graphics into digital form for processing and storage in a computer.

## Basic Features

Scanners can save untold hours of tedious manual retyping. Suppose you wrote a book a long time ago, and have lost the digital files (or used an old-fashioned typewriter, and never had any digital files!). The hard copy of that book sits in your basement, awaiting the massive editing that can turn it into a great novel. It needs the power of your computer's word processor. But you can't deal with the prospect of retyping its 1000 pages. A scanner, equipped with *optical character recognition* (OCR), can do away with most of the hard labor involved in getting a hard-copy manuscript onto disk.

A good scanner can be had for a couple of hundred dollars. But the value of optical scanning is hard to measure. For many entrepreneurs, it can make the difference between staying afloat and going bankrupt. It can do the work of one or two full-time typists for a tiny fraction of the long-term cost. Big companies can save money, too. Lawyers and doctors find scanners invaluable for backing up files of all kinds. Aside from storing text, scanners can make digital copies of vital papers, records, and receipts, which can be easily backed up on CD-R or CD-RW media.

A typical scanner can render color images, text, photographs, and everything else needed to make a complete, accurate digital record of any document. Color scanners use three different light beams (red, blue, and green) to get three different images, which are processed and combined in much the same way as a color television camera works. The image resolution of a scanner is measured in dots per inch (dpi), just as is done with printers. The higher the dpi specification, the more detail the scanner can see.

For reliable scanning of text and most images, a resolution of at least 300 dpi is recommended. Virtually all scanners meet this requirement. For images, greater detail translates into more memory consumed. Color increases the amount of memory or storage that an image takes up, if the image resolution remains constant.

## Configurations

Scanners come in three basic configurations. The scanner that's best for you will depend on what you want to do with it, and on how much money you're willing to spend for it.

The cheapest type of scanner is a *handheld scanner*. It looks something like a miniature vacuum-cleaner head, or one of the bar-code readers in retail stores. You roll the unit over the paper containing the text and/or graphics you want to scan. Because the unit is not as wide as most pages, you'll have to make two or three passes over the page. Handheld scanners are preferred by people who scan small images, such as snapshots. They are light in weight, and need almost no desk space. One potential problem is that you might try to scan too fast. Some handheld scanners have speed indicators that tell you if you're going too fast. Another potential difficulty is not getting a straight-line scan. Most handheld units have built-in guides (like miniature rolling pins) that minimize this problem.

If you want to scan a book or magazine, a *flatbed scanner* is much easier to use than a handheld scanner. The unit looks something like a photocopier. Using a flatbed scanner is similar to working a small photocopy machine. You lay the page, photo, or sheet down on a clear glass, and the scanning head moves past it, picking up the image. Flatbed scanners consume desk space, which, if you have a couple of printers and a fax machine, might already be at a premium.

A *sheet scanner*, also called a *feedthrough scanner*, resembles a fax machine (and in fact, many of these units can do double duty as fax machines). As its name implies, this type of scanner pulls sheets of paper through, one by one. You can stack several pages, one on top of the other, and the machine will automatically feed and scan them. However, you can't scan bound books or magazines as you can with a flatbed scanner—unless you're willing to rip out individual pages.

## Precautions

Even the best scanners make some mistakes when used with OCR. This is especially true if text contains nonstandard symbols. Highly technical material presents the worst problems. Some mathematical symbols are so esoteric that the average person (let alone a machine) is befuddled by them. Ink spots, stray markings, and smudges on a page can cause scanning errors, in much the same way as background noise confuses a speech recognition system.

A human reader can often tell what a printed letter should be, even when it is severely mutilated. But computers lack human intuition. To some extent this can be corrected by a built-in *spell checker*. Some OCR programs have spell checking, but this introduces its own set of problems because it, too, is imperfect.

If a scanner doesn't recognize a character, it will usually print a "tag"—a blank space, underline, or default symbol such as @ or #. Scanned text must always be carefully proofread, and corrections made with word-processing software, after the data has been stored on the hard drive.

## The Modem

The term *modem* is a contraction of *modulator/demodulator*. A modem interfaces a computer to a telephone line, *digital subscriber line* (DSL), cable system, fiber-optic network, wireless network, or radio transceiver, allowing you to communicate with other computer users and to "surf the Internet."

### Data Speed

Modems work at various speeds, usually measured in kilobits per second (kbps) or megabits per second (Mbps). Sometimes you'll hear about speed units called the *baud* and *kilobaud*. (A kilobaud is 1000 baud.) Baud and bits per second are almost the same units, but they are not identical. People sometimes use the term *baud* when they really mean *bps*.

Modem speeds, particularly in cable, fiber-optic, and wireless networks, keep increasing as computer communications technology advances. Modems are rated according to the highest data speed they can handle. A typical *telephone modem* works at about 56 kbps. Modems for more advanced connections operate much faster.

### Slow Modems

A computer works with binary *digital* signals, which are rapidly fluctuating direct currents. For digital data to be conveyed over a telephone or radio circuit, the data must be converted to *analog* form. In a telephone modem or radio-transceiver modem, this is done by changing the digit 1 into an audio tone, and the digit 0 into another tone with a different pitch. The result is an extremely fast back-and-forth alternation between the two tones.

In *modulation*, digital data from the computer is changed into analog data for transmission over the telephone line or radio medium. The modulator is therefore a digital-to-analog converter (D/A converter or DAC). *Demodulation* changes the analog signals from the telephone line or radio medium back to digital signals that a computer can understand. The demodulator is thus an analog-to-digital converter (A/D converter or ADC). The highest practical speed for this type of modem is approximately 56,000 bits per second (bps), or 56 kilobits per second (kbps).

Amateur radio operators use a variety of digital modes at considerably slower speeds than 56 kbps for Internet-like communications. The principal advantage of amateur radio lies in the fact that it can work when all else fails. Such communications are not fast, but no infrastructure is

needed. This makes it possible for amateur radio operators to provide emergency communications into and out of areas stricken by natural disasters that destroy the conventional communications infrastructure.

### Internal versus External Modems

An *internal modem* is a printed-circuit board, also called a *card*. Virtually all computers, both desktop and notebook, are sold with internal modems that can interface with the telephone line at speeds of up to 56 kbps.

An *external modem* is a self-contained unit. It has a cord that runs to either the computer's *serial data port* (also called the *communications port*) or one of the USB ports, and another cord that runs to the telephone line, cable system, or satellite dish. If you want to use a DSL, cable, or satellite Internet connection, you will usually have to buy or rent an external modem from the service provider.

## The Internet

The *Internet* is a worldwide system, or *network*, of computers. It got started in the late 1960s, originally conceived as a network that could survive nuclear war. Back then it was called *ARPAnet*, named after the Advanced Research Project Agency (ARPA) of the United States federal government.

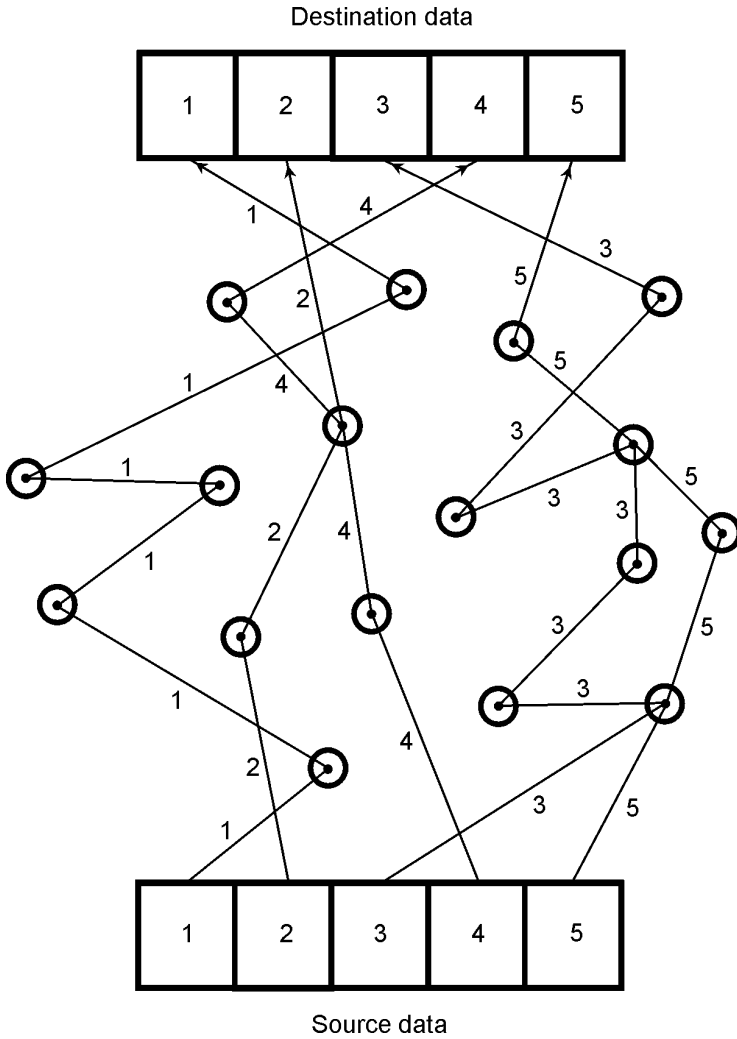
### Protocol and Packets

When people began to connect their computers into ARPAnet, the need became clear for a universal set of standards, called a *protocol*, to ensure that all the machines “speak the same language.” The modern Internet is such that you can use any computer to take advantage of all the network's resources.

Internet activity consists of computers “talking” to one another. This occurs in machine language. However, the situation is vastly more complicated than when data goes from one place to another within a single computer. In the Internet (often called simply *the Net*), data must often go through several different computers to get from the transmitting or *source* computer to the receiving or *destination* computer. These intermediate computers are called *nodes*, *servers*, *hosts*, or *Internet service providers* (ISPs).

Millions of people are simultaneously using the Net. The most efficient route between a given source and destination can change from moment to moment. The Net is set up in such a way that signals always try to follow the most efficient route. If you are connected to a distant computer, say a machine at the National Hurricane Center, the requests you make of it, and the data it sends back to you, are broken into small units called *packets*. Each packet coming to you has, in effect, your computer's name written on it. But not all packets necessarily travel the same route through the network. Ultimately, all the packets are reassembled into the data you want (such as the infrared satellite image of a hurricane), even though they do not arrive in the same order they were sent.

Figure 33-4 is a simplified drawing of Internet data transfer for a hypothetical file containing five packets transferred during a period of *heavy Net traffic* (meaning that a lot of people are using the Internet at that particular time, so it is operating at near capacity). Nodes are shown as black dots surrounded by circles. In this example, some packets pass through more nodes, and/or over a much greater physical distance, than others. If Net traffic were light, all the packets might follow the same route, or pass through fewer nodes. This is why it takes longer to acquire data on the Net during peak hours of use, as compared with times when there are comparatively few people connected



**33-4** Internet data flows in packets from the source to the destination.

into it, no matter how fast your connection happens to be. A file cannot be completely reconstructed until all the packets have arrived and the destination computer has ensured that there are no errors.

### E-mail and Newsgroups

For many computer users, communication by means of *electronic mail (e-mail)* and/or *newsgroups* has practically replaced the postal service. You can leave messages for, and receive them from, friends and relatives scattered throughout the world.

To effectively use e-mail or newsgroups, everyone must have an *Internet address*. These tend to be arcane. An example is `sciencewriter@tahoe.com`. The first part of the address, before the @ symbol, is the *username*. The word after the @ sign and before the period (or dot) represents the *domain*

*name*. The three-letter abbreviation after the dot is the *domain type*. In this case, “com” stands for “commercial.” Tahoe.com is thus a commercial provider. Other common domain types include “net” (network), “org” (organization), “edu” (educational institution), and “gov” (government). In recent years, country abbreviations have been increasingly used at the ends of Internet addresses, such as “us” for United States, “de” for Germany, “uk” for United Kingdom, and “jp” for Japan. Other abbreviations are constantly coming into common usage, such as “info” for “informational site” and “biz” for “business.”

## Internet Conversations

You can carry on a “teletype-style” (real-time text) conversation with other computer users using the Internet. When done among users within a single service provider, this is called *chat*. When done among people connected to different service providers, it is called *Internet relay chat* (IRC). Typing messages to, and reading them from, other people in real time is more personal than letter writing, because your addresses get their messages immediately. But it’s less personal than talking on the telephone, because you cannot hear, or make, vocal inflections.

It is possible to digitize voice signals and transfer them via the Internet. This has given rise to hardware and software schemes that claim to provide virtually toll-free long-distance telephone communications. As of this writing, this is similar to cellular telephone communications in terms of reliability and quality of connection. When Net traffic is light, such connections can be good. But when Net traffic is heavy, the quality is marginal. Audio signals, like any other form of Internet data, are broken into packets. All, or nearly all, the packets must be received and reassembled before a good signal can be heard. This takes variable time, depending on the route each packet takes through the Net. If many of the packets arrive disproportionately late, the destination computer can only do its best to reassemble the signal. In the worst case, the signal does not get through at all.

## Getting Information

One of the most important features of the Internet is the fact that it can connect you with millions upon millions of sources of information (and misinformation). Data is transferred among computers by means of a protocol that allows the files on the hard drives of distant computers to become available exactly as if the data were stored on your own computer’s hard drive, except the access time is slower. You can also store files on distant computers’ hard drives. When using the Internet for obtaining information, you should be aware of the time at the remote location, and avoid, if possible, accessing files during the peak hours at the remote computer. Peak hours usually correspond to working hours, or approximately 8:00 a.m. to 5:00 p.m. local time, Monday through Friday. You must take time differences into account if you’re not in the same time zone as the remote computer.

The *World Wide Web* (also called *WWW* or *the Web*) is one of the most powerful information servers you will find on the Internet, and in common usage, the terms *Internet* and *Web* have become almost synonymous. The outstanding feature of the Web is *hypertext*, a user-friendly scheme for cross-referencing of documents. In fact, the names of Web sites generally begin with the four letters “http,” which stands for *hypertext transfer protocol*. Certain words, phrases, and images make up *links*. When you select a link in a *Web page* or *Web site* (a document containing text, graphics, and often other types of files), your computer is transferred to another site dealing with the same or a related subject. This site will usually also contain numerous links. Before long, you might find yourself surfing the Web for hours, going from site to site. The word *surfing* derives from the similarity of this activity to television “channel surfing.”

## Quiz

Refer to the text in this chapter if necessary. A good score is 18 correct. Answers are in the back of the book.

1. One megabyte is the same amount of data as
  - (a) 1024 bytes.
  - (b) 1024 KB.
  - (c) 1024 GB.
  - (d) 1/1024 KB.
2. The Web should be expected to work fastest for a user in New York City at
  - (a) 2:00 a.m. local time on a Tuesday.
  - (b) 4:00 p.m. local time on a Wednesday.
  - (c) 12:30 p.m. local time on a Thursday.
  - (d) any time; it doesn't matter.
3. The sharpness of the image on a computer display can be specified in terms of
  - (a) refresh rate.
  - (b) interlace rate.
  - (c) wavelength.
  - (d) dot pitch.
4. The term *cylinder* refers to
  - (a) a set of tracks in a hard drive.
  - (b) a particular type of memory chip.
  - (c) a drum-shaped data storage medium.
  - (d) the spindle that turns a CD-R or CD-RW.
5. An example of a mass-storage device is a
  - (a) hard drive.
  - (b) microprocessor.
  - (c) modem.
  - (d) read-write head.
6. The character string *stangibilisco@rushmore.com* would most likely represent
  - (a) a Web site.
  - (b) the location of data in a computer's memory.
  - (c) an e-mail address.
  - (d) a computer's serial number.
7. The megabit per second (Mbps) is a common unit that expresses
  - (a) the memory capacity of a chip.
  - (b) the storage capacity of an external hard drive.



- (c) the image resolution of a printer.
  - (d) the speed of an Internet connection.
8. A platter is
- (a) part of a hard drive.
  - (b) a unit of memory.
  - (c) an element of a digital image.
  - (d) a semiconductor chip.
9. Protocol ensures that
- (a) a hard drive runs smoothly.
  - (b) a display or printer reproduces color accurately.
  - (c) a display or printer generates a clear image.
  - (d) computers can exchange data.
10. A packet is
- (a) a computer memory module.
  - (b) a unit of  $2^{10}$  bytes.
  - (c) a piece of data sent over the Net.
  - (d) a picture element in a display.
11. The main microprocessor in a computer is located
- (a) on the motherboard.
  - (b) in the external hard drive.
  - (c) in the memory chip.
  - (d) in the power supply.
12. Cross-referencing among Web pages is done with
- (a) digital signal processing.
  - (b) an analog-to-digital converter.
  - (c) Internet relay chat.
  - (d) hypertext links.
13. Web page addresses usually begin with the letters “http,” which signifies a form of
- (a) digital signal processing.
  - (b) protocol.
  - (c) modem configuration.
  - (d) color rendition.
14. A telephone modem contains
- (a) an internal CR-R or CD-RW drive.
  - (b) a microprocessor.
  - (c) an A/D converter.
  - (d) an image resolver.

15. An advantage of a flatbed scanner over a feedthrough scanner is the fact that
  - (a) a flatbed scanner can be combined with a fax machine, but a feedthrough scanner cannot.
  - (b) a flatbed scanner can be used to scan magazines or books intact, but a feedthrough scanner cannot.
  - (c) a flatbed scanner can reproduce color drawings and photographs, but a feedthrough scanner cannot.
  - (d) a flatbed scanner takes up less desktop space than a feedthrough scanner.
16. Which of the following types of external storage provide the fastest access time, provided no software conflicts occur?
  - (a) A magnetic tape drive
  - (b) A CD-R or CD-RW drive
  - (c) A flash memory module
  - (d) An external hard drive
17. Which of the following is a serial-access medium?
  - (a) Magnetic tape
  - (b) A diskette
  - (c) A hard drive
  - (d) A CD-R or CD-RW
18. Which of the following character strings represents the proper format for an e-mail address?
  - (a) <http://www.sciencewriter.net>
  - (b) [www.mcgraw-hill.com](http://www.mcgraw-hill.com)
  - (c) [blackhills.com](http://blackhills.com)
  - (d) None of the above
19. Which of the following devices is best suited for animated graphics work involving fast motion, such as high-end gaming?
  - (a) A laser printer
  - (b) An external hard drive or flash memory module
  - (c) Hypertext transfer protocol
  - (d) A display that does not use interlacing
20. A thermal printer might be an ideal choice for
  - (a) a salesperson who is on the road.
  - (b) someone working with animated graphics.
  - (c) an author who needs to print a huge text document.
  - (d) a photographer who needs top-quality color printouts.

# Monitoring, Robotics, and Artificial Intelligence

ELECTRONIC AND COMPUTER SYSTEMS FIND APPLICATIONS IN MANY SCIENTIFIC AND TECHNOLOGICAL fields. Increasingly, devices and systems formerly known only to geeks and techies are becoming commonplace in the consumer and hobby market.

## Keeping Watch

Most *monitoring systems* consist of radio transmitters and receivers. Some systems employ lasers at IR or visible wavelengths. A few employ wire or cable links. Some have mechanical hardware, such as robots, that are controlled by the signals reaching the receiver. Here are a few specific examples of monitoring systems.

### Baby Monitor

A short-range AM or FM radio transmitter and receiver can be used to listen at a distance to the sounds in an infant's room. The transmitter contains a sensitive microphone, a whip antenna, and a power supply. The receiver is battery-powered and portable. It has a short antenna, similar to the antennas on cordless telephone sets. The receiver can pick up signals from the transmitter at distances of up to about 50 m (165 ft). The signals pass easily through the walls, ceilings, and floors in frame houses.

A so-called baby monitor is subject to interference from other units that might be operating nearby on the same channel. Some baby monitors have multiple, selectable channels to help combat this problem. If interference occurs, the channel can be changed. Communications privacy and security are not a concern.

### Smoke Detector

Smoke and fire change the characteristics of the atmosphere. Smoke consists of solid particles, and fire burns away oxygen and produces other gases such as carbon dioxide, carbon monoxide, and sulfur dioxide. These changes can be sensed, and alarms set off if the changes exceed certain limits.

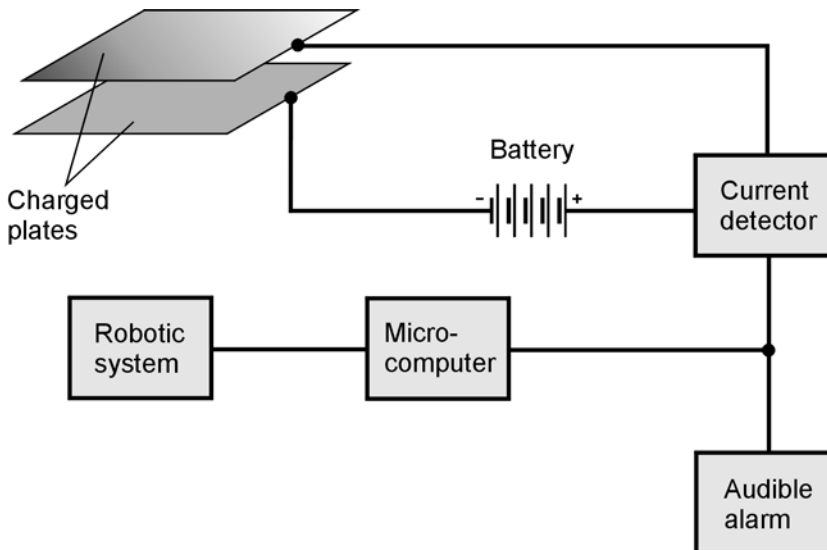
A *smoke detector* senses changes in two characteristics of the air: the *dielectric constant* (an expression of the extent to which pollutants increase the capacitance of the air) and the *ionization potential* (an expression of the extent to which pollutants change the voltage necessary to produce a spark that jumps through the air). Smoke and fire almost always affect one or both of these parameters.

Figure 34-1 is a functional diagram of a smoke detector that operates by sensing changes in the dielectric constant of the air. Two electrically charged plates are positioned a small distance apart. The plates, and the air between them, form a capacitor. A source of dc is connected to the plates. Normally, the plates retain a constant charge, and the current in the circuit is zero. If the dielectric constant of the air increases, the capacitance changes, causing a small, momentary electric current to flow. This current can be detected, and the resulting signal sets off an alarm. The signal can also actuate a robotic system, such as a group of water sprinklers.

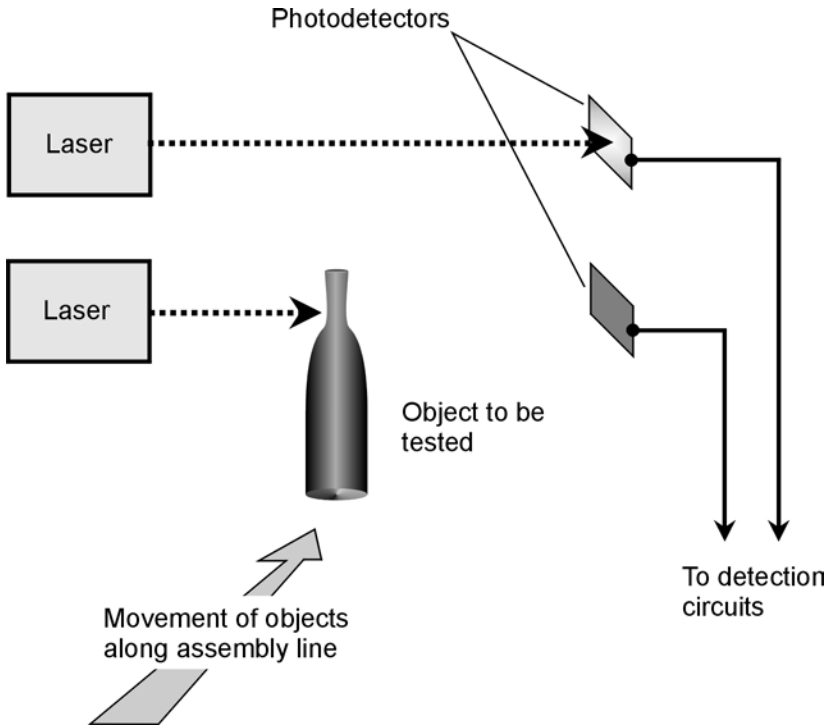
### Quality Control

Lasers are useful in industrial monitoring and control applications. An example is the *quality control* (QC) checking of bottles for height as they move along an assembly line. A laser/robot combination can find and remove bottles that are not of the correct height. The principle is shown in Fig. 34-2. If a bottle is too short, both laser beams reach the photodetectors. If a bottle is too tall, neither laser beam reaches the photodetectors. In either of these situations, a *robot arm*, equipped with a *grripper*, picks the faulty bottle off the line and discards it. Only when a bottle is within a narrow range of heights (the acceptable range) does the top laser reach its photodetector while the bottom laser is blocked. Then the bottle is allowed to pass.

Of course, proper operation of the QC machine shown in Fig. 34-2 depends on the reliability of the lasers and photodetectors. If, for example, the lower laser burns out and there is no way for the system operator to know about it, the machine will pass all bottles, whether or not they are of the correct height.



34-1 Simplified functional diagram of a smoke detector.



34-2 A method of quality control (QC) for manufactured items.

## Tracking People

Suppose that a person is sentenced to house arrest. Compliance can be monitored by having the person carry a conventional beeper (pager). A police or probation officer can page the person at random times; the person must then call the officer within a couple of minutes. The call can be traced, and the location of the telephone verified. This is a simple method of electronically tracking the whereabouts of a person.

A more secure method of ascertaining that a person is at a certain place, at a certain time, is by means of a short-range radio transmitter and receiver. The person wears the transmitting unit. Tamper-proof receiving units are placed at the convict's home, in the car, and at the place of work. The transmitter range is similar to that of a baby monitor. Receiver signals are sent to a central monitoring point. The signals are encoded so the monitoring personnel (or computers) know whether the person is at home, in the car, or at work. Any deviation from normal patterns can be detected.

*Radiolocation* provides another way to keep track of people. A *transponder* can be carried or worn by the person to be tracked; continuous signals can be sent to the unit asking for a position fix, and the unit can respond through a wireless network such as the cellular telephone system.

## Electronic Bug

An *electronic bug* consists of a tiny radio transmitter that can be hidden in a room, placed in a shirt pocket, or planted in a car. The antenna is a length of thin, almost invisible wire. A receiver can be located nearby. The device operates at a low RF power level (on the order of a few milliwatts) to con-

serve battery energy. If the transmitter is near a wireless repeater that connects into a larger system, eavesdropping can be done anywhere within the coverage of the wireless system. With the advent of low-earth-orbit (LEO) satellite systems, it is theoretically possible to bug a room on the other side of the world.

Wireless electronic bugs can be detected by means of a device called an *RF field-strength meter*. This instrument consists of a microammeter connected to a short whip antenna through a semiconductor diode. The diode rectifies the RF signal, producing dc that shows up as an indication on the meter. When the meter is close to the bug, the current rises. When the bug is within a few centimeters of the meter, the needle may go to full-scale.

The presence of RF fields in your house does not necessarily mean you are being bugged. Many appliances and electronic devices produce RF fields, including computers, radio-operated remote-control units, cell phones, cordless phone sets, and certain medical devices. Even radio receivers and TV sets emit some RF energy.

### **Electric Eye**

The simplest device for detecting an unwanted visitor is an *electric eye*. Narrow beams of IR or visible light are shone across all reasonable points of entry, such as doorways and window openings. A photodetector receives energy from each beam. If, for any reason, the photodetector stops receiving its assigned beam, an alarm is actuated. A person breaking into a property cannot avoid interrupting at least one beam if every possible point of entry has a large enough number of electric eyes spaced at suitable intervals.

### **IR Motion Detector**

A common intrusion alarm device employs an *IR motion detector*. Two or three wide-angle IR pulses are transmitted at regular intervals; these pulses cover most of the room in which the device is installed. A receiving transducer picks up the returned IR energy, normally reflected from the walls, floor, ceiling, and furniture. The intensity of the received pulses is noted by a microprocessor. If anything in the room changes position, there is a change in the intensity of the received energy. The microprocessor detects this change and triggers an alarm (Fig. 34-3). This type of device consumes very little power in regular operation, so batteries can serve as the power source.

### **Radiant Heat Detector**

Certain devices can detect changes in the indoor environment by directly sensing the IR energy (often called *radiant heat*) emanating from objects. Humans, and all warm-blooded animals, emit IR. So does fire. A simple IR sensor, in conjunction with a microprocessor, can detect rapid or large increases in the amount of radiant heat in a room. The time threshold can be set so that gradual or small changes will not trigger the alarm, while significant changes, such as are caused by a person entering the room, will trigger it. The temperature-change threshold can be set so that a small animal will not actuate the alarm, while a full-grown person will. This type of device, like the IR motion detector, can operate from batteries.

The main limitation of radiant-heat detectors is the fact that they can be fooled. False alarms are a risk. The sun, coming out on an overcast day, might suddenly shine directly on the sensor and trigger the alarm. It is also possible that a person clad in a winter jacket, thermal pants, insulated boots, hood, and face mask, entering from a cold outdoor environment, will fail to set off the alarm. For this reason, radiant-heat sensors are used more often as fire-alarm actuators than as intrusion detectors.



34-3 An IR motion detector.

### Ultrasonic Motion Detector

Motion in a room can be detected by sensing the changes in the relative phase of acoustic waves. An *ultrasonic motion detector* employs a set of transducers that emit acoustic waves at frequencies above the range of human hearing (higher than 20 kHz). Another set of transducers picks up the reflected acoustic waves, whose wavelength is on the order of a few millimeters. If anything in the room changes position, the relative phase of the waves, as received by the various acoustic pickups, will change. This data is sent to a microprocessor, which can trigger an alarm and/or notify the police.

## Robot Generations and Laws

Some researchers have analyzed the evolution of robots, marking progress according to so-called robot generations. One of the first engineers to make formal mention of robot generations was the Japanese engineer *Eiji Nakano*.

### First Generation

According to Nakano, a *first-generation robot* is a simple mechanical arm. Such machines have the ability to make precise motions at high speed, many times, for a long time. They have found widespread industrial application and have been in existence since the middle of the twentieth century. These are the fast-moving systems that install rivets and screws in assembly lines, that solder connections on printed circuits, and that, in general, have taken over tedious, mind-numbing chores that would otherwise have to be done by humans.

First-generation robots can work in groups if their actions are synchronized. The operation of these machines must be constantly watched, because if they get out of alignment and are allowed to keep operating anyway, the result can be a series of bad production units.

## Second Generation

A *second-generation robot* has some level of *artificial intelligence* (AI), also called *machine intelligence*. Such a machine is equipped with various sensors that keep it informed about goings-on in the *work environment*. A computer called a *robot controller* processes the data from the sensors and adjusts the operation of the robot accordingly. The earliest second-generation robots came into common use around 1980.

Second-generation robots can stay synchronized with each other, without having to be overseen constantly by a human operator. Periodic checking is needed, however, because things can always go wrong. In fact, as a system becomes more complex, the number of ways in which it can malfunction increases. This is why human beings will never find themselves out of work because of robots. Someone has to make sure the robots keep working properly!

## Third Generation

Nakano gave mention to *third-generation robots*, but in the years since the publication of his original paper, some things have changed. Two major avenues are developing for advanced robot technology. These are the *autonomous robot* and the *insect robot*. An autonomous robot is a single machine that works on its own. It contains a controller and can do things largely without supervision, either by an outside computer or by a human being. A good example of this type of third-generation robot is the *personal robot* about which technophiles dream. An insect robot is one of a set of several (or many) identical units that act together to perform a specific task.

## Fourth Generation and Beyond

Nakano did not write about anything past the third generation of robots. But we might mention a *fourth-generation robot*: a machine of a sort yet to be deployed. An example is a fleet or population of robots that can reproduce themselves, and perhaps even a system that can evolve to meet changing conditions in its work environment. Past that, we might say that a *fifth-generation robot* is something humans haven't even imagined yet.

## Asimov's Three Laws

In one of his early science-fiction stories, the famous author Isaac Asimov first mentioned the word *robotics*, along with three fundamental rules that, in his opinion, all robots ought to obey. These rules were first coined in the 1940s, but *Asimov's three laws of robotics* are still considered valid today:

- First law: A robot must not injure, or allow the injury of, any human being.
- Second law: A robot must obey all orders from humans, except orders that would contradict the first law.
- Third law: A robot must protect itself, except when to do so would contradict the first law or the second law.

## Robot Arms

A robot arm, in conjunction with an *end effector* (hand, gripper, or tool), is called a *manipulator*. Some robots, especially industrial robots, are nothing more than sophisticated manipulators. A robot arm can be categorized according to its geometry. Some manipulators resemble human arms. The joints in these machines can be given names like “shoulder,” “elbow,” and “wrist.” Other manipulators are so much different from human arms that these names don't make sense.



## Degrees of Freedom

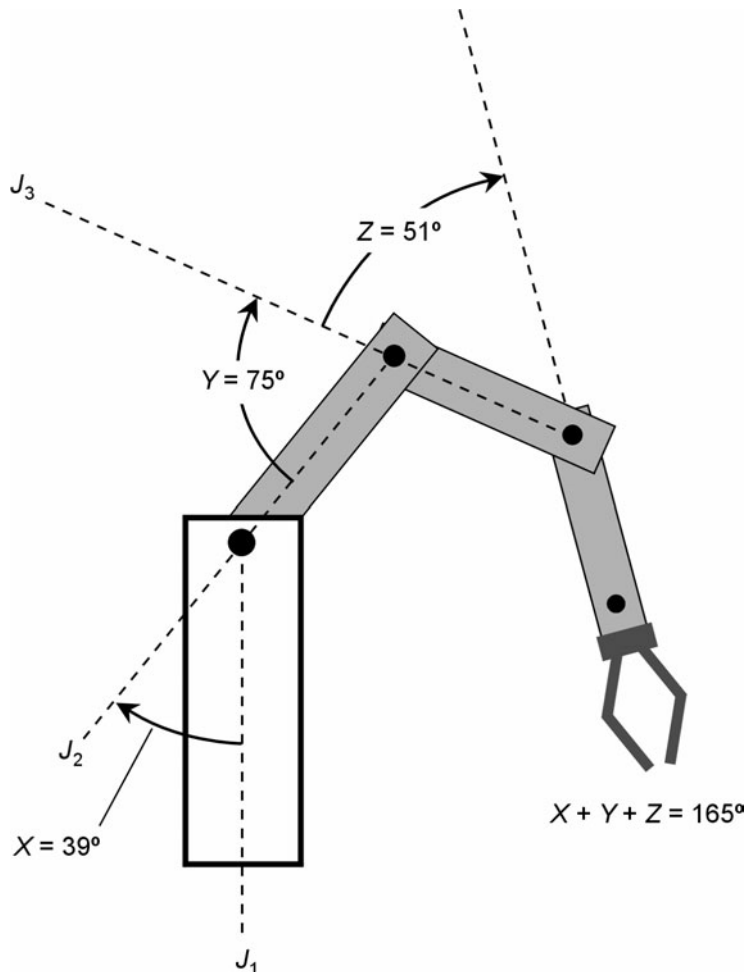
The term *degrees of freedom* refers to the number of different ways in which a robot manipulator can move. Most manipulators move in three dimensions, but often they have more than three degrees of freedom.

You can use your own arm to get an idea of the degrees of freedom that a robot arm might have. Extend your right arm straight out toward the horizon. Extend your index finger so it is pointing. Keep your arm straight, and move it from the shoulder. You can move your shoulder joint in three ways. Up-and-down movement is called *pitch*. Movement to the right and left is called *yaw*. You can rotate your whole arm from the shoulder (albeit to a limited extent); this motion is called *roll*. Your shoulder therefore has three degrees of freedom: pitch, yaw, and roll.

Now move your arm from the elbow only. Holding your shoulder in the same position constantly, you will see that your elbow joint has the equivalent of pitch in your shoulder joint. But that is all. Your elbow, therefore, has one degree of freedom.

Extend your arm toward the horizon again. Now move only your arm below the elbow. Your forearm and wrist can bend up and down, side to side, and it can also twist. Your lower arm has three degrees of freedom.

**34-4** An example of degrees of rotation in articulated geometry. Angles  $X$ ,  $Y$ , and  $Z$  are measured relative to axes  $J_1$ ,  $J_2$ , and  $J_3$ , respectively.



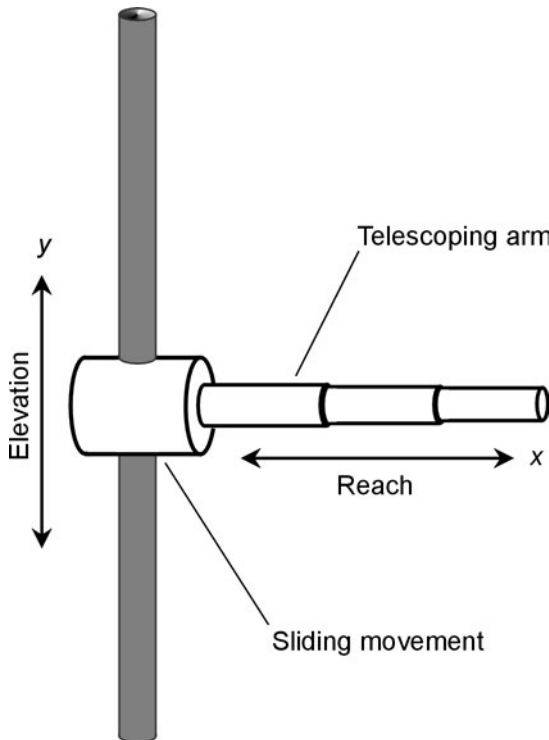
In total, your arm has seven degrees of freedom: three in the shoulder, one in the elbow, and three in the arm below the elbow.

It is tempting to suppose that a robot should never need more than three degrees of freedom, because, after all, space has only three dimensions. But the extra possible motions, provided by multiple joints, give a robot arm (and a human arm) versatility that it could not have with only three degrees of freedom.

### Degrees of Rotation

The term *degrees of rotation* refers to the extent to which a robot joint, or a set of robot joints, can turn clockwise or counterclockwise with respect to a prescribed linear axis. Some reference point is always used, and the angles are given in degrees with respect to that joint. Rotation in one direction (usually clockwise) is represented by positive angles; rotation in the opposite direction is specified by negative angles. Thus, for example, if angle  $X = 58^\circ$ , it refers to a rotation of  $58^\circ$  clockwise with respect to the reference axis. If angle  $Y = -74^\circ$ , it refers to a rotation of  $74^\circ$  counterclockwise.

Figure 34-4 shows a robot arm with three joints. The reference axes are  $J_1$ ,  $J_2$ , and  $J_3$  for rotation angles  $X$ ,  $Y$ , and  $Z$ , respectively. The individual angles add together. To move this robot arm to a certain position within its *work envelope* (the region in space that the arm can reach and actually manipulate things), the operator enters data into a computer. This data includes the measures of angles  $X$ ,  $Y$ , and  $Z$ . The operator has specified  $X = 39^\circ$ ,  $Y = 75^\circ$ , and  $Z = 51^\circ$ .



**34-5** Cartesian coordinate geometry in two dimensions.

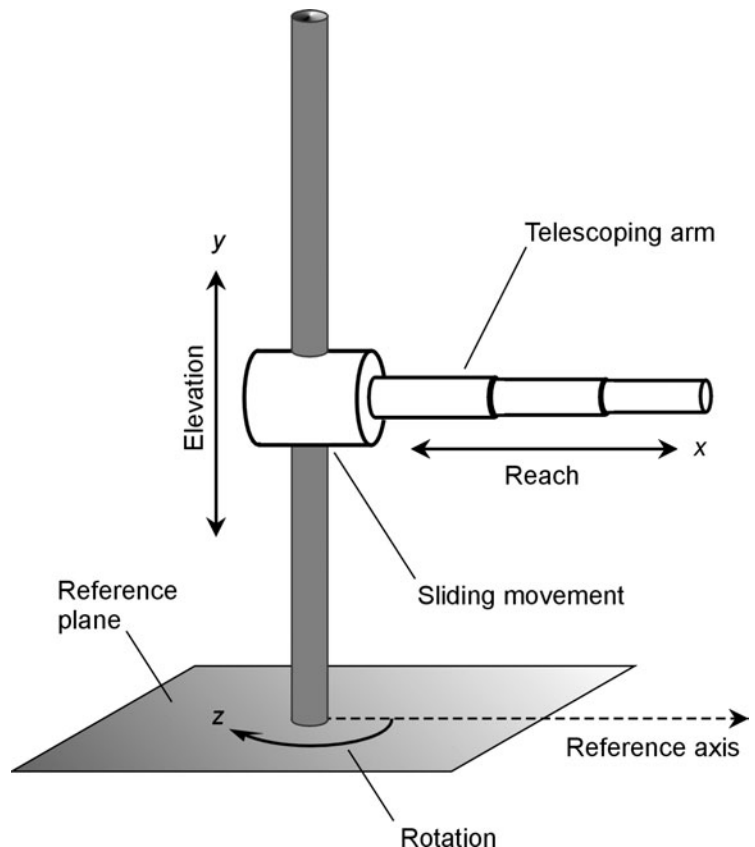
## Articulated Geometry

The word *articulated* means “broken into sections by joints.” A robot arm with *articulated geometry* bears some resemblance to the arm of a human. The versatility is defined in terms of the number of degrees of freedom. For example, an arm might have three degrees of freedom: base rotation (the equivalent of azimuth), elevation angle, and reach (the equivalent of radius). If you’re a mathematician, you might recognize this as a system of *spherical coordinates*. There are several different articulated geometries for any given number of degrees of freedom. Figure 34-4 is a simplified drawing of a robot arm that uses articulated geometry.

## Cartesian Coordinate Geometry

Another mode of robot arm movement is known as *Cartesian coordinate geometry* or *rectangular coordinate geometry*. This term comes from the *Cartesian coordinate system* often used for graphing mathematical functions. The axes are always perpendicular to each other. Variables are assigned the letters  $x$  and  $y$  in a two-dimensional *Cartesian plane*, or  $x$ ,  $y$ , and  $z$  in *Cartesian three-space*. The dimensions are called *reach* for the  $x$  variable, *elevation* for the  $y$  variable, and *depth* for the  $z$  variable. Figure 34-5 is a simplified rendition of a robot arm capable of moving in two dimensions using Cartesian coordinate geometry.

**34-6** Cylindrical coordinate geometry in three dimensions.



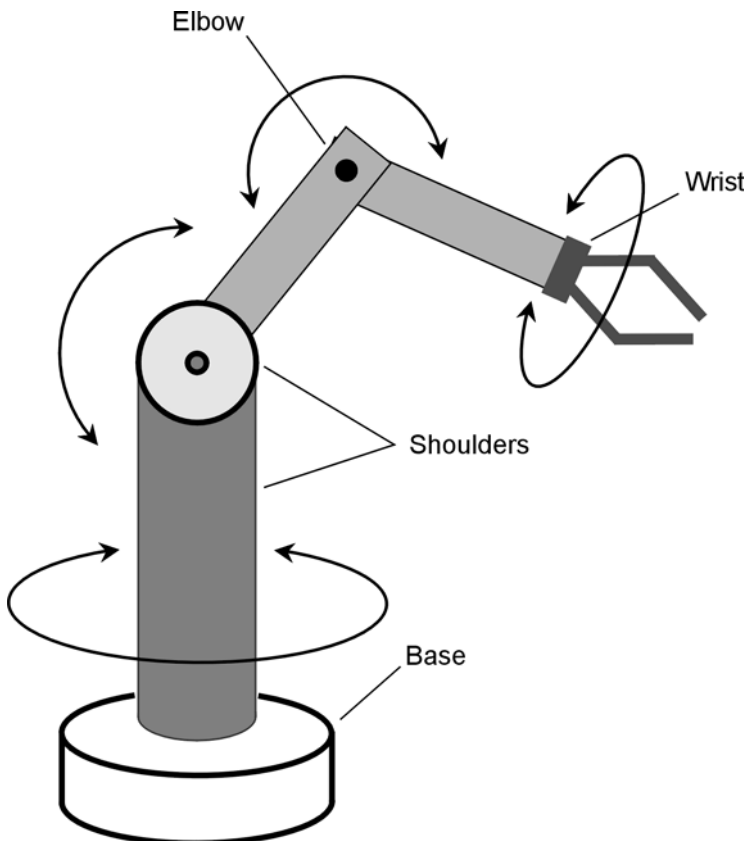
### Cylindrical Coordinate Geometry

A robot arm can be guided by means of a two-dimensional *navigator's polar coordinate system* with an elevation dimension added (Fig. 34-6). This is known as *cylindrical coordinate geometry*. In this system, a *reference plane* is used. An origin point is chosen in this plane. A *reference axis* is defined, running away from the origin in the reference plane. In the reference plane, the position of any point can be specified in terms of reach  $x$ , elevation  $y$ , and rotation  $z$ . The rotation is defined as the angle that the reach arm subtends relative to the reference axis. In this example, it is in the clockwise sense. Note that this is just like the situation for two-dimensional Cartesian coordinate geometry shown in Fig. 34-5, except that the sliding movement is also capable of rotation.

The rotation angle  $z$  can range from  $0^\circ$  to  $360^\circ$  clockwise from the reference axis. In some systems, the range is specified as  $0^\circ$  to  $+180^\circ$  (up to a half circle clockwise from the reference axis), and  $0^\circ$  to  $-180^\circ$  (up to a half circle counterclockwise from the reference axis).

### Revolute Geometry

A robot arm capable of moving in three dimensions using *revolute geometry* is shown in Fig. 34-7. The whole arm can rotate through a full circle ( $360^\circ$ ) at the base point, or *shoulder*. There is also an elevation joint at the base that can move the arm through  $90^\circ$ , from horizontal to vertical. A joint in the middle of the robot arm, at the *elbow*, moves through  $180^\circ$ , from a straight position to dou-



34-7 Revolute geometry in three dimensions.

bled back on itself. There might be, but is not always, a *wrist* joint that can flex like the elbow and/or twist around and around.

A 90° elevation revolute robot arm can reach any point within a half sphere. The radius of the half sphere is the length of the arm when its elbow and wrist (if any) are straightened out. A 180° elevation revolute arm can be designed that will reach any point within a fully defined sphere, with the exception of the small obstructed region around the base.

## Robot Hearing and Vision

Machine hearing involves detection of acoustic waves, along with amplification and analysis of the resulting audio signals. Machine vision involves the interception of visible, infrared (IR), or ultraviolet (UV) radiation, and translating this energy into electronic images. Machine hearing and vision can allow robots to locate, and in some cases classify or identify, objects in the environment.

### Binaural Hearing

Even with your eyes closed, you can usually tell from which direction a sound is coming. This is because you have *binaural hearing*. Sound arrives at your left ear with a different intensity, and in a different phase, than it arrives at your right ear. Your brain processes this information, allowing you to locate the source of the sound, with certain limitations. If you are confused, you can turn your head until the direction becomes apparent.

Robots can be equipped with binaural hearing. Two acoustic transducers are positioned, one on either side of the robot's head. A microprocessor compares the relative phase and intensity of signals from the two transducers. This lets the robot determine, within certain limitations, the direction from which sound is coming. If the robot is confused, it can turn until the confusion is eliminated and a meaningful bearing is obtained. If the robot can move around and take bearings from more than one position, a more accurate determination of the source location is possible if the source is not too far away.

### Visible-Light Vision

A visible-light robotic *vision system* must have a device for receiving incoming images. This is usually a *charge coupled device* (CCD) video camera, similar to the type used in home video cameras. The camera receives an analog video signal. This is processed into digital form by an ADC. The digital signal is clarified by means of DSP. The resulting data goes to the robot controller.

The moving image, received from the camera and processed by the circuitry, contains an enormous amount of information. It's easy to present a robot controller with a detailed and meaningful moving image. But getting the robot controller to know what's happening, and to determine whether or not these events are significant, is another problem altogether.

### Optical Sensitivity and Resolution

*Optical sensitivity* is the ability of a machine vision system to see in dim light or to detect weak impulses at invisible wavelengths. In some environments, high optical sensitivity is necessary. In others, it is not needed and might not be wanted. A robot that works in bright sunlight doesn't need to be able to see well in a dark cave. A robot designed for working in mines, pipes, or caverns must be able to see in dim light, using a system that might be blinded by ordinary daylight.

*Optical resolution* is the extent to which a machine vision system can differentiate between objects that are close together in the field of vision. The better the optical resolution, the keener the vision. Human eyes have excellent optical resolution, but machines can be designed to have superior resolution.

In general, the better the optical resolution, the more confined the field of vision must be. To understand why this is true, think of a telescope. The higher the magnification, the better its optical resolution will be, up to a certain maximum useful magnification. Increasing the magnification reduces the angle, or field, of vision. Zeroing in on one object or zone is done at the expense of other objects or zones.

Optical sensitivity and resolution are interdependent. If all other factors remain constant, improved sensitivity causes a sacrifice in resolution. Also, the better the optical resolution, the more incident light it requires to function well. In this case, a good analogy is camera film (the old-fashioned kind). The fastest films require more light than slow ones. The corollary to this is the fact that if you want excellent detail in a photograph, you will have to expose the film for a comparatively long period of time.

### **Invisible and Passive Vision**

Robots have an advantage over people when it comes to vision. Machines can see at wavelengths to which humans are blind.

Human eyes are sensitive to EM waves whose length ranges from 390 to 750 nanometers (nm). The nanometer is a thousand-millionth ( $10^{-9}$ ) of a meter. The longest visible wavelengths look red. As the wavelength gets shorter, the color changes through orange, yellow, green, blue, and indigo. The shortest waves look violet. Infrared (IR) energy is at wavelengths somewhat longer than 750 nm. Ultraviolet (UV) energy is at wavelengths somewhat shorter than 390 nm.

Machines, and most nonhuman living species, can see energy in a range of wavelengths that differs somewhat from the range of wavelengths to which human eyes respond. For example, insects can see UV that humans cannot, but insects are blind to red and orange light that humans can see. (Have you used orange bug lights when camping to keep the flying pests from coming around at night, or those UV devices that attract bugs and then zap them?)

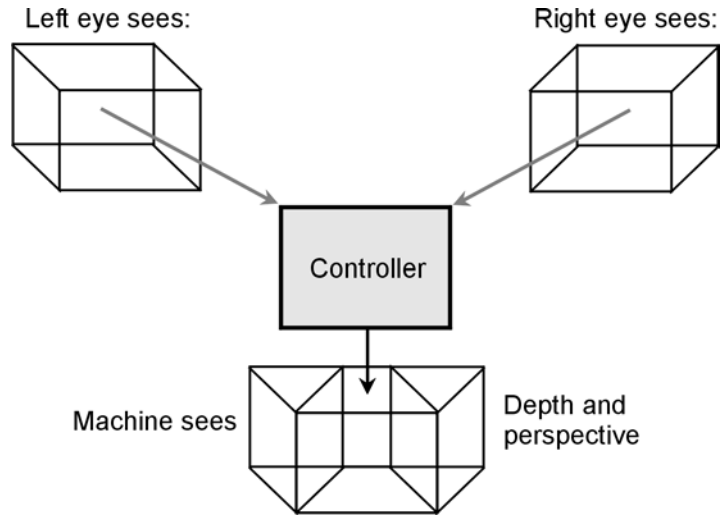
A robot can be designed to see IR and/or UV, as well as (or instead of) visible light, because video cameras can be sensitive to a range of wavelengths much wider than the range humans can see. Robots can be made to see in an environment that is dark and cold, and that radiates too little energy to be detected at any electromagnetic wavelength. In these cases the robot provides its own illumination. This can be a simple lamp, a laser, an IR device, or a UV device. Radar and sonar can also be used.

### **Binocular Vision**

*Binocular machine vision* is the analog of binocular human vision. It is sometimes called *stereo vision* or *stereoscopic vision*. In humans, binocular vision allows perception of depth. With only one eye—that is, with monocular vision—you can infer depth only to a limited extent, and that perception is entirely dependent on your knowledge of the environment or scene you are observing. Almost everyone has had the experience of being fooled when looking at a scene with one eye covered or blocked. A nearby pole and a distant tower might seem to be adjacent, when in fact they are a city block apart.

Figure 34-8 shows the basic concept of binocular robot vision. High-resolution video cameras, and a sufficiently powerful robot controller, are essential components of such a system.

**34-8** Binocular machine vision. Two different views of the same object are combined to achieve a sense of depth and perspective.



## Color Sensing

Robot vision systems often function only in grayscale, like old-fashioned 1950s television. But color sensing can be added, in a manner similar to the way it is added to television systems. Color sensing can help a robot with AI figure out what an object is. Is that horizontal surface a parking lot, or is it a grassy yard? Sometimes, objects have regions of different colors that have identical brightness as seen by a grayscale system. Such objects, obviously, can be evaluated in more detail with a color-sensing system than with a vision system that sees only shades of gray.

In a typical color-sensing vision system, three grayscale cameras are used. Each camera has a color filter in its lens. One filter passes red light, another passes green light, and another passes blue light. These are the three *primary colors*. All possible hues, levels of brightness, and levels of saturation are made up of these three colors in various ratios. The signals from the three cameras are processed by a microcomputer, and the result is fed to the robot controller.

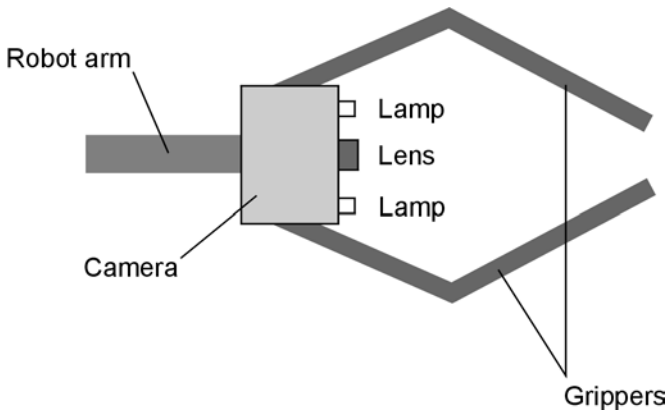
## The Eye-in-Hand System

In order to assist a robot gripper (hand) in finding its way, a camera can be placed in the mechanism. The camera must be equipped for work at close range, from about 1 m down to 1 mm or less. The positioning error must be kept as small as possible. To be sure that the camera gets a good image, lamps are included in the gripper along with the camera (Fig. 34-9). This so-called eye-in-hand system can be used to precisely measure the distance between the gripper and the object it is seeking. It can also make positive identification of the object.

The eye-in-hand system takes advantage of the properties of a *servo*. The robot is equipped with, or has access to, a computer that processes the data from the camera and sends instructions back to the gripper. Most eye-in-hand systems use visible light for guidance and manipulation. Infrared (IR) can be used when it is necessary for the robot gripper to sense differences in temperature.

## The Flying Eyeball

In environments hostile to humans, robots find many uses, from manufacturing to exploration. One such device, especially useful underwater, has been called a *flying eyeball*. A cable, containing



**34-9** A robotic eye-in-hand system.

the robot in a special launcher housing, is dropped from a boat. When the launcher gets to the desired depth, it lets out the robot, which is connected to the launcher by a tether. The tether and the drop cable convey data back to the boat.

In some cases, the tether for a flying eyeball can be eliminated, and a wireless link can be used to convey data from the robot to the launcher. The link is usually in the IR or visible red portion of the spectrum. The robot contains a video camera and one or more lamps to illuminate the underwater environment. It also has a set of thrusters (jets or propellers) that let it move around according to control commands sent from the boat. Human operators on board the boat watch the images and guide the robot.

## Robot Navigation

Mobile robots must get around in their environment without wasting motion, without running into things, and without tipping over or falling down a flight of stairs. The nature of a *robot navigation system* depends on the size of the work area, the type of robot used, and the sorts of tasks the robot is required to perform. In this section, we'll look at four common methods of robot navigation.

### Clinometer

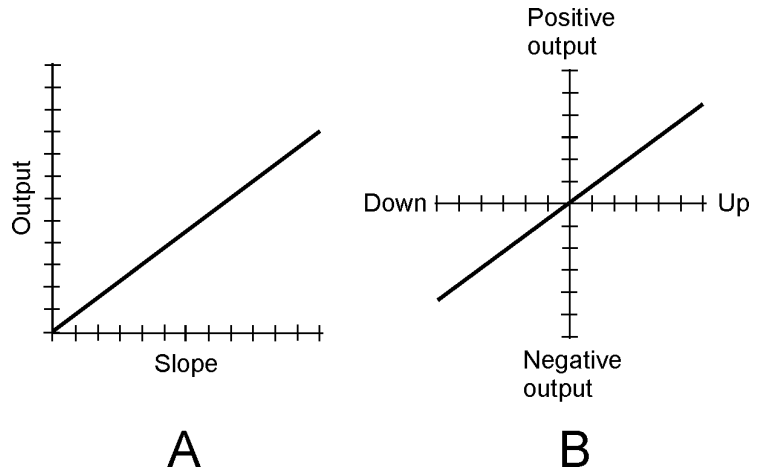
A *clinometer* is a device for measuring the steepness of a sloping surface. Mobile robots use clinometers to avoid inclines that might cause them to tip over or that are too steep for them to ascend while carrying a load.

The floor in a building is almost always horizontal. Thus, its incline is zero. But sometimes there are inclines such as ramps. A good example is the kind of ramp used for wheelchairs, in which a very small elevation change occurs. A rolling robot cannot climb stairs, but it can use a wheelchair ramp, provided the ramp is not so steep that it would upset the robot's balance or cause it to lose its payload.

In a clinometer, a transducer produces an electrical signal whenever the device is tipped from the horizontal. The greater the angle of incline, the greater the electrical output, as shown in the graph of Fig. 34-10A. A clinometer might also show whether an incline goes down or up. A downward slope might cause a negative voltage at the transducer output, and an upward slope a positive voltage, as shown in the graph at Fig. 34-10B.



**34-10** Response functions for clinometers. At A, the output voltage is a simple function of slope. At B, up-slope causes positive output voltage; downslope causes negative output voltage.



## Edge Detection

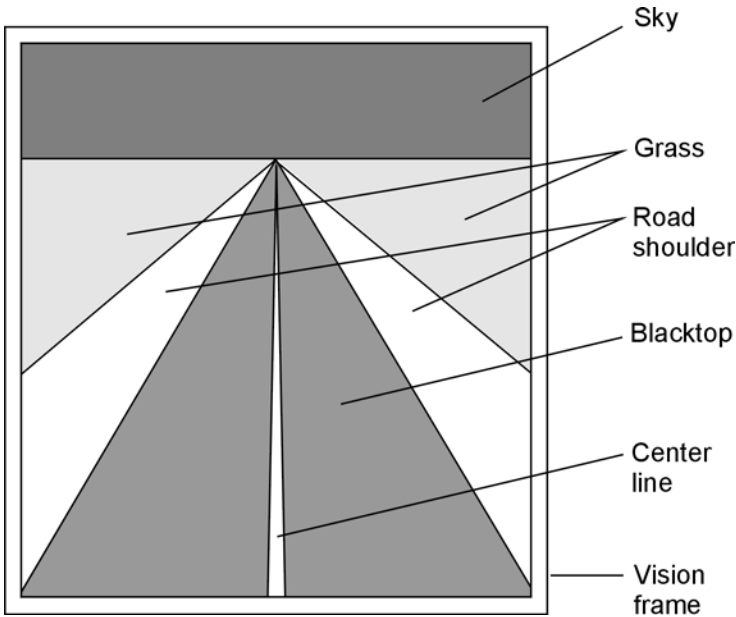
The term *edge detection* refers to the ability of a robot vision system to locate boundaries. It also refers to the robot's knowledge of what to do with respect to those boundaries. A robot car, bus, or truck can use edge detection to see the edges of a road and use the data to keep itself on the road. But it must stay a certain distance from the right-hand edge of the pavement to avoid crossing into the lane of oncoming traffic (Fig. 34-11). It also must stay off the road shoulder. It must be able to tell the difference between pavement and other surfaces, such as gravel, grass, sand, and snow.

The interior of a home or business contains straight-line edge boundaries of all kinds, and each boundary represents a potential point of reference for a mobile robot's edge detection system. The controller in a personal home robot must be programmed to know the difference between, say, the line where carpet ends and tile begins, and the line where a flight of stairs begins. The vertical line produced by the intersection of two walls would present a different situation than the vertical line produced by the edge of a doorway, even though they might appear identical. Thus, edge detection cannot function very well without a certain amount of AI in the robot controller.

## Embedded Path

An *embedded path* is a means of guiding a robot along a specific route. This scheme is commonly used by a mobile robot called an *automated guided vehicle* (AGV). A common embedded path consists of a buried, current-carrying wire. The current in the wire produces a magnetic field that the robot can follow. This method of guidance has been suggested as a way to keep a car on a highway, even if the driver isn't paying attention. The wire needs a constant supply of electricity for this guidance method to work. If this current is interrupted for any reason, the robot will lose its way unless some backup navigation method (or human control) is substituted.

Alternatives to wires, such as colored or reflective paints or tapes, do not need a supply of power, and this gives them an advantage. Tape is easy to remove and put somewhere else; this is difficult to do with paint and practically impossible with wires embedded in concrete. However, tape is obscured by snowfall; and at night, glare from oncoming headlights might be confused for reflections from the tape.



**34-11** In edge detection, a robot uses visual boundaries to facilitate navigation. This is what the controller might see in a robot car driving down the center line of a highway.

### Range Sensing and Plotting

*Range sensing* is the measurement of distances to objects in a robot's environment in a single dimension. *Range plotting* is the creation of a graph of the distance (range) to objects, as a function of the direction in two or three dimensions.

In linear or one-dimensional (1D) range sensing, a signal is sent out, and the robot measures the time it takes for the echo to come back. This signal can be sound, in which case the device is sonar. Or it can be a radio wave; this constitutes radar. Laser beams can also be used. Close-range, one-dimensional range sensing is known as *proximity sensing*.

Two-dimensional (2D) range plotting involves mapping the distance to various objects, as a function of direction in a geometric plane. The echo return time for a sonar signal, for example, might be measured every few degrees around a complete circle in the horizontal plane, resulting in a set of range points. A better plot would be obtained if the range were plotted every degree, every tenth of a degree, or even every hundredth of a degree. But no matter how detailed the direction resolution, a 2D range plot is done in only one plane, such as the floor level in a room, or some horizontal plane above the floor. The greater the number of echo samples in a complete circle (that is, the smaller the angle between samples), the more detail can be resolved at a given distance from the robot, and the greater the distance at which a given amount of detail can be resolved.

Three-dimensional (3D) range plotting is done in spherical coordinates: azimuth (compass bearing), elevation (degrees above the horizontal), and range (distance). The distance must be measured for a large number of diverse orientations. In a furnished room, a 3D sonar range plot would show ceiling fixtures, things on the floor, objects on top of a desk, and other details not visible with a 2D plot. The greater the number of echo samples in a complete sphere surrounding the robot, the more detail can be resolved at a given distance, and the greater the range at which a given amount of detail can be resolved.

# Telepresence

*Telepresence* is a refined, advanced form of robot remote control. The robot operator gets a sense of being “on location,” even if the remotely controlled machine, or *telechir*, and the operator are miles apart. Control and feedback are done by means of *telemetry* sent over wires, optical fibers, or radio.

## What It's Like

Here is an example of a telepresence scenario. The robot is autonomous and has a humanoid form. The control station consists of a suit that you wear or a chair in which you sit with various manipulators and displays. Sensors can give you feelings of pressure, sight, and sound. You wear a helmet with a viewing screen that shows whatever the robot camera sees. When your head turns, the robot head, with its vision system, follows, so you see an image that changes as you turn your head, as if you were in a space suit or diving suit at the location of the robot. Binocular machine vision provides a sense of depth. Binaural machine hearing allows you to perceive sounds. Special vision modes let you see UV or IR; special hearing modes let you hear ultrasound or infrasound.

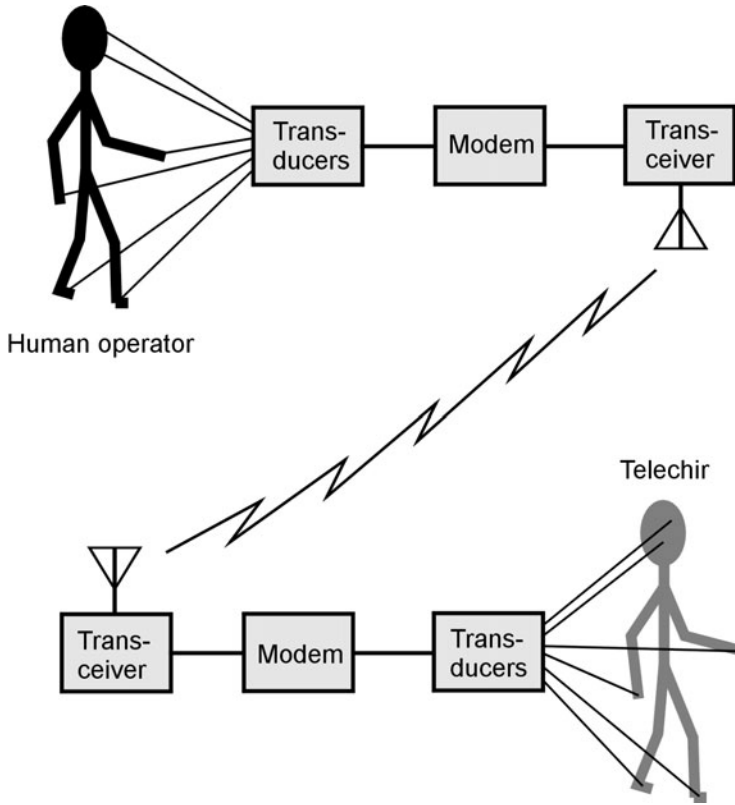
A telechir can be propelled by means of a *track drive* (similar to those used by military tanks), a *wheel drive*, or *robot legs*. If the propulsion uses legs, you propel the telechir by literally walking around a room in your telepresence suit! In the case of track drive or wheel drive, you sit in a chair and drive the robot like a tank or a car. The telechir, which is a form of android, has two arms, each with grippers resembling human hands. When you want to pick something up, you go through the normal motions with your own hands. Back-pressure sensors and position sensors let you feel (to a limited extent) what's going on. If an object masses 1 kg, it feels as if it masses 1 kg because of resistance provided by the back-pressure sensors. But it will be as if you're wearing thick gloves; you won't be able to feel texture. You might throw a switch, and something that masses 10 kg feels as if it masses only 1 kg. This might be called “strength  $\times 10$ ” mode. If you switch to “strength  $\times 100$ ” mode, a 100-kg object seems to mass only 1 kg.

## Functional Description

Figure 34-12 is a simplified diagram of a robot telepresence system using an android and a wireless control link.

At the control end of the system (where the human operator is), electromechanical transducers convert the human operator's movements into electrical signals. The modem converts these signals into analog form. The analog signals modulate the RF produced by the transceiver, which in turn transmits commands by wireless to the telechir end of the system. Radio signals from the telechir arrive at the operator-end transceiver. These signals are converted into analog impulses, and are translated by the modem into electrical signals that power transducers, giving the operator a sense of what is going on.

At the telechir end of the system, the transceiver receives control signals from the human operator. These signals are translated by the modem into impulses that drive electromechanical transducers, propelling or manipulating the telechir. Data obtained by the telechir, such as its visual sense of position, environmental sounds, or the apparent mass of a lifted object, are converted by the modem into electrical signals. These signals modulate the RF energy produced by the transceiver. Signals are thus sent back to the operator end by wireless.



**34-12** Telepresence combines remote monitoring and control, giving the operator the feeling of being onsite at a distant location.

## Applications

Here are a few potential applications for a telepresence system using a human operator and an android:

- Working in extreme heat or cold
- Working under high pressure, such as on the sea floor
- Working in a vacuum, such as in space
- Working where there is dangerous radiation
- Disarming bombs
- Handling toxic substances
- Serving in high-risk police or military situations

Of course, the robot must be able to survive conditions at its location. Also, it must have some way to recover if it falls or gets knocked over.

## Limitations

The technology for telepresence has existed for some time. But certain logistical problems have always bedeviled engineers committed to developing such systems.

The most serious limitation is the fact that telemetry cannot, and never will, travel faster than the speed of light in free space. This is not a problem over short-range links (a few hundred kilome-

ters or less), but it is slow in outer space, and nigh impractical in interplanetary operations. The moon is approximately 1.3 light-seconds away from the earth; that means that any command sent to a telechir from the earth to the moon takes 1.3 s to get there, and any data from the telechir takes another 1.3 s to get back. For planets in the solar system, the delay is on the order of several minutes to several hours. On an interstellar scale, telepresence is out of the question. The nearest stars are at distances of several light-years.

Another problem is the resolution of the robot's vision. A human being with good eyesight can see things in considerable detail. To send images in real-life detail, at reasonable speed, requires a signal with broad bandwidth. There are engineering problems (and cost problems) that go along with this. However, if one is willing to deal with the cost and accommodate the signal bandwidth, robot vision systems can be designed that offer optical resolution superior to human eyesight.

Still another limitation is best put as a question: How will a robot be able to feel something and transmit these impulses to the human brain? For example, an apple feels smooth, a peach feels fuzzy, and an orange feels shiny yet bumpy. How can this sense of texture be realistically transmitted to the human brain?

## The Mind of the Machine

A simple electronic calculator doesn't have AI. But a machine that can learn from its mistakes, or that can show reasoning power, does. Between these extremes, there is no precise dividing line. As computers become more powerful, people tend to set higher standards for what they call AI. Things that were once thought of as AI are now ordinary. Things that seem fantastic now will someday be humdrum. There is a tongue-in-cheek axiom: We can call "computer intelligence" true AI only as long as it remains a little bit mysterious.

### Robotics and AI

Robotics and AI complement each other. Scientists have dreamed for more than a century about building *smart androids*: robots that look like people, act like people, and can even reason like people. Androids exist, but they aren't very smart. Powerful computers exist, but they lack mobility.

If a machine has the ability to move around under its own power, to lift things, and to move things, it seems reasonable that it should do so with some degree of intelligence if it is to accomplish anything worthwhile. Conversely, if a computer is to manipulate anything, it must be able to cause a machine to do physical work according to a precise program.

### Expert Systems

The term *expert systems* refers to a method of reasoning in AI. Sometimes this scheme is called the *rule-based system*. Expert systems are used in the control of smart robots.

The heart of an expert system is a set of facts and rules. In the case of a robotic system, the facts consist of data about the robot's work environment, such as a factory, an office, or a kitchen. The rules are statements of the logical form "If X, then Y," similar to many of the statements in high-level programming languages. An *inference engine* decides which logical rules should be applied in various situations and instructs the robot to carry out certain tasks. But the operation of the system can only be as sophisticated as the data supplied by human programmers.

Expert systems can be used in computers to help people do research, make decisions, and make forecasts. A good example is a program that assists a physician in making a diagnosis. The

computer asks questions and arrives at a conclusion based on the answers given by the patient and doctor.

One of the biggest advantages of expert systems is the fact that reprogramming is easy. As the environment changes, the robot can be taught new rules, and supplied with new facts.

### **How Smart a Machine?**

Experts in the field of AI have been disappointed in the past few decades. Computers have been designed that can handle tasks no human could ever contend with, such as navigating a space probe. Machines have been built that can play board games well enough to compete with human masters. Modern machines can understand, as well as synthesize, much of any spoken language. But these abilities, by themselves, don't count for much in the dreams of scientists who hope to create *artificial life*.

The human mind is incredibly complicated. A circuit that would have occupied, and used all the electricity in, a whole city in 1940 can now be housed in a box the size of a vitamin pill and run by a battery. Imagine this degree of miniaturization happening again, then again, and then again. Would that begin to approach the level of sophistication in your body's nervous system?

Is the human brain nothing more than an amazingly complicated digital switching network? Or is there something more to the human mind? No electronic device yet constructed has come anywhere near human intelligence in every respect. Some experts think that a machine will someday be built that is smarter than its creators. Others insist that the very idea is ridiculous.

It is tempting to extrapolate: If technological trends of the past few decades continue indefinitely, will the only limit on machine intelligence be defined by human imagination?

## **Quiz**

Refer to the text in this chapter if necessary. A good score is 18 correct. Answers are in the back of the book.

1. An android takes the form of
  - (a) an insect.
  - (b) the human body.
  - (c) a simple robot arm.
  - (d) a stereo vision system.
2. According to Asimov's three laws, under what circumstances is it all right for a robot to injure a human being?
  - (a) Under no circumstances
  - (b) When the human being specifically requests it
  - (c) In case of an accident
  - (d) In case the robot controller is infected with a computer virus
3. An RF field strength meter can be used to
  - (a) test the performance of a binaural hearing system.
  - (b) detect the presence of ionized air.

- (c) measure the dielectric constant of the air.
  - (d) detect the presence of a wireless bugging system.
4. The extent to which a machine vision system can differentiate between two objects that are close together is called the
- (a) optical magnification.
  - (b) optical sensitivity.
  - (c) optical selectivity.
  - (d) optical resolution.
5. A robot car or truck can best keep itself traveling down a specific lane of traffic by means of
- (a) stereoscopic machine hearing.
  - (b) epipolar navigation.
  - (c) edge detection.
  - (d) proximity sensing.
6. A rule-based system is also known as
- (a) a logic gate.
  - (b) an expert system.
  - (c) a back-pressure sensor.
  - (d) a telechir.
7. Suppose you are using a battery-powered, multichannel baby monitor, and you hear one end of a two-way radio conversation on the receiver. You check the baby's room, and it is quiet. How might this problem be resolved?
- (a) Put the receiver in a different location.
  - (b) Switch the monitor to a different channel.
  - (c) Interchange the transmitting and receiving units.
  - (d) Use ac power instead of battery power.
8. In robotics, the term *manipulator* refers to
- (a) a robot propulsion system.
  - (b) a robot arm, and the device at its end (such as a gripper).
  - (c) the system used to remotely control a telechir.
  - (d) a computer that guides a fleet of mobile robots.
9. A device with an IR sensor can be used to detect the presence of
- (a) slow-moving objects.
  - (b) RF signals.
  - (c) ionized air.
  - (d) warm or hot objects.

10. Proximity sensing is most closely akin to
  - (a) direction measurement.
  - (b) edge detection.
  - (c) range plotting.
  - (d) binaural machine hearing.
11. A telechir is always used in conjunction with a specialized system of
  - (a) track drive.
  - (b) wheel drive.
  - (c) remote control.
  - (d) ionization potential measurement.
12. A limit to the distance over which telepresence is practical is imposed by
  - (a) the speed of EM wave propagation.
  - (b) the image resolution of the vision system.
  - (c) the ability of a robot to determine texture.
  - (d) all of the above.
13. The ionization potential of the air can be determined in order to
  - (a) detect smoke.
  - (b) plot distances and directions.
  - (c) measure slope.
  - (d) detect boundaries.
14. Two-dimensional range plotting
  - (a) takes place along a single geometric line.
  - (b) takes place in a single geometric plane.
  - (c) is done using spherical coordinates.
  - (d) requires an ultrasonic sonar system.
15. Spherical coordinates can uniquely define the position of a point in up to
  - (a) one dimension.
  - (b) two dimensions.
  - (c) three dimensions.
  - (d) four dimensions.
16. The total number of ways in which a robot arm can move is known as
  - (a) functional orientation.
  - (b) degrees of freedom.
  - (c) dimensional versatility.
  - (d) coordinate geometry.



17. The region in space throughout which a robot arm can accomplish tasks is called its
- (a) coordinate geometry.
  - (b) reference axis.
  - (c) reference frame.
  - (d) work envelope.
18. A robot arm that moves along three independent axes, each of which is straight and perpendicular to the other two, employs
- (a) revolute geometry.
  - (b) spherical coordinate geometry.
  - (c) Cartesian coordinate geometry.
  - (d) cylindrical coordinate geometry.
19. Fill in the blank to make the following sentence true: "A color vision system can use three grayscale cameras, equipped with filters that pass \_\_\_\_ light."
- (a) red, yellow, and blue
  - (b) blue, red, and green
  - (c) white, black, and gray
  - (d) orange, green, and violet
20. A robot typically determines the steepness of a slope by means of
- (a) an epipolar navigation system.
  - (b) a clinometer.
  - (c) an end effector.
  - (d) a proximity sensor.

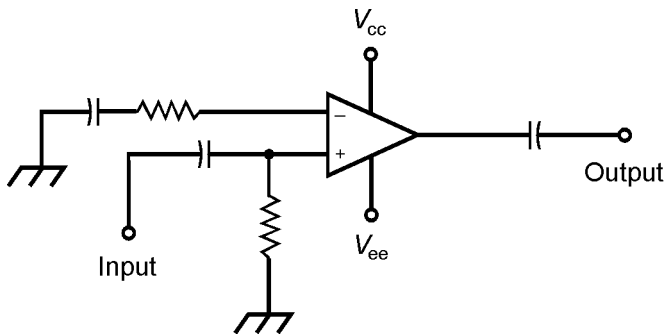
# Test: Part 4

DO NOT REFER TO THE TEXT WHEN TAKING THIS TEST. A GOOD SCORE IS AT LEAST 37 CORRECT. Answers are in the back of the book. It's best to have a friend check your score the first time, so you won't memorize the answers if you want to take the test again.

1. Which of the following tasks can a displacement transducer perform?
  - (a) It can detect the peak amplitude of an ac signal and display it on a screen.
  - (b) It can convert acoustic energy into radio signals.
  - (c) It can detect radiated IR energy and display IR images on a screen.
  - (d) It can convert an electrical signal into mechanical rotation through a defined angle.
  - (e) It can be used to measure the frequency of a complex wave.
  
2. Which of the following is *not* an advantage of an IC, compared to a circuit built with discrete components (individual resistors, capacitors, inductors, diodes, and transistors)?
  - (a) Compactness
  - (b) Reliability
  - (c) Ease of maintenance
  - (d) Low power consumption
  - (e) Unlimited power handling capacity
  
3. Computer memory is typically measured in
  - (a) kilobits, megabits, and gigabits.
  - (b) kilobits per second, megabits per second, and gigabits per second.
  - (c) kilobytes, megabytes, and gigabytes.
  - (d) kilobytes per second, megabytes per second, and gigabytes per second.
  - (e) any of the above.

4. A mode of communications in which either party can hear while talking, thus allowing one party to instantly interrupt the other at any time, is called
- (a) half simplex.
  - (b) full simplex.
  - (c) half duplex.
  - (d) full duplex.
  - (e) nothing, because there is no such mode.
5. An FM stereo tuner generally covers a frequency range of
- (a) 535 kHz to 1605 kHz.
  - (b) 88 MHz to 108 MHz.
  - (c) 3 MHz to 30 MHz.
  - (d) 9 kHz to 300 MHz.
  - (e) 144 MHz to 148 MHz.
6. Suppose an antenna has a radiation resistance of  $35 \Omega$  and a loss resistance of  $15 \Omega$ . What is the efficiency of the antenna?
- (a) 20 percent
  - (b) 30 percent
  - (c) 43 percent
  - (d) 70 percent
  - (e) Impossible to calculate from this information
7. The peak power output from a multicavity Klystron
- (a) can be much greater than the average power output.
  - (b) can be slightly greater than the average power output.
  - (c) is about the same as the average power output.
  - (d) is slightly less than the average power output.
  - (e) is much less than the average power output.
8. To which of the following does the term *platter* apply?
- (a) An individual disk in a hard drive
  - (b) An individual disk in CD-R media
  - (c) An individual disk in CD-RW media
  - (d) An individual disk in flash memory
  - (e) An individual disk in RAM
9. Why can signals in the so-called shortwave band sometimes propagate for thousands of kilometers without the need for satellites, repeaters, or any other infrastructure?
- (a) Because EM waves at these frequencies are often returned to the earth by the ionosphere
  - (b) Because waves at these frequencies can propagate through the ground or the ocean as electric currents

- (c) Because EM waves at these frequencies are reflected back to the earth by the moon
  - (d) Because EM waves at these frequencies propagate along the boundary between the troposphere and the stratosphere
  - (e) Forget it! Signals in the shortwave band cannot propagate beyond a line of sight without satellites, repeaters, or other human-made systems.
10. What is the purpose of the filament in a vacuum tube?
- (a) It prevents secondary electrons from causing excessive screen grid current.
  - (b) It causes holes to flow more easily in the collector.
  - (c) It heats the cathode, thereby increasing the electron emission.
  - (d) It glows, making the tube easy to locate when servicing is necessary.
  - (e) Forget it! Vacuum tubes do not have filaments.
11. Figure Test4-1 is a schematic diagram of

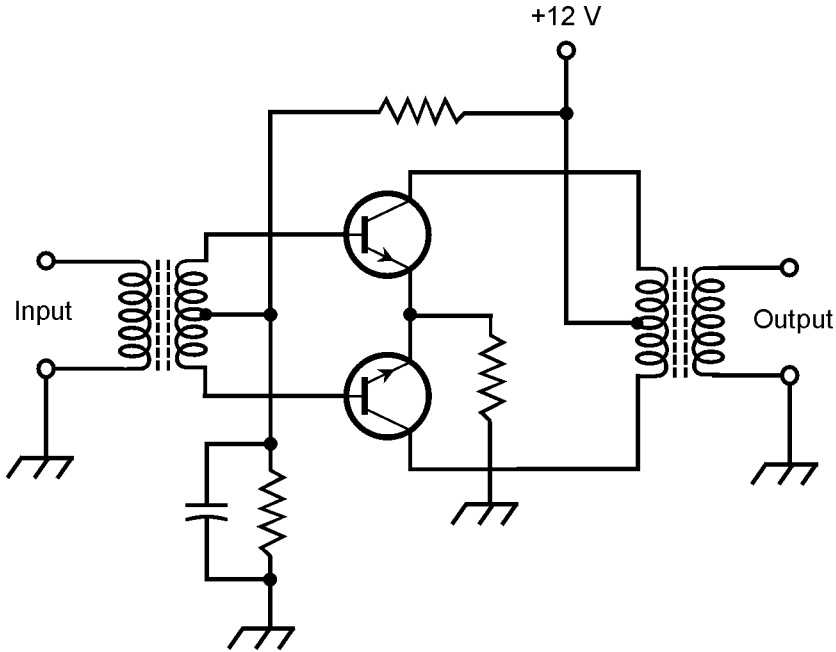


**Test 4-1** Illustration for Part 4 Test Questions 11 and 12.

- (a) a closed-loop op amp circuit with positive feedback.
  - (b) a closed-loop op amp circuit with negative feedback.
  - (c) an open-loop op amp circuit.
  - (d) an op amp oscillator circuit.
  - (e) an op amp modulator circuit.
12. Suppose that, in the circuit of Fig. Test4-1, a resistor is connected between the output and the inverting input. What effect will this have?
- (a) It will increase the gain.
  - (b) It will decrease the gain.
  - (c) It will increase the frequency.
  - (d) It will decrease the frequency.
  - (e) It will have no effect.
13. Which of the following is an example of a serial-access storage medium for computer data?
- (a) An external hard drive
  - (b) An internal hard drive

- (c) A CD-R
  - (d) A CD-RW
  - (e) A magnetic tape
14. Which of the following is an undesirable characteristic of ELF radiation?
- (a) It can cause radiation burns like those produced by overexposure to X rays.
  - (b) It can increase the susceptibility of a CRT to jitter and poor resolution.
  - (c) It can reduce the efficiency of an antenna system.
  - (d) It can result in undesirable modulation in a radio transmitter.
  - (e) None of the above apply.
15. A zepp antenna measuring  $\lambda/2$  can be oriented vertically, and the feed line placed so it lies in the same line as the radiating element. This antenna is known as
- (a) a vertical dipole.
  - (b) a Yagi.
  - (c) a collinear array.
  - (d) an end-fire array.
  - (e) a J pole.
16. The GPS can be used to determine the position of a point in
- (a) one dimension.
  - (b) two dimensions.
  - (c) three dimensions.
  - (d) four dimensions.
  - (e) five dimensions.
17. In a component-type hi-fi system, the cables connecting the various devices should
- (a) be resonant at the intended operating frequency.
  - (b) consist of single conductors only.
  - (c) be connected in parallel.
  - (d) consist of two parallel wires whenever possible.
  - (e) be shielded whenever practicable.
18. In a radar display, the azimuth of a target is equal or proportional to
- (a) the distance of the blip representing the target from the center of the display.
  - (b) the clockwise angle between a radial line from the center toward geographic north, and a radial line from the center through the blip representing the target.
  - (c) the diameter of a circle whose center coincides with the center of the display, and that passes through the blip representing the target.

- (d) the distance of the blip representing the target from the outer edge of the display, or from a circle whose diameter is equal to the diameter of the display.
  - (e) Forget it! A radar set cannot determine the azimuth of a target.
19. Which of the following is an example of wireless technology?
- (a) The Global Positioning System
  - (b) A fiber-optic communications system
  - (c) Digital-to-analog conversion
  - (d) Serial-to-parallel conversion
  - (e) All of the above
20. A smoke detector can function by sensing a change in the
- (a) temperature of the air.
  - (b) relative humidity of the air.
  - (c) barometric pressure of the air.
  - (d) ionization potential in the air.
  - (e) amount of nitrogen in the air.
21. Computer disk storage is typically measured in
- (a) kilobits, megabits, and gigabits.
  - (b) kilobits per second, megabits per second, and gigabits per second.
  - (c) kilobytes, megabytes, and gigabytes.
  - (d) kilobytes per second, megabytes per second, and gigabytes per second.
  - (e) any of the above.
22. Which of the following antennas would most likely be used in space communications?
- (a) A half-wave dipole antenna
  - (b) A ground-plane antenna
  - (c) A helical antenna
  - (d) A small loop or loopstick antenna
  - (e) A coaxial antenna
23. The use of a cell phone may be prohibited in a commercial aircraft in flight because
- (a) it can interfere with other people's computers.
  - (b) it can interfere with other people's cell phones.
  - (c) the interior of the aircraft produces a dangerous RF resonant cavity.
  - (d) the area of coverage will not be wide enough to be of any use.
  - (e) it can interfere with flight instruments and communications.
24. The circuit shown in Fig. Test4-2 shows two NPN transistors connected in
- (a) reverse series.
  - (b) cascade.



**Test 4-2** Illustration for Part 4 Test Questions 24 and 25.

- (c) inverse parallel.
  - (d) push-push.
  - (e) push-pull.
25. The circuit in Fig. Test4-2 would be a bad choice for use as a medium-power audio amplifier because
- (a) bipolar transistors do not work well in audio applications.
  - (b) transformers should not be used in audio amplifiers.
  - (c) this type of circuit invariably produces a lot of distortion.
  - (d) this type of circuit cannot produce appreciable output power.
  - (e) Hold it! The circuit in Fig. Test4-2 would be a good choice for use as a medium-power audio amplifier.
26. In an electrostatic CRT, the electron beam can be diverted up and down by
- (a) synchronization pulses applied to the deflecting coils.
  - (b) ELF radiation from the accelerating anodes.
  - (c) charged plates that attract or repel moving electrons.
  - (d) modulation of the signal on the control grid.
  - (e) variable input to the electron gun.

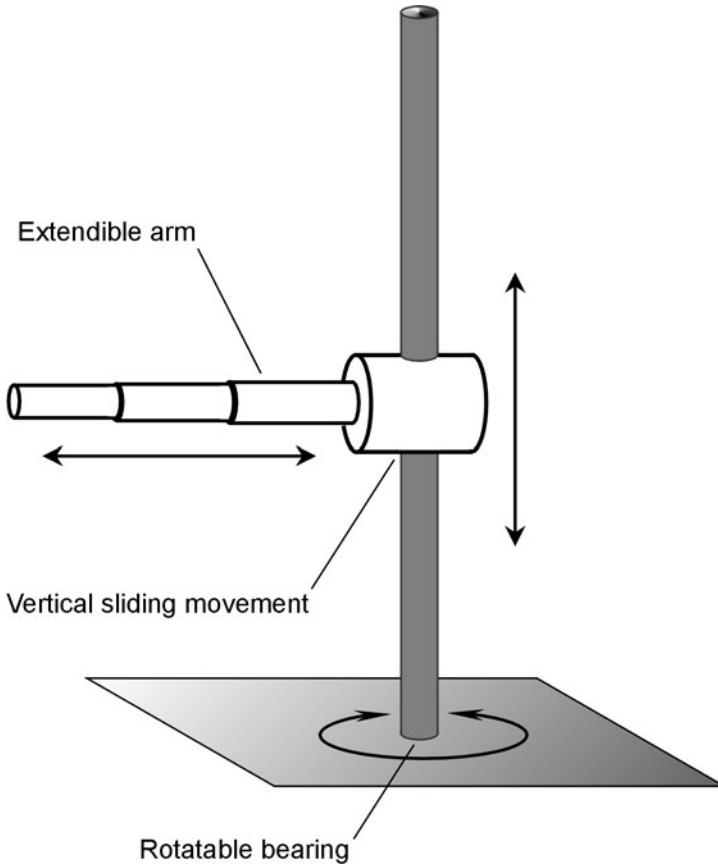
27. A device that translates 500-Hz sound waves into ac by means of the interaction between a permanent magnet and a coil of wire is an example of
- (a) an electromechanical loudspeaker.
  - (b) a passive pressure sensor.
  - (c) a dynamic transducer.
  - (d) an inductive proximity sensor.
  - (e) a piezoelectric transducer.
28. Which of the following considerations is most important in a hi-fi audio power amplifier?
- (a) Optimizing the efficiency
  - (b) Minimizing the operating voltage
  - (c) Maximizing the linearity
  - (d) Maximizing the driving power
  - (e) Biasing beyond the cutoff point
29. A loopstick antenna, oriented so its ends point toward the eastern and western horizons, exhibits
- (a) nulls in the response to signals coming from the east and west.
  - (b) peaks in the response to signals coming from the east and west.
  - (c) nulls in the response to signals coming from the north and south.
  - (d) an omnidirectional response in three-dimensional space.
  - (e) a peak in the response to signals coming from any horizontal direction, and a null in the response to signals coming from the zenith (directly overhead).
30. Fill in the blank to make the following sentence true: “An \_\_\_\_\_ functions by sensing variations in the relative phase of acoustic waves reflected from objects in the environment.”
- (a) IR transducer
  - (b) acoustic transponder
  - (c) electrostatic loudspeaker
  - (d) ultrasonic motion detector
  - (e) acoustic pressure sensor
31. Which of the following is an advantage of a large LCD over a large CRT display?
- (a) Lower cost
  - (b) Lighter weight
  - (c) Greater ELF radiation
  - (d) The use of electron beams and deflecting coils
  - (e) All of the above
32. An op amp can produce or facilitate
- (a) a low signal-to-noise ratio in an oscillator.
  - (b) improved efficiency in an antenna system.



- (c) high signal gain over a wide range of frequencies.
  - (d) a match between a feed line and an antenna.
  - (e) none of the above.
33. Which of the following is an application of a gas-filled tube?
- (a) A high-frequency RF power amplifier
  - (b) A microwave oscillator
  - (c) An audio oscillator or amplifier
  - (d) Decorative lighting
  - (e) Impedance matching in an antenna system
34. For a communications satellite with an elliptical orbit around the earth, the point at which the altitude is lowest is the
- (a) geominimum.
  - (b) proxima.
  - (c) perigee.
  - (d) approach.
  - (e) optimum.
35. Suppose an antenna has a radiation resistance of  $35 \Omega$  and its feed line has a characteristic impedance of  $50 \Omega$ . What is the efficiency of the antenna?
- (a) 15 percent
  - (b) 41 percent
  - (c) 59 percent
  - (d) 70 percent
  - (e) Impossible to calculate from this information
36. Which of the following is a characteristic of a stepper motor?
- (a) It rotates in defined increments, not continuously.
  - (b) Its turning power increases as its speed increases.
  - (c) It cannot rotate faster than approximately 1 rpm.
  - (d) It is not suitable for use in mechanical devices, but only in sensors.
  - (e) It operates without an external source of power.
37. With respect to ICs, terms such as MSI, LSI, and VLSI define the
- (a) maximum number of transistors on the chip.
  - (b) maximum frequency at which the chip can operate.
  - (c) maximum power output the chip can produce.
  - (d) maximum gain the chip can produce.
  - (e) maximum number of switching operations the chip can perform per second.

38. In a tetrode vacuum tube amplifier, the screen grid
- (a) reduces the capacitance between the control grid and the plate.
  - (b) serves as an auxiliary output in case the plate circuit is overloaded.
  - (c) conducts holes away from the control grid.
  - (d) prevents undesirable reverse bias between the cathode and the control grid.
  - (e) Forget it! Tetrode tubes do not have screen grids.
39. Fill in the blank in the following sentence to make it true: "On the Web, data is transmitted by means of a \_\_\_\_ that allows the files on distant computers to appear as if they are on your own computer."
- (a) protocol
  - (b) memory chip
  - (c) microprocessor
  - (d) CD-R
  - (e) tape drive
40. A robot arm with articulated geometry
- (a) has only one degree of freedom.
  - (b) has joints, similar to those in a human arm.
  - (c) can function in only two dimensions.
  - (d) can rotate and move up and down, but cannot bend.
  - (e) can perform only coarse movements.
41. The instantaneous output voltage of a differentiator is proportional to
- (a) the peak instantaneous input voltage.
  - (b) the rate at which the instantaneous input voltage changes.
  - (c) the accumulated input voltage as a function of time.
  - (d) the difference between the instantaneous voltages of the two input signals.
  - (e) the sum of the instantaneous voltages of the two input signals.
42. What types of devices require particular care in handling, so they are not destroyed by electrostatic discharges that can build up on a technician's body?
- (a) Vacuum tubes
  - (b) Ferrite resistors
  - (c) Ceramic capacitors
  - (d) MOS components
  - (e) Rectifier diodes
43. According to Isaac Asimov's rules for the behavior of robots, a robot must not injure, or allow the injury of, any human,
- (a) unless the robot is forced to prevent its own destruction.
  - (b) unless the human is committing a crime.

- (c) unless the human is trying to turn off the robot's power supply.
  - (d) unless it is the result of an accident.
  - (e) under any circumstances.
44. Computer data speed is typically measured in
- (a) kilobits, megabits, and gigabits.
  - (b) kilobits per second, megabits per second, and gigabits per second.
  - (c) kilobytes, megabytes, and gigabytes.
  - (d) kilobytes per second, megabytes per second, and gigabytes per second.
  - (e) any of the above.
45. Doppler radar is useful for measuring or estimating
- (a) the frequency of a sine wave that is modulated with a complex signal.
  - (b) the wind speed in the funnel cloud of a tornado.
  - (c) the distance between a robot gripper and a tool or object to be manipulated.
  - (d) the depth of the ocean at a particular location.
  - (e) the direction from which a thunderstorm is coming.
46. A device that translates a 30-kHz ac voltage into ultrasonic waves by producing stress on a crystal, thereby causing the crystal to vibrate, is an example of
- (a) an electromechanical loudspeaker.
  - (b) a passive pressure sensor.
  - (c) a dynamic transducer.
  - (d) an inductive proximity sensor.
  - (e) a piezoelectric transducer.
47. Fill in the blank in the following sentence to make it true: "A \_\_\_\_ antenna is a vertical radiator, usually measuring  $\lambda/4$ , elevated above the surface of the earth, and operated against a system of  $\lambda/4$  horizontal or slightly drooping radials."
- (a) dipole
  - (b) coaxial
  - (c) collinear
  - (d) ground-plane
  - (e) vertizontal
48. Figure Test4-3 shows a robotic system that employs
- (a) Cartesian coordinate geometry.
  - (b) spherical coordinate geometry.
  - (c) polar coordinate geometry.
  - (d) cylindrical coordinate geometry.
  - (e) Riemannian coordinate geometry.



**Test 4-3** Illustration for Part 4 Test Question 48.

49. A satellite in a LEO system is usually placed in
- (a) an equatorial orbit.
  - (b) a geosynchronous orbit.
  - (c) a retrograde orbit.
  - (d) a polar orbit.
  - (e) an orbit midway between the earth and the orbit of the moon.
50. Sound waves in the atmosphere consist of
- (a) rotating magnetic fields.
  - (b) alternating electric fields.
  - (c) variable electron beams.
  - (d) moving molecules.
  - (e) rotating electron orbits within atoms.

# Final Exam

DO NOT REFER TO THE TEXT WHEN TAKING THIS EXAM. A GOOD SCORE IS AT LEAST 75 CORRECT OUT OF 100. Answers are in the back of the book. It's best to have a friend check your score the first time, so you won't memorize the answers if you want to take the exam again.

1. A wire can be coiled around a magnetic compass to make
  - (a) an ac wattmeter.
  - (b) an ac voltmeter.
  - (c) a dc wattmeter.
  - (d) a dc ohmmeter.
  - (e) a dc galvanometer.
  
2. The inductive reactance of a fixed air-core coil of wire, assuming the wire has zero dc resistance (it conducts perfectly),
  - (a) does not change as the frequency changes.
  - (b) increases as the frequency increases.
  - (c) decreases as the frequency increases.
  - (d) alternately increases and decreases as the frequency increases.
  - (e) is zero at all frequencies.
  
3. *Power* is defined as
  - (a) the rate at which current flows in a circuit.
  - (b) the product of voltage and resistance in a circuit.
  - (c) the rate at which energy is radiated or dissipated.
  - (d) the accumulation of energy over time.
  - (e) the amount of heat generated in a circuit.

4. In an amplifier that employs a P-channel JFET, the device can usually be replaced with an N-channel JFET having similar specifications, provided that
- (a) all the resistors are reversed in polarity for the circuit in question.
  - (b) the power supply polarity is reversed for the circuit in question.
  - (c) the drain, rather than the source, is placed at signal ground.
  - (d) the input is supplied to the drain, rather than to the gate.
  - (e) the output is taken from the source, rather than from the drain.
5. At the exact moment a 60-Hz ac sine wave is at its positive peak voltage, the instantaneous rate of change in the voltage is
- (a) large and positive.
  - (b) small and positive.
  - (c) large and negative.
  - (d) small and negative.
  - (e) zero.
6. In order for a bipolar transistor to conduct under conditions of no signal input, the bias must be
- (a) in the forward direction at the emitter-base (E-B) junction, sufficient to cause forward breakover.
  - (b) in the reverse direction at the E-B junction, but not sufficient to cause avalanche effect.
  - (c) such that the application of a signal would cause the transistor to go into a state of cutoff.
  - (d) such that the application of a signal would cause the transistor to go into a state of saturation.
  - (e) such that the application of a signal would cause the transistor to become nonlinear.
7. The risk of electrocution to personnel who service “live” electrical or electronic equipment increases in proportion to the
- (a) maximum amount of power, in watts, that the equipment demands.
  - (b) maximum amount of energy, in watt-hours, that the equipment consumes.
  - (c) maximum amount of current, in amperes, that the equipment draws.
  - (d) maximum potential difference, in volts, that exists in the equipment.
  - (e) maximum resistance, in ohms, that exists in the equipment.
8. In a logical AND gate with four inputs, the output is high if and only if
- (a) none of the inputs are high, and all four are low.
  - (b) one of the inputs is high, and the other three are low.
  - (c) two of the inputs are high, and the other two are low.
  - (d) three of the inputs are high, and the other one is low.
  - (e) all four of the inputs are high, and none are low.
9. Fill in the blank in the following sentence to make it true: “A \_\_\_\_ receiver derives its audio output by mixing incoming signals with the output of a tunable local oscillator.”
- (a) crystal set
  - (b) direct conversion

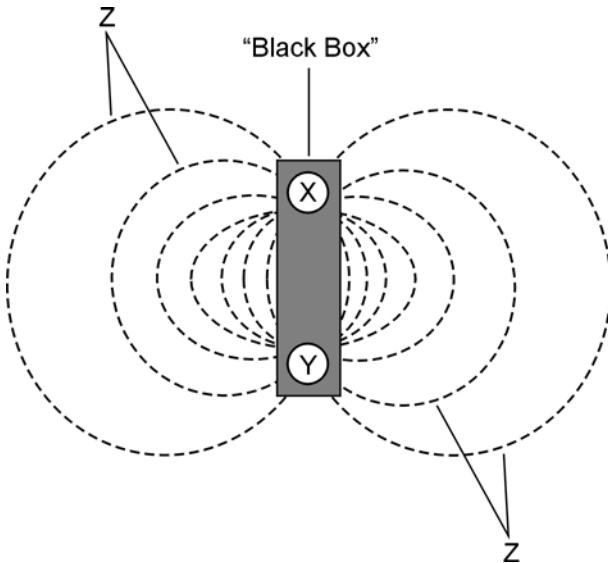
- (c) phase-locked
  - (d) very low frequency
  - (e) frequency modulation
10. In a power supply filter, the term *L section* refers to a combination of
- (a) a capacitor in parallel with the rectifier output, and a choke in series.
  - (b) a capacitor in parallel with the rectifier output, and another capacitor in series.
  - (c) a choke in parallel with the rectifier output, and a capacitor in series.
  - (d) a diode in parallel with the transformer output, and another diode in series.
  - (e) any of the above.
11. What is the frequency of a pure sine wave that has a period of 50 nanoseconds? (A nanosecond is equal to one thousand-millionth, or  $10^{-9}$ , of a second.)
- (a) 2.0 MHz
  - (b) 20 MHz
  - (c) 0.20 GHz
  - (d) 2.0 GHz
  - (e) 20 GHz
12. Fill in the blank in the following sentence to make it correct: "The peak inverse voltage rating of a rectifier diode is the maximum instantaneous reverse-bias voltage that it can withstand without \_\_\_\_ taking place."
- (a) forward breakover
  - (b) avalanche effect
  - (c) harmonic generation
  - (d) oscillation
  - (e) thermal destruction
13. Suppose a coil and capacitor are connected in series, with  $jX_L = j50$  and  $jX_C = -j90$ , and there is no resistance. What is the complex impedance of this combination?
- (a)  $0 - j40$
  - (b)  $0 + j40$
  - (c)  $40 + j50$
  - (d)  $50 - j90$
  - (e) It cannot be determined without more information.
14. In an IC, the term *linear* arises from the fact that
- (a) all of the etchings on the semiconductor material are straight lines.
  - (b) the amplification factor generally decreases in direct proportion to the instantaneous input amplitude.
  - (c) the amplification factor generally increases in direct proportion to the instantaneous input amplitude.
  - (d) the amplification factor is generally constant as the instantaneous input amplitude varies.
  - (e) the device never exhibits distortion under any circumstances.

15. Suppose a dc circuit has  $10\text{ k}\Omega$  of resistance, and a 12-V battery is connected to it. How much current is drawn from the battery?

- (a)  $1.2\ \mu\text{A}$
- (b)  $12\ \mu\text{A}$
- (c)  $1.2\ \text{mA}$
- (d)  $12\ \text{mA}$
- (e) None of the above

16. The high input impedance of a MOSFET makes this type of device ideal for use in

- (a) weak-signal amplifiers.
- (b) high-power oscillators.
- (c) high-current rectifiers.
- (d) antenna tuning networks.
- (e) graphic equalizers.



**Exam-1** Illustration for Final Exam Question 17.

17. In Fig. Exam-1, the dashed lines labeled (Z) can represent

- (a) the electric lines of flux in the vicinity of two electrically charged objects (X and Y inside the black box) having opposite polarity.
- (b) the magnetic lines of flux in the vicinity of two electrically charged objects (X and Y inside the black box) having opposite polarity.
- (c) the electric lines of flux in the vicinity of two electrically charged objects (X and Y inside the black box) having the same polarity.
- (d) the magnetic lines of flux in the vicinity of two electrically charged objects (X and Y inside the black box) having the same polarity.
- (e) any of the above.



18. The drain of a JFET is the analog of the
- (a) plate of a vacuum tube.
  - (b) emitter of a bipolar transistor.
  - (c) cathode of a diode.
  - (d) positive electrode in a solar cell.
  - (e) substrate of a MOSFET.
19. Imagine a circuit with  $21\ \Omega$  of inductive reactance and  $28\ \Omega$  of resistance. What is the absolute-value impedance of this circuit?
- (a)  $7\ \Omega$
  - (b)  $24\ \Omega$
  - (c)  $35\ \Omega$
  - (d)  $49\ \Omega$
  - (e) None of the above
20. Suppose you want to obtain a 12-V battery by connecting 1.5-V size D flashlight cells together. This can be done by
- (a) connecting 8 cells together in parallel, making sure to always connect the terminals so they are plus-to-plus and minus-to-minus.
  - (b) connecting 8 cells together in series, making sure to always connect the terminals so they are plus-to-minus.
  - (c) either connecting 8 cells together in parallel as described in (a), or connecting 8 cells together in series as described in (b).
  - (d) connecting two sets of 4 cells together in series, making sure to always connect the terminals so they are plus-to-minus, and then connecting these sets together in parallel, making sure to connect the terminals so they are plus-to-plus and minus-to-minus.
  - (e) connecting two sets of 4 cells together in parallel, making sure to always connect the terminals so they are plus-to-plus and minus-to-minus, and then connecting these sets together in series, making sure to connect the terminals so they are plus-to-minus.
21. One of the technical limitations of capacitive proximity sensors is the fact that they
- (a) are not very sensitive to objects that are poor electrical conductors.
  - (b) are insensitive to objects that reflect light.
  - (c) are insensitive to metallic objects.
  - (d) cannot be used with oscillators.
  - (e) require extreme voltages in order to function properly.
22. A flute sounds different than a violin, even if the two instruments are played at the same pitch, because of a difference in the
- (a) phase.
  - (b) chamber length.
  - (c) frequency.
  - (d) waveform.
  - (e) bias.

23. In a personal computer, the internal mass-storage medium is nearly always
- (a) a CD (compact disc) in some form.
  - (b) a diskette drive.
  - (c) a tape drive.
  - (d) read-only memory (ROM).
  - (e) none of the above.
24. A device to which signals can be sent asking for data, and that responds with the requested information by means of radio or a wireless network, is known as a
- (a) transceiver.
  - (b) transponder.
  - (c) repeater.
  - (d) duplexer.
  - (e) signal reflector.
25. Suppose you want to use a transformer to obtain exactly 30 V rms ac from an electrical outlet that supplies exactly 120 V rms ac. The primary-to-secondary turns ratio of the transformer should be
- (a) exactly 1:16.
  - (b) exactly 1:4.
  - (c) variable.
  - (d) exactly 4:1.
  - (e) exactly 16:1.
26. Fill in the blank in the following statement to make it true: “Two problems can occur with the \_\_\_\_ type of intrusion alarm. First, the sun, suddenly emerging from clouds or from shadow, might shine on the sensor and trigger the device when there is no threat. Second, someone with malicious intent, wearing winter clothing and entering on a cold day, might fail to set off the alarm in the presence of a real threat.”
- (a) IR radiation detector
  - (b) electromagnetic detector
  - (c) ultrasonic detector
  - (d) ultraviolet detector
  - (e) fluxgate detector
27. Suppose that five resistors, each one having the same ohmic value as all the others, are connected in series, and the entire combination is connected to a 12-V battery. How does the voltage across any one of the resistors compare with the battery voltage, assuming no other loads are connected to the circuit?
- (a) The voltage across any one of the resistors is 5 times the battery voltage.
  - (b) The voltage across any one of the resistors is the same as the battery voltage.
  - (c) The voltage across any one of the resistors is  $\frac{1}{5}$  of the battery voltage.
  - (d) The voltage across any one of the resistors is  $\frac{1}{25}$  of the battery voltage.
  - (e) It is impossible to answer this without more information.

28. When the carrier and one of the sidebands are removed from an amplitude-modulated signal, the result is
- (a) a frequency-modulated signal.
  - (b) a phase-modulated signal.
  - (c) a frequency-shift-keyed signal.
  - (d) a pulse-modulated signal.
  - (e) a single-sideband signal.
29. In log polar navigation, a computer converts a polar coordinate image to
- (a) a Riemannian coordinate image.
  - (b) a spherical coordinate image.
  - (c) a cylindrical coordinate image.
  - (d) a curvilinear coordinate image.
  - (e) a rectangular coordinate image.
30. The charged particles in the nucleus of an atom are
- (a) electrons.
  - (b) protons.
  - (c) positrons.
  - (d) neutrons.
  - (e) negatrons.
31. Which of the following functions (a), (b), (c), or (d), if any, is a semiconductor diode never used to perform?
- (a) Rectification
  - (b) Detection
  - (c) Frequency multiplication
  - (d) Signal mixing
  - (e) A semiconductor diode can perform any of the above functions.
32. The reactance of a fixed capacitor, assuming it has no leakage conductance (that is, it has a theoretically infinite dc resistance),
- (a) does not change as the frequency changes.
  - (b) increases negatively as the frequency increases.
  - (c) decreases negatively (approaches zero) as the frequency increases.
  - (d) alternately becomes negative and positive as the frequency increases.
  - (e) is theoretically infinite at all frequencies.
33. What is the dc resistance of a component with a dc conductance of  $10^{-4}$  S?
- (a)  $0.0001 \Omega$
  - (b)  $0.01 \Omega$
  - (c)  $100 \Omega$
  - (d)  $10 \text{ k}\Omega$
  - (e) None of the above

34. A common-collector transistor circuit is often used
- (a) to provide high gain and sensitivity over a wide range of frequencies.
  - (b) to match a high impedance to a low impedance.
  - (c) as a high-fidelity audio power amplifier.
  - (d) as an oscillator at microwave frequencies.
  - (e) as the rectifier in a dc power supply.
35. The internal conductance (expressed in siemens) of an ammeter is generally
- (a) low.
  - (b) directly proportional to the current.
  - (c) inversely proportional to the current.
  - (d) high.
  - (e) any value; it doesn't matter.
36. A half-wave length of wire, fed at one end with a quarter-wave section of parallel-wire transmission line, is known as
- (a) a zeppelin antenna.
  - (b) a dipole antenna.
  - (c) a Yagi antenna.
  - (d) a random wire antenna.
  - (e) an end-fire antenna.
37. A twin T oscillator is commonly used for generating
- (a) AF signals.
  - (b) high-frequency RF signals.
  - (c) microwave RF signals.
  - (d) powerful bursts of RF energy.
  - (e) FM signals.
38. Suppose that five resistors, each one having the same ohmic value as all the others, are connected in series, and the entire combination is connected to a 12-V battery. How does the current through any one of the resistors compare with the current drawn from the battery?
- (a) The current through any one of the resistors is 5 times the current drawn from the battery.
  - (b) The current through any one of the resistors is the same as the current drawn from the battery.
  - (c) The current through any one of the resistors is  $\frac{1}{5}$  as great as the current drawn from the battery.
  - (d) The current through any one of the resistors is  $\frac{1}{25}$  as great as the current drawn from the battery.
  - (e) It is impossible to answer this without more information.

39. When a nonzero net reactance exists in an ac circuit, the volt-ampere (VA) power, also known as the *apparent power*, is

- (a) zero.
- (b) equal to the true power.
- (c) less than the true power.
- (d) greater than the true power.
- (e) infinite.

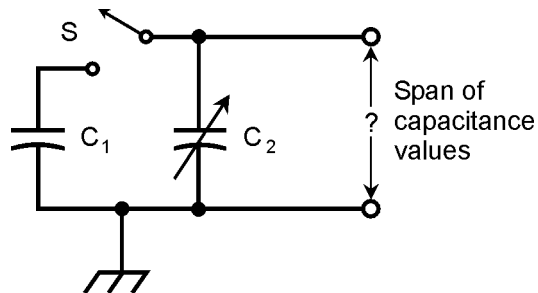
40. When an impurity, also called a *dopant*, is added to a semiconductor material, it can contain either an excess of electrons or a deficiency of electrons. A dopant with an excess of electrons, also known as a *donor impurity*, results in the production of

- (a) an NPN transistor.
- (b) a PNP transistor.
- (c) a P-N junction.
- (d) a P type semiconductor.
- (e) an N type semiconductor.

41. Suppose an audio amplifier produces 100 W rms output when the input is 5.00 W rms. This represents a power gain of

- (a) 1.30 dB.
- (b) 2.60 dB.
- (c) 13.0 dB.
- (d) 20.0 dB.
- (e) 26.0 dB.

**Exam-2** Illustration for  
Final Exam  
Questions 42  
and 43.



42. In Fig. Exam-2, suppose the value of the fixed capacitor,  $C_1$ , is 330 pF, and the range of the variable capacitor,  $C_2$ , is 10 to 365 pF. What is the span of capacitance values (minimum to maximum) that can be obtained with this parallel combination, with switch S either open or closed, as desired?

- (a) 10 to 365 pF, over a continuous range
- (b) 10 to 181 pF, but with a gap in the range
- (c) 10 to 181 pF, over a continuous range

- (d) 10 to 695 pF, but with a gap in the range
- (e) 10 to 695 pF, over a continuous range

43. In Fig. Exam-2, suppose the value of the fixed capacitor,  $C_1$ , is 220 pF, and the range of the variable capacitor,  $C_2$ , is 5 to 100 pF. What is the span of capacitance values (minimum to maximum) that can be obtained with this parallel combination, with switch S either open or closed, as desired?

- (a) 5 to 100 pF, over a continuous range
- (b) 5 to 320 pF, but with a gap in the range
- (c) 5 to 320 pF, over a continuous range
- (d) 5 to 69 pF, but with a gap in the range
- (e) 5 to 69 pF, over a continuous range

44. Which of the following units is most often used for specifying data speed in digital communications systems?

- (a) Megabytes per second
- (b) Megabauds per second
- (c) Megabits per second
- (d) Megahertz per second
- (e) Any of the above units are used equally often when specifying data speed in digital communications systems.

45. Imagine four 100- $\mu$ H inductors connected in a  $2 \times 2$  series-parallel combination. Suppose there is no mutual inductance among them. What is the net inductance of this matrix?

- (a) 25  $\mu$ H
- (b) 50  $\mu$ H
- (c) 100  $\mu$ H
- (d) 200  $\mu$ H
- (e) 400  $\mu$ H

46. In the output of a dc power supply, voltage regulation can be obtained by

- (a) connecting a Zener diode in series, paying attention to the polarity.
- (b) connecting a choke in series and an electrolytic capacitor in parallel, paying attention to the polarity.
- (c) connecting a capacitor in parallel and a Zener diode in series, paying attention to the polarity.
- (d) connecting a Zener diode in parallel, paying attention to the polarity.
- (e) connecting a resistor in parallel and another resistor in series.

47. Imagine an inductor and resistor connected in series, such that the inductive reactance and the resistance are both equal to 400  $\Omega$  at a frequency of 100 MHz. How are the instantaneous current and the instantaneous voltage related?

- (a) They are in phase.
- (b) They are 180° out of phase.

- (c) The current leads the voltage by  $90^\circ$ .
- (d) The current lags the voltage by  $90^\circ$ .
- (e) None of the above are true.

48. Suppose you want to use a transformer to match the output of an audio amplifier to a speaker. The amplifier has a purely resistive output impedance of exactly  $128\ \Omega$ . The speaker has a purely resistive impedance of exactly  $8\ \Omega$ . The primary-to-secondary turns ratio of the transformer should be

- (a) exactly 1:16.
- (b) exactly 1:4.
- (c) variable.
- (d) exactly 4:1.
- (e) exactly 16:1.

49. Imagine a string of 20 holiday lights, all connected in parallel and plugged into a 120-V rms ac wall outlet. If one of the bulbs is short-circuited, what will happen?

- (a) All of the remaining 19 bulbs will shine a little more brightly than before.
- (b) All of the remaining 19 bulbs will shine a little less brightly than before.
- (c) The fuse or circuit breaker will blow, and all of the remaining 19 bulbs will go out.
- (d) Most of all of the remaining 19 bulbs will burn out, and some may explode.
- (e) The remaining 19 bulbs will continue to shine exactly as before—no more or less brightly.

50. Suppose a coil and capacitor are connected in series, with  $jX_L = j40$  and  $jX_C = -j70$ , and the coil has an internal resistance of  $10\ \Omega$ . Suppose the frequency of operation is 12.5 MHz. This circuit will exhibit resonance at

- (a) some frequency below 12.5 MHz.
- (b) some frequency above 12.5 MHz.
- (c) 12.5 MHz; it is resonant under the conditions stated.
- (d) any and all frequencies.
- (e) no frequency, because of the internal resistance.

51. Imagine a transmission line consisting of two wires, each having a diameter of exactly 1 mm, and uniformly spaced exactly 10 mm apart with nothing but air as the dielectric between them. If the two wires are moved so they are uniformly spaced exactly 50 mm apart, and the dielectric between them is still nothing but air, what happens to the characteristic impedance?

- (a) It does not change.
- (b) It increases.
- (c) It decreases.
- (d) It is impossible to say without knowing what is connected to the line.
- (e) It is impossible to say without knowing the frequency.

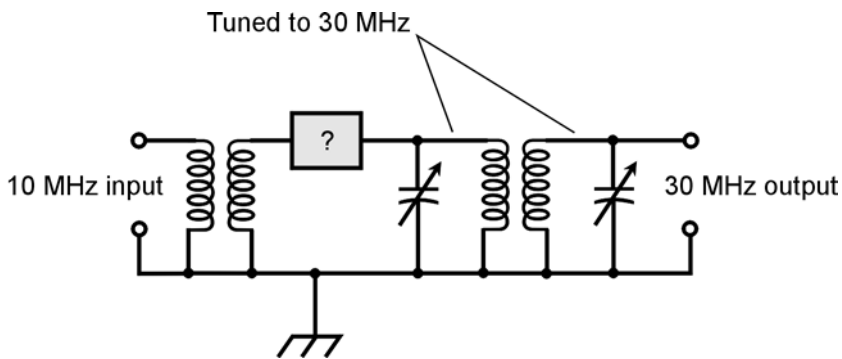
52. In an FM signal, deviation is

- (a) the difference between the highest instantaneous carrier frequency and the lowest instantaneous carrier frequency.
- (b) the maximum extent to which the instantaneous carrier frequency differs from the unmodulated-carrier frequency.

- (c) the bandwidth of the signal, expressed in hertz, kilohertz, or megahertz.
  - (d) the rate at which the phase of the signal varies, expressed in hertz per second, kilohertz per millisecond, or megahertz per microsecond.
  - (e) None of the above is true.
53. A typical angular speed for an old-fashioned vinyl disk in a hi-fi sound system is
- (a) 12 rpm.
  - (b) 24 rpm.
  - (c) 33 rpm.
  - (d) 60 rpm.
  - (e) 90 rpm.
54. An autotransformer can be recognized by the fact that it has
- (a) a continuously adjustable turns ratio.
  - (b) a solenoidal core that can be moved in and out of a coil.
  - (c) a single, tapped winding.
  - (d) a special set of terminals for use with an automotive battery.
  - (e) a toroidal core.
55. One of the purposes of the suppressor grid in a pentode vacuum tube amplifier is to
- (a) keep the cathode from emitting too many electrons, which could drive the control grid current to excessive levels.
  - (b) ensure that the plate is maintained at a constant negative voltage, so electrons will not be attracted to it.
  - (c) prevent the screen grid from shorting out to the plate, which could cause the power supply to burn out.
  - (d) repel secondary electrons emitted by the plate, keeping the screen grid current from getting too high.
  - (e) prevent distortion caused by overdrive in the control grid, screen grid, or cathode circuits.
56. Consider two sine waves having identical frequency. Suppose that one wave leads the other by exactly  $\frac{1}{20}$  of a cycle. What is the phase difference between the two waves?
- (a)  $9^\circ$
  - (b)  $12^\circ$
  - (c)  $18^\circ$
  - (d)  $20^\circ$
  - (e)  $24^\circ$
57. In a cellular communications system, a technology in which cells overlap, but signals do not interfere with each other because every phone set is assigned a unique signal code, is known as
- (a) code-division multiple access.
  - (b) diversity reception.
  - (c) a phase-locked loop.



- (d) analog signal mixing.  
 (e) digital diversity encryption.
58. A steady magnetic field can be produced by
- a straight wire carrying a constant direct current.
  - a loop of wire carrying a constant direct current.
  - a coil of wire carrying a constant direct current.
  - a constant-intensity stream of protons in free space, moving in a straight line.
  - any of the above.



**Exam-3** Illustration for Final Exam Question 59.

59. In Fig. Exam-3, what type of component can the box with the query symbol (?) contain in order for the circuit to function as indicated?
- A capacitor
  - A resistor
  - A battery
  - An inductor
  - A diode
60. The binary number 1001101 represents which decimal number?
- 44
  - 55
  - 66
  - 77
  - 88
61. Suppose a dc circuit has  $10\text{ k}\Omega$  of resistance, and a 12-V battery is connected to it. How much power is drawn from the battery?
- $1.2\ \mu\text{W}$
  - $12\ \mu\text{W}$

- (c) 1.2 mW
- (d) 12 mW
- (e) None of the above

62. In which of the following applications would you never find an op amp used as the main active circuit component?

- (a) A differentiator
- (b) An integrator
- (c) A highpass audio filter
- (d) A lowpass audio filter
- (e) A microwave oscillator

63. Imagine a circuit with  $21 \Omega$  of inductive reactance and  $28 \Omega$  of resistance. What is the complex impedance of this circuit?

- (a)  $7 \Omega$
- (b)  $24 \Omega$
- (c)  $35 \Omega$
- (d)  $49 \Omega$
- (e) None of the above

64. In a circuit containing conductance, inductive susceptance, and capacitive susceptance, a condition of resonance exists if and only if

- (a) the conductance is zero.
- (b) the conductance is infinite.
- (c) the inductive susceptance is infinite and the capacitive susceptance is zero.
- (d) the inductive and capacitive susceptances are both zero.
- (e) the inductive and capacitive susceptances cancel out.

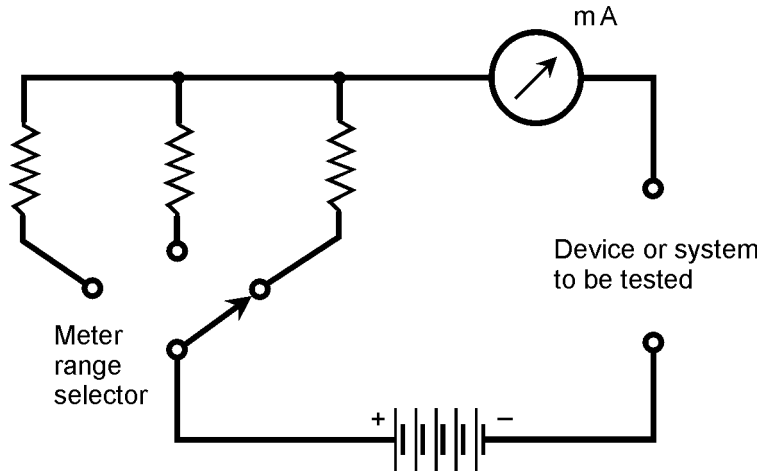
65. Imagine a perfect square wave with no dc component. That means the positive and negative portions have equal amplitudes and equal durations. Suppose you are told that the voltage is 10.0 V pk-pk. What is the average voltage?

- (a) 0.00 V
- (b) 5.00 V
- (c) 7.07 V
- (d) 10.0 V
- (e) It is impossible to calculate this without more information.

66. Fill in the blank in the following statement to make it true: "It takes a certain minimum \_\_\_\_\_ voltage, known as the *forward breakover voltage*, for conduction to occur through a P-N junction."

- (a) RF
- (b) AF
- (c) forward bias
- (d) negative
- (e) positive

**Exam-4** Illustration for  
Final Exam  
Question 67.



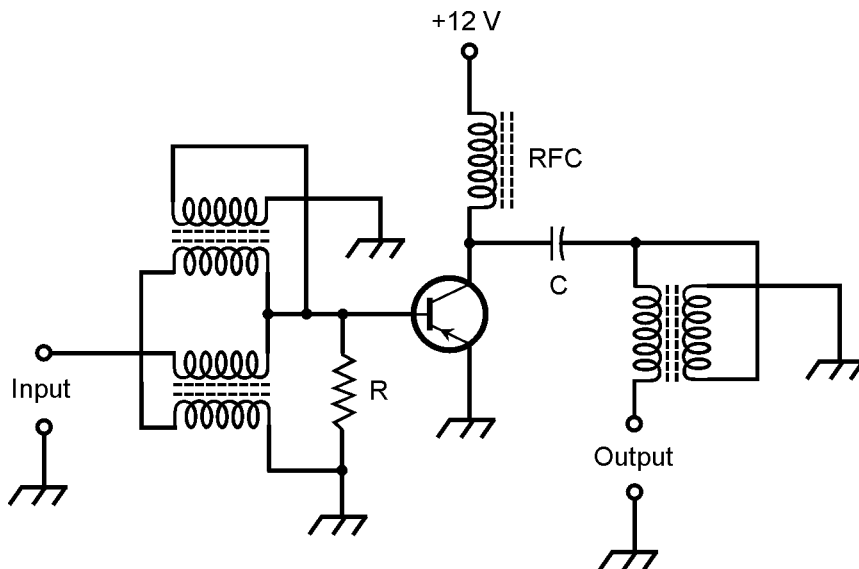
67. Figure Exam-4 is a generic diagram of a device that can be used to measure
- dc voltage.
  - dc resistance.
  - dc power.
  - ac voltage.
  - ac power.
68. Suppose an audio amplifier produces 8.0 V rms output with 80 mV rms input. Suppose the amplifier input impedance is the same as the output load impedance. What is the gain of this amplifier in this situation?
- 20 dB
  - 40 dB
  - 100 dB
  - 200 dB
  - We must know the actual input and output impedance values, in ohms, in order to determine the gain in this case.
69. The maximum gain obtainable with an amplifier that employs a given bipolar transistor
- gradually increases as the frequency increases.
  - stays the same as the frequency increases.
  - gradually decreases as the frequency increases.
  - alternately increases and decreases as the frequency increases.
  - remains constant up to a certain frequency, and then abruptly drops to zero above that frequency.
70. A phase comparator would most likely be found in
- an AF oscillator.
  - an AM detector.

- (c) an RF power amplifier.
  - (d) a PLL frequency synthesizer.
  - (e) An antenna tuning network.
71. An atom with 7 protons and 5 electrons is an example of
- (a) a positive isotope.
  - (b) a negative isotope.
  - (c) a positive ion.
  - (d) a negative ion.
  - (e) a neutral ion.
72. Electrical energy can be derived from hydrogen on a small scale in a device called
- (a) an electrolytic cell.
  - (b) an alkaline cell.
  - (c) a fission cell.
  - (d) a fusion cell.
  - (e) a fuel cell.
73. The output wave of a common-gate amplifier circuit with a pure sine-wave input
- (a) is in phase with the input wave.
  - (b) lags the input wave by  $90^\circ$  of phase.
  - (c) leads the input wave by  $90^\circ$  of phase.
  - (d) is  $180^\circ$  out of phase with the input wave.
  - (e) is inverted with respect to the input wave.
74. Which of the following capacitor types is polarized?
- (a) Electrolytic
  - (b) Paper
  - (c) Ceramic
  - (d) Mica
  - (e) Air variable
75. The maximum radiation and response from a full-wave loop antenna, assuming there are no nearby conductors or obstructions to distort the pattern, occurs
- (a) in the plane of the loop.
  - (b) perpendicular to the plane of the loop.
  - (c) at  $45^\circ$  angles to the plane of the loop.
  - (d) parallel to the horizon, in all directions of the compass.
  - (e) in all directions in three dimensions; it is a true omnidirectional antenna.
76. Suppose that five resistors, each having a different ohmic value from any of the others, are connected in series, and the entire combination is connected to a 12-V battery. How does the current through any one of the resistors compare with the current drawn from the battery?

- (a) The current through any one of the resistors is more than the current drawn from the battery.
- (b) The current through any one of the resistors is the same as the current drawn from the battery.
- (c) The current through any one of the resistors is less than the current drawn from the battery.
- (d) It depends on the relative ohmic values of the resistors.
- (e) None of the above statements are true.

77. Suppose a battery-powered circuit has  $1.00\ \Omega$  of net resistance, and  $15.0\ \text{A}$  of current flows through it. How much power is demanded from the battery?

- (a)  $225\ \text{W}$
- (b)  $150\ \text{W}$
- (c)  $22.5\ \text{W}$
- (d)  $66.7\ \text{mW}$
- (e) It is impossible to calculate this, based on the information given.



**Exam-5** Illustration for Final Exam Question 78.

78. Figure Exam-5 shows a generic circuit that can be used for broadband RF power amplification. What is wrong with the circuit as shown?

- (a) The resistor,  $R$ , should be replaced with a capacitor.
- (b) The capacitor,  $C$ , should be replaced with a resistor.
- (c) The  $+12\text{-V}$  power source should be replaced with a  $-12\text{-V}$  power source.

- (d) The choke, RFC, should be replaced with a capacitor.
- (e) Nothing is wrong with this circuit.

79. Imagine a string of 20 holiday lights, all connected in parallel and plugged into a 120-V rms ac wall outlet. If one of the bulbs is unscrewed and removed from the circuit, what will happen?

- (a) All of the remaining 19 bulbs will shine a little more brightly than before.
- (b) All of the remaining 19 bulbs will shine a little less brightly than before.
- (c) The fuse or circuit breaker will blow, and all of the remaining 19 bulbs will go out.
- (d) Most of all of the remaining 19 bulbs will burn out, and some may explode.
- (e) The remaining 19 bulbs will continue to shine exactly as before—no more or less brightly.

80. When a volatile RAM chip is used in a computer,

- (a) the contents of the memory can be easily overwritten.
- (b) the contents of the memory remain intact even if power is removed.
- (c) data in memory cannot be transferred to or from external media.
- (d) the contents of memory never last more than a few minutes.
- (e) excessive ELF radiation is produced.

81. A nickel-cadmium (NICAD) battery should never be discharged until it is totally dead because

- (a) the battery may explode.
- (b) the battery will have excessive voltage after it is recharged.
- (c) the battery will develop unwanted inductance after it is recharged.
- (d) the polarity of one or more of the cells may reverse, ruining the battery.
- (e) Forget it! There is no problem with discharging a NICAD battery until it is totally dead.

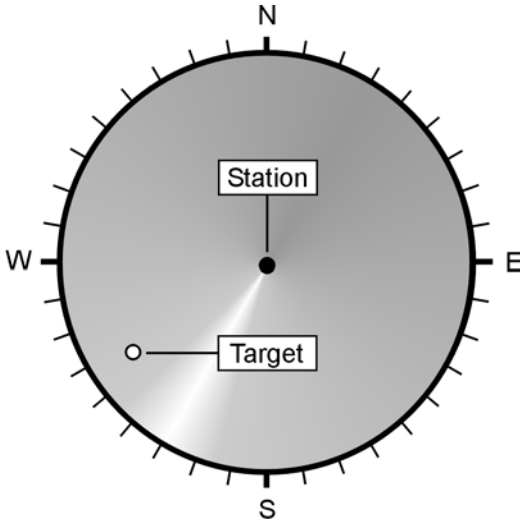
82. Which of the following is an advantage of a CRT display over an LCD for a computing workstation?

- (a) You can get a bigger screen for a lower price.
- (b) The CRT emits less ELF energy.
- (c) The CRT is less massive.
- (d) The CRT takes up less space on the desk.
- (e) All of the above are true.

83. The gauss is a unit of

- (a) charge carrier flow speed.
- (b) magnetic flux density.
- (c) electrostatic field strength.
- (d) electromagnetic field intensity.
- (e) electrical charge quantity.

84. Imagine an inductor and resistor connected in series, such that the inductive reactance is  $300\ \Omega$  and the resistance is  $500\ \Omega$  at a particular frequency. If the frequency is decreased, what happens to the relative phase of the instantaneous current and the instantaneous voltage?
- (a) Their relative phase does not change.
  - (b) They become more nearly in phase.
  - (c) They become more out of phase.
  - (d) They go alternately into and out of phase.
  - (e) It is impossible to say anything about this without more information.
85. Imagine two sine waves, one with a frequency of  $50.00\ \text{Hz}$  and the other with a frequency of  $60.00\ \text{Hz}$ . The phase difference between these two waves is
- (a) not definable.
  - (b)  $10.00^\circ$ .
  - (c)  $10.00\%$ .
  - (d)  $16.67\%$ .
  - (e)  $\frac{5}{6}$  of a radian.
86. Consider a variable capacitor connected in parallel with a fixed toroidal inductor. Suppose a signal of constant frequency is applied to the combination. If the variable capacitor is adjusted so its value changes from  $200\ \text{pF}$  to  $100\ \text{pF}$ , what happens to the reactance of the inductor?
- (a) It doubles.
  - (b) It becomes half as great.
  - (c) It quadruples.
  - (d) It becomes  $\frac{1}{4}$  as great.
  - (e) It does not change.
87. When two inductors having the same value are connected in series with a small opposing mutual inductance (say the coefficient of coupling is 5 percent or so), the total reactance of the combination at a specific, constant frequency is
- (a) equal to the reactance of either inductor.
  - (b) slightly greater than the reactance of either inductor.
  - (c) slightly less than the reactance of either inductor.
  - (d) slightly greater than twice the reactance of either inductor.
  - (e) slightly less than twice the reactance of either inductor.
88. The power factor in an ac circuit is defined as
- (a) the actual power divided by the maximum power the circuit can handle.
  - (b) the ratio of the real power to the imaginary power.
  - (c) the ratio of the apparent power to the true power.
  - (d) the ratio of the true power to the apparent power.
  - (e) the ratio of the imaginary power to the apparent power.



**Exam-6** Illustration for  
Final Exam  
Question 89.

89. Figure Exam-6 shows a radar display and a single blip, indicating the position of a target. What is the approximate azimuth of this target?
- 122°
  - 212°
  - 238°
  - 148°
  - More information is needed to answer this.
90. Which of the following is an e-mail address in the proper format?
- www.sciencewriter.net
  - members.authorsguild.net/stangib
  - sciencewriter@tahoe.com
  - http://google.com
  - All of the above
91. A robot can be guided along a specific route by means of a current-carrying wire beneath the surface. The robot detects, and follows, the magnetic field produced by the current, thus following the wire. This is called
- an epipolar system.
  - a fluxgate magnetometer.
  - a servo system.
  - an electromagnetic router.
  - an embedded path.



92. Which of the following substances is sometimes used as the semiconductor material in junction field-effect transistors (JFETs)?

- (a) Gallium arsenide
- (b) Mica
- (c) Glass
- (d) Polystyrene
- (e) Porcelain

93. Suppose that five resistors, each one having the same ohmic value as all the others, are connected in parallel, and the entire combination is connected to a 12-V battery. How does the current through any one of the resistors compare with the current drawn from the battery?

- (a) The current through any one of the resistors is 5 times the current drawn from the battery.
- (b) The current through any one of the resistors is the same as the current drawn from the battery.
- (c) The current through any one of the resistors is  $\frac{1}{5}$  as great as the current drawn from the battery.
- (d) The current through any one of the resistors is  $\frac{1}{25}$  as great as the current drawn from the battery.
- (e) It is impossible to answer this without more information.

94. Figure Exam-7 illustrates a method of

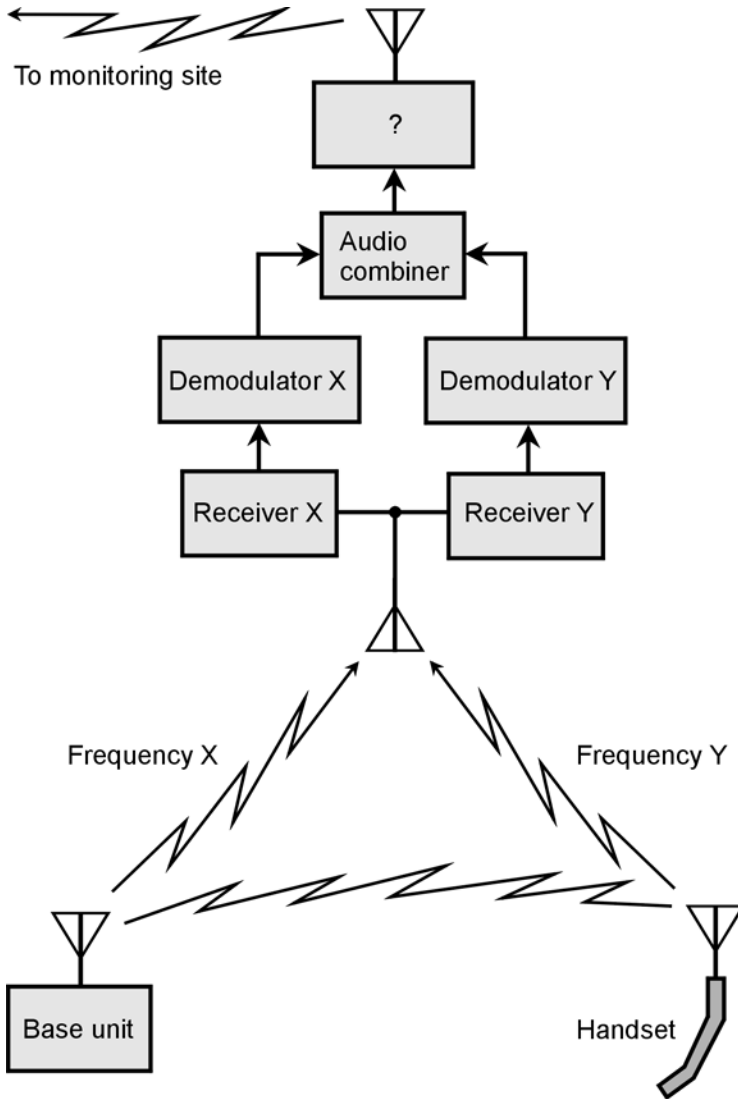
- (a) wireless-only encryption.
- (b) end-to-end encryption.
- (c) wireless tapping of a cell phone set.
- (d) wireless tapping of a cordless phone set.
- (e) dual-diversity shortwave reception.

95. In Fig. Exam-7, the box at the top contains a query symbol (?). What type of device should this be, in order for the system to perform its intended task?

- (a) A low-frequency receiver
- (b) A microwave transponder
- (c) A video and data receiver
- (d) A wireless transmitter
- (e) A GPS receiver

96. Some substances cause magnetic lines of flux to bunch closer together than they would be if the magnetic field existed in a vacuum. This property is known as

- (a) electromagnetism.
- (b) diamagnetism.
- (c) flux magnification.
- (d) flux constriction.
- (e) ferromagnetism.



**Exam-7** Illustration for Final Exam Questions 94 and 95.

97. In digital audio tape recording and reproduction, the noise can be practically eliminated because
- noise is analog in nature, and can be nulled or filtered out without degrading the digital signals.
  - noise is digital in nature, and can be eliminated by means of specialized computer programs.
  - music and voices are digital in nature, and can be reproduced multiple times without distortion.
  - music and voices are analog in nature, while noise is digital and can therefore be canceled out.
  - Forget it! In a digital audio tape recording system, noise is a more serious problem than it is in an analog system.

98. Imagine a capacitor and resistor connected in series, such that the capacitive reactance is equal to  $-200 \Omega$  and the resistance is equal to  $200 \Omega$  at a frequency of 70 MHz. How are the instantaneous current and the instantaneous voltage related?
- (a) They are in phase.
  - (b) They are  $90^\circ$  out of phase.
  - (c) The current leads the voltage by  $45^\circ$ .
  - (d) The current lags the voltage by  $45^\circ$ .
  - (e) None of the above are true.
99. The amount of current that a silicon photodiode can deliver in direct sunlight depends on
- (a) the forward breakover voltage.
  - (b) the thickness of the substrate.
  - (c) the surface area of the P-N junction.
  - (d) the applied voltage.
  - (e) the reverse bias.
100. Fill in the blank to make the following sentence true: "Electrical current can be expressed in terms of the number of \_\_\_\_ passing a given point per unit time."
- (a) neutrons
  - (b) webers or gauss
  - (c) charge carriers
  - (d) wave cycles
  - (e) isotopes

*This page intentionally left blank*

**A**  
APPENDIX

# Answers to Quiz, Test, and Exam Questions

## Chapter 1

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. b  | 2. c  | 3. d  | 4. d  | 5. d  |
| 6. a  | 7. b  | 8. c  | 9. a  | 10. b |
| 11. c | 12. a | 13. d | 14. c | 15. c |
| 16. a | 17. c | 18. b | 19. c | 20. d |

## Chapter 2

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. b  | 2. c  | 3. d  | 4. b  | 5. c  |
| 6. a  | 7. c  | 8. c  | 9. a  | 10. d |
| 11. d | 12. a | 13. b | 14. b | 15. c |
| 16. d | 17. b | 18. c | 19. a | 20. c |

## Chapter 3

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. b  | 2. a  | 3. c  | 4. c  | 5. a  |
| 6. d  | 7. c  | 8. a  | 9. d  | 10. c |
| 11. a | 12. b | 13. a | 14. c | 15. d |
| 16. b | 17. d | 18. a | 19. c | 20. b |

## Chapter 4

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. a  | 2. c  | 3. d  | 4. a  | 5. b  |
| 6. d  | 7. c  | 8. a  | 9. c  | 10. a |
| 11. d | 12. b | 13. d | 14. b | 15. a |
| 16. c | 17. c | 18. a | 19. d | 20. b |

**Chapter 5**

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. b  | 2. d  | 3. c  | 4. c  | 5. a  |
| 6. b  | 7. d  | 8. a  | 9. b  | 10. d |
| 11. b | 12. c | 13. b | 14. c | 15. a |
| 16. c | 17. b | 18. c | 19. b | 20. c |

**Chapter 6**

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. c  | 2. a  | 3. a  | 4. c  | 5. d  |
| 6. b  | 7. b  | 8. c  | 9. b  | 10. c |
| 11. d | 12. c | 13. a | 14. c | 15. a |
| 16. a | 17. b | 18. c | 19. b | 20. a |

**Chapter 7**

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. c  | 2. c  | 3. b  | 4. c  | 5. d  |
| 6. b  | 7. a  | 8. b  | 9. a  | 10. d |
| 11. b | 12. a | 13. b | 14. c | 15. d |
| 16. c | 17. c | 18. c | 19. a | 20. d |

**Chapter 8**

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. c  | 2. a  | 3. d  | 4. b  | 5. c  |
| 6. b  | 7. b  | 8. c  | 9. a  | 10. c |
| 11. c | 12. d | 13. b | 14. a | 15. d |
| 16. d | 17. c | 18. b | 19. b | 20. a |

**Test: Part 1**

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. a  | 2. a  | 3. c  | 4. b  | 5. a  |
| 6. b  | 7. a  | 8. e  | 9. e  | 10. c |
| 11. c | 12. c | 13. b | 14. c | 15. d |
| 16. b | 17. a | 18. a | 19. e | 20. e |
| 21. d | 22. d | 23. d | 24. c | 25. a |
| 26. c | 27. d | 28. a | 29. c | 30. c |
| 31. c | 32. d | 33. e | 34. d | 35. e |
| 36. b | 37. d | 38. d | 39. c | 40. a |
| 41. a | 42. d | 43. c | 44. e | 45. c |
| 46. c | 47. d | 48. a | 49. e | 50. c |

### Chapter 9

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. c  | 2. c  | 3. a  | 4. c  | 5. d  |
| 6. d  | 7. a  | 8. b  | 9. c  | 10. b |
| 11. d | 12. b | 13. b | 14. d | 15. d |
| 16. a | 17. c | 18. c | 19. d | 20. a |

### Chapter 10

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. b  | 2. a  | 3. d  | 4. b  | 5. b  |
| 6. c  | 7. d  | 8. c  | 9. a  | 10. d |
| 11. b | 12. d | 13. a | 14. a | 15. d |
| 16. b | 17. a | 18. c | 19. a | 20. d |

### Chapter 11

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. d  | 2. a  | 3. a  | 4. a  | 5. c  |
| 6. d  | 7. b  | 8. a  | 9. d  | 10. d |
| 11. b | 12. c | 13. b | 14. a | 15. b |
| 16. c | 17. d | 18. a | 19. b | 20. b |

### Chapter 12

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. c  | 2. a  | 3. b  | 4. b  | 5. d  |
| 6. b  | 7. d  | 8. c  | 9. a  | 10. b |
| 11. b | 12. d | 13. c | 14. a | 15. c |
| 16. b | 17. d | 18. c | 19. d | 20. c |

### Chapter 13

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. c  | 2. a  | 3. b  | 4. d  | 5. c  |
| 6. a  | 7. b  | 8. c  | 9. d  | 10. c |
| 11. d | 12. d | 13. a | 14. b | 15. d |
| 16. b | 17. d | 18. c | 19. b | 20. a |

### Chapter 14

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. b  | 2. d  | 3. a  | 4. c  | 5. b  |
| 6. a  | 7. b  | 8. a  | 9. a  | 10. d |
| 11. c | 12. b | 13. b | 14. b | 15. c |
| 16. b | 17. c | 18. d | 19. a | 20. c |

### Chapter 15

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. b  | 2. b  | 3. a  | 4. d  | 5. b  |
| 6. c  | 7. a  | 8. c  | 9. d  | 10. a |
| 11. a | 12. a | 13. c | 14. a | 15. d |
| 16. b | 17. d | 18. c | 19. c | 20. b |

### Chapter 16

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. b  | 2. a  | 3. d  | 4. d  | 5. d  |
| 6. b  | 7. c  | 8. c  | 9. a  | 10. c |
| 11. b | 12. b | 13. a | 14. b | 15. b |
| 16. d | 17. c | 18. a | 19. b | 20. a |

### Chapter 17

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. c  | 2. c  | 3. a  | 4. d  | 5. a  |
| 6. b  | 7. c  | 8. d  | 9. b  | 10. c |
| 11. c | 12. a | 13. d | 14. a | 15. c |
| 16. b | 17. d | 18. d | 19. c | 20. a |

### Chapter 18

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. c  | 2. a  | 3. d  | 4. b  | 5. b  |
| 6. a  | 7. c  | 8. d  | 9. a  | 10. d |
| 11. b | 12. b | 13. c | 14. d | 15. c |
| 16. a | 17. c | 18. d | 19. b | 20. b |

### Test: Part 2

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. e  | 2. e  | 3. a  | 4. a  | 5. e  |
| 6. d  | 7. c  | 8. d  | 9. a  | 10. e |
| 11. c | 12. a | 13. d | 14. c | 15. c |
| 16. e | 17. e | 18. d | 19. c | 20. b |
| 21. b | 22. d | 23. c | 24. c | 25. e |
| 26. c | 27. b | 28. a | 29. e | 30. c |
| 31. a | 32. a | 33. a | 34. a | 35. b |
| 36. e | 37. c | 38. d | 39. c | 40. b |
| 41. b | 42. c | 43. c | 44. c | 45. d |
| 46. d | 47. e | 48. e | 49. a | 50. a |



### Chapter 19

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. b  | 2. d  | 3. c  | 4. a  | 5. c  |
| 6. a  | 7. d  | 8. b  | 9. c  | 10. b |
| 11. d | 12. b | 13. b | 14. a | 15. b |
| 16. c | 17. a | 18. c | 19. d | 20. c |

### Chapter 20

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. b  | 2. d  | 3. c  | 4. b  | 5. a  |
| 6. a  | 7. d  | 8. c  | 9. b  | 10. d |
| 11. b | 12. a | 13. c | 14. a | 15. c |
| 16. d | 17. b | 18. a | 19. c | 20. a |

### Chapter 21

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. c  | 2. d  | 3. b  | 4. a  | 5. b  |
| 6. c  | 7. a  | 8. b  | 9. d  | 10. c |
| 11. a | 12. d | 13. d | 14. b | 15. c |
| 16. a | 17. c | 18. b | 19. a | 20. c |

### Chapter 22

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. d  | 2. a  | 3. b  | 4. d  | 5. d  |
| 6. c  | 7. b  | 8. c  | 9. b  | 10. d |
| 11. c | 12. a | 13. c | 14. d | 15. a |
| 16. d | 17. b | 18. c | 19. b | 20. b |

### Chapter 23

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. b  | 2. d  | 3. a  | 4. b  | 5. c  |
| 6. a  | 7. b  | 8. c  | 9. d  | 10. b |
| 11. c | 12. a | 13. d | 14. a | 15. c |
| 16. d | 17. b | 18. c | 19. a | 20. a |

### Chapter 24

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. a  | 2. b  | 3. b  | 4. c  | 5. d  |
| 6. c  | 7. d  | 8. c  | 9. d  | 10. b |
| 11. c | 12. b | 13. d | 14. c | 15. c |
| 16. d | 17. b | 18. a | 19. b | 20. c |

**Chapter 25**

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. c  | 2. c  | 3. c  | 4. c  | 5. a  |
| 6. d  | 7. b  | 8. a  | 9. c  | 10. d |
| 11. b | 12. a | 13. a | 14. b | 15. d |
| 16. a | 17. a | 18. c | 19. a | 20. d |

**Chapter 26**

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. b  | 2. a  | 3. b  | 4. c  | 5. c  |
| 6. a  | 7. d  | 8. d  | 9. a  | 10. d |
| 11. c | 12. a | 13. a | 14. d | 15. a |
| 16. d | 17. c | 18. d | 19. b | 20. c |

**Test: Part 3**

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. e  | 2. b  | 3. a  | 4. c  | 5. a  |
| 6. c  | 7. a  | 8. e  | 9. e  | 10. d |
| 11. d | 12. e | 13. c | 14. c | 15. d |
| 16. c | 17. c | 18. e | 19. a | 20. d |
| 21. c | 22. e | 23. a | 24. c | 25. d |
| 26. a | 27. b | 28. e | 29. a | 30. b |
| 31. c | 32. c | 33. b | 34. b | 35. a |
| 36. d | 37. a | 38. b | 39. b | 40. e |
| 41. c | 42. a | 43. d | 44. b | 45. a |
| 46. b | 47. b | 48. c | 49. e | 50. b |

**Chapter 27**

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. b  | 2. a  | 3. b  | 4. d  | 5. c  |
| 6. a  | 7. b  | 8. d  | 9. c  | 10. b |
| 11. c | 12. b | 13. d | 14. a | 15. b |
| 16. a | 17. d | 18. c | 19. c | 20. d |

**Chapter 28**

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. c  | 2. c  | 3. a  | 4. d  | 5. b  |
| 6. d  | 7. c  | 8. a  | 9. d  | 10. c |
| 11. a | 12. c | 13. b | 14. a | 15. c |
| 16. a | 17. c | 18. d | 19. d | 20. b |

### Chapter 29

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. c  | 2. d  | 3. a  | 4. b  | 5. c  |
| 6. d  | 7. c  | 8. c  | 9. d  | 10. d |
| 11. a | 12. a | 13. b | 14. a | 15. c |
| 16. c | 17. a | 18. a | 19. b | 20. d |

### Chapter 30

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. c  | 2. a  | 3. a  | 4. c  | 5. b  |
| 6. d  | 7. b  | 8. a  | 9. d  | 10. d |
| 11. d | 12. a | 13. b | 14. c | 15. a |
| 16. a | 17. d | 18. b | 19. c | 20. a |

### Chapter 31

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. b  | 2. a  | 3. d  | 4. b  | 5. c  |
| 6. b  | 7. d  | 8. a  | 9. d  | 10. a |
| 11. c | 12. a | 13. b | 14. d | 15. a |
| 16. c | 17. a | 18. b | 19. a | 20. b |

### Chapter 32

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. c  | 2. c  | 3. d  | 4. c  | 5. b  |
| 6. a  | 7. d  | 8. d  | 9. a  | 10. c |
| 11. b | 12. b | 13. d | 14. d | 15. a |
| 16. b | 17. a | 18. c | 19. a | 20. c |

### Chapter 33

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. b  | 2. a  | 3. d  | 4. a  | 5. a  |
| 6. c  | 7. d  | 8. a  | 9. d  | 10. c |
| 11. a | 12. d | 13. b | 14. c | 15. b |
| 16. c | 17. a | 18. d | 19. d | 20. a |

### Chapter 34

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. b  | 2. a  | 3. d  | 4. d  | 5. c  |
| 6. b  | 7. b  | 8. b  | 9. d  | 10. c |
| 11. c | 12. a | 13. a | 14. b | 15. c |
| 16. b | 17. d | 18. c | 19. b | 20. b |

**Test: Part 4**

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1. d  | 2. e  | 3. c  | 4. d  | 5. b  |
| 6. d  | 7. a  | 8. a  | 9. a  | 10. c |
| 11. c | 12. b | 13. e | 14. e | 15. e |
| 16. c | 17. e | 18. b | 19. a | 20. d |
| 21. c | 22. c | 23. e | 24. e | 25. e |
| 26. c | 27. c | 28. c | 29. a | 30. d |
| 31. b | 32. c | 33. d | 34. c | 35. e |
| 36. a | 37. a | 38. a | 39. a | 40. b |
| 41. b | 42. d | 43. e | 44. b | 45. b |
| 46. e | 47. d | 48. d | 49. d | 50. d |

**Final Exam**

- |       |       |       |       |        |
|-------|-------|-------|-------|--------|
| 1. e  | 2. b  | 3. c  | 4. b  | 5. e   |
| 6. a  | 7. d  | 8. e  | 9. b  | 10. a  |
| 11. b | 12. b | 13. a | 14. d | 15. c  |
| 16. a | 17. a | 18. a | 19. c | 20. b  |
| 21. a | 22. d | 23. e | 24. b | 25. d  |
| 26. a | 27. c | 28. e | 29. e | 30. b  |
| 31. e | 32. c | 33. d | 34. b | 35. d  |
| 36. a | 37. a | 38. b | 39. d | 40. e  |
| 41. c | 42. e | 43. b | 44. c | 45. c  |
| 46. d | 47. e | 48. d | 49. c | 50. b  |
| 51. b | 52. b | 53. c | 54. c | 55. d  |
| 56. c | 57. a | 58. e | 59. e | 60. d  |
| 61. e | 62. e | 63. e | 64. e | 65. a  |
| 66. c | 67. b | 68. b | 69. c | 70. d  |
| 71. c | 72. e | 73. a | 74. a | 75. b  |
| 76. b | 77. a | 78. c | 79. e | 80. a  |
| 81. d | 82. a | 83. b | 84. b | 85. a  |
| 86. e | 87. e | 88. d | 89. c | 90. c  |
| 91. e | 92. a | 93. c | 94. d | 95. d  |
| 96. e | 97. a | 98. c | 99. c | 100. c |

**B**  
APPENDIX

# Schematic Symbols

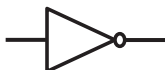
ammeter

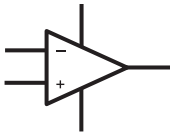


amplifier, general




amplifier, inverting




amplifier, operational 


AND gate 


antenna, balanced 

antenna, general 

antenna, loop 


antenna, loop, multiturn 


battery, electrochemical 

capacitor, feedthrough 

capacitor, fixed 

capacitor, variable 

capacitor, variable, split-rotor 

capacitor, variable, split-stator 

cathode, electron-tube, cold



cathode, electron-tube, directly heated



cathode, electron-tube, indirectly heated



cavity resonator



cell, electrochemical



circuit breaker



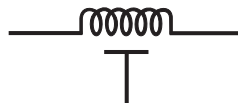
coaxial cable



crystal, piezoelectric



delay line



diac



diode, field-effect



diode, general
















diode, Gunn

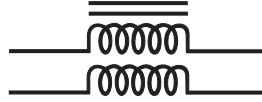


diode, light-emitting	
diode, photosensitive	
diode, PIN	
diode, Schottky	
diode, tunnel	
diode, varactor	
diode, Zener	
directional coupler	
directional wattmeter	
exclusive-OR gate	
female contact, general	
Ferrite bead	
filament, electron-tube	
fuse	

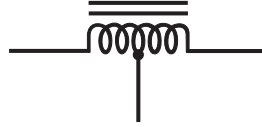


galvanometer	
grid, electron-tube	
ground, chassis	
ground, earth	
handset	
headset, double	
headset, single	
headset, stereo	
inductor, air core	
inductor, air core, bifilar	
inductor, air core, tapped	
inductor, air core, variable	
inductor, iron core	

inductor, iron core, bifilar



inductor, iron core, tapped



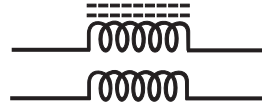
inductor, iron core, variable



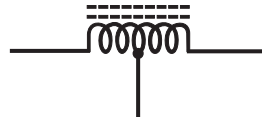
inductor, powdered-iron core



inductor, powdered-iron core, bifilar



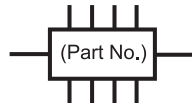
inductor, powdered-iron core, tapped



inductor, powdered-iron core, variable



integrated circuit, general



jack, coaxial or phono








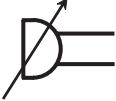





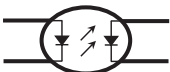


jack, phone, 2-conductor



jack, phone, 3-conductor



key, telegraph	
lamp, incandescent	
lamp, neon	
male contact, general	
meter, general	
microammeter	
microphone	
microphone, directional	
milliammeter	
NAND gate	
negative voltage connection	
NOR gate	
NOT gate	
optoisolator	

OR gate



outlet, 2-wire, nonpolarized



outlet, 2-wire, polarized



outlet, 3-wire



outlet, 234-volt



plate, electron-tube



plug, 2-wire, nonpolarized



plug, 2-wire, polarized



plug, 3-wire



plug, 234-volt



plug, coaxial or phono



plug, phone, 2-conductor



plug, phone, 3-conductor



positive voltage connection



potentiometer



probe, radio-frequency



or



rectifier, gas-filled



rectifier, high-vacuum



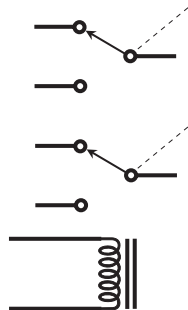
rectifier, semiconductor

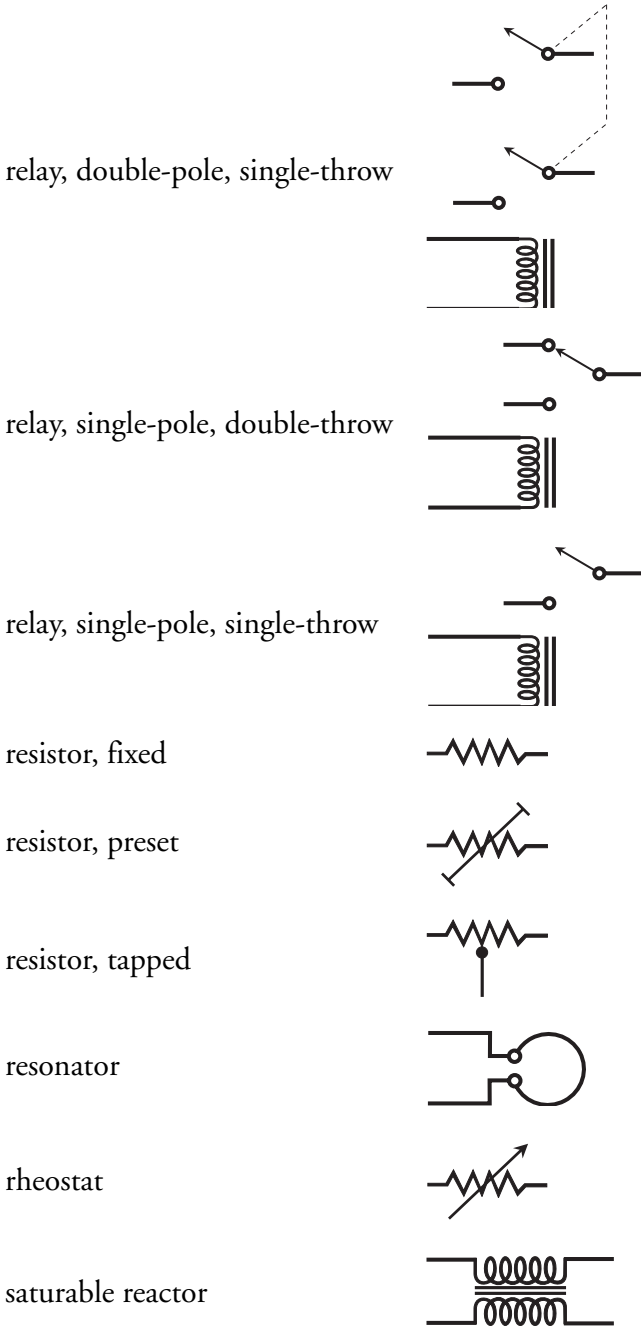


rectifier, silicon-controlled



relay, double-pole, double-throw

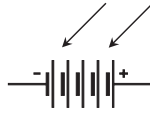




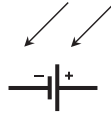
signal generator



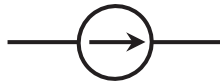
solar battery



solar cell



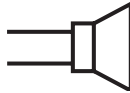
source, constant-current



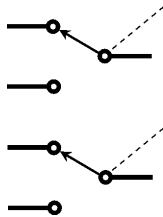
source, constant-voltage



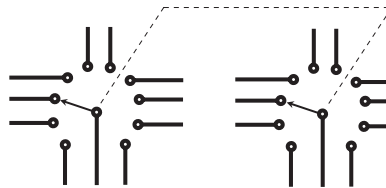
speaker



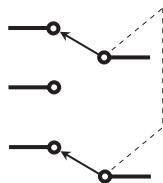
switch, double-pole, double-throw



switch, double-pole, rotary



switch, double-pole, single-throw



switch, momentary-contact



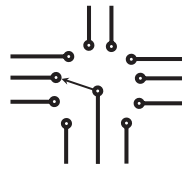
switch, silicon-controlled



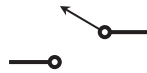
switch, single-pole, double-throw



switch, single-pole, rotary



switch, single-pole, single-throw



terminals, general, balanced



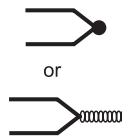
terminals, general, unbalanced



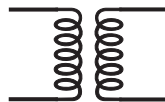
test point



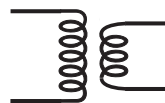
thermocouple



transformer, air core

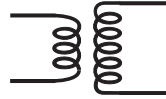


transformer, air core, step-down

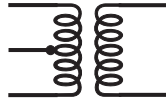




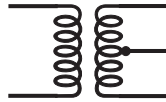
transformer, air core, step-up



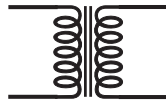
transformer, air core, tapped primary



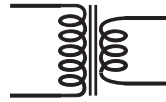
transformer, air core, tapped secondary



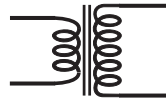
transformer, iron core



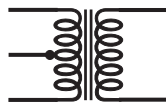
transformer, iron core, step-down



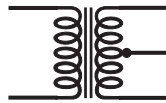
transformer, iron core, step-up



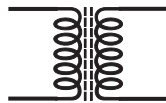
transformer, iron core, tapped primary



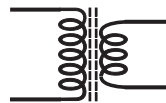
transformer, iron core, tapped secondary



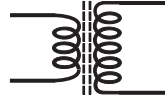
transformer, powdered-iron core



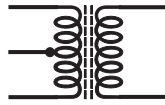
transformer, powdered-iron core, step-down



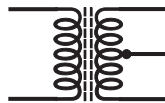
transformer, powdered-iron core, step-up



transformer, powdered-iron core, tapped primary



transformer, powdered-iron core, tapped secondary



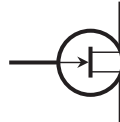
transistor, bipolar, NPN



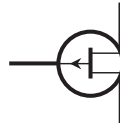
transistor, bipolar, PNP



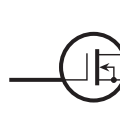
transistor, field-effect, N-channel



transistor, field-effect, P-channel



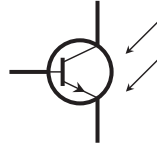
transistor, MOS field-effect, N-channel



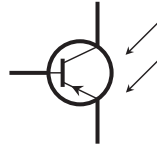
transistor, MOS field-effect, P-channel



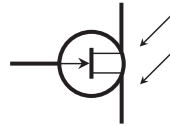
transistor, photosensitive, NPN



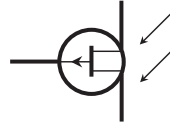
transistor, photosensitive, PNP



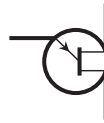
transistor, photosensitive, field-effect, N-channel



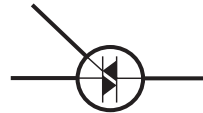
transistor, photosensitive, field-effect, P-channel



transistor, unijunction



triac



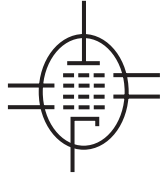
tube, diode



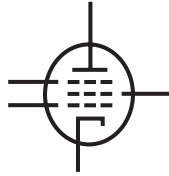
tube, heptode



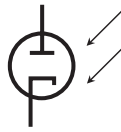
tube, hexode



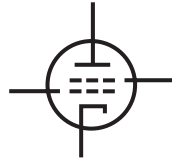
tube, pentode



tube, photosensitive



tube, tetrode



tube, triode



unspecified unit or component



voltmeter



wattmeter



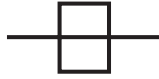
waveguide, circular



waveguide, flexible



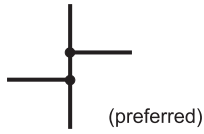
waveguide, rectangular



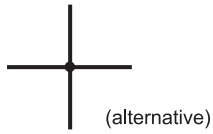
waveguide, twisted



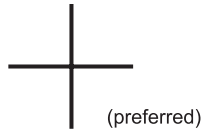
wires, crossing, connected



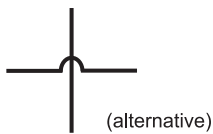
or



wires, crossing, not connected



or



*This page intentionally left blank*

# Suggested Additional Reading

- Brindley, K. *Starting Electronics*, 3d ed. Oxford, England: Newnes, 2004.
- Cutcher, D. *Electronic Circuits for the Evil Genius*. New York: McGraw-Hill, 2005.
- Gibilisco, S. *Electricity Demystified*. New York: McGraw-Hill, 2005.
- Gibilisco, S. *Electronics Demystified*. New York: McGraw-Hill, 2005.
- Horn, D. T. *Basic Electronics Theory with Projects and Experiments*, 4th ed. New York: McGraw-Hill, 1994.
- Long, L. *Personal Computing Demystified*. New York: McGraw-Hill, 2004.
- McComb, G. *Electronics for Dummies*. Hoboken, NJ: John Wiley & Sons, 2005.
- Scherz, P. *Practical Electronics for Inventors*. New York: McGraw-Hill, 2000.
- Slone, G. R. *TAB Electronics Guide to Understanding Electricity and Electronics*, 2d ed. New York: McGraw-Hill, 2000.
- Slone, G. R. *The Audiophile's Project Sourcebook*. New York: McGraw-Hill, 2002.
- Wise, E. *Robotics Demystified*. New York: McGraw-Hill, 2004.

*This page intentionally left blank*



# Index

## A

- ac. *See* alternating current
- acceptor impurity, 317–318
- acoustic direction finder, 529–531
- acoustic transducer, 597
- acoustics, 538–540
- ADC. *See* analog-to-digital converter
- A/D. *See* analog-to-digital
- admittance
  - complex, 250–253, 255–256
  - definition of, 240
  - vector representation of, 241, 252–253
- AGV. *See* automated guided vehicle
- AI. *See* artificial intelligence
- air-core coil, 57, 166
- alkaline cell, 105–106
- alpha cutoff frequency, 357–358
- alternating current
  - definition of, 27–28, 143
  - electromagnet, 121–122
  - source, schematic symbol, 57
- AM. *See* amplitude modulation
- amateur radio, 559–560, 580–581
- American Radio Relay League, 560–561
- American Red Cross, 30
- American Wire Gauge (AWG), 20
- ammeter, 39–41
- ampere, 10–11, 19–20
- ampere-hour, 103–104
- ampere-turn, 32, 40, 120
- amplification
  - audio, 389–391
  - factor, 357
  - radio-frequency, 391–393
- amplifier
  - audio, 389–391, 426
  - class A, 383
  - class AB, 383–384
  - class B, 384–386
  - class C, 385
  - current, 355–356
  - distortion in, 388–389
  - drive, 385, 388–389
  - driving power, 385
  - efficiency, 387–388
  - flat topping in, 388–389
  - hi-fi, 543
  - linear, 385
  - operational, 402
  - overdrive, 388–389
  - power, 384–388, 391–393
  - push-pull, 385–386
  - voltage, 369
  - weak-signal, 391
- amplitude
  - effective, 153
  - instantaneous, 151

amplitude (*Cont.*)

- limiting, 329
  - modulation, 149–150, 409–410, 429–430
  - negative peak, 152
  - peak, 151–152
  - peak-to-peak, 152
  - positive peak, 152
  - root-mean-square, 153
- analog audio tape, 548
- analog meter, 47
- analog-to-digital
- conversion, 415–416, 449–450, 547
  - converter, 523, 547, 580

## AND

- gate, 445–446
- operation, 443

## android, 605

## angle of lag, 207–210

## angle of lead, 221–224

## angular displacement transducer, 519

## angular frequency, 151

## anode, 318–319, 505, 509–510

## antenna

- billboard, 481
- broadside, 481–482
- coaxial, 475–476
- corner reflector, 485–486
- dipole, 281, 471–472
- directivity plots for, 478–480
- dish, 484–485
- doublet, 471–472
- driven element in, 480
- effective radiated power of, 478
- efficiency of, 471–472
- end-fire, 480–481
- feed point, 475
- folded dipole, 473–474
- forward gain for, 479–480
- front-to-back ratio for, 479–480
- front-to-side ratio for, 480
- ground systems for, 477–478
- ground-mounted vertical, 475
- ground-plane, 475
- half-wave, 473–474
- helical, 485

antenna (*Cont.*)

- horn, 485–486
  - isotropic, 478
  - longwire, 481
  - loop, 476–477
  - loopstick, 476
  - parasitic, 482–483
  - phased, 480–481
  - power gain of, 478
  - quad, 482–483
  - reference, 478
  - safety in working with, 486
  - tuner, 299, 474
  - vertical, 473–476
  - Yagi, 482–483
  - zepp, 473–474
- antinode, acoustic, 540
- apogee, 558
- apparent power, 255–267
- arctangent, 109
- armature coil, 126
- Armstrong oscillator, 394–395
- ARRL. *See* American Radio Relay League
- articulated geometry, 593–595
- artificial intelligence, 605–606
- ASCII, 446
- Asimov, Isaac, 592
- Asimov's three laws of robotics, 592
- asymmetrical square wave, 145
- atmospheric noise, 391
- atom, 3
- atom smasher, 27
- atomic number, 3–4
- atomic weight, 4
- attenuator, 381, 546
- audio
- amplifier, 389–391, 426
  - attenuator, 546
  - band-rejection filter, 433
  - bandpass filter, 433
  - filtering, 433
  - mixer, 544–545
  - notch filter, 433
  - oscillator, 400–402
  - splitter, 546
  - tape, 548–549
  - waveform, 400

audio-frequency-shift keying, 408  
 audio-taper potentiometer, 91–92  
 aurora borealis, 31, 422  
 auroral propagation, 422  
 automated guided vehicle, 601  
 automotive battery, 107  
 autotransformer, 292  
 avalanche, 321, 325, 328–330  
 average forward current, 338  
 AWG. *See* American Wire Gauge  
 axial leads, 96  
 azimuth, 527–528, 602

## B

baby monitor, 587  
 background noise, 541  
 back-pressure sensor, 524–525, 603  
 baffle, 539–540  
 balance control, 543–544  
 balanced load, 295  
 balanced modulator, 411–412  
 balanced transmission line, 295  
 balun, 295  
 bandpass filter, 411, 433  
 band-rejection filter, 433  
 band-stop filter, 546  
 bar-graph meter, 50  
 base  
   current, 353–354  
   of transistor, 352  
 bass, 389, 538  
 battery  
   automotive, 107  
   defined, 14  
   ideal, 104–105  
   internal resistance, 104  
   lantern, 106  
   lead-acid, 107–108  
   maximum deliverable current,  
     104  
   mercuric-oxide, 106–107  
   mercury, 106–107  
   nickel-based, 108  
   schematic symbol, 56  
   shelf life, 104  
   silver-oxide, 106

battery (*Cont.*)  
   storage capacity, 103–104  
   transistor, 106  
 baud, 580  
 Baudot, 446  
 beat frequency, 327, 425  
 beat-frequency oscillator, 425  
 bell ringer, 124  
 belt drive, 549  
 bias  
   for amplification, 355–357  
   in bipolar transistor, 353–357  
   definition of, 85, 320, 330  
   in field-effect transistor, 367–370  
 billboard antenna, 481  
 binary data, 446  
 binary number system, 442  
 binaural hearing, 597  
 binocular vision, 598–599  
 binomial, 231  
 biometric security measures, 565  
 bipolar transistor, 352–364  
 birdie, 426  
 bit, 448, 570  
 bits per second, 448–449, 570–571  
 black box, 57, 445  
 blackout, 293  
 blanking pulse, 417  
 bleeder resistor, 87–88  
 body capacitance, 523  
 boom box, 546–547  
 Boolean algebra, 443–444  
 bps. *See* bits per second  
 bridge  
   impedance, 258  
   rectifier circuit, 339–341, 347  
 brightness, in color image, 453  
 British thermal unit, 27  
 broadband  
   power amplifier, 391–392  
   transformer, 297  
 broadside antenna, 481–482  
 brownout, 293  
 bubble memory, 128  
 buffer, 450  
 button cell, 108  
 byte, 448, 570

**C**

## cable

- coaxial, 171, 423
- data transmission, 423
- modem, 559

## camera tube, 511–513

## candela, 49

## capacitance

- body, 523
- definition of, 175–176
- interelectrode, 84, 508
- junction, 320–321, 326
- mutual, 177
- in parallel, 178–179
- in series, 177–178
- units of, 177

## capacitive coupling, 389

## capacitive pressure sensor, 523

## capacitive proximity sensor, 524–525

## capacitive reactance

- definition of, 214
- frequency and, 216–217
- RC plane and, 218–219

## capacitor

- alternating current and, 215–216
- ceramic, 180
- coaxial, 183–184
- dielectric in, 176–177
- electrolytic, 181, 342
- filter, 87–88
- fixed, 179–182
- integrated-circuit, 181–182
- leakage resistance of, 218
- mica, 180, 399
- padder, 279
- paper, 179
- plastic-film, 180–181
- polyethylene, 180
- polystyrene, 180, 399
- practical, 176–177
- schematic symbol, 57
- semiconductor, 181–182
- silver-mica, 399
- tantalum, 181–182
- temperature coefficient, 184
- tolerance, 184

capacitor (*Cont.*)

- trimmer, 183
- tubular, 180
- variable, 182–184

## capacitor-input filter, 342–344

## carbon-composition resistor, 88–89

## carbon-film resistor, 89–90

## carrier

- definition of, 407
- frequency, 149–150, 407
- power consumed by, 409–410

## Cartesian coordinate geometry,

594–595

## Cartesian coordinate plane, 231

## Cartesian three-space, 595

## cascade, 342

## Cassegrain dish feed, 484–485

## cat whisker, 326

## catastrophic interference, 437

## cathode, 318–319, 505–506

## cathode-ray tube, 50, 416, 509–511

## cavity resonator, 280

CCD. *See* charge-coupled deviceCD. *See* compact discCDMA. *See* code-division multiple accessCD-R. *See* compact disc recordableCD-RW. *See* compact disc rewritable

## cell

- alkaline, 105–106
- button, 108
- in communications, 554–556
- cylindrical, 108
- dry, 14
- electrochemical, 56, 102–108
- flooded, 108
- fuel, 110–111
- ideal, 104–105
- internal resistance, 104
- lithium, 107
- maximum deliverable current, 104
- mercuric-oxide, 106–107
- mercury, 106–107
- nickel-based, 103, 108
- phone, 554–556, 564–565
- photovoltaic, 13, 48–49, 109–110
- primary, 102–103
- schematic symbol, 56

- cell (*Cont.*)
  - secondary, 102–103
  - shelf life, 104
  - silver-oxide, 106
  - spacecraft, 108
  - storage capacity, 103–104
  - Weston standard, 103–104
  - wet, 14
  - zinc-carbon, 105
- cellular communications, 554–556
- central processing unit, 569–570
- ceramic capacitor, 180
- channel, in field-effect transistor, 365
- character, in Morse code, 407–408
- characteristic curve, 369–370
- characteristic impedance, 236–238, 274–275, 297–299, 484
- charge carrier, 10, 18, 317–318
- charge-coupled device, 597
- chassis ground, 57, 340
- chemical energy, 14
- chime, 124
- choke-input filter, 342–344
- chopping wheel, 523
- circuit breaker, 347
- circular polarization, 421
- circular waveguide, 483–484
- circular-motion model of sine wave, 190–192
- Clapp oscillator, 396–397
- class A amplifier, 383
- class AB amplifier, 383–384
- class B amplifier, 384–386
- class C amplifier, 385
- clinometer, 600–601
- clipping, 329
- clock, 447
- cloning, cell phone, 565
- closed-loop configuration, 493
- cluster, in hard drive, 571–572
- CMOS. *See* complementary metal-oxide-semiconductor logic
- coaxial antenna, 475–476
- coaxial cable, 171
- coaxial capacitor, 183–184
- coaxial transmission line, 170
- code-division multiple access, 555
- coefficient of coupling, 164
- coherent communications, 435
- coil
  - air-core, 57
  - iron-core, 57
- cold cathode, 506
- collector
  - current, 354
  - power input, dc, 386–387
  - of transistor, 352
- color
  - code, resistor, 96–98
  - sensing, in vision system, 599
- Colpitts oscillator, 395–397
- commercial-transaction security, 562
- common base circuit, 359–360
- common collector circuit, 359–361
- common drain circuit, 375
- common emitter circuit, 358–359, 382
- common gate circuit, 374–375
- common ground, 171
- common source circuit, 373–374, 382–383
- communications
  - port, 581
  - security, 561–565
- commutator, 126
- compact disc
  - in audio recording, 547–548
  - recordable, 128, 573
  - rewritable, 128, 573
- compact hi-fi system, 541–542
- comparator, 496
- complementary metal-oxide-semiconductor logic, 497–498
- complex admittance, 250–253, 255–256
- complex impedance, 229–238, 245–248, 255–256, 259
- complex number
  - absolute value of, 232–233
  - addition, 230
  - definition of, 229–230
  - multiplication, 231
  - plane, 231–232
  - subtraction, 230
  - vector, 232

complex wave, 147  
 component  
   density, 317, 498  
   hi-fi system, 542  
 composite, 151  
 compound, 6–7  
 compression, data, 450–451  
 computer  
   central processing unit, 569–570  
   diskette, 573  
   display, 575–577  
   external storage, 573  
   flash memory, 573  
   gaming, 576  
   hard drive, 128, 571–572  
   Internet and, 581–583  
   logic board, 569  
   map, 529  
   microprocessor, 447, 569–570  
   modem, 580–581  
   motherboard, 569  
   printer, 577–578  
   random-access memory, 574–575  
   scanner, 578–580  
   tape drive, 573  
 conductance, 22–23, 238  
 conductance/inductance/capacitance circuit,  
   253–255, 277  
 conductance/susceptance plane,  
   240–241  
 conductivity, 22  
 conductor  
   defined, 8–9  
   schematic symbol, 55  
 conservation of current, 76  
 conservation of voltage, 77  
 continuous duty, 95  
 continuous wave, 430  
 contrast, 417  
 control grid, 507  
 conventional current, 72  
 conventional dish feed, 484–485  
 cordless phone, 564–565  
 core method, 289–290  
 corner reflector, 485–486  
 coulomb, 10–11  
 coulombs per second, 10–11

counter, frequency, 47, 448  
 counterpoise, 477  
 coupling  
   capacitive, 389  
   coefficient of, 164  
   transformer, 389–391  
   tuned-circuit, 390–391  
 CPU. *See* central processing unit  
 CRT. *See* cathode-ray tube  
 crystal-controlled oscillator, 398  
 crystal set, 326  
 current, 10–11, 18–19  
 current  
   amplifier, 355–356  
   average forward, 338  
   calculations, 59  
   conservation of, 76  
   conventional, 72  
   equalizing resistor, 346  
   gain, 357, 380  
   induced, 286  
   lag, 205–210  
   limiting, 85–86  
   loop, 276  
   node, 276  
   surge, 342, 346  
   theoretical, 72  
 cutoff, 384–385  
 CW. *See* continuous wave  
 cycles per second, 143  
 cylinder, in hard drive, 571–572  
 cylindrical cell, 108  
 cylindrical coordinate geometry,  
   595–596

## D

D/A. *See* digital-to-analog  
 DAC. *See* digital-to-analog converter  
 D layer, 421  
 dark noise, 512  
 D'Arsonval movement, 39–40  
 dash, in Morse code, 407–408  
 data  
   compression, 450–451  
   speed, 570–571  
 dc. *See* direct current

- dead zone, acoustic, 540
- decay, in sawtooth wave, 146
- decibel
  - for current, 380
  - gain, 93–94
  - impedance and, 381
  - loss, 93–94
  - meter, 47–48
  - for power, 380–381
  - for sound, 93–94
  - threshold of hearing and, 94
  - for voltage, 380
- decimal number system, 442
- declining discharge curve, 106–107
- decoder, 436
- decryption, 563–564
- decryption key, 563
- deflecting coil, 510
- deflecting plate, 511
- degrees of freedom, 593–594
- degrees of phase, 150
- degrees of rotation, 593–594
- demodulation
  - in modem, 580
  - in radio receiver, 326, 426, 429–432
- demodulator, 326
- demultiplexing, 496
- dependent variable, 148
- depletion
  - mode, 372–373
  - region, 320, 330, 367
- depth, 595
- derating function, 276
- derivative of waveform 189
- descrambling, 547
- desensitization, 427
- detection, 326, 429–432
- detector, 326, 426, 429–432
- diamagnetism, 122
- dielectric, 9, 170, 176–177
- dielectric constant, 177, 588
- differentiator, 494–495
- digital audio tape, 549
- digital cable television, 434
- digital encryption, 563–565
- digital integrated circuit, 496–498
- digital meter, 46–47
- digital satellite television, 418
- digital signal processing, 418, 435–436, 547–549
- digital-to-analog
  - conversion, 418, 449–450, 547
  - converter, 547, 580
- diode
  - amplitude limiting, 329
  - Esaki, 331
  - IMPATT, 331
  - infrared-emitting, 331
  - light-emitting, 331
  - PIN, 328
  - photodiode, 519
  - photosensitive, 332–333, 519
  - rectifier, 325–326, 338
  - semiconductor, 318–319
  - signal mixing, 327
  - switching, 328
  - tube, 504–506
  - tunnel, 331
  - uses of, 325–336
  - varactor, 321
  - Zener, 328–329
- dipole
  - antenna, 281, 471–472
  - magnetic, 119
- direct current
  - feedback, 399
  - power input, 386–387
  - pulsating, 28–30, 153–154
  - superimposed, 153–154
- direct drive, 549
- direct-conversion receiver, 425, 432
- direct-current electromagnet, 120–121
- direction
  - finder, acoustic, 529–531
  - finding, radio, 476, 530–531
  - of vector, 191
- directivity plot, antenna, 478–480
- directly heated cathode, 505–506
- director, 482
- discharge curve
  - declining, 106–107
  - flat, 104–105
- discriminator, 432

dish antenna, 484–485  
 diskette, 573  
 displacement transducer, 519–523  
 display  
   cathode-ray-tube, 50, 416, 509–511, 575  
   interlacing, 576  
   liquid crystal, 50, 331, 416, 511,  
     575–576  
   plasma, 511, 576  
   refresh rate, 576  
   screen size, 576  
 dissector tube, 512–513  
 distortion, 357, 388–389  
 Dolby, 541  
 domain  
   of function, 148  
   name, in Internet, 582–583  
   type, in Internet, 583  
 donor impurity, 317–318  
 dopant, 317  
 doping, 317  
 Doppler radar, 528  
 dot  
   in Morse code, 407–408  
   pitch, in video display, 576  
 dot-matrix printer, 578  
 dots per inch, 578–579  
 double-conversion receiver, 426  
 double sideband, 411  
 doublet antenna, 471–472  
 doubling, 548  
 drain  
   of field-effect transistor, 365  
   power input, dc, 386–387  
 drive, 87, 385, 388–389  
 driven element, in antenna, 480  
 driving power, 385  
 droop angle, of antenna radials, 475  
 dry cell, 14  
 DSB. *See* double sideband  
 DSP. *See* digital signal processing  
 dual-diversity reception, 434–435  
 dummy antenna, 87  
 dwell  
   frequency, 437  
   time, 437

dynamic  
   current amplification, 356  
   microphone, 517  
   mutual conductance, 370  
   range, 425, 541  
   speaker, 517  
   transducer, 517–519  
 dynode, 512

## E

earth ground, 57  
 earth-moon-earth, 423  
 eavesdropping, 561  
 EB. *See* exabyte  
 ECL. *See* emitter-coupled logic  
 E core, 289–290  
 E layer, 422  
 echo, 527–528  
 eddy current, 288  
 edge  
   detection, 601–602  
   triggering, 447  
 Edison, Thomas, 155  
 EEPROM. *See* electrically erasable  
   programmable read-only  
   memory  
 effective amplitude, 153  
 effective power, 266  
 effective radiated power, 478  
 effective voltage, 27, 29  
 efficiency  
   of antenna, 471–472  
   of power amplifier, 387–388  
 EHF. *See* extremely high frequency  
 elastomer pressure sensor, 524  
 elbow, 596–597  
 electric eye, 590  
 electric field, 17–18, 419  
 electric generator, 13, 521–522  
 electric lines of flux, 17–18, 119  
 electric monopole, 119  
 electric motor, 13, 126, 520  
 electrical ground, 340, 477  
 electrically erasable programmable read-only  
   memory, 499



- electrochemical
  - cell, 56, 102–108
  - energy, 102–105
- electrolyte, 181
- electrolytic capacitor, 181, 342
- electromagnet, 31, 40, 117, 120–122
- electromagnetic cathode-ray tube, 510
- electromagnetic deflection, 36–37
- electromagnetic field, 419–420
- electromagnetic induction, 286
- electromagnetic interference, 477, 549–550
- electromagnetic radiation, 144
- electromagnetic spectrum, 420
- electromagnetic wave polarization, 421
- electromagnetic wave propagation, 421–423
- electromagnetism, 13
- electromotive force, 12–13, 41
- electron
  - definition of, 4–5
  - gun, 509–510
  - secondary, 512
  - shell, 4–5
  - volt, 27
- electron tube
  - anode in, 504, 509–510
  - camera, 511–513
  - cathode in, 504
  - cathode-ray, 50, 416, 509–511
  - control grid in, 507
  - diode, 504–506
  - dissector, 512–513
  - dynode in, 512
  - electrodes in, 505–509
  - filament in, 505
  - forms of, 504–505
  - gas-filled, 504–505
  - heptode, 507–508
  - hexode, 507–508
  - image dissector, 512–513
  - image orthicon, 512
  - interelectrode capacitance in, 508
  - Klystron, 513–514
  - magnetron, 513
  - mercury-vapor, 506
  - neon-bulb, 504–505
  - electron tube (*Cont.*)
    - pentode, 507–508
    - photomultiplier, 512–513
    - plate in, 505–506
    - size of, 315
    - triode, 507
    - vacuum, 504
    - vidicon, 511
  - electronic bug, 589–590
  - electronic fingerprinting, 565
  - electronic mail, 582–583
  - electroscope, 38–39
  - electrostatic cathode-ray tube, 511
  - electrostatic charge, 12
  - electrostatic deflection, 38–39, 42
  - electrostatic emitter, 518
  - electrostatic meter, 38, 42
  - electrostatic pickup, 517–518
  - electrostatic transducer, 517–518
  - electrostatic voltmeter, 42
  - element, 3
  - elevation
    - in robot arm, 595–596
    - in spherical coordinates, 602
  - ELF. *See* extremely low frequency
  - elliptical orbit, 557
  - elliptical polarization, 421
  - embedded path, 601
  - EME. *See* earth-moon-earth
  - EMI. *See* electromagnetic interference
  - emitter
    - follower, 360–361
    - of transistor, 352
  - emitter-coupled logic, 496–497
  - encoder, 436
  - encryption, 563–565
  - end effector, 592
  - end-fire antenna, 480–481
  - end-to-end encryption, 563–564
  - energy
    - chemical, 14, 102
    - definition of, 25
    - electrochemical, 102–105
    - nonelectrical, 13–14
    - units, 25–27
  - enhancement mode, 373

envelope detector, 429–430  
 epipolar navigation, 532–533  
 EPROM. *See* erasable programmable read-only memory  
 erasable programmable read-only memory, 499  
 erase head, 548  
 eraser-head pointer, 520  
 erg, 26–27  
 ERP. *See* effective radiated power  
 error voltage, 398–399  
 Esaki diode, 331  
 exabyte, 449, 570  
 exclusive OR gate, 445–446  
 expert system, 605–606  
 external modem, 581  
 external storage, 573  
 extremely high frequency, 424  
 extremely low frequency, 511, 577  
 eye-in-hand system, 599–600

## F

F layer, 422  
 facsimile, 416  
 farad, 177  
 fast-scan television, 417–418, 433–434  
 fax, 416  
 feed point, of antenna, 475  
 feedback  
   dc, 399  
   negative, 393  
   positive, 393–394  
   at single frequency, 394  
 feedthrough scanner, 579  
 ferromagnetic core, 166–167, 288–289  
 ferromagnetic material, 31, 116–117, 122  
 ferromagnetism, 122  
 FET. *See* field-effect transistor  
 FET voltmeter, 44–45  
 fiber optics, 332, 424  
 field coil, 126  
 field-effect transistor, 44–45, 365–378  
 field strength  
   magnetic, 120  
   radio-frequency, 479–480  
 field-strength meter, 590

fifth-generation robot, 592  
 filament, 505  
 film-type resistor, 89–90  
 filter  
   audio, 433  
   bandpass, 411, 433  
   band-rejection, 433  
   capacitor, 87–88  
   capacitor-input, 342–344  
   choke, 169  
   choke-input, 342–344  
   notch, 433  
   power-supply, 342–344  
   section, 342–344  
 firmware, 499  
 first-generation robot, 591  
 fixed capacitor, 179–182  
 flare angle, 485  
 flash memory, 573  
 flat discharge curve, 104–105  
 flatbed scanner, 579  
 flat-topping, 388–389  
 flip-flop, 446–448  
 flooded cell, 108  
 flux density, 32  
 flux lines, 117–118  
 fluxgate magnetometer, 531–532  
 flying eyeball, 599–600  
 FM. *See* frequency modulation  
 folded dipole antenna, 471–472  
 foot-pound, 27  
 formatting, of hard drive, 571  
 forward  
   bias, 319, 354–355  
   breakover, 319, 325, 329  
   gain, 479–480  
   power, 267  
 fossil-fuel source, 111  
 four-channel stereo, 547  
 fourth-generation robot, 592  
 frame, 416–418  
 frequency, 28, 143–144  
 frequency  
   alpha cutoff, 357–358  
   angular, 151  
   counter, 47, 448

- frequency (*Cont.*)  
 definition of, 28, 143–144  
 domain, 148, 327  
 drift, 399–400  
 dwell, 437  
 fundamental, 280  
 harmonic, 280  
 hopping, 437  
 mark versus space, 430  
 modulation, 411–414, 429, 432  
 multiplication, 326–327, 413  
 offset, 430–432  
 spectrum, 148–150  
 sweeping, 437  
 synthesizer, 398–399  
 versus wavelength, 419  
 frequency-division multiplexing, 436  
 frequency-shift keying, 408, 430–432  
 frequency-spreading function, 437  
 front end, 426–428  
 front-to-back ratio, 479–480  
 front-to-side ratio, 480  
 FSTV. *See* fast-scan television  
 fuel cell, 110–111  
 full-scale deflection, 40  
 full-wave rectification, 28–30, 339–343, 347  
 full-wave voltage doubler, 341  
 fundamental frequency, 148–149, 280  
 fuse  
   quick-break, 347  
   slow-blow, 347  
 fusion, 4
- G**
- gain  
 in decibels, 379–381  
 definition of 93–94, 357  
 forward, 479–480  
 negative, 379  
 gain bandwidth product, 357–358  
 galena, 326  
 gallium arsenide, 316, 425  
 galvanism, 36  
 galvanometer, 36–37  
 gaming, 576
- gas, 7–8  
 gas-filled tube, 504–505  
 gate  
   of field-effect transistor, 365  
   voltage, 365  
 gated flip-flop, 447  
 gauss, 32, 120  
 GB. *See* conductance/susceptance, gigabyte  
 gbps. *See* gigabit per second  
 generator  
   defined, 154–155  
   efficiency of, 155  
   electric, 521–522  
 geographic poles, 115  
 geomagnetic declination, 116  
 geomagnetic equator, 116  
 geomagnetic field, 13, 115–117  
 geomagnetic inclination, 116  
 geomagnetic lines of flux, 115  
 geomagnetic poles, 115  
 geomagnetic storm, 117  
 geostationary satellite, 424, 556–557  
 germanium, 316  
 gigabit, 448  
 gigabit per second, 571  
 gigabyte, 449, 570  
 gigacycle, 143  
 gigahertz, 28, 143–144  
 gigawatt, 265  
 gilbert, 32, 120  
 GLC. *See* conductance/inductance/capacitance  
 Global Positioning System, 533–534  
 GPS. *See* Global Positioning System  
 graphic equalizer, 389, 545–546  
 grayscale, 416  
 gripper, 588–589  
 ground  
   bus, 477–478  
   chassis, 57, 340  
   earth, 57, 345  
   electrical, 340, 477  
   loop, 477–478  
   loss, 421  
   for power supply, 345  
   radial system, 475, 477  
   radio-frequency, 477

ground (*Cont.*)  
 signal, 358  
 systems for antennas, 477–478  
 grounded-cathode circuit, 508–509  
 grounded-grid circuit, 508–509  
 ground-mounted vertical antenna, 475  
 ground-plane antenna, 475  
 Gunn diode, 330–331  
 Gunn effect, 330  
 Gunnplexer, 331

## H

hacker, 562–563  
 half-wave  
 antenna, 473–474  
 rectification, 28–29, 325–326, 338–340,  
 342–343  
 ham radio, 559–560  
 handheld scanner, 579  
 handoff, 554  
 hand-print recognition, 565  
 hard drive, 128, 571–572  
 harmonic, 148–149, 280, 326–327, 539  
 Hartley oscillator, 394–395  
 HDTV. *See* high-definition television  
 headphone radio, 546  
 headset, 543  
 heatsink, 338  
 helical antenna, 485  
 helium, 4  
 henry, 161  
 heptode tube, 507–508  
 hertz, 27–28, 143–144  
 heterodyne, 327  
 hexadecimal number system, 443  
 hexode tube, 507–508  
 HF. *See* high frequency  
 high-definition television, 417–418  
 high frequency, 424  
 highpass response, 493–494  
 high-tension transmission line, 155, 293  
 hole, 10, 317–318  
 horizontal polarization, 421  
 horn antenna, 485–486  
 hot-wire meter, 39  
 hue, in color image, 453

hydrogen, 3–4, 110–111  
 hydrogen fuel cell, 110–111  
 hypertext, 583  
 hysteresis loss, 288–289

## I

IC. *See* integrated circuit.  
 ideal battery, 104–105  
 ideal cell, 104–105  
 IF. *See* intermediate frequency  
 illumination meter, 48–49  
 image  
 compression, 451  
 dissector, 512–513  
 orthicon, 512  
 resolution, 513, 576, 578  
 in superheterodyne receiver, 426, 428  
 transmission, video, 416–418  
 imaginary number  
 definition of, 229  
 line, 229–230  
 unit, 229  
 IMD. *See* intermodulation distortion  
 IMPATT diode, 331  
 impedance  
 absolute value of, 235  
 characteristic, 236–238, 274–275, 297–299,  
 484  
 complex, 229–238, 248, 255–256, 259  
 matching, 88, 238, 275, 294–299, 474  
 measurement of, 258  
 mismatch, 275  
 nonreactive, 235  
 purely resistive, 235, 259  
 step-down transformer, 296  
 step-up transformer, 296  
 surge, 236  
 transfer ratio, 296  
 vector representation of, 234, 247–248  
 independent variable, 148  
 indirectly heated cathode, 505–506  
 induced current, 286  
 inductance  
 definition of, 160–161  
 mutual, 161–162, 164–166, 168  
 in parallel, 163–164

- inductance (*Cont.*)
  - in series, 161–162
  - stray, 171
  - unit of, 161
  - unwanted, 171
- inductance/capacitance, 277, 523
- inductive reactance
  - definition of, 201
  - frequency and, 202–203
  - RL plane and, 203–205
- inductor
  - at audio frequencies, 169
  - definition of, 161
  - practical, 161
  - at radio frequencies, 169–170
  - roller, 396
  - transmission lines as, 169–170
- inference engine, 605–606
- infrared
  - definition of, 13
  - intensity measurement, 49
  - motion detector, 590–591
  - transducer, 519
- infrared-emitting diode, 331
- inkjet printer, 577
- instantaneous amplitude, 143, 151
- instantaneous maximum, 27
- instantaneous power, 266
- instantaneous rate of change, 189
- instantaneous voltage, 29, 189
- insulated-gate field-effect transistor, 371
- insulator, 8–9
- integrated circuit
  - advantages of, 491–492
  - capacitor, 181–182
  - comparator, 496
  - complementary metal-oxide-semiconductor, 497–498
  - demultiplexer, 496
  - digital, 496–498
  - emitter-coupled logic, 495–496
  - limitations of, 492
  - linear, 492–496
  - memory, 499
  - metal-oxide-semiconductor, 497–498
- integrated circuit (*Cont.*)
  - multiplexer, 496
  - operational amplifier, 402, 493–495
  - regulator, 344
  - resistor, 90
  - timer, 496
  - transistor-transistor logic, 496–497
  - voltage regulator, 495
- integrator, 495
- interactive photovoltaic system, 110
- interelectrode capacitance, 184, 508
- interference
  - in communications, 437
  - pattern, 513
- interlacing, 576
- intermediate frequency, 426, 428
- intermodulation distortion, 427
- internal modem, 581
- internal resistance, 40, 41, 104
- Internet
  - address, 582
  - cell phones and, 555–556
  - domain name, 582–583
  - domain type, 583
  - electronic mail and, 582–583
  - newsgroups, 583–584
  - origin of, 581
  - packets and, 581
  - protocol, 581
  - service provider, 581
  - World Wide Web and, 583
- interwinding capacitance, 289–290
- intrinsic semiconductor, 328
- inverse tangent, 209
- inversion, logical, 444
- inverter, logical, 444–446
- ion, 6
- ionization potential, 588
- ionosphere, 421–422
- ionospheric propagation, 421–422
- interlacing, 418
- IRED. *See* infrared-emitting diode
- iris-print recognition, 565
- iron-core coil, 56
- irregular wave, 147–148
- isotope, 4

isotropic antenna, 478  
iterative process, 71

**J**

j operator, 204  
JFET. *See* field-effect transistor  
jitter, 418  
J-K flip-flop, 447–448  
joule, 25–27  
joystick, 520  
junction capacitance, 320–321, 326, 330

**K**

KB. *See* kilobyte  
kbps. *See* kilobit per second  
keying  
  audio-frequency-shift, 408  
  frequency-shift, 408  
  on/off, 407–408  
kilobaud, 580  
kilobit, 448  
kilobit per second, 571  
kilobyte, 449, 570  
kilocycle, 143  
kilohertz, 28, 143–144  
kilohm, 20  
kilosiemens, 22  
kilovolt, 17  
kilowatt, 24, 265  
kilowatt-hour, 25–27, 46  
kilowatt-hour meter, 46  
Kirchhoff's Current Law, 76  
Kirchhoff's First Law, 75–77  
Kirchhoff's Second Law, 77–78  
Kirchhoff's Voltage Law, 77  
Klystron, 513–514

**L**

lag, angle of, 207–210  
lagging phase, 194  
lamp, incandescent, 57  
LAN. *See* local area network  
lantern battery, 106

large-scale integration, 498  
laser beam, 588–589  
laser printer, 577  
latency, 556–557  
LC. *See* inductance/capacitance  
LCD. *See* liquid-crystal display  
lead-acid battery, 107–108  
leading phase, 194  
leakage resistance, 218  
LED. *See* light-emitting diode  
LEO. *See* low-earth-orbit satellite  
LF. *See* low frequency  
light meter, 48–49  
light-emitting diode, 331  
lightning, 12  
limiter, 329, 432  
linear amplifier, 385  
linear displacement transducer, 519  
linear function, 492–493  
linear integrated circuit, 492–496  
linearity, 541  
linear-taper potentiometer, 90–91  
lines of flux, 117–119  
lines of force, 118  
liquid, 7–8  
liquid-crystal display, 50, 331, 416,  
  511  
lithium battery, 107  
lithium cell, 107  
LO. *See* local oscillator  
load, 237, 252, 295, 397, 400  
loading control, 393  
local area network, 558–559  
local node, 452  
local oscillator, 425  
location system, 527–531  
lock-in time, 47  
lodestone, 115–116  
logarithmic-taper potentiometer,  
  91–92  
logic  
  binary, 443–444  
  board, 569  
  emitter-coupled, 496–497  
  gates, 444–446  
  metal-oxide-semiconductor, 497–498

logic (*Cont.*)  
 negative, 444–445  
 positive, 444–445  
 transistor-transistor, 496–497

log-log graph, 358

log-polar navigation, 532–533

log-taper potentiometer, 91–92

longwire antenna, 481

loop, 276

loop antenna, 476–477

loopstick antenna, 476

loran, 532

loss

in decibels, 93–94, 379–381

ohmic, 293

resistance, 471–472

lossless image compression, 451

lossy image compression, 451

loudness, 540

loudness meter, 47–48

low frequency, 424

low-earth-orbit satellite, 424, 556–557

lower sideband, 409–411

lowpass response, 493–494

LSB. *See* lower sideband

LSI. *See* large-scale integration

lumen, 49

## M

machine

hearing, 597

language, 575

vision, 529, 597–600

magnet, permanent, 117, 124

magnetic compass, 155–116

magnetic dipole, 119

magnetic disk, 128

magnetic field, 30, 120, 419

magnetic field strength, 120

magnetic force, 117

magnetic flux density, 32

magnetic levitation, 117

magnetic lines of flux, 30–31, 117–118

magnetic pole, 31, 119

magnetic tape, 127

magnetic units, 32

magnetism, 30–32, 115–140

magnetomotive force, 32, 40

magnetron, 513

magnitude of vector, 191

majority carrier, 10, 318

make-before-break, 555

manipulator, 592

mark, 408, 430

master/slave flip-flop, 447

matrix, 64

maximum deliverable current, 104

maxwell, 32, 120

MB. *See* megabyte

mbps. *See* megabit per second

medium-earth-orbit satellite, 557–558

medium frequency, 424

medium-scale integration, 498

megabit, 448

megabit per second, 571

megabyte, 449, 570

megacycle, 143

megahertz, 28, 143–144

megasiemens, 22

megavolt, 17

megawatt, 24, 265

megohm, 20

memory

backup, 499

capacity, 575

channels, in hi-fi tuner, 543

drain, in nickel-based cell, 108

flash, 573

integrated-circuit, 499

nonvolatile, 499

random-access, 499, 574–575

read-only, 499

volatile, 499

MEO. *See* medium-earth-orbit satellite

mercuric-oxide battery, 106–107

mercuric-oxide cell, 106–107

mercury

battery, 106–107

cell, 106–107

mercury-vapor tube, 506

metal-film resistor, 89–90

metal oxide, 317  
 metal-oxide semiconductor, 370–373  
 metal-oxide-semiconductor field-effect transistor, 370–373  
 meteor scatter propagation, 422–423  
 meteor shower, 422  
 meter, schematic symbol, 56  
 meter shunt, 40–41  
 methane, in fuel cell, 111  
 methanol, in fuel cell, 111  
 MF. *See* medium frequency  
 mica capacitor, 180, 399  
 micro fuel cell, 111  
 microammeter, 41  
 microampere, 19  
 microfarad, 177  
 microhenry, 161  
 microprocessor, 447, 569–570  
 microsiemens, 22  
 microvolt, 17  
 microwatt, 24, 265  
 microwave band, 513  
 midrange  
   sound, 538  
   speaker, 543  
 military-level security, 562–563  
 military specifications, 562  
 milliamper, 19  
 millihenry, 161  
 milliohm, 22  
 millisiemens, 22  
 millivolt, 17  
 milliwatt, 24, 265  
 minority carrier, 10, 318  
 mixer  
   audio, 544–545  
   radio-frequency, 426  
 mixing product, 327  
 mobile hi-fi system, 546  
 mobile operation, in amateur radio, 559  
 modem, 408, 449, 555–556, 559, 580–581  
 modular construction, 492  
 modulation  
   amplitude, 149–150, 409–410  
   definition of, 407

modulation (*Cont.*)  
   envelope, 266  
   index, 414  
   in modem, 580  
   phase, 413–414  
   pulse, 414–415  
   reactance, 413  
   waveform, 385  
 modulator, balanced, 411–412  
 modulator/demodulator. *See* modem  
 molecule, 7–8  
 monitoring system, 587–591  
 monopole, electric, 119  
 moonbounce, 423  
 Morse code, 385, 407–408, 446  
 MOS. *See* metal-oxide semiconductor  
 MOSFET. *See* metal-oxide-semiconductor field-effect transistor  
 motherboard, 569  
 motion detector  
   infrared, 590–591  
   ultrasonic, 591  
 motor, 13, 126, 520–521  
 motor/generator, 521–522  
 mouse, 520  
 M/S flip-flop, 447  
 MSI. *See* medium-scale integration  
 multimeter, 40–41, 44  
 multiplexing, 436, 496  
 multivibrator oscillator, 401–402  
 Murray code, 446  
 music synthesizer, 400  
 mutual inductance, 161–162, 164–166, 168

## N

NAND gate, 445–446  
 nanoampere, 19  
 nanohenry, 161  
 nanowatt, 265  
 narrowband FM, 413  
 navigation, robot, 600–602  
 navigational methods, 531–534  
 negative feedback, 393  
 negative peak amplitude, 152  
 negative resistance, 233, 330



- negative temperature coefficient, 96
- network
  - of computers, 449
  - local area, 558–559
  - wireless, 558–559
- neutron, 3–4
- newsgroups, Internet, 583–584
- nickel-based battery, 108
- nickel-based cell, 103, 108
- node
  - acoustic, 540
  - in communications, 451–452
- noise
  - atmospheric, 391
  - background, 541
  - figure, 391, 425
  - white, 391
- nonelectrical energy, 13–14
- nonlinear scale, 43
- nonlinearity, 326–327, 357
- nonreactive impedance, 235
- nonvolatile memory 499
- nonvolatility, 128, 499
- NOR gate, 445–446
- north geomagnetic pole, 115
- north magnetic pole, 119
- NOT
  - gate, 445–446
  - operation, 443
- notation, 24–25
- notch filter, 433
- NPN transistor, 352
- N-channel field-effect transistor, 365–366
- N-type semiconductor, 10, 109–110, 317–318
- N-type silicon, 109–110
- nucleus, 3
- number system
  - binary, 442
  - decimal, 442
  - hexadecimal, 443
  - octal, 443
- O**
- OCR. *See* optical character recognition
- octal number system, 443
- octet, 448
- ohm, 9, 20–22
- ohm per foot, 20
- ohm per kilometer, 20
- ohm per meter, 20
- ohmic loss, 293
- ohmic value, 95
- ohmmeter, 43–44
- Ohm's Law
  - for ac circuits, 259–261
  - for dc circuits, 12–13, 58–61
- on/off keying, 407–408
- op amp, 402, 493–495
- open-loop configuration, 493
- operating point, 356
- operational amplifier, 402, 493–495
- optical character recognition, 579–580
- optical encoder, 522–523
- optical resolution, 598
- optical sensitivity, 597–598
- optoisolator, 332–333
- OR
  - gate, 445–446
  - operation, 443
- oscillator
  - Armstrong, 394–395
  - audio, 400–402
  - beat-frequency, 425
  - Clapp, 396–397
  - Colpitts, 395–397
  - crystal-controlled, 398
  - definition of, 393–394
  - Hartley, 394–395
  - load impedance of, 397, 400
  - local, 425
  - multivibrator, 401–402
  - reference, 398
  - reliability of, 400
  - stability of, 396–397, 399–400
  - twin T, 401
  - voltage-controlled, 330, 397
- oscilloscope, 50, 327, 510
- overdrive, 87, 356–357, 388–389
- overloading, 427
- overshoot, in meter, 39

oxygen, 7  
ozone, 7

- P**
- packet
    - radio, 559
    - wireless, 451–452
  - padding capacitor, 279
  - palette, 453
  - paper capacitor, 179
  - paraboloidal reflector, 484
  - parallel circuit
    - power distribution in, 74–75
  - parallel data transmission, 450
  - parallel resistances, 63, 72–74
  - parallel resonance, 252
  - parallel-to-serial conversion, 450–451
  - parallel-wire transmission line, 169–170, 423
  - parasitic antenna, 482–483
  - parasitic element, 482
  - particle accelerator, 27
  - passband, 425
  - PB. *See* petabyte
  - PC card, 573
  - PCMCIA card, 573
  - P-channel field-effect transistor, 365–366
  - peak
    - amplitude, 151–152
    - inverse voltage, 338
    - power, 266
  - peak-to-peak amplitude, 152
  - pen recorder, 49–50
  - pentode tube, 507–508
  - perigee, 557–558
  - period, 143–144, 146
  - periodic ac wave, 143
  - permanent magnet, 117, 124
  - permeability, 122–123, 166–167
  - permeability tuning, 167, 396
  - personal computer. *See* computer
  - personal identification number, 565
  - petabyte, 449, 570
  - phase
    - angle, 192–195, 207–210, 221–224, 269–270
    - coincidence, 192–193
    - phase (*Cont.*)
      - comparator, 398
      - degrees of, 150
      - difference, 151, 192–195
      - lagging, 194
      - leading, 194
      - modulation, 413–414
      - opposition, 193
      - quadrature, 193
      - radians of, 151
      - of sound wave, 540
    - phased antenna, 480–481
    - phase-locked loop, 398–399
    - phosphor, 509–510
    - photocathode, 512
    - photodetector, 416, 588–589
    - photodiode, 519
    - photoelectric proximity sensor, 525–526
    - photomultiplier, 512–513
    - photosensitive diode, 332–333, 519
    - photovoltaic cell, 13, 48–49, 109–110, 333, 519
    - photovoltaic system, interactive, 110
    - photovoltaic system, stand-alone, 109
    - picket fencing, 421
    - pickup head, 127
    - picofarad, 177
    - piezoelectric crystal, 280, 518–519
    - piezoelectric transducer, 518–519
    - PIN diode, 328
    - pinchoff, 367, 384–385
    - pip, 148
    - pitch
      - of dots in video display, 576
      - of motion in robot arm, 593
      - of sound, 539
    - pixel, 576
    - plain old telephone service, 416
    - plasma display, 511, 576
    - plastic-film capacitor, 180–181
    - plate, of electron tube, 505–506
    - plate voltage, 509
    - platter, in hard drive, 571
    - playback
      - head, 548
      - mode, 548
    - PM. *See* pulse modulation

P-N junction, 318–321  
 PNP transistor, 352–353  
 polar orbit, 557  
 polarization, wave, 421  
 polarizer, 105  
 polyethylene capacitor, 180  
 polystyrene capacitor, 180, 399  
 portable hi-fi system, 546–547  
 portable operation, in amateur radio, 559  
 position sensor, 603  
 positive feedback, 393–394  
 positive peak amplitude, 152  
 positive temperature coefficient, 96  
 pot core, 168, 291–292, 401  
 potential, 13  
 potential difference, 13, 17, 21, 41  
 potentiometer  
   audio-taper, 91–92  
   linear-taper, 90–91  
   logarithmic-taper, 91–92  
   log-taper, 91–92  
   schematic symbol, 55  
   slide, 90  
 POTS. *See* plain old telephone service  
 power  
   amplification, 384–388  
   amplifier, 391–393  
   apparent, 265–267  
   calculations, 61–62  
   defined, 23–24, 265  
   dissipation, 86–87  
   division of, 64  
   effective, 266  
   effective radiated, 478  
   factor, 269–273  
   forward, 267  
   gain, 357, 380–381  
   input, dc, 386–387  
   instantaneous, 266  
   output, 387  
   peak, 266  
   rating, of resistor, 95–96  
   reactive, 267–273  
   reflected, 267  
   supply, 337–351  
   transformer, 292–294, 337–338

power (*Cont.*)  
   transmission, 273–276  
   true, 267–275, 281  
   units of, 265  
   VA, 268–274  
 prefix multiplier, 24  
 preamplifier, 427  
 preselector, 425, 428  
 pressure sensor  
   back, 524–525  
   capacitive, 523  
   elastomer, 524  
 primary cell, 102–103  
 primary colors, 599  
 primary winding, 286–288  
 printer  
   dot-matrix, 578  
   image resolution of, 578  
   inkjet, 577  
   laser, 577–578  
   thermal, 578  
 product detector, 431–432  
 programmable multiplier/divider, 398  
 propagation, wave, 421–423  
 propane, in fuel cell, 111  
 proton, 3–4  
 proximity sensor  
   capacitive, 524–525  
   photoelectric, 525–526  
 P-type semiconductor, 10, 109–110,  
   317–318  
 P-type silicon, 109–110  
 pulsating direct current, 28  
 pulse modulation, 414–415  
 purely resistive impedance, 235, 259  
 push-pull amplifier, 385–386

## Q

quad antenna, 482–483  
 quad stereo, 547  
 quadraphonic sound, 547  
 quality control, 588–589  
 quarter-wave vertical antenna, 474–476  
 quick-break fuse, 347  
 quiet room, 540

**R**

- rack-mounted hi-fi system, 542
- radar, 527–528
- radial leads, 96
- radial system, ground, 475, 477
- radian, 151
- radians of phase, 151
- radiant-heat detector, 590
- radiation resistance, 281, 471–472
- radio direction finding, 476, 530–531
- radio-frequency
  - amplification, 391–393
  - bands, 423–424
  - field strength, 479–480
  - ground, 477
  - interference, 549–550
  - spectrum, 420
  - transformer, 297–299
- radiolocation, 589
- radioteletype, 408
- RAM. *See* random-access memory
- random-access memory, 499
- range
  - plotting, 602
  - in radar, 528
  - sensing, 602
  - in spherical coordinates, 602
- raster, 418
- ratio detector, 432
- RC. *See* resistance/capacitance
- RDF. *See* radio direction finding
- reach, 595–596
- reactance
  - capacitive, 214–228, 239
  - inductive, 200–213, 239
  - modulation, 413
  - power consumption and, 268
- reactive power, 267–273
- read-only memory, 499
- read-write
  - head, in hard drive, 571
  - memory, 499
- receiver
  - direct-conversion, 425, 432
  - double-conversion, 426
  - receiver (*Cont.*)
    - dynamic range, 425
    - noise figure, 425
    - selectivity, 425
    - sensitivity, 425
    - single-conversion, 426
    - superheterodyne, 425–427
- reception
  - dual-diversity, 434–435
  - wireless, 424–437
- record mode, 548
- recording head, 127, 548
- rectangular coordinate geometry, 595
- rectangular response, 428
- rectangular wave, 145
- rectangular waveguide, 483–484
- rectification
  - full-wave, 28–30, 339–343
  - half-wave, 28–30, 325–326, 338–340, 342–343
- rectifier
  - diode, 325–326, 338
  - full-wave bridge, 339–314
  - full-wave center-tap, 339–340
  - half-wave, 325–326, 338–340
  - selenium, 338
  - silicon, 338
- red/green/blue, 453–454
- reference antenna, 478
- refresh rate, 576
- reflected power, 267
- reflector
  - corner, 485–486
  - in dish antenna, 484–485
  - in helical antenna, 485
  - paraboloidal, 484
  - in parasitic antenna, 482
  - spherical, 484–485
- reference oscillator, 398
- relay
  - normally closed, 125
  - normally open, 125
- remanence, 122–123
- repeater, 554–555
- residual magnetism, 122–123

- resistance
  - calculations, 60–61
  - definition of, 9, 20–22
  - leakage, 218
  - loss, 471–472
  - negative, 233, 330
  - parallel, 63, 72–74
  - radiation, 281, 471–472
  - series, 62, 69–71
  - series-parallel, 64–65
- resistance/capacitance
  - circuit, 523
  - phase angle, 221–224
  - plane, 218–220
- resistance/inductance
  - phase angle, 207–210
  - plane, 203–205
- resistance/inductance/capacitance circuit, 248–250, 257–258, 277
- resistivity, 20
- resistor
  - bleeder, 87–88
  - carbon-composition, 88–89
  - carbon-film, 89–90
  - color code, 96–98
  - coupling, 88
  - current-equalizing, 346
  - defined, 9, 21
  - film-type, 89–90
  - fixed, 88–90
  - impedance-matching, 88
  - integrated-circuit, 90
  - metal-film, 89–90
  - ohmic value of, 95
  - power rating of, 95–96
  - schematic symbol, 55
  - temperature-compensated, 96
  - tolerance, 95–96
  - uses of, 85–88
  - wirewound, 89
- resolution
  - in high-definition television, 418
  - image, 513, 576
  - in slow-scan television, 471
  - optical, 598
  - of printer, 578
- resolution (*Cont.*)
  - sampling, 415–416
  - of scanner, 579
- resonance
  - definition of, 247, 252, 276
  - parallel, 252, 277–278
  - series, 247, 276–278
- resonant devices, 280–282
- resonant frequency, 276–278
- resonant notch, 493–494
- resonant peak, 493–494
- response time, in meter, 39
- retarding field, 513–514
- retentivity, 122–123
- retracing, 417
- reverse bias, 320, 330, 354–355
- revolute geometry, 596–597
- RF. *See* radio-frequency
- RFI. *See* radio-frequency interference
- RGB. *See* red/green/blue
- rheostat, 92–93
- rim drive, 549
- ringer, 124
- ripple frequency, 340
- rise, in sawtooth wave, 146
- robot
  - arm, 588–589, 592–597
  - generations, 591–592
  - hearing, 597
  - leg, 603
  - navigation, 600–602
  - vision system, 597–600
- robotics, Asimov's three laws of, 592
- roll, of motion in robot arm, 593
- roller inductor, 396
- ROM. *See* read-only memory
- root-mean-square amplitude, 153
- rotating vector, 190–192
- rotation, 596
- rotor, in variable capacitor, 182
- “rounding-off bug,” 71
- R-S flip-flop, 447–448
- R-S-T flip-flop, 447–448
- RTTY. *See* radioteletype
- RX plane, 233–236

**S**

- safety, in electrical work, 30
- sampling
  - rate, 416
  - resolution, 415–416
- satellite
  - geostationary, 424, 556–557
  - low-earth-orbit, 424, 556–557
  - medium-earth-orbit, 557–558
  - radio, 542
  - systems, 424
- saturation
  - of bipolar transistor, 354–356
  - in color image, 453
  - of ferromagnetic core, 166
- sawtooth wave, 146–147
- scan
  - converter, 434
  - mode, 543
- scanner
  - basic features, 579
  - configurations, 579–580
  - optical character recognition and, 579–580
- schematic diagram, 55–57
- schematic symbols, 55–56
- scrambling, 547
- secondary cell, 102–103
- secondary electron, 512
- secondary winding, 286–288
- second-generation robot, 592
- sector, in hard drive, 571–572
- security
  - in communications, 561–565
  - levels of, 561–563
- seek mode, 543
- selective squelching, 433
- selectivity, of receiver, 425
- selenium
  - chemical element, 316
  - rectifier, 338
- selsyn, 521
- semiconductor
  - capacitor, 181–182
  - definition of, 9–10
  - diode, 318–319
  - intrinsic, 328
  - semiconductor (*Cont.*)
    - materials, 316–317
    - N-type, 10, 109–110, 317–318
    - P-type, 10, 109–110, 317–318
- sensitivity
  - optical, 597–598
  - receiver, 425
- sensor
  - back-pressure, 524–525, 603
  - capacitive, 523, 524–525
  - definition of, 523
  - elastomer, 524
  - photoelectric, 525–526
  - position, 603
  - proximity, 524–526
  - texture, 526–527
- sequencing code, 437
- sequential logic gate, 446
- serial
  - data port, 580
  - data transmission, 450
- serial-access storage, 573
- serial-to-parallel conversion, 450–451
- series circuit
  - power distribution in, 74
- series-parallel resistances, 64–65
- series resistances, 62, 69–71
- series resonance, 247
- series-tuned Colpitts oscillator, 396–397
- servo, 599
- shape factor, 428
- sheet scanner, 579
- shelf life, 104
- shell, electron, 4–5
- shell method, 289–290
- SHF. *See* superhigh frequency
- shortwave
  - band, 560
  - listening, 560–561
- shoulder, 596
- shunt resistance, 40–41
- sideband, 409–412
- signal
  - comparison, 529–531
  - mixing, 327
- signal-plus-noise-to-noise ratio, 425
- signal-to-noise ratio, 425

- significant figures, 60  
 silicon  
   chemical element, 316  
   rectifier, 338  
   steel, 288  
 silver-mica capacitor, 399  
 silver-oxide battery, 106  
 silver-oxide cell, 106  
 sine function, 144  
 sine wave, 28, 144–145, 188–199, 400–401, 539  
 single-conversion receiver, 426  
 single sideband, 410–412, 432  
 sky wave, 421–422  
 slide potentiometer, 90  
 slope detection, 429, 432  
 slow-blow fuse, 347  
 slow-scan television, 416–417  
 small-scale integration, 498  
 smoke detector, 587–588  
 solar cell, 48–49, 333  
 solar flare, 117  
 solar panel, 108, 109, 333  
 solar sail, 115  
 solar wind, 115–116  
 solenoid, 92, 124, 168  
 solenoidal core, 290  
 solid, 7–8  
 sonar, 529–530  
 sound-level meter, 48  
 sound power, 94  
 source  
   of field-effect transistor, 365  
   follower, 375  
 south geomagnetic pole, 115  
 south magnetic pole, 119  
 space, 408, 430  
 spacecraft cell, 108  
 spark coil, 161  
 speaker, 543  
 spectral display, 409–410  
 spectrum analyzer, 327  
 spell checker, 580  
 spherical coordinates, 602  
 spherical reflector, 484–485  
 spread-spectrum communications, 437  
 square wave, 145  
 squelch, 433  
 SSB. *See* single sideband  
 SSI. *See* small-scale integration  
 SSTV. *See* slow-scan television  
 stand-alone photovoltaic system, 109  
 static current amplification, 356  
 static electricity, 11–12  
 static triggering, 447  
 stator, in variable capacitor, 182  
 step angle, 520  
 step-down transformer, 286  
 stepper motor, 520–521  
 step-up transformer, 286  
 stereoscopic vision, 598  
 storage capacity, 103–104  
 standing waves, 275–276  
 stray inductance, 171  
 streaming audio, 542  
 stroke, 12  
 subscriber, 452  
 substrate, 109–110, 365–366  
 superheterodyne receiver, 425–427  
 superhigh frequency, 424  
 superimposed dc, 153–154  
 supply reel, 548  
 surface wave, 421  
 surge  
   current, 342, 346  
   impedance, 236  
   protector, 346  
 susceptance  
   capacitive, 238–239  
   definition of, 238  
   inductive, 238–239  
 switching diode, 328  
 synchro, 521–522  
 synchronized communications, 435  
 synchronous flip-flop, 447–448
- T**  
 T flip-flop, 447  
 tachometer, 522  
 take-up reel, 548  
 tantalum capacitor, 181–182

## tape

- analog audio, 548
- digital audio, 549
- drive, for computer, 573
- hiss, 541
- magnetic, 127
- player, 542
- recording, 127

TB. *See* terabyte

## telechir, 603–605

## telemetry, 603–605

## teleoperation, 521

## telephone modem, 580

## telepresence, 603–605

## television

- broadcast channels, 433
- digital cable, 434
- digital satellite, 418
- fast-scan, 417–418, 433–434
- high-definition, 417–418
- reception, 433–434
- slow-scan, 416–417

## temperature coefficient

- of capacitor, 184
- negative, 96, 184
- positive, 96, 184
- zero, 184

## temperature-compensated resistor, 96

## temperature compensation, 96

## terabyte, 449, 570

## terahertz, 144

## terminal node controller, 451–452

## terminal unit, 408

## terminals, schematic symbol, 57

## tesla, 32, 120

## tetrode tube, 507

## texture sensor, 526–527

## theoretical current, 72

## thermal heating, 39

## thermal printer, 578

## thermocouple, 39

## thermocouple principle, 39

## third-generation robot, 592

## three-wire ac system, 345

## threshold detector, 319

threshold of hearing, 48, 94,  
540

## timbre, 539

## time domain, 148, 194, 327

## time-division multiplexing, 436

## timer, 496

TNC. *See* terminal node controller

## tolerance, 43, 95–96, 184

## tone control, 389, 544

## toroid, 92, 167–168, 290–291

## toroidal core, 290–291, 402

## touch pad, 520

## track, in hard drive, 571–572

## trackball, 520

## tracking, in receiver tuning, 428

## transceiver, 554–555

## transconductance, 370

## transducer

- acoustic, 597
- displacement, 519–523
- dynamic, 517–519
- electrostatic, 517–518
- infrared, 519
- piezoelectric, 518–519
- visible-light, 519
- wave, 517–519

## transformer

- audio-frequency, 294
- balanced-to-unbalanced, 295
- balun, 295
- broadband, 297
- core method of winding, 289–290
- coupling, 295, 389–391
- E-core, 289–290
- ferromagnetic cores in, 288–289
- for impedance matching, 275, 294–299
- interwinding capacitance of, 289–290
- iron, 288
- pot-core, 291–292
- power, 292–294, 337–338
- primary winding of, 286–288
- principle of, 286–289
- radio-frequency, 297–299
- ratings, 337–338
- secondary winding of, 286–288
- shell method of winding, 289–290
- solenoidal-core, 290
- step-down, 286, 337
- step-up, 286, 337



transformer (*Cont.*)  
 toroidal, 290–291  
 transmission-line, 297–299  
 turns ratio of, 287–288  
 unbal, 295  
 unbalanced-to-balanced, 295  
 volt-ampere capacity of, 337–338

transient  
 definition of, 346–347  
 suppressor, 346

transistor  
 amplifier, basic, bipolar, 381–382  
 amplifier, basic, field-effect, 382–383  
 base of, 353  
 battery, 106  
 biasing, 353–357, 367–370  
 bipolar, 352–364, 425  
 collector of, 352  
 drain of, 365  
 emitter of, 352  
 field-effect, 365–378, 425  
 gate of, 365  
 NPN, 352  
 PNP, 352–353  
 power, 344–345  
 size of, 315  
 source of, 365

transistor-transistor logic, 496–497

transmatch, 299, 474

transmission line  
 balanced, 295  
 characteristic impedance of, 236–238  
 coaxial, 170, 236–237, 423  
 dielectric in, 170  
 as inductor, 169–170  
 mismatched, 275  
 parallel-wire, 169–170, 236–237, 423  
 power measurement in, 274  
 section, 280  
 standing waves in, 275–276  
 transformer, 297–299  
 unbalanced, 295  
 velocity factor of, 280  
 waveguide, 423, 483–484

transmission, wireless, 407–418

transponder, 589

treble, 389, 538

triangular wave, 146–147

trimmer capacitor, 183

triode tube, 507

tropospheric propagation, 422

true power, 267–275, 281

TTL. *See* transistor-transistor logic

TU. *See* terminal unit

tube. *See* electron tube

tubular capacitor, 180

tuned-circuit coupling, 390–391

tuned power amplifier, 391–393

tuner, 543

tuning control, 393

tunnel diode, 331

turns ratio, 287–288

turntable, 542

TV. *See* television

tweeter, 543

twin T oscillator, 401

twinlead, 295

two-wire ac system, 345

## U

UHF. *See* ultrahigh frequency

ULSI. *See* ultra-large-scale integration

ultrahigh frequency, 417, 424, 513

ultra-large-scale integration, 498

ultrasonic motion detector, 591

ultraviolet, 49

unbal, 295

unbalanced load, 295

unbalanced transmission line 295

uninterruptible power supply, 107

Universal Serial Bus, 573

unwanted inductance, 171

upper sideband, 410–411

UPS. *See* uninterruptible power supply

USB. *See* upper sideband, Universal Serial Bus

user identification, 565

username, 582–583

## V

VA power, 268–274

vacuum tube, 504

valve. *See* electron tube

Van de Graaff generator, 11  
 varactor, 321, 413  
 variable capacitor, 182–184  
 vector  
   admittance, 241  
   in complex-number plane, 232  
   definition of, 191  
   diagram, 195  
   direction of, 191, 232  
   impedance, 234  
   magnitude of, 191, 232  
   in RC plane, 219–220  
   rendition of sine wave, 195  
   in RL plane, 205  
   rotating, 190–192  
 velocity factor, 170, 280, 419–420  
 vertical antenna, 473–476  
 vertical polarization, 421  
 very high frequency, 417, 424  
 very low frequency, 424  
 very-large-scale integration, 498  
 vestigial sideband, 417  
 vidicon, 511  
 VHF. *See* very high frequency  
 vinyl disk, 549  
 visible-light transducer, 519  
 vision system, 529, 597–600  
 VLF. *See* very low frequency  
 VLSI. *See* very-large-scale integration  
 voice-pattern recognition, 565  
 volatile memory, 499  
 volt, 17–18  
 voltage  
   calculations, 60  
   conservation of, 77  
   divider network, 78–80  
   doubler, 341  
   gain, 357, 380  
   instantaneous, 29, 189  
   loop, 276  
   negative-going, 188–189  
   node, 276  
   peak inverse, 338  
   positive-going, 188–189  
   regulation, 328–329, 341, 344–345,  
   495

voltage (*Cont.*)  
   spike, 346  
   transient, 346–347  
   Zener, 319  
 voltage-controlled oscillator, 330, 397  
 volt-ampere, 45, 265–266, 268–273, 337–338  
 voltmeter, 41–42, 44–45  
 volt-ohm-milliammeter, 44  
 volume  
   control, 389  
   of sound, 540  
   unit, 47–48  
 volume unit, 47–48  
 volume-unit meter, 47–48  
 VU. *See* volume unit

## W

watt, 23–24, 265  
 watt rms, 47  
 watt-hour, 25–27, 46  
 watt-hour meter, 46  
 wattmeter, 45, 387  
 wave  
   complex, 147–148  
   composite, 151  
   cosine, 189  
   derivative of, 189  
   irregular, 147–148  
   polarization, 421  
   propagation, 421–423  
   rectangular, 145  
   sawtooth, 146–147, 539  
   sine, 28, 144–145, 188–199, 400–401,  
   539  
   sky, 421–422  
   square, 145, 539  
   surface, 421  
   transducer, 517–519  
   triangular, 146–147, 539  
 waveform, 539  
 waveguide, 423, 483–484  
 wavelength versus frequency, 419  
 weak-signal amplifier, 391  
 weber, 32, 120  
 Weston standard cell, 103–104

wet cell, 14  
white noise, 391  
wideband FM, 413  
wire-equivalent security, 561–562  
wireless  
  eavesdropping, 561  
  network, 558–559  
  router, 559  
  tap, 561, 564–565  
wireless-only encryption, 563–564  
wiredtapping, 561  
wirewound resistor, 89  
wiring diagram, 56–57  
woofer, 543  
word, in Morse code, 408  
work envelope, 594  
World Wide Web, 583  
wrist, 596–597

**X**

XOR gate, 445–446

**Y**

Yagi antenna, 482–483  
yaw, of motion in robot arm, 593

**Z**

Zener diode, 328–329, 344–345  
Zener voltage, 329  
zepp antenna, 473–474  
zero beat, 425  
zero bias, 353–354  
zero temperature coefficient,  
  184  
zinc-carbon cell, 105

*This page intentionally left blank*

## About the Author

Stan Gibilisco is one of McGraw-Hill's most prolific and popular authors. His clear, reader-friendly writing style makes his books accessible to a wide audience, and his experience as an electronics engineer, researcher, and mathematician makes him an ideal editor for reference books and tutorials. Stan has authored several titles for the McGraw-Hill Demystified library of home-schooling and self-teaching volumes, along with more than 20 other books and dozens of magazine articles. His work has been published in several languages. Booklist named his *McGraw-Hill Encyclopedia of Personal Computing* one of the "Best References of 1996," and named his *Encyclopedia of Electronics* one of the "Best References of the 1980s."