03-Shyles-45161.qxd 1/5/2007 7:30 PM Page 33

Chapter 3

How Television Works

Long before computers became common to home and workplace, the telegraph, telephone, radio, television, and communication satellites, among other media marvels, put people instantly in touch with one another around the globe. Working together, these devices make global communication possible; in combination, they enable us to encode, transmit, store, retrieve, and display information in the forms of data, text, live-action images, and high-fidelity sounds, often in real time, thereby enriching communication. In contrast, the computer by itself is merely a data storage, retrieval, and processing device, utterly incapable of providing the communication functions we have come to expect from our media systems.

This chapter explains the development and function of traditional broadcasting systems that have enabled us to communicate instantly through radio and television across continents for more than a century. The chapter also explains how these systems are linked with computers through our telecommunications network, resulting in an infrastructure that makes streaming video, program sharing, and distribution possible through the Internet and e-mail channels. To understand what lies ahead in the video production field, one must understand how traditional broadcasting works and also how broadcasting is becoming integrated with computer and telecommunication systems to form our digital media network.

BROADCASTING AND THE SIGNIFICANCE OF CODE

The desire to communicate from afar is part of human nature. Long before radio and television enabled us to broadcast sounds and images instantly around the globe, we invented less powerful means to send messages to distant places. For example, we invented the megaphone, which extends the reach of the human voice, but not greatly. Other methods of communicating long-distance have included beacons, semaphore flags, drums, smoke signals, and telegraphy.

What these systems share is that they all rely on prior agreements (or codes) between senders and receivers about what various signals will mean. For example, anyone who does not know the sounds or letters associated with flag positions in semaphore will not get the message even if the flags can be clearly seen. It is the code or pattern of intelligence conveyed by the flags, not the view of the flags themselves (the physical carrier), that makes it possible to convey messages. In short, clear reception of the carrier is a necessary but not a sufficient condition for successful transmission of meaning. Successful communication relies on both unimpeded reception of a message's physical component and accurate decoding of the pattern of information (or intelligence) it contains. Of course, it is still possible to misinterpret messages after they are received, but cultural issues of meaning are not even considered until an encoded message is received and decoded.

How is intelligence carried in a message system? A common feature of communication is the need to vary some aspect of a signal to encode information. A pattern of some kind must be crafted into some physical form for a message to be generated, stored, transmitted, received, and consumed. And all patterns require some form of variation or change.

On a simpler level, consider again communication using semaphore flags. If the sender of semaphore flag signals fails to move the flags (no encoding), no message is sent even if the flags can be clearly seen. Similarly, in Morse code, if the telegrapher were to send nothing but dots at regular intervals, there would be no information to decode since there is nothing in Morse code associated with an endless series of dots.

In the technical jargon of radio and television broadcasting, the term for creating a pattern of intelligence through variation is modulation. The term modulation is synonymous for imposing a message (pattern, change, or variation) on a carrier.

Among the most pervasive, rapid, and successful systems ever developed for communicating at a distance are radio and television broadcasting. To explain the process of sending audio/visual messages via broadcasting, I first describe the physical nature of radio energy, which makes broadcasting possible, and then describe how audio signals and televised scenes are encoded, transmitted, received, and decoded. After describing traditional broadcasting, I focus on significant developments over the past half century that have extended its reach and range, including satellite broadcasting and cable television, both of which have advanced traditional broadcasting without changing its analog nature. I then describe how the more recent transition from analog to digital platforms, integrating broadcasting with computers and telecommunication networks, has brought a cornucopia of new video products, services, and opportunities to both producers and consumers.

THE PHYSICAL NATURE OF RADIO ENERGY

Among the physical phenomena that make broadcasting possible is the propagation of radio waves, or electromagnetic radiation, through space. At the simplest level, rotating a loop of copper wire in a magnetic field generates radio energy. Such rotation induces an electric current in the wire. As the wire passes through each full rotation, the intensity and direction of the flow of electrons varies in an orderly manner called a sine wave (see Figure 3.1). Figure 3.1 indicates that sine waves produced by continuous rotation feature several characteristics, which, we will see later, are also present in sound and light waves. These include **frequency** (the number of cycles per second, or cps), **period** (the time it takes for one cycle to occur), **amplitude** (the magnitude of voltage at its greatest intensity), **wavelength** (the length of one cycle in meters), and **phase** (the difference between the same points on different waves). In a vacuum, radio waves travel at the speed of light, about 186,000 miles per second, or 300,000,000 meters per second.

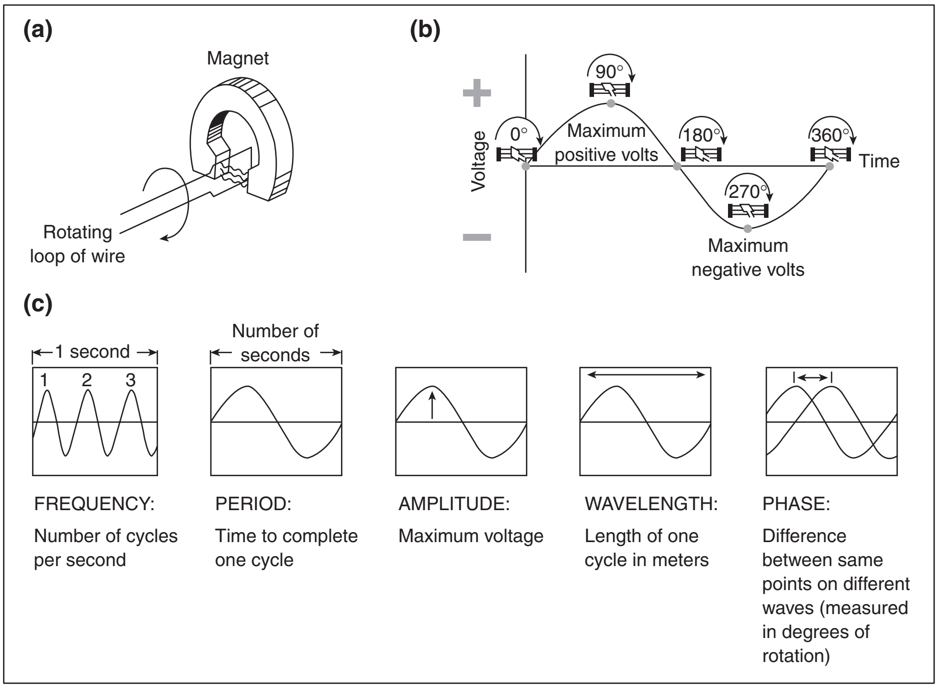

Figure 3.1 The basic sine wave of radio energy. (a) The wave is produced by a loop of wire rotating in a magnetic field. (b) One cycle of a sine wave, as the loop goes through a full (360-degree) rotation. (c) Properties of the sine wave.

PROPAGATING RADIO WAVES

In 1819, the Danish scientist Hans Oersted, while experimenting with electrical effects of magnetic fields, discovered that magnetism and electricity were related. By 1831, Michael Faraday had discovered induction, the ability of an electric current in a wire to create a similar current in a nearby wire without physical contact between them. Based on Faraday's discovery, Joseph Henry developed the first efficient electromagnet.

In 1837, Samuel F. B. Morse used Henry's discoveries about electromagnets to patent a long-distance telegraph system using electrical signals to encode messages. This method was a significant improvement over optical telegraphy systems in use at the time, which depended on telescopes and clear weather to send messages. Morse's electrical telegraphy system was more powerful and reliable than the optical systems then in use. It worked under more varied weather conditions and could send messages farther, more quickly, and more reliably than its predecessors.

As electrical telegraphy developed, it was observed that some "leakage" of electricity from telegraph wires appeared to magnetize some nearby metallic objects. This phenomenon was explained in 1865 by the English physicist James Clerk Maxwell, who presented evidence that electrical impulses emitted from wires traveled through space in a manner similar in form and speed to light waves. Maxwell called them **electromagnetic waves.** Thomas Edison tried to capitalize on this leakage phenomenon to send telegrams to people aboard moving trains. Unfortunately, the waves sent into the atmosphere by the telegraph wires were a chaotic mixture of signals leaking from other wires in the area, making the patterned dots and dashes from any particular message unintelligible.

The problem of how to separate electromagnetic waves from one another was solved by the German scientist Heinrich Hertz. In 1887, Hertz demonstrated that an electromagnetic wave using an oscillating circuit could be propagated and detected amid other waves. An oscillating circuit produces an electric current that changes direction at a stable frequency. An example of an oscillating circuit (albeit a relatively slow one compared to radio frequencies) is that found in a typical American household electrical outlet, which supplies alternating current (AC) at 60 cycles per second. In honor of Hertz's discovery, the unit called a hertz (abbreviated Hz) was adopted in the 1960s as a synonym for "cycles per second."

It was soon confirmed that a radio wave, when propagated at a stable frequency, does not mix with waves of other frequencies. In 1895, the Italian scientist Guglielmo Marconi sent the first wireless telegraph message. These early wireless messages were in the form of Morse code, using the simplest modulation technique—namely, an interrupted continuous wave (ICW). In this method of radio modulation, a continuous, alternating current, made up of a succession of identical sine waves, is broken into a series of pulses corresponding to the dots and dashes of Morse code. This is done simply by opening and closing a circuit for relatively short or long periods to turn the radio wave on or off. Thus, radio energy was used for the first time as the physical material to carry a pattern of intelligence to encode information. Although ICW is still widely used, it is limited in that it does not vary enough to carry sounds, such as music or speech. Eventually, advances in digital technology would make it possible to store enough pulses of information in binary code (patterns of 0s and 1s) to render sounds and/or images on CDs, videodisks, and computers. In Marconi's day, however, further advances were needed to permit broadcasting of audio signals.

CONVERTING SOUND INTO ELECTRICAL ENERGY

Alexander Graham Bell made possible the advance from Morse code to the sending of an electrical replica of the human voice (voice modulation). In 1876, Bell invented the telephone, which makes a current of electricity vary with changing sound waves generated by the human voice. The telephone transmits a pattern of electricity that faithfully matches a pattern of sound waves made by speech. How does this happen?

A telephone mouthpiece uses a microphone to convert sound waves (vibrations in the air) into a matching pattern of electric current. To do this, sound waves created by the voice are directed onto a thin metal diaphragm, which vibrates according to a pattern of sound waves imposed on it. A metal diaphragm (typically a thin disk of aluminum) forms the top of a cylinder containing carbon particles that can conduct electricity. When sound waves enter the mouthpiece, they cause the aluminum to vibrate so that the carbon particles are rapidly squeezed and loosened. When electricity flows through the cylinder, the current increases and decreases as the carbon particles are squeezed and released. Loud sounds cause sound waves to press hard on the diaphragm, compressing the carbon particles tightly, making it easier for electric current to flow, thus increasing the amount of electricity passing through the circuit. When the sound is low, less pressure is exerted on the carbon particles, allowing them to remain more loosely packed and making it harder for current to pass, resulting in a smaller current. In this way, the current passing through the circuit matches the pattern of sound waves striking the diaphragm. If it is a close match, an accurate replica, we call it **high fidelity** (*fidelity* means faithfulness to the original). This process of changing (modulating) sound waves into patterns of electricity is termed *transduction,* and the telephone is therefore a **transducer.**

At the receiving end, how is the electrical pattern transformed back into sound (demodulated)? The telephone is equipped with an earpiece that has a diaphragm that can freely vibrate in and out. In the center of the diaphragm is a coil of wire acting as an electromagnet. A permanent magnet surrounds the electromagnet, supplying a force against which the electromagnet pulls. As the incoming current varies in strength, so does the magnetic force of the electromagnet. Magnetic forces surrounding the diaphragm cause it to vibrate at the same rate, vibrating the surrounding air. The sound waves generated by this motion create a replica of the original sound. Figure 3.2 diagrams this process.

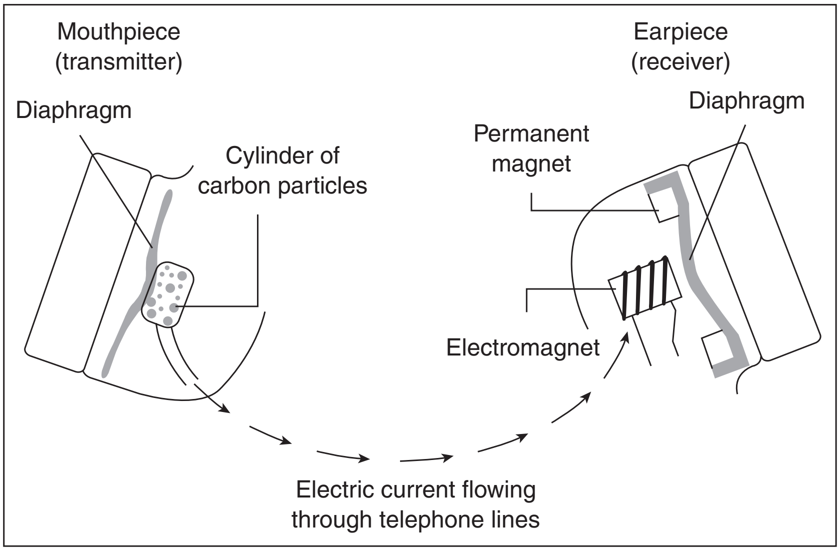

Figure 3.2 Operation of the telephone, a simple transducer. Sound entering the mouthpiece vibrates a metal diaphragm atop a cylinder of carbon particles through which an electric current is passing. This vibration produces a pattern of electric current that replicates a pattern of the sound waves. At the receiver end, the incoming current creates variations in the strength of the earpiece's electromagnet. These variations cause the receiver diaphragm to vibrate, reproducing the original sound.

The encoding and decoding processes in microphones and loudspeakers work essentially the same way as in the telephone. Standard radio and television microphones, though, are sensitive to a fuller range of the audio spectrum and therefore have higher fidelity than those found in telephones. Likewise, radio and television speakers have more power and fidelity than telephone earpieces.

MODULATING RADIO WAVES WITH AUDIO SIGNALS

The telephone makes it possible to project an electrical version of the human voice through long distances over wires and then to recover a replica of the original sound from the transmitted electricity. It soon became possible to modulate radio waves in a similar way without wires. This change resulted from the work of two electrical engineers, England's Sir John Ambrose Fleming and America's Lee De Forest.

Attenuation and Amplification

Sound waves, like radio waves, naturally dissipate as they move farther away from their source. As distance increases, the strength of a wave decreases. This phenomenon is called *attenuation.* To picture this process, imagine the effect of dropping a stone into a pond of still water. The stone causes circular waves of water to move away from the point where it hits, and as the waves move outward, they weaken. At some distance, the original disturbance attenuates to such a degree that the water remains undisturbed by the original splash.

In Fleming's day, it was already well known that electron motion produces current in a closed circuit. In the language of electrical theory, Fleming knew that a voltage applied to a metal wire conducts electrons. What Fleming discovered, however, was that an electrode inside an evacuated heated filament lamp (a glass vacuum tube) could also conduct an electric current. Fleming noticed a one-directional current between the heated filament (called the *cathode*) and the positive electrode (known as an anode or *plate*). Because it contained two elements, Fleming called the device a *diode.*

De Forest extended Fleming's work by interposing a thin metal open-meshed *grid* between the heated filament and the anode. When a separate voltage was fed to the grid, De Forest could control the magnitude of electricity flowing from the cathode to the plate. With a grid, De Forest obtained a large voltage change at the plate from just a small voltage change on the grid. Thus, by introducing a third element to Fleming's diode, De Forest's *triode* made it possible to amplify weak radio signals received from distant radio transmitters. Figure 3.3 diagrams the triode vacuum tube, the original heart of radio amplifiers. Since the 1950s, successive

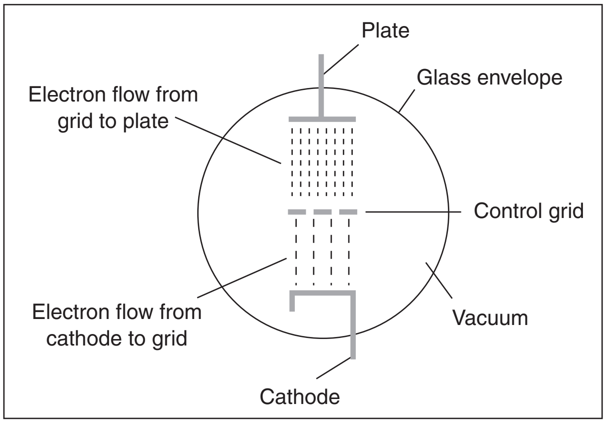

Figure 3.3 A triode vacuum tube solved the problem of amplifying radio signals. Small voltage changes on the control grid modify the electrical flow from the cathode to the plate.

generations of solid-state technologies (transistors, semiconductors, integrated circuits, and microprocessors) have replaced vacuum tubes, but the principles of amplification are the same in both tube and solid-state technologies.

Modulating the Carrier

By feeding an electrical signal converted from sound waves to the grid of a triode, relatively weak audio signals could be amplified enough to be used for radio transmissions. However, before sound waves could be transmitted to distant points without wires, the amplified audio signal had to be superimposed onto a radio frequency (RF) carrier. This is because sound waves are pressure waves and do not propagate across space at the speed of light like electromagnetic radio waves.

The RF carrier is created with an *oscillator,* an electronic circuit that produces a sine wave at a specific frequency. The RF carrier may then be modulated or made to vary by an audio signal (voice or other information) superimposed on it. In other words, the pattern imposed on the RF carrier is sound, converted into an electrical signal, supplied by a microphone or some other audio source (e.g., a CD or cassette tape).

The two most common techniques of modulating a radio wave are amplitude and frequency modulation. When an audio signal modulates the amplitude of a carrier, the process is called **amplitude modulation (AM).** When the audio signal modulates the frequency of a carrier, the process is called **frequency modulation (FM).** In AM radio, the carrier consists of a sine wave whose amplitude is made to copy the variations of an audio source. In FM radio, it is the frequency of the carrier wave that is changed by an audio source. Figure 3.4 illustrates these two common types of voice modulation in radio broadcasting.

As it turns out, FM modulation is superior to AM because it produces better fidelity with much higher noise immunity. For example, auto ignition noises and high-tension lines can cause hum and static on AM signals because those disturbances can adversely affect the amplitude of the received carrier. By contrast, FM signals are generally not affected by such impulse noises in the atmosphere.

Transmitting the Carrier

Audio signals imposed on RF carriers may be further amplified. Finally, they are fed from a transmitter to an antenna for propagation. In standard AM transmission, the range of frequencies used for radio carriers is between 535 and 1,705 kilohertz (abbreviated kHz, meaning thousands of hertz). Each channel is allocated a frequency range (or bandwidth) of 9 kHz to operate in. This means there is enough space in the radio spectrum allocated for 130 AM radio channels in any given area.

Roughly speaking, radio waves propagate in all directions unless they are intentionally altered from this pattern. The effective coverage area can radiate for

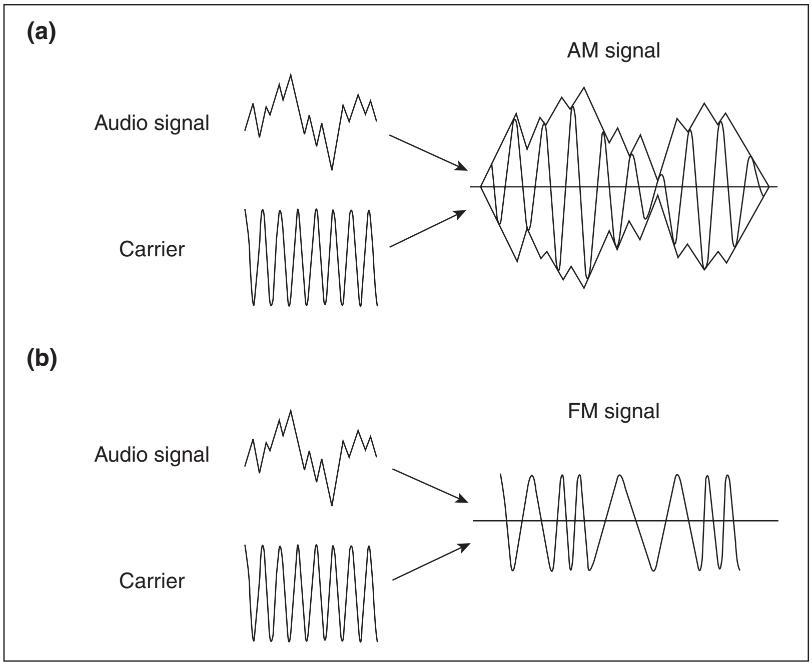

Figure 3.4 AM and FM signals. (a) In AM transmission, the audio signal modulates (varies) the amplitude of the carrier wave. (b) In FM transmission, the audio signal modulates the frequency of the carrier wave.

miles surrounding the transmitting antenna, making it possible for millions of radio sets in a coverage area to receive a signal. However, because radio signals attenuate as distance increases, they must be amplified at the receiver to make them strong enough to drive a speaker.

Demodulating the Carrier

The function of a radio receiver is to tune into a particular frequency from among those available, detect the modulated carrier operating at that frequency, and remove the audio signal from the carrier. This part of the process is known as **demodulation.** The isolated audio signal is then amplified and directed to a speaker so that the original audio information can be heard. A block diagram of the demodulation process is presented in Figure 3.5.

So far, we have provided a basic model of how audio information is transmitted via radio energy to distant points and then recovered. But how does radio energy broadcast motion images? Some preliminary facts set the stage for an explanation of the process of video transmission.

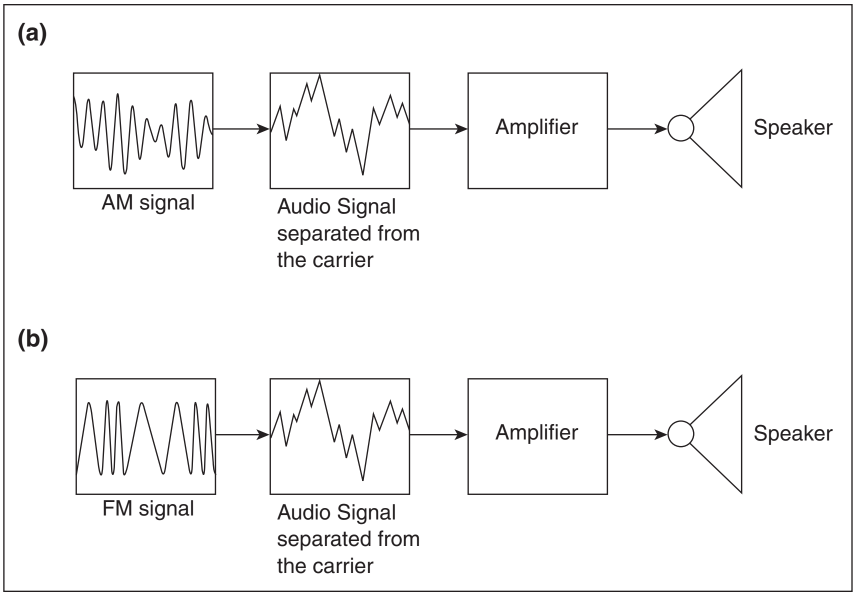

Figure 3.5 Demodulation of (a) AM signals and (b) FM signals.

CHANNEL SPACE

In using radio energy to transmit sound plus full-motion images, a greater portion of the radio spectrum (bandwidth) is needed than for sound alone. This is because there is a lot more information present in motion images plus audio than in audio alone. The need for greater bandwidth to transmit greater amounts of information is analogous to a fire department using larger diameter hoses than those used by homeowners in their gardens to deliver a greater amount of water per given unit of time.

To accommodate television's need for greater bandwidth, whereas American broadcasting allocates 9 kHz per channel for standard AM radio, television bandwidth is more than 660 times larger, or 6 MHz (6,000 kHz) per video channel. This means that one television channel contains enough bandwidth to accommodate more than 600 AM radio stations.

Determining how much radio spectrum would be allocated for each television station was done after a great deal of technical debate and testing by the National Television System Committee (NTSC). The NTSC's first objective was to suggest technical standards that would permit an acceptable level of picture quality or **resolution.** With enough resolution, the video image would be clear, convincing, and aesthetically pleasing. However, the NTSC also wanted to conserve spectrum space, using no more than necessary for each channel assignment.

The NTSC rightly viewed the radio spectrum as a limited natural resource, which it continues to be today, even though technological developments have increased its usable range. Despite these increases, the race for bandwidth by new technologies (satellites, cell phones, digital applications, etc.) is unrelenting, making it essential to allocate its use wisely.

The job of the NTSC was tricky because any increase in image detail requires a commensurate increase in bandwidth for each channel. Unfortunately, every increase in channel bandwidth reduces the total number of channels in a given portion of the spectrum.

As it turned out, the standard bandwidth for each television channel, adopted in 1941, was 6 MHz. This allowed 4.5 MHz for the AM-modulated video signal, a complex video waveform (explained later) including synchronization, scanning, blanking, and, eventually, color information. The remaining 1.5 MHz provided a guard band or buffer between adjacent channels operating in the same geographic area, to reduce interference, and space for transmitting the FM-modulated audio portion of the television signal. Figure 3.6 diagrams these original features of the television channel. Over time, ancillary signals have been embedded into existing television channels to provide supplementary services (e.g., closed-captioning for the hearing impaired, and so on). In addition, further portions of the radio spectrum have been allocated to accommodate satellite transmissions, digital video, and a spate of data, text, and interactive services.

It is interesting to note that amplitude modulation is used for the video portion and frequency modulation for the audio portion of the television signal. This is

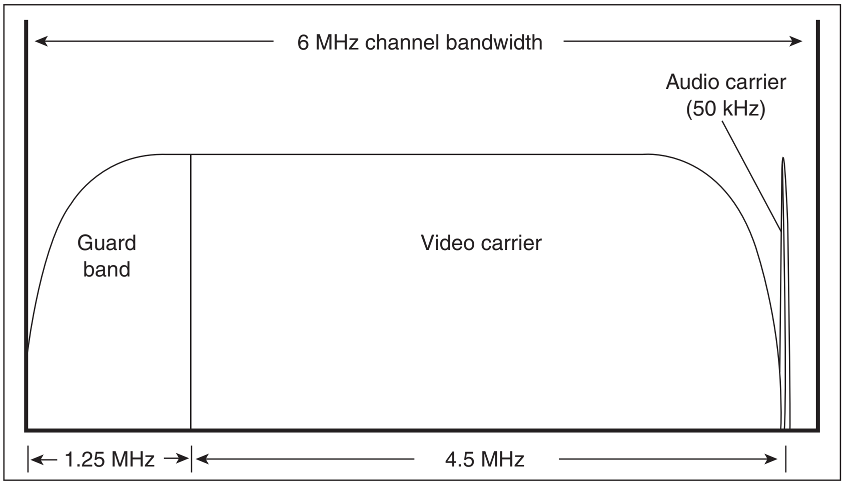

Figure 3.6 Audio and video portions of a standard 6-MHz television channel.

because FM is less subject to noise and interference than AM, making it less subject to static and therefore more suitable for audio reception. Furthermore, AM is better suited for video transmission because it exhibits fewer problems caused by multipath reception of the signal. Multipath reception occurs when the same signal reflects from obstacles such as buildings and bridges, reaching a receiving antenna from more than one path. Because the distance traveled by multipath signals is usually different, different parts of the signal arrive at the antenna at the same time. For AM signals, this causes less severe interference at the television receiver than would occur if the signals were FM.

CONVERTING LIGHT INTO ELECTRICAL ENERGY

Just as telephone and radio technologies harness natural qualities of electricity and electromagnetic radiation to transmit voice-modulated audio signals, television relies on natural phenomena of photoelectric effects, including photoconductivity and photoemissive effect, to convert light into, and back from, electrical energy.

Photoconductivity

To change light into electricity, video depends on **photoconductivity,** which occurs when light on some metals increases the flow of electricity in those metals. One of the earliest examples of photoconductivity was observed in 1873 with the metal selenium. When selenium was used in an electrical circuit, the current through it increased during exposure to light. Unfortunately for video applications, selenium responds too slowly to light to be useful for replicating natural motions. But luckily, cesium silver and other silver-based materials are excellent for such applications.

Photoemissive Effect

In the **photoemissive effect,** discovered by Hertz in 1887, visible light results from some materials' exposure to energy that may not be visible to the eye. Sources of such energy include streams of electrical energy or photons of higher-than-visible light energy, such as ultraviolet rays or X-rays. The photoemissive effect is similar to that seen in radium dials once used to make watch faces glow in the dark.

In the picture tube of a television receiver, the inside of the screen is coated with fluorescent material. When a stream of electrons strikes the screen, it glows because of the photoemissive effect. As the stream of electrons is made stronger, the portion of the screen struck by the electron stream glows more brightly. When the stream is made weaker, the glow decreases. If the stream can be modulated in accordance with the darker and brighter portions of a scene focused by the lens of a television camera, that scene can be rendered on the screen. If the re-creation process can be done quickly enough, then smooth motion can be rendered convincingly.

In monochrome (black-and-white) television receivers, the fluorescent material needs only to be able to glow with a range of brightness roughly proportional to the intensity of the stream of electrons hitting it; color is of no consequence—only brightness variations are important. However, in color television, materials that glow with different colors when streams of electrons hit them must be used. To understand this process, let us begin with the major components of the monochrome television system.

MONOCHROME VIDEO

Television cameras (see Figure 3.7) use a lens system to focus light from a scene into a **pickup tube** or, in microprocessor systems, a **charge-coupled device (CCD).** The pickup tube or CCD is the place where light reflected from a scene is converted into an electrical signal. The output is then amplified and fed to external circuits for recording, routing to closed-circuit locations, broadcast from a transmitter, or transmission via cable or satellite.

Within a studio complex featuring more than one camera, each camera is connected to a **camera control unit (CCU).** The CCU enables a technician to adjust and match camera operation for all cameras to eliminate jarring differences in how they render the same scene. In a television studio, camera operators can immediately view the video signal routed to the **viewfinder** of each camera.

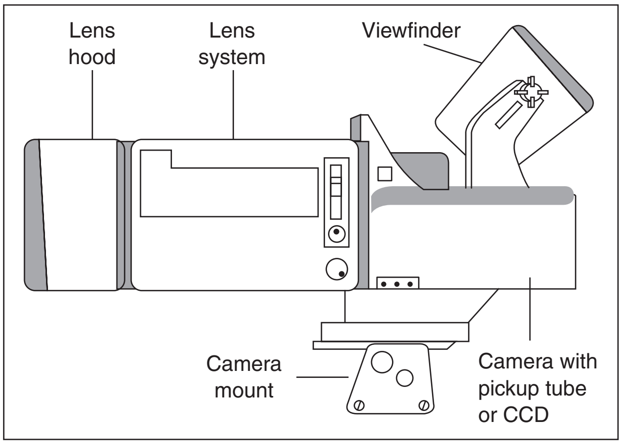

Figure 3.7 The most basic parts of a video camera.

SCANNING

Transmitting all the details of a given picture simultaneously over the same circuit would lead to a chaotic mixing of signals in a single output, resulting in an unintelligible product similar to what jigsaw puzzles tend to look like when they are dumped from their boxes. Such visual chaos is analogous to what Edison faced when he tried to send intermixed wireless telegraph signals to receivers aboard moving trains.

To maintain the fidelity of the original image seen by the camera when it is received by a television set, small areas of the picture are converted into discrete magnitudes of electric current matching the brightness information present in each portion, and then each is sent out in order. This is done so each picture element (**pixel**) can be received and converted into light without being confused with any others.

In theory, we could create a separate circuit for each area of the screen and then send all of the information at once. But such a method is impractical because it would require hundreds of thousands of separate circuits for just one video channel of NTSC video (one for each pixel) (see Photo 3.1 of the Bell Telephone receiver of

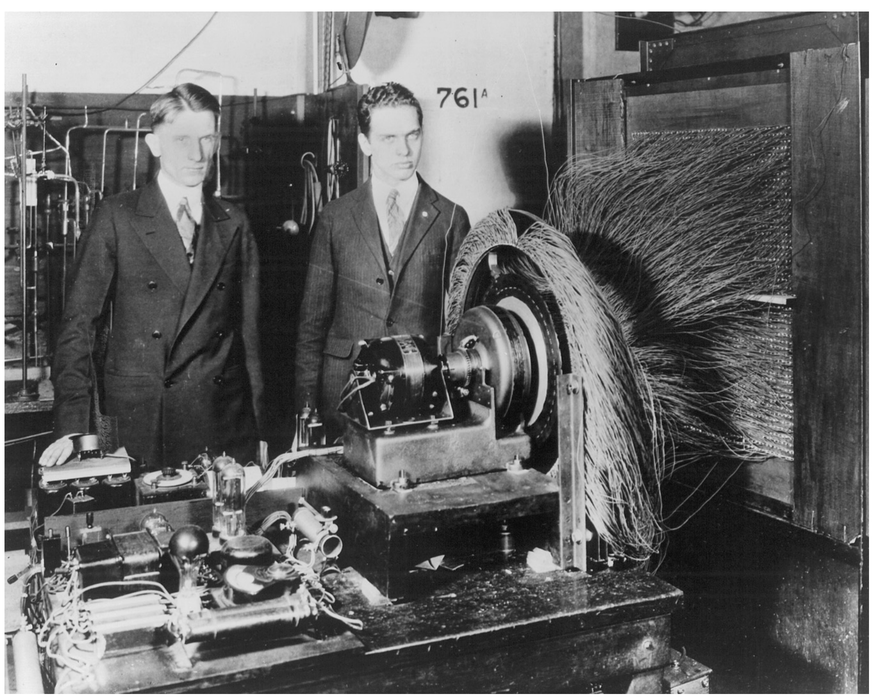

Photo 3.1 The Bell Telephone television receiver of 1927, which used thousands of separate circuits to compose a picture. What a nightmare! The impracticality of such a device prompted development of the electronic scanning method shown in Figure 3.8.

SOURCE: Property of AT&T archives. Reprinted with permission of AT&T.

1927). Instead, a *scanning* method is used to transmit the brightness information for each pixel in turn. Scanning makes it possible to use just one circuit per channel.

The original monochrome video system converted picture information into electrical signals by focusing light onto a mosaic of pixels, each composed of an individual cesium silver globule. In such a system, when a scene to be televised was focused on the mosaic, electrons became stored in each pixel in magnitudes roughly proportional to the intensity of light focused on each one. Stored electrons were then instantly attracted by an anode in the camera tube, leaving the mosaic with a copy of the original scene in the form of varying amounts of electrical charge.

In American broadcasting, the traditional NTSC video mosaic is currently composed of 525 horizontal lines, containing about 211,000 pixels. An electron gun is used to scan each line from left to right, top to bottom, in an orderly fashion. As the electron beam passes each pixel, it replaces electrons lost to the anode, enabling the video signal to exist in an external circuit. This signal is then coupled to video amplifiers for immediate transmission.

Interlaced Scanning

The human visual system detects flicker, a source of severe eye fatigue, below about 45 image presentations per second. To defeat flicker problems, the film industry has adopted a standard film speed of 24 frames per second, each illuminated twice, for a rate of 48 presentations of picture information per second. For television, a system called interlaced scanning is used to avoid flicker problems (see Figure 3.8).

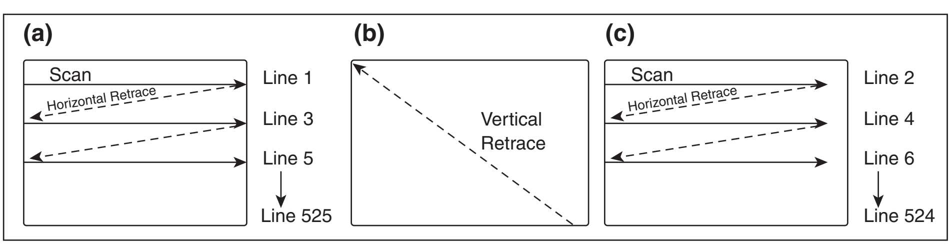

Figure 3.8 The NTSC video image is composed of 525 lines of picture information delivered 30 times per second using *interlaced scanning.* In interlaced scanning, (a) an electron beam first scans the 262.5 odd-numbered lines of the screen. (b) The beam then scans the 262.5 even-numbered lines. (c) After completing all 525 lines, a vertical retrace brings the beam to the top-left position to start the entire process all over again. During retrace, the electron beam is turned off to eliminate spurious illuminations of the screen. Each scan of 262.5 lines is called a *field* and takes 1/60 of a second. Each scan of all 525 lines is called a *frame* and takes 1/30 of a second. Therefore, NTSC scanning delivers 30 complete pictures per second of interlaced video.

Interlaced scanning takes advantage of **persistence of vision** or the tendency for an image to persist for a short period of time after a stimulus is no longer physically present to our eyes. In interlaced scanning, instead of having 525 successive sweeps of the screen, two separate scans of 262.5 lines are used. An electron beam alternately scans the odd-numbered lines of the 525 and then the even-numbered lines, thus creating the illusion of covering the entire field twice. This arrangement defeats the flicker problem, resulting for practical purposes in the appearance of smooth motion.

Each successive scan of 262.5 lines is called a **field.** Because line frequency (normal wall current, or AC power) in the United States is 60 Hz, it is convenient to scan each field in 1/60 of a second. As a result, 60 fields per second are televised, a rate fast enough to eliminate flicker. Each complete scan of all 525 lines, or two successive fields, is called a **frame.** Thus, in the traditional NTSC video system, 30 frames per second are televised.

Electromagnet coils surrounding a fixed cathode ray tube (CRT) inside the camera control the scan of the electron beam across each line. As the gun projects a stream of electrons at the tube face, varying magnetic forces generated within the coils bend it along its path. In this way, the camera performs its work without using any mechanical parts. This makes the scanning process extremely reliable.

In general terms, each time the beam finishes a line, it returns to the extreme-left position, but shifted downward to the next odd or even line, to begin scanning again. This move back is called *horizontal retrace.* When the beam finishes scanning the last line, it returns to the top-left position to begin the entire process over again. This move back is called *vertical retrace.* During each retrace, the electron beam is turned off to eliminate spurious illuminations. The signal to turn off the electron beam is called the *blanking* signal. The vertical blanking interval (VBI) and *retrace* signals, along with the synchronization information needed to keep the receiver precisely in step with the transmitter, are embedded in the overall television signal.

In reality, the VBI reduces picture detail, such that only 483 lines of the 525 transmitted are ultimately delivered with viewable picture information. However, it is during the VBI and in some other parts of the video signal that additional text and information services (e.g., closed-captioning) have found a home since the NTSC established technical standards for American television.[1]

RECEIVER OPERATION

A television set receives video, audio, and all ancillary signals needed to replicate the original televised scene and audio information. It has a loudspeaker, a phosphor-coated picture tube, an electron gun, circuits for synchronization and scanning purposes, and currently, with increasing frequency, additional equipment for receiving specialized services (i.e., set-top boxes, translators, converters). Regardless of tube size, the standard NTSC **aspect ratio** of tube height to width is three units by four units (4:3), respectively. As with the television camera, the neck of the picture tube is fitted with magnetic deflection coils that control the direction of an electron beam. The beam scans horizontal paths across the picture tube's phosphor coating.

When a television signal is received, the sound component (transmitted as FM) is routed to circuits where it is demodulated and sent to a loudspeaker. The video or AM portion of the signal is routed to the picture tube, where it directs the electron beam to emit electrons in amounts roughly in proportion to the brightness levels of the original scene. As the electron beam sweeps across the face of the picture tube, its varying intensities cause variations in the brightness of the phosphors, replicating the original scene. To synchronize the video signal so that pixels can be reassembled without mixing them up, deflection coils around the neck of the picture tube are fed horizontal and vertical sync pulses from the original video signal. These pulses control the deflection of the electron beam across the screen, thus keeping the receiver in step with the original signal.

COLOR TRANSMISSION AND RECEPTION

Color television broadcasting began after the monochrome system was already in place and millions of black-and-white sets were in use. This made it desirable to find a color system compatible with monochrome technology (hence economic and marketing constraints were at work on even the most basic engineering decisions from the beginning). To make color television compatible with monochrome transmission, color information was added to the monochrome signal without changing the 6-MHz bandwidth set aside for each TV channel. In addition, both black-and-white and color receivers were made capable of receiving both monochrome and color signals (a requirement called *backward compatibility*). This meant transmission had to be virtually identical for both monochrome and color systems.

Chrominance, Luminance, and Saturation

To transmit **chrominance** (color or hue) information, the color camera's optical system separates the light entering it into three primary colors: red, blue, and green. It is a fortunate characteristic of human vision that virtually any color can be reproduced from these additive primary colors. Furthermore, any colored light can be specified with only two additional qualities: **luminance** (or brightness) and **saturation** (or vividness). Saturation can be thought of as the degree to which a color is free from impurities, such as dilution by white light. Low-saturation colors are paler, whereas highly saturated colors are more pure and vivid.

In early color cameras, light was broken into its primary color components with filters and a set of dichroic mirrors. A dichroic mirror passes light at one wavelength while reflecting light at other wavelengths. Today, most color cameras use a prism block called a *beam splitter* to break light into its primary colors.

Once the light has been split, the separate light beams are directed into three separate pickup tubes for processing into video signals. When a CCD microprocessor is used, a silicon lattice absorbs different wavelengths of light at different depths to distinguish colors. In either case, the patterns of electrical voltage generated in an external circuit match the levels of the original pattern of light received by the camera.

Some cameras use a single imaging element with a filter to separate incoming light into its component values. Others use filters to separate light into only two colors, as well as additional microprocessors to assign values to the third color needed to reproduce the colors the camera is seeing.

In color cameras, video signals from the three pickup tubes or the CCD are combined to produce a signal containing all of the picture information to be transmitted. Signals are combined using a phase-shifting technique so that they can be transmitted in one video channel and then retrieved without confusion. The overall signal contains the audio and picture information as well as blanking and synchronization pulses, and so on. This colorplexed video signal modulates the video carrier for transmission to receivers.

Black-and-white television sets treat the color portion of the video signal as if it were part of the intended monochrome transmission. To avoid degraded reception, the scanning motions are used to mask the chrominance signal. This way, any pixels brightened by interference during one line scan are made to darken by an equal amount in the next line scan. The net effect of chrominance signal interference over successive scans is thus virtually eliminated.

The tube in the color receiver contains three electron guns that project separate beams, which deflect simultaneously in the standard interlaced scanning pattern over the face of the picture tube. One of the guns projects the red color signal, one projects blue, and the third projects green. The screen of the receiver is coated with phosphor dots that glow red, blue, or green when struck by a stream of electrons. The phosphors are uniformly distributed over the face of the picture tube, arranged in adjacent groups of three dots that form tiny triangles, each containing a phosphor dot for each color. The dots are so small that a single one cannot be distinguished by the viewer's eye. The color of any one triangle is the additive function of the varying intensities with which each dot in the triangle is made to glow by the strength of the electron beam hitting it. Virtually any color may be rendered with this method. If electrons from all three guns strike their respective dots in a triangle with the right intensity, the color of that triangle will appear white. If no electrons strike a trio of dots in a triangle, the color of that triangle will be black. In this way, black-and-white images are possible on a color receiver.

To ensure that the electron beams from the red, blue, and green guns hit only phosphor dots that glow red, blue, and green, respectively, a metal plate called a *shadow mask* is inserted close to the phosphor coating between the electron guns and screen (see Figure 3.9). The plate is pierced with more than 200,000 holes and is positioned so that it masks two of the dots in each triangle from being hit by unwanted electrons. In this way, the electron beams are confined to the phosphor dots of the proper color.

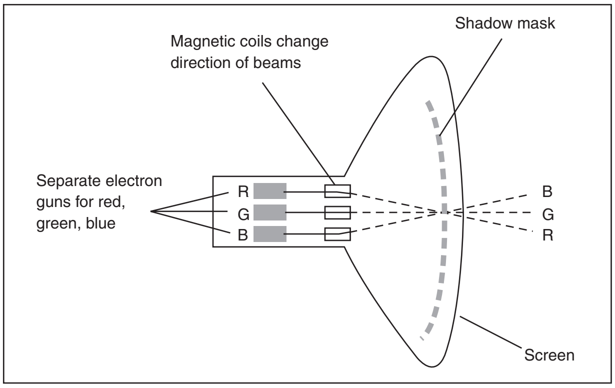

Figure 3.9 Diagram of the traditional color television receiver, showing how the shadow mask keeps the separate electron beams targeted at the proper points on the screen.

SIGNAL TRANSMISSION

Television signals may be propagated over the air from terrestrial antennas for distribution within local television markets, or they may be distributed across the country via telephone long lines using coaxial or fiber-optic cables. When such facilities are not available, convenient, or cost-effective, microwave relay links may be used to distribute television programs. Sometimes microwave delivery is not feasible due to distance, power, terrain, or other limitations; in such cases, video transmissions can be sent using satellite uplinks (covered later in this chapter). Since the 1960s, the development and growth of microwave, cable, and satellite technologies have made live television transmission and reception possible from almost any location on Earth.

The camcorder and desktop editor have more recently put production capability into the hands of the audience. Videotape recorders, computer memory chips, and video servers now enable receivers to store programs for later use and permit audiences the luxury of watching programs repeatedly and at their convenience. Special effects and digital graphics processing permit virtually endless enhancement and manipulation of video images. High-quality audio and multiple speaker configurations offer stereo and surround sound for consumers' home entertainment systems. Interactive systems enable users to engage in dialogue with program providers and with one another. Projection and big-screen video as well as flat-screen displays, including both plasma display panels (PDPs) and liquid crystal displays (LCDs), now influence homebuilders and realtors to feature entertainment theater space as a selling point in marketing homes. The remote control continues to influence the way we watch television, as well as influencing the way programmers think about how to capture our attention.

Since the advent of television, technical developments have continued to make television more engaging than it was in its infancy. Yet the core of the system still uses radio energy to broadcast television signals. However, the picture is changing right before our eyes. We are witnessing a profound transition from analog to digital platforms, as well as a convergence of broadcasting, computer, and telecommunication technologies. What is the result? An explosion of interactive information and entertainment services to American households and beyond. The new order enables us to originate our own programming if we wish. The age of interactive **digital television (DTV)** is upon us. What are some of the implications of this change?

THE ADOPTION OF A DIGITAL TELEVISION (DTV) STANDARD

On December 24, 1996, the Federal Communications Commission (FCC) announced its decision to adopt a digital television standard for a free, universally available digital broadcast television service for America. Originally, the main goal of developing an advanced television system was to provide America with higher quality video images, known as **high-definition television (HDTV or HD)**. However, rapid development of digital technologies has expanded the objectives of public interest groups, computer and television manufacturers, telecommunication providers, cable and satellite interests, filmmakers, broadcasters, and others interested in enhanced video services.

Beyond HD, the DTV age now implies expanded applications to include movies on demand, telephone and computer data delivery, interactive programming, distance learning, paging systems, home shopping and e-commerce applications, video production and editing, and so on.

Understanding the public policy and economic interests behind DTV development is key to understanding the technical configuration of the new system. Among the goals of the FCC was to adopt a "world leading digital broadcast television technology" that would

1. put more video program choices into the hands of American consumers,
2. provide better quality audio and video than that available with the NTSC system,
3. provide innovative services due to data transmission capability, and
4. offer compatibility (interoperability) between video and computer technology to spur innovation and competition.

To achieve these goals, the FCC began inquiries in 1987 into the potential for advanced television (ATV) services. At that time, industry research teams suggested more than 20 systems. In February 1993, after determining that any system to be adopted would be fully digital, four such systems were considered.

In May 1993, seven companies and institutions representing the four remaining systems formed a "grand alliance" to develop a "best of the best" single system to present to the FCC for approval. Over the next two and a half years, a final digital system was developed, tested, documented, and recommended to the FCC. In December 1996, the system was approved.

The Advanced Television System Committee (ATSC), composed of a 54-member group including television workers, television and film producers, trade associations, television and equipment manufacturers, and segments of the academic community, has endorsed the newly adopted DTV system as "the best digital broadcast television system in the world." The ATSC has characterized the system as having unmatched flexibility and ability to incorporate future improvements. However, some industry parties have voiced objections about having the government impose the standard. Some at the time questioned whether it might be better to allow market forces to dictate standards rather than have the government intervene. Some suggested having the government issue standards only for spectrum allocation, transmission, and reception, to avoid interference problems, but to leave all other conditions (e.g., frame rates, number of scanning lines, aspect ratio of the screen) open. Ultimately, the FCC decided that letting market forces determine standards would lead to the development of incompatible systems that would be too costly to consumers who might have to invest in several different receivers to gain access to different programs. Incompatible systems might also require the use of set-top boxes, translation devices, and other interface hardware and software that might slow down encoding, transmission, and decoding of data streams, thus degrading the efficiency of the entire system. In addition, the FCC reasoned that a government-mandated standard would be the best way to guarantee universal access to broadcasting services for all Americans. The FCC viewed broadcasting as unique, free, and available to nearly every American who relies on it as a primary source of information and entertainment. Because of these characteristics, the FCC reasoned that the goals of certainty and reliability take on a special significance and strengthen the case for the adoption of a government-imposed DTV standard. Finally, the FCC reasoned that allowing different standards to develop might make the conversion process from the current analog system to a fully digital service more difficult. For these reasons, letting the market drive the selection of a standard was rejected.

To make the DTV standard as industry-friendly as possible, the FCC invited standards to be developed by industry parties. In this way, it was believed the DTV system that developed would better reflect industry needs. For this reason, the standard is called "voluntary."

Characteristics of the New Standard

Like the NTSC television format, the new DTV standard calls for each television channel to occupy a 6-MHz bandwidth. To fit the more complex digital signal demands of DTV (at times with many times the picture resolution of the current NTSC format) into the same space used for current analog signals, digital compression techniques (described in more detail later) are used. However, unlike the NTSC format that uses only interlaced scanning of 525 lines at 30 frames per second on a screen three units high by four units wide, DTV remains relatively flexible on these dimensions.

For example, to promote compatibility (interoperability) with other services, including rerunning archives of NTSC programs, newer telecommunication and computer-based media, and film formats, the DTV standard can broadcast and receive both interlace-scanned programs and those produced in a new noninterlaced scanning format called **progressive scanning.** In progressive scanning, each line is scanned in order, with no skipping, at a maximum rate of 60 frames per second (double the current NTSC frame rate).

The new system accommodates both the traditional NTSC horizontal line format as well as some newer ones. Currently, the NTSC format is fixed at 525 lines of pixels distributed in a rectangular arrangement. In this design, the distance between pixels is greater horizontally than vertically. However, in the new system, a maximum of 1,080 horizontal lines of pixels will be featured in a square arrangement; that is, pixels will be equally spaced in horizontal and vertical directions. As a result, new receivers will be compatible with both NTSC format programs as well as many computer displays.

The two new line formats in the DTV standard include one with 720 horizontal lines per frame and one with 1,080 lines per frame in a 16:9 aspect ratio of width to height.

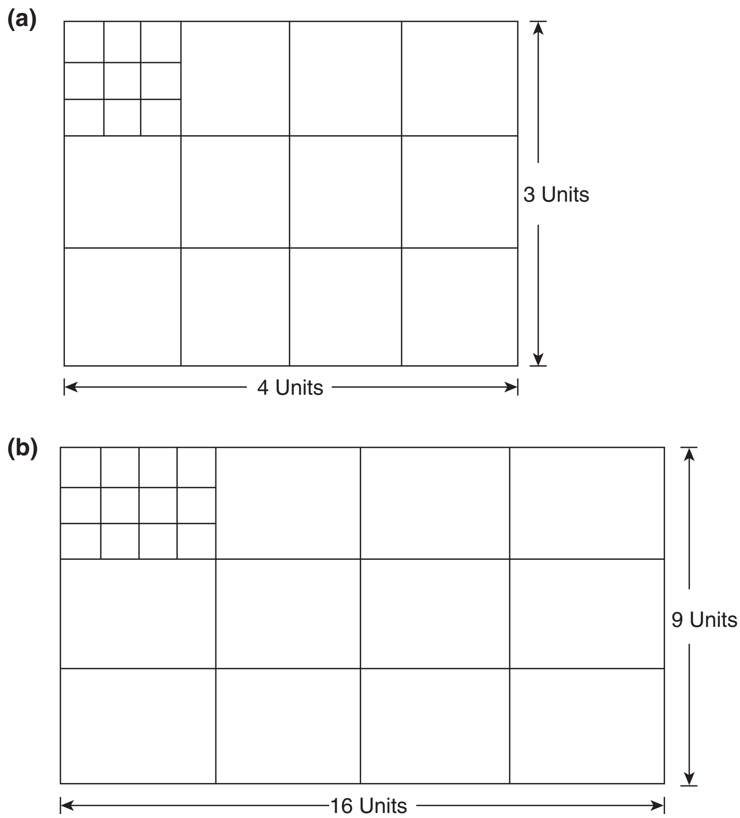

Figure 3.10 Traditional (NTSC) and DTV aspect ratios of the television screen. (a) The traditional NTSC 4 × 3 aspect ratio (4 units wide by 3 units high), also referred to as 1.33:1 units of width to height. (b) The expanded 16 × 9 screen size, also referred to as 1.78:1 units of width to height. Notice that the expanded size of the DTV screen is the square of the NTSC standard ($4^2 \times 3^2$). The expanded width is nearly equal to the 1.85:1 screen format used in the movie industry.

The 16:9 aspect ratio and square pixel arrangement means that when 720 horizontal lines are being scanned (not counting those lost to blanking and retrace), 1,280 vertical lines of pixels are used, for a total of 921,600 pixels potentially contributing to the overall video image. Similarly, when 1,080 horizontal lines are used, 1,920 vertical lines of pixels are used, for a total of 2,073,600 pixels potentially contributing to the overall image. These numbers are 5 to 10 times greater than those associated with the NTSC format and help convey the added picture resolution available from the new system. Several frame rates are also available under the DTV standard, including 24, 30, and 60 frames per second, making DTV more compatible with film, NTSC video, and computers.[2]

Finally, the 16:9 aspect ratio provided by DTV is more compatible with the format used in many films produced throughout the world, and with flexible letterboxing capability, presenting programs produced in aspect ratios different from the 16:9 format becomes easy. **Letterboxing** is a technique used to preserve the original aspect ratio of a film by blacking out portions of the screen, usually at the top and bottom. Film content is not cut from the frame. With letterboxing, the complete frame is transmitted, and no parts of the picture are left out.

In addition to these characteristics, new system capabilities include the following:

1. Layering of video and audio signals that enables multiplexing and transport of different programs simultaneously over the same channel. For example, layering makes it possible to broadcast two HD programs at the same time or "multicast" (transmit multiple data streams) several standard definition television (SDTV) programs at a visual quality better than that currently available. Current estimates claim that more than five such programs or dozens of CD-quality audio signals can be multicast simultaneously.
2. RF transmission.
3. Rapid delivery of large amounts of data (e.g., the contents of the daily newspaper could be sent in less than two seconds).
4. Capability for interactive transmission of educational materials.
5. Provision for universal closed-captioning for the deaf.

With all of these developments, it is clear that DTV will continue to expand the power, pervasiveness, and influence of television. As new configurations become available, it will become increasingly important for message makers and consumers to understand how new devices may be used to reach and influence audiences and how audiences will use them for their own ends.

Satellite communication, cellular telephones, and microwave links carrying faxes, e-mails, and computer databases all depend on wireless transmission of modulated radio signals to connect distant users. Without these infrastructures, millions now on the wireless network would be isolated from one another. As long as we use radio energy to send data representing sounds and images across space at the speed of light, it will be necessary to know how broadcasting works to have a full understanding of digital media. Understanding how broadcasting systems operate is one of the essentials.

COMPUTERS AND TELECOMMUNICATIONS IN VIDEO

In addition to understanding how broadcasting systems operate, it is also essential for you to understand how computers and telecommunications systems contribute to the video production process. Perhaps most obvious is the role of computers in creating special effects in postproduction editing (for both audio and video). Simply put, virtually all video editing is now done on computers, from network shops to small independent producers; even Hollywood has adopted digital editing for first-run films.

Less obvious is the role of these systems in video distribution, another phase of postproduction. In this application, distant users can share computer files of video programs, over both wire and wireless channels, with a computer, a fiber or phone connection, and an e-mail address. While this raises important copyright issues and underscores the vulnerability exhibitors face from media pirates, the convenience of such options makes the use of computers (combined with telecommunication channels) for rapid distribution of media products too attractive to ignore.

Computers are now also central to shooting video. For example, in the production phase, digital cameras record sounds and images as digital data, through an encoding process known as **sampling** and **quantizing** (explained later). In addition, because digital cameras record video and sound as data files, programs can be sent to computer servers immediately from multiple locations either for broadcast or further processing (i.e., editing) without losing signal quality; files can also be sent over any size channel (**broadband** or **narrowband**) either in one continuous transmission or in discrete bursts, again without signal loss. Analog video either falls short or can't do any of these things.

Besides offering videographers greater reliability, digital cameras are also lighter in weight than their analog counterparts, have longer battery life than ever before, and require lower light levels to record air-quality material. In marketing terms (a concern during every phase of production), digital encoding offers additional advantages, including the ability to embed extra information about programs regarding production dates, ratings, personnel involved, additional content (i.e., outtakes stored as separate chapter content), and other information designed to attract or advise consumers.

Clearly, computers and telecommunications channels are involved in every phase of video production from shooting and editing to distribution and marketing, making it critical to know as much as possible about how these technologies contribute to the process. The remainder of this chapter explains how computers and telecommunications infrastructures work, in both conceptual and physical terms relevant to the production setting.

Computers

How do computers enable users to create (*encode*) and display (*decode*) video information? In computers, the main substance acted on by electrical signals for coding and storing video information is silicon, a naturally occurring element.[3] When treated with other materials, silicon can be used to encode, store, and manipulate any kind of information, including video images and sound. To explain how computers make this possible, you should know first how **binary code** may be used to represent any kind of information and then, from both conceptual and physical perspectives, how computers may be used to make, store, and retrieve television programs.

Using binary code to render and store information predates modern digital computers by centuries. Simply put, binary code uses just two symbols to record information. Remember, it is essential to *vary* some aspect of a signal for it to carry a message. This is because some variation or modulation is required to impose a pattern of intelligence on any medium; otherwise, no message can be encoded.[4]

The Binary Number System

Around 1900, telegraphers began using Morse code to communicate to distant places without wires (*wireless telegraphy*) by propagating patterns of radio energy in long and short bursts to represent the letters of the alphabet. This was one of the earliest uses of binary code in a broadcast setting.

Beyond language functions, binary code can also be used to express any numerical value and perform mathematical calculations. The ability of binary code to capture both verbal and mathematical ideas is extraordinarily important because it shows how *any type and amount of information that can be expressed may be rendered into numerical code using only 0s and 1s.* In other words, although the binary number system uses only two symbols, it is completely versatile. For example, decimal numbers (so called because they use 10 symbols, from 0 through 9) may be expressed in binary form, as shown in Figure 3.11.

Notice that when the supply of digits runs out, decimal numbers move one place to the left, where they are used all over again in a new column at an increased power of 10. In binary, the same practice is followed, but with one difference: When the numbers move over, they increase in value by powers of only 2. So, as Figure 3.11 shows, the number 10 in the decimal system is expressed as 1010 in binary, where the digit “0” on the far right tells us there are no 1s, the digit next to it tells us there is one 2, the next that there are no 4s, and the last on the far left that there is one 8, for a total of 10 (decimal), and so on.

Notice that as places move from right to left in the decimal system, decimal values increase tenfold (i.e., place values go from 1s to 10s to hundreds to thousands, etc.), whereas when places move in the binary system, values increase by powers of only 2 (i.e., from 1s to 2s to 4s to 8s, etc.). Nevertheless, using binary code, it is possible to represent any number.

Binary code is difficult for human beings to use because long strings of repeating 0s and 1s can quickly challenge our perceptual systems, as the binary number representing the number 256 illustrates (see Figure 3.11). But it is just the opposite

Decimal		**Binary**
1 (decimal)	=	1 (binary)
2	=	10
3	=	11
4	=	100
5	=	101
6	=	110
7	=	111
8	=	1000
9	=	1001
10	=	1010
16	=	10000
32	=	100000
64	=	1000000
100	=	1100100
256	=	100000000
Etc.		

Figure 3.11 A sample of decimal numbers with their binary equivalents in each row.

for computers, which are highly compatible with a two-state system of variation, in part because electrical impulses may easily be turned on and off with the flick of a switch, just like a light bulb.

In addition to using circuits with on-off electrical signals to represent alphanumeric expressions, we can also use them to represent audiovisual content in the form of *bi*nary digi*ts* (or *bits,* as known in the computer industry). Two principles are of interest here: First, information can be processed as a series of yes-no choices in terms of binary digits (0s and 1s); second, such information can be simulated in an electrical circuit. In terms of technical advancement of computers during the early 20th century, as fate would have it, the start of World War II precipitated intense, independent computer technology development programs on both sides of the Atlantic.

How is binary code used to encode text and video? I focus first on the encoding of alphanumeric characters because so much content is available in textual form in television and other media industries (i.e., streaming video content on Web sites and, obviously, print media). Then I describe how video content may be rendered into computer code, as well as the hardware that enables today's computers to store, retrieve, and display such content.

ASCII: Why "1" Is a Beautiful Number in the Computer World

At any given moment, a binary digit (bit) has the capacity to store only one of two pieces of information, expressed as a 0 or a 1. To increase capacity, more bits are needed. So, for example, if you want to record the flavor of a cake as either chocolate or vanilla, one bit is all you need: a *0* could stand for *chocolate,* a *1* for *vanilla.* But to characterize the cake further, say in terms of both flavor (chocolate or vanilla) and type of icing (i.e., butter vs. butter free), two bits of code are required. That's because four states or conditions are in play, and that is the number that can be accommodated with two bits of code: The first bit can be 0 while the second bit is 0 (say, a chocolate cake with butter icing), the first bit can be 1 while the second bit is 0 (vanilla with butter icing), the first bit can be 0 while the second bit is 1 (chocolate/butter free), or both can be 1s (vanilla/butter free). In like manner, if three bits are used, eight conditions can be accommodated; in other words, the capacity to characterize a unique combination of conditions jumps to eight with three bits of code, as shown in the eight rows of Figure 3.12 (perhaps the column on the far left designates sugar vs. sugar free).

0	0	0
0	0	1
0	1	0
0	1	1
1	0	0
1	0	1
1	1	0
1	1	1

Figure 3.12 Eight unique combinations are possible with three bits of code.

Notice the trend. The number of unique possibilities or sets of conditions that can be characterized as the number of bits increases is equal to $2n$, where n is the number of bits. So, if there are three bits in use, the number of unique conditions that can be handled (the *capacity* of three bits) is 2^3 or 8. With four bits, 16 unique conditions can be captured, and so on.

The Implications of This Trend for Accommodating Text

The relationship between the number of bits of code and the capacity to code text is critical since many graphical characters must be displayed, including 26 letters in the English alphabet (52 if you count uppercase and lowercase), 10 digits, and a variety of punctuation marks (commas, periods, dollar signs, etc.).

In addition, a number of control commands provide syntax to control the layout of the graphical content, including spacing between words and sentences, tab and backspace commands, and so on. Clearly, four bits of code are not enough to handle all that since more than 16 characters are included in just one case of the alphabet alone. How many bits should be set aside for alphanumeric text?

In 1967, this question was answered in the United States when the American Standard Code for Information Interchange, or ASCII (rhymes with "gas key"), was developed. It calls for a seven-bit code standard for alphanumeric characters. With seven bits available, you can manipulate 128 unique items of information, which is enough capacity to accommodate all the letters of the alphabet (both uppercase and lowercase), all 10 digits, punctuation marks, and so on.

Adopting a single standard simplifies information exchange among computers. By using ASCII code, computers can share information without translation (a huge advantage). That is why the heading of this section calls 1 a beautiful number in the computer industry: *An agreed-on standard simplifies information sharing.*

At the time ASCII was developed, computer memory was both limited and costly, but it was obvious that a six-bit code lacked the capacity to handle alphanumeric text. So a seven-bit architecture was adopted. As it turns out, however, almost all computer systems today are based on an eight-bit architecture (each eight-bit chunk is called a *byte*); that is, information is stored in chunks of eight-bit bytes, even ASCII code.

Capturing Sound and Video With Binary Code

ASCII code is great for coding text, but it is too limited for handling sound or images. How is binary code used to capture sound?

Sound is an analog phenomenon; that is, it is a continuous stream of information, usually made up of waves of compressed air that cause the generating element of a microphone or your eardrums to vibrate sympathetically. In analog recording, physical sound vibrations are converted into a pattern of electrical signals matching the original stimulus (a process called *transduction*).

By contrast, in digital recording, or in the conversion of the voice, say, for cellular telephone transmission, binary numbers are used to represent the varying fluctuations of electricity generated by sound waves through a process of *sampling* and *quantization.* To do this, the pattern of electricity matching sound signals is converted into a digital data stream using an *analog-to-digital converter.* How does this happen?

In the sampling phase of the analog-to-digital conversion, a circuit captures instants of sound at rapid intervals on the order of thousands of times per second (see Figure 3.13). Each unit captured (each *sample*) is then converted to a number according to its amplitude at that particular moment. The number associated with

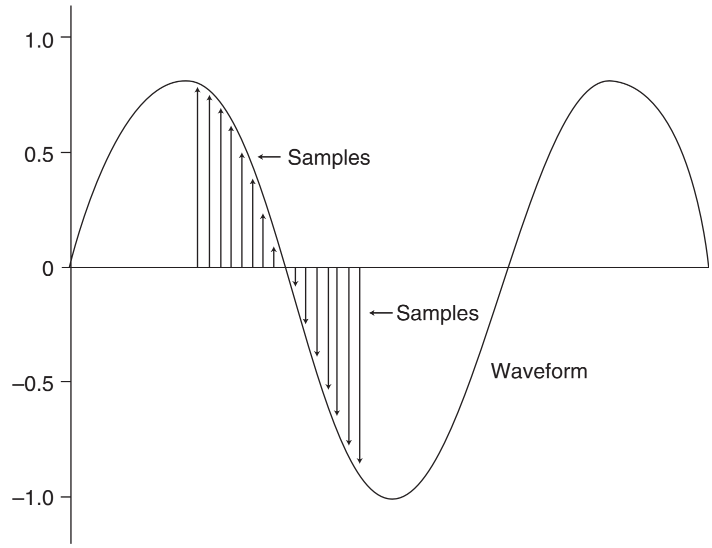

Figure 3.13 Illustration of the sampling phase of the analog-to-digital conversion.

each sample is then stored in binary form. Figure 3.13 illustrates this part of the analog-to-digital conversion process.

After the sound is sampled, another circuit takes each sampled value and quantizes it—that is, each unit is assigned a value of amplitude nearest the one that has been captured from an array of available choices. Thus, each sample is represented by the nearest allowable level of voltage the circuit is programmed to assign (see Figure 3.14). Any actual value seen to lie between two quantization steps is assigned the value closest to it.

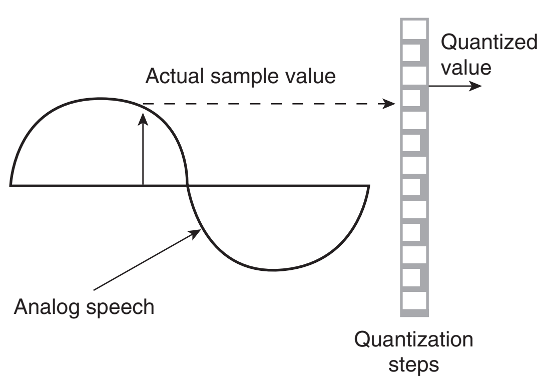

Figure 3.14 Illustration of the quantization phase of the analog-to-digital conversion.

In the case of compact disk (CD) recordings, where fidelity is of paramount importance, and transmission and bandwidth matters are not an issue, a greater number of samples quantized more finely may be used than, say, in the cell phone industry, where transmission and bandwidth issues are critical.

In all cases, once samples have been quantized, each bit of audio information is placed into a sequence called a pulse train consisting of series of 0s and 1s representing the original sound, which is, at the receiving end, converted back into electrical signals through the use of a digital-to-analog converter. It is this signal that is finally fed to your earpiece or amplifiers connected to the speakers of your stereo system or television entertainment center.

Digital image capture for both still and motion visuals is similar to the method just described for audio, but with a much larger data capacity. In the case of images, the analog information (i.e., light focused by a camera lens on a scene of interest) is converted into an electrical signal that is then coded as a digital data stream. To accomplish this, a CCD is the transducer instead of a microphone (see Figure 3.15). In a digital camera, the lens focuses light onto a CCD panel consisting of several hundred thousand (or several million) tiny light-sensitive diodes called *pixels.* Each pixel measures the amount of light hitting it, translating the brighter stimuli into higher electrical charges and the darker stimuli into lower

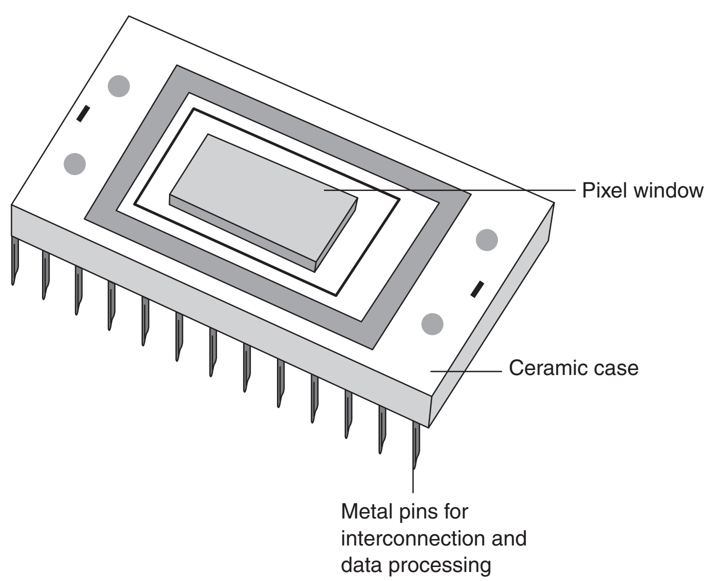

Figure 3.15 Schematic diagram of a charge-coupled device (CCD). A CCD these days can contain millions of pixels. Each one converts the light energy hitting it into an electric charge.

electrical charges. In this way, a mosaic of light intensities renders the original scene, creating a faithful black-and-white image of it.

To add color to the picture, a beam splitter (see Figure 3.16) is used to separate the light entering the camera into varying levels of red, green, and blue light (discussed earlier in this chapter). In some digital cameras (called three-chip cameras), a separate chip is used for each of these colors, from which the full color spectrum is reconstructed through an overlay process. In cameras using only one chip, selected pixels are fitted with permanent color filters. In such cameras, another computer inside the camera determines the true color of the light arriving at each pixel by interpolating information from the surrounding area. Using only one chip results in less accurate rendering of the scene to be captured but saves cost.

Digital video cameras work essentially the same way as digital still cameras, but with an additional sensor layer behind the image panel, allowing each image

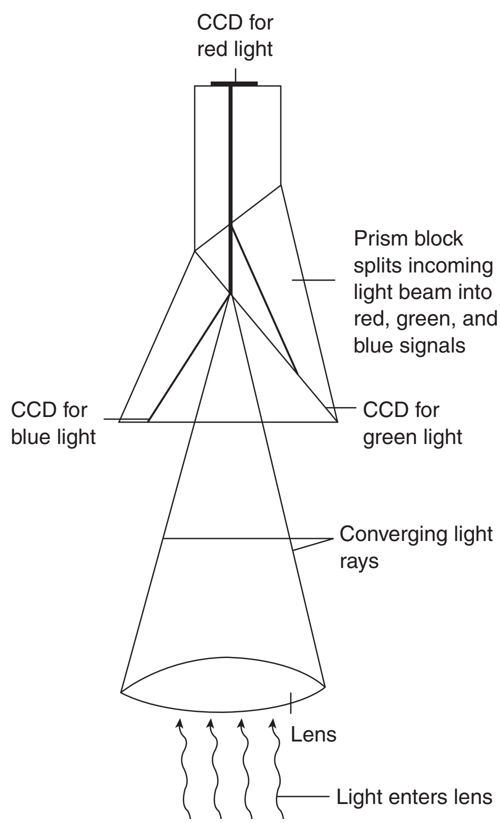

Figure 3.16 Schematic diagram of a beam splitter, a prism block used to break incoming light into three separate light beams (red, blue, and green) and to direct each one to a CCD for signal processing as a video image.

to be transferred to the second layer so that the first layer can refresh itself in order to capture a series of images in rapid succession. This process happens many times per second, creating the illusion of motion when replayed. Finally, the analog visual images are digitized essentially the same way as sound, through a process of sampling, quantizing, and coding. In the sampling stage, a number of selected instants of the analog signal (measured in MHz rather than kHz) are captured. Then each is assigned a quantized value from among an array of choices. Then, the values are coded into binary number equivalents composed of sequences of 0s and 1s. In recovering the information to make it viewable again, a reversal of this process is accomplished through a digital-to-analog conversion.

Standard Computer Architecture: ALU, CCU, Memory, Input, and Output

Once created, binary code representing content of whatever kind must be accessed properly for the content to be kept intact. If information is processed incorrectly, if sequences are apprehended out of order, or if parts of units are chunked with parts of adjacent units rather than with the ones they were originally framed with, the pattern of intelligence could be confounded, resulting in a mish-mash of incomprehensible output. To avoid this, rules of syntax are imposed on the information to keep it intact.

To process code successfully, computer architecture contains five components that are necessarily segregated from one another, including a central arithmetic logic unit (ALU) to carry out mathematical calculations, a central control unit to integrate operations, a *memory* to store information and permanent instructions about how to process it, a component for entering data (an *input* component), and an *output* component for displaying data to make content accessible to users (see Figure 3.17).

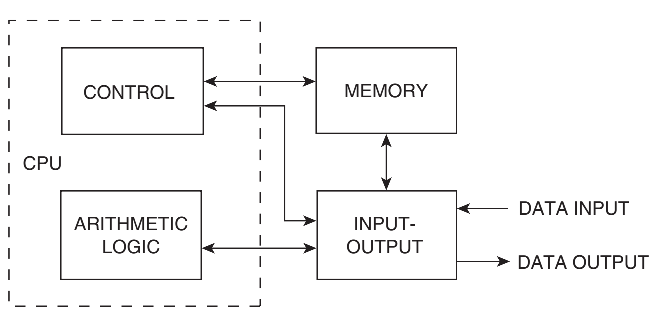

Figure 3.17 Five components of computer architecture provide accurate information processing.

Input components include such items as the keyboard and computer mouse, but those are not the only ones. Others include pressure-sensitive touch screens, such as those seen in restaurants, and plastic pens used to enter data on handheld PCs. Common output components include RGB video monitors and computer screens, printers, audio speakers, telephone earpieces, and headphones, among others.

The five-component computer architecture is still in use today. In addition, the computer's operations are performed in sequence one at a time (an imposition of temporal order to guard further against confusion), using a clock chip to order operations.

Computer systems may be programmed to control complicated and varied sets of conditions, exhibiting what has come to be called **artificial intelligence (AI).** In such systems, binary code may be used to execute an array of preset instructions designed to alter incoming information with a wide variety of practical benefits.

One application of such technology in the television industry is seen in the use of the *V-Chip,* a computer device required by FCC rules to be installed on all television sets 13 inches or larger manufactured after January 1, 2000. Thanks to digital text embedded as *header information* in most television programs, the V-Chip can read the ratings information associated with a show and then block the program from display if deemed undesirable by the user based on a set of instructions.

From Tubes to Transistors

As mentioned earlier, the ability of triode vacuum tubes to perform signal-processing functions made it a versatile electronics tool, especially in broadcasting, telecommunication industries, and computer technology. But as mentioned, vacuum tubes were not perfect; they were large, hot, electricity hogs and unreliable—they quickly burned out. When thousands of them were crammed in a box, as was often the case in early computer manufacturing, the temperature inside could soon exceed 120 degrees F. Such conditions required too much maintenance. Something better was needed. Something cooler. More reliable. Smaller. Less piggish.

By the late 1940s, alternative computer technology in the form of smaller, lighter, more reliable **semiconductors** was under development at AT&T and Bell Laboratories. Semiconductors could conduct an electric current, performing the same work as vacuum tubes but without the heat and with far less electricity consumption. Semiconductor technology gave rise to the first reliable solid-state *transistors* (called so because, depending on the electrical conditions presented to them, they acted as either a *trans*mitter or a res*istor* of electric current).

Transistors could do everything vacuum tubes could do but without any of the tube's shortcomings—no overheating, no breakable glass or filaments, no overconsumption of electricity. Transistors were everything vacuum tubes were not—they were small, cool, light, and reliable. Their tiny size meant their electric signals had to travel only a small distance to reach their destinations; that meant a great increase in data-processing speed and efficiency.

The trend toward miniaturization put pressure on manufacturers to find ways of connecting more and more transistors for increased data processing. Manufacturers also wanted to develop ways of mass-producing transistors. Rather than making them one at a time, companies began etching them into large silicon wafers using a photoengraving process. Soon entire sets of transistors consisting of amplifiers, resistors, and capacitors were being produced together on a single substrate of semiconductor material (see Photo 3.2).

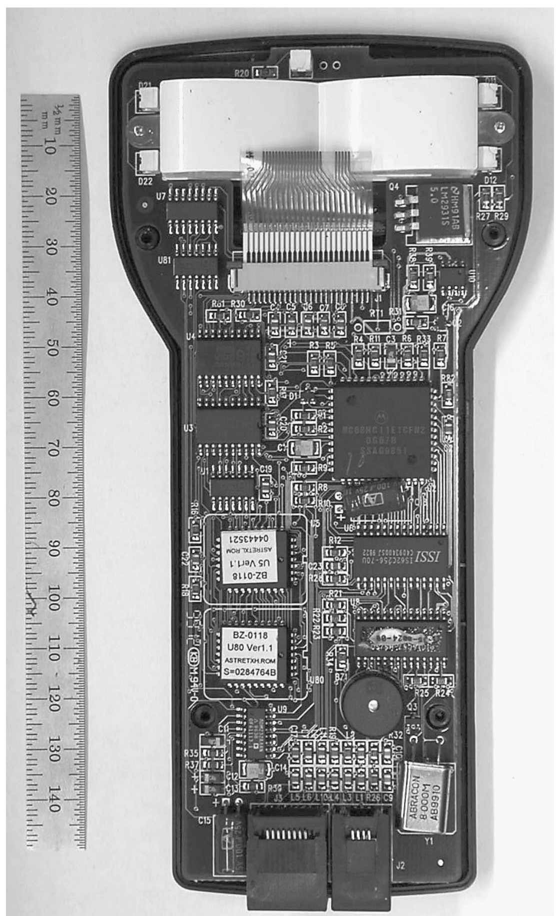

Photo 3.2 Integrated circuits such as this one, with millions of transistors, enable it to process digital information quickly.

SOURCE: Meade Autostar circuit board reprinted with permission of Richard Seymour and John Amsbaugh.

Chip Manufacturing

The leader in the field of miniaturization was Texas Instruments, which, through the efforts of its employee, Jack Kilby, conceived of a plan to produce multiple components of semiconductor material simultaneously on the same wafer. Kilby's invention, the first integrated circuit (IC), made its debut in 1959. By 1962, mass production of ICs (now called *chips*) was in full swing. Since that time, the size of each generation of transistor has decreased as their number in a single chip has increased (see Photo 3.3).

Recording television programs on computer chips requires vast amounts of data-processing capability. Without these breakthroughs in chip development and miniaturization, capturing live images with natural motion and high-fidelity sound as digital data files on computers would likely still be a dream.

Several types of computer chips are used for processing video content. Some chips are *memory* chips, designed to store information. One specific kind of memory chip, called a ROM (read-only memory) chip, acts as a permanent store

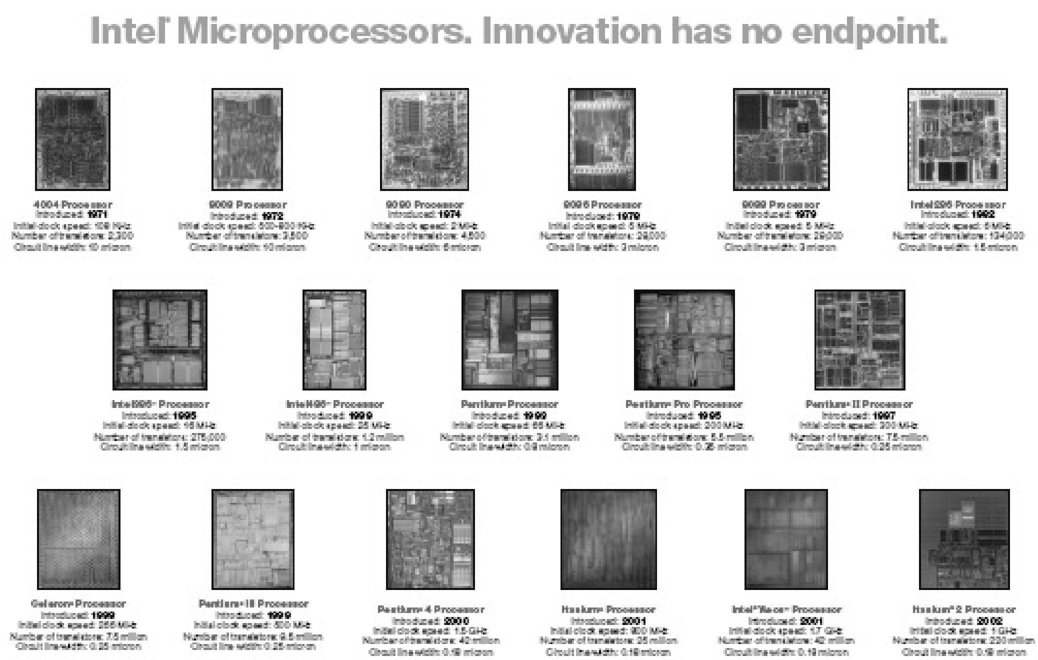

Photo 3.3 Several generations of integrated circuits show how the number of transistors on a single chip has increased over the past 20 years, enabling great increases in signal processing capability, making digital audio and video possible, among many other applications.

SOURCE: Reprinted with permission of Intel, www.intel.com/museum.

for binary code—that is, the transistor switches are set to react to electricity flowing through them the same way every time. It is like a file cabinet that simply holds the information (programs) inside.

Another type of chip is called a RAM (random-access memory) chip. RAM chips allow new information to be encoded and then deleted when no longer needed. RAM chips allow transistor switches to create different patterns of electric signals to flow, representing different types of information.

Yet another critically important type of chip is that which performs calculations and makes logical decisions that control a computer's activities in different parts of the machine. This chip is called the CPU (central processing unit) chip. Another is a clock chip that uses a quartz crystal to time operations so that all instructions are carried out one at a time in proper sequence. Together, the various chips coordinate and execute computer operations to ensure that operations go as planned.

The modern digital computer, through the use of binary code and silicon chip technology, has made it possible for individuals to *encode* a wide variety of textual and audiovisual messages that were once produced only by publishing houses, recording companies, and television studios. It is now a critical element of video production.

But the ability to *transmit* such messages is not the result of computers. Rather, it is the integration of computers with broadcasting and *telecommunication* infrastructures that permits program sharing and distribution to audiences. Without the telephone network and broadcasting technologies, computers would be incapable of sharing information across wide areas. The next section explains how computers are integrated with these older technologies to provide transmission capability for the television industry and independent producers.

Telecommunications

Video transmission occurs through the integration of computers and other digital devices with broadcasting and telecommunication technologies. Without the public switched telephone network (PSTN) and similar (i.e., private) infrastructures and broadcasting technologies, the immediate sharing of television programs among distant users would be impossible.

This section explains how video programs are transmitted. A central topic is **bandwidth,** or the capacity of a channel to move digital information from one place to another in a given period of time. I describe some of the communication technologies that transfer messages between points, including wire and cable facilities (copper, coaxial, and fiber-optic), as well as wireless facilities using radio energy (terrestrial microwave and satellite transmitters); all are currently used for transmitting broadcast programming. I also describe how digitalization increases the capacity and flexibility of telecommunication channels. The topics include packet switching, multiplexing, and signal compression, all designed to make more efficient use of available bandwidth. I also outline some of the advantages and disadvantages of digital versus analog modulation, as well as quality control issues (i.e., error correction features) designed to ensure that programs received are the same as those that are sent.

Because analog television receivers are so plentiful in the United States, it will likely be several years before we see a full transition to digital broadcasting. This is because set owners are often reluctant to invest in expensive new systems that may fail to become standard. Frequently, the adoption process is further slowed by incompatibilities across competing systems.

Of course, the digital revolution is still a work in progress. For example, twisted-pair copper wire, part of 19th-century telephone technology originally designed for analog voice traffic, is still in use today for connecting households to telephone exchange offices. At the same time, the network's switching technology has been continually upgraded (from electromechanical relays to vacuum tubes to digital solid-state equipment) to increase efficiency. Since around 1990, however, telephone companies began carrying digital content in addition to analog voice traffic through the adoption of 21st-century technologies not even dreamed of when the phone system first developed, including *T-1* and *T-3* lines, optical fiber for long-distance transmission, and digital subscriber lines (DSL), among others.

Today, the only part of the PSTN that has remained analog is the local loop (from homes to telephone exchange offices). And although existing copper wire has been successfully adapted for many digital functions (e.g., both computer data transmission and digital voice calls originating from cell phones must still be capable of reaching wire line phones through the local loop), the local infrastructure is not yet suitable for everything (i.e., broadband applications such as television-quality streaming video).

As a result, the current infrastructure still cannot deliver to most households high-speed service for real-time interactive video and other products and services requiring high capacity. Nevertheless, the eventual transition to a fully digital system for broadcasting is inevitable.

With exceptions, digital transmission is a relatively recent phenomenon. For decades, our major commercial television networks distributed thousands of hours of programming to hundreds of affiliate stations coast to coast in the form of analog video signals via high-bandwidth coaxial cable provided by the telephone company. The first such transcontinental distribution began in 1951, using AT&T coaxial cable to carry analog video signals.

But since the early 1990s, we have seen a transition from analog to digital transmission in both wire and wireless technologies. Why the change? What are the advantages of digital transmission that make it an attractive alternative for the television industry over traditional analog service?

Digital transmission was really fostered by the development and success of digital computers—as you already know, binary code is the only kind of information computers can understand. However, the conversion from analog to digital platforms in the television industry is not merely the result of pressure to conform to the needs of computers; rather, it derives from the promise of a quantum leap in *capacity* (and, by implication, profits) that comes with digital delivery of products and services.

Analog Versus Digital Signals

In the beginning of this section, I defined digital bandwidth as a measure of the amount of information that could pass through a communication channel in a given period of time. However, in the early days of broadcasting (when the term *bandwidth* was first used), bandwidth was defined not as a measure of information capacity but simply as the frequency range assigned to a specific broadcasting channel or service (e.g., short-wave radio, AM and FM radio, and, later, television).

Analog bandwidths were allocated according to both the physical limitations of the broadcasting technologies available at the time they entered service and the intended functions of those services. Some services needed wider bandwidths than others to deliver acceptable signals. For example, television channels are bigger than radio channels because they have to deliver both audio and video signals; therefore, they require greater bandwidth. Once assigned, analog spectrum allocations were fixed; that is, analog radio channels provided just radio programs, television channels provided just television programs, and so on.

Today, we live in a very different world, with deep implications for the notion of bandwidth. Digital transmission permits a more flexible and efficient use of bandwidth in both the electromagnetic spectrum and hardwire systems. First, digital transmission is more *flexible* than analog because binary code, unlike analog signals, can represent voice, text, data, images, and/or video, either alone or in combination with one another in any size channel and in any part of the frequency spectrum. Therefore, the bandwidth used to carry digital information can be any size, need not be restricted to just one type of service, and need not be limited to just one part of the spectrum. Second, digital transmission can be sent *without noise or loss of fidelity,* preserving original signal quality no matter how many times a program is transmitted. Third, digital transmissions are more *efficient* than analog because, unlike analog signals, binary code can undergo data compression without signal loss before being sent, allowing more information to be transmitted over the same channel in a given period of time.

Furthermore, analog signals require a dedicated circuit large enough to accommodate the signal being sent and must remain intact during transmission to be received successfully. By contrast, digital signals may be sent over any type of line, transmitted in parts, received out of order, and then reassembled into the original message at the receiver. For example, the digital code representing a high-bandwidth one-hour television program may take several hours to send over a low-bandwidth telephone line, but it can still be retrieved in its original form; analog signals can't survive such treatment. For all these reasons, using radio spectrum for analog signals is nowadays considered wasteful. The implications of this for the television industry, put simply, are that with increased capability, flexibility, and efficiency, *digital transmission can make greater use of available bandwidth, potentially resulting in greater profits.*

Packet Switching

As mentioned, digital signals can be broken into small bundles, intermixed with one another, transmitted out of order, and then successfully reassembled later with no loss of signal quality. They do not even require dedicated lines to carry them. This form of data handling is called **packet switching.**

Packet switching is a method of transmitting digital information in manageable units over available transmission lines. Each unit consists of a segment of the information you are sending, with *header information* or *meta-code*, which is information about the bundle itself, telling, among other things, where each should go (i.e., via e-mail) and how each should be handled once it gets there.

Packet switching allows transmission lines to be shared by millions of users, thus reducing quiescent periods, making the entire system more efficient. Messages do not need to occupy the same circuit from start to finish to survive transmission (in the language of the telephone company, packet-switched digital messages do not require *dedicated lines*). Furthermore, different messages can be sent through the same line sequentially without concern that they will be confounded with one another.

Both the traditional PSTN and the newer Integrated Services Digital Network (ISDN), introduced in the United States in 1992, use the same twisted-pair copper wire infrastructure between homes and central offices. Both were designed for continuous voice carriage. However, ISDN makes the local loop digital by connecting directly with digital technology at the central office. This means that unlike traditional analog telephone service, ISDN lines eliminate the need for digital content to undergo analog-to-digital conversions (i.e., via a modem) to be transmitted.

Unfortunately, both the traditional local loop and ISDN are circuit-switched connections that function exactly like a dedicated line—that is, when engaged, they serve just the parties connected to that line, to the exclusion of all others, even when no data are being sent. This arrangement is wasteful whenever the connection is used for data transfer rather than voice transmission (i.e., e-mailing video content or downloading Web pages) because in such instances, unlike continuous voice communication, data may be sent in bursts (or *packets*) during periods when the line connecting them is available. For this reason, data transmissions are characterized as *bursty,* meaning that transmission times may be intermixed with times of no activity. During such periods, other network users *could* be using that line for data transfer, but because it is tied up or dedicated in a circuit-switched arrangement, it is unavailable. For bursty data, it is a more efficient use of network assets to use transmission lines in a packet-switched arrangement to accommodate multiple users.

Packet switching as a mode of data transfer is generally attributed to a 1964 U.S. Department of Defense project by the Advanced Research Project Agency (ARPA), whose charge it was to develop a computer network called ARPAnet, the precursor of the Internet. By the 1980s, other networks sprang up based on the ARPAnet model. By the mid-1990s, these were absorbed into a network of networks, now called the Internet. By 1998, the number of computers connected to the Internet grew from 4 million to more than 30 million (Lu, 1998, p. 24). During that time, local-area networks (LANs) were brought on-line to connect computers in a local area to allow sharing of both data (including binary code of television programs) and hardware.

The reason the telephone company is the most desirable infrastructure for computer traffic is simple: The PSTN is the most ubiquitous system currently available for such functions. Among its most valuable assets is that it has hundreds of millions of telephones already in place providing both wire and wireless interactive (two-way) connections for information exchange.

Unfortunately, because computers process only digital code, and the phone company was originally designed to handle only analog voice traffic, many technical challenges have had to be overcome to reconcile these two essentially incompatible modes of communication. Nevertheless, the PSTN has, with adjustments, proven to be a feasible conduit for a lot of digital traffic. What makes it especially attractive is that it is the least expensive and most pervasive network available for such service. The alternative—namely, building a separate digital network of comparable magnitude—is less attractive because, among other reasons, it would be prohibitively expensive to build one with coverage comparable to that of the PSTN.

The best solution, therefore, is to develop technology to augment and adapt the extant telephone system for digital functions. And that is exactly what is being done. One way to send digital information over the analog local loop is by converting digital signals into analog sounds that can be carried over telephone lines. The first device built for this function was the modem. The term *modem* stands for modulator-demodulator. When placed between a computer terminal and a telephone line, it produces tones from carrier signals modulated by binary data, performing translation functions between them. Because the telephone system was set up for two-way communication, modems perform both modulation and demodulation processing of the tones. By way of a computer, a modem, and a telephone, users can access and share a vast array of binary-coded messages, including television programs.

Throughput

Digital bandwidth is measured by **throughput,** a measure of the amount of bandwidth carrying *meaningful* data compared to the overall information capacity of a device or channel. In digital channels, some available bandwidth is always used for nondata carriage functions (e.g., header information telling receiving computers the number and order of packets that should be received and other quality control functions). Therefore, throughput is always less than maximum capacity.

In addition, the speed of a channel is conditioned by the capacity of the receiver equipment it is "speaking" to. That is, if a fast system is sending data to a slower one, they must operate at a mutually acceptable level, which is the *fall-back rate* of the slower system. To do this, systems are equipped to monitor this aspect of their interaction, and this negotiation at the beginning of their interaction (called a *handshake*) determines the data rate that can be used. Digital systems also feature a method of checking for errors in transmission, known as a *checksum* calculation, with further cost to throughput.

In digital systems where interactivity is not possible (e.g., in compact disks or DVDs), a receiving computer cannot request that information be sent again if an error is detected. In such cases, a *forward error correction* method is used to maintain quality. In such systems, redundant information is sent for purposes of comparison. If discrepancies are detected, the system can then substitute code for data that fail to arrive intact. Some digital video satellite systems use forward error correction methods to maintain quality.

For all these reasons, raw bandwidth should not be taken as the true measure of channel capacity. Rather, quality control features in the form of meta-code must be taken into account to get a true picture of signal capacity in a digital channel.

Digital Compression

Meta-code, error correction procedures, checksum calculations, and so on are not the only issues that can alter channel capacity in digital transmission systems; the ability to compress data files before they are sent can also have an influence. But unlike the items mentioned above, which all reduce throughput, data compression actually increases it. For example, some high-bandwidth digital content (e.g., a video teleconference) can exceed a channel's (i.e., a phone line's) transmission capability. In such cases, a practical solution is to translate long strings of repeating binary code into shorter strings before they are transmitted and then include metacode telling the receiving computer how to retrieve the original content from the shorter strings. This strategy of data handling is called *digital compression.* Digital compression takes two forms: **lossy compression** and **lossless compression.** Lossy compression refers to digital compression that results in some loss of information when a data file is retrieved.

As an example of lossy compression, imagine a video teleconference featuring a static image of a group of executives seated around a table with a gray wall in the background. In this case, the video image may feature the same code for color and location for 90% of the pixels comprising the image for a significant period of time. Because much of the image is static for long periods, it makes sense to employ an image algorithm that reduces the code specifying the refresh rate of the video transmission for those parts of the signal that do not change, thus reducing significantly the bandwidth required to maintain an acceptable picture. When a motion occurs that interrupts the status of the pixels (e.g., an executive moves to the podium to give a report), the portion of the video image affected by the motion can be refreshed once again. In this example, some video information at the beginning of the new motion is lost, but the loss is insignificant. Therefore, the compression algorithm used is called *lossy* because losing certain parts of the signal is acceptable.

However, sometimes *any* signal loss resulting from compression is unacceptable. In such cases, the compression algorithm used must be capable of retrieving the original signal in its entirety. Such compression algorithms are called *lossless.*

An example of lossless compression is when a replacement algorithm specifies a short code to be used in place of a longer code, as when the letter *X* is substituted for a long and complex calculus formula appearing repeatedly in a math book. After transmission of the text file, an algorithm reverses the process so that every appearance of *X* is replaced once again by the calculus formula. Such a compression algorithm would be lossless because none of the information in the original message is lost. Lossless compression is necessary whenever any loss of the original data could be fatal to the meaning of the message.

Two compression standards have been widely adopted in the television industry: One, called *JPEG* (pronounced "jay-peg"), is named after the organization that developed both lossless and lossy versions of compression for storing and transmitting still images, the Joint Photographic Experts Group; the other, called *MPEG-2* (pronounced "em-peg 2"), is named after the organization that developed this lossy compression technique for motion visuals for both the television and film industries, the Moving Picture Experts Group.

MPEG-2 compression is similar to the video teleconference example given above. It is a form of lossy compression designed to increase throughput of video signals. However, while successfully reducing bandwidth requirements, it has negative implications for editing.

Put simply, MPEG was not developed to be edited; it was designed for preedited video intended for playback only. What makes the editing of MPEG video difficult is that compressed frames cannot be used as edit points. To offset this shortcoming, the MPEG-2 standard periodically sends a full frame of video (called a "reference frame") that can be edited. The problem is that the reference frame is generally sent only two or three times per second, which means that editors are forced make cuts fully a third or half second away from where they may want to make them. For some work, such as inserting additional footage between a commercial and a program segment, this limitation may be acceptable, but for cutting between speakers, it may cause serious synchronization problems.

To overcome this limitation, some systems allow reference frames to be recalled at selected locations. Other solutions include using a system of compression called *intraframe compression*, where each frame of video is compressed while still allowing any frame to be used as an edit point.

Video Distribution and Delivery

Since the inception of the television industry, new technologies have been added to expand distribution. Both wire and wireless technologies increase available bandwidth for distributing products and services. Cable connections include coaxial and fiber-optic cable; wireless connections include terrestrial microwave and satellite radio transmitters. Since the debut of these technologies, improvements have increased their capabilities at an impressive rate. Here I briefly describe some of these technologies and their role in video distribution and delivery.

Coaxial Cable. Coaxial cable or "coax" (rhymes with "no tax") uses two electronic conductors. One is a center copper wire sheathed in a protective layer of insulating material called a dielectric. The second is a wire mesh made of copper or aluminum wrapped around the dielectric. An outer layer of plastic shielding forms the outer body of the cable, providing insulation (see Figure 3.18). The diameter of coax cable varies from as small as .8 mm to more than 2 inches. Coaxial cable is named such because it has two wires with a common axis.

Coaxial cable makes it possible to transmit long distances with little signal loss. Its shielding reduces interference usually present in standard wiring. Its structure also makes it possible for it to handle signals well into the GHz range while minimizing spurious radiation. To offset signal loss that does occur, amplifiers (called repeaters) are added at regular intervals throughout the distribution system.

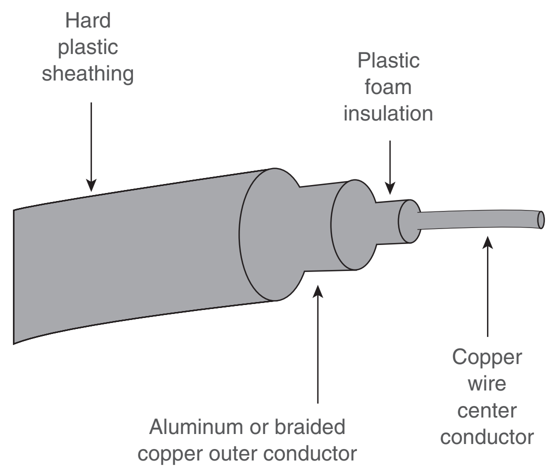

Figure 3.18 Coaxial cable. Exposed view of coaxial cable used in transmitting signals for telephone and video traffic.

For decades, U.S. television networks delivered programs to their affiliates through the use of coaxial cable provided by AT&T, the nation's long-line telephone company. One reason coaxial cable has been used for this service is because it delivers high-frequency signals more efficiently than standard copper wire. Coaxial cable is now being replaced by fiber-optic cable.

Optical Fiber. Fiber-optic cable uses flexible glass in place of copper or aluminum, as well as light pulses in place of electrical signals, to transmit digital information (see Photo 3.4).

The basic principle on which fiber-optic cable operates is to use a pipe to carry light. Signals through fiber-optic cable may be encoded using either a laser light or light-emitting diode (LED) and decoded by a photodiode at the receiving end.

There are many advantages to using optical fiber in telecommunications applications. First, glass is lighter, cheaper, and more plentiful than copper or aluminum. Second, and more important, brute capacity for encoding information with light is unsurpassed by any other conduit used for message transmission. For example, because light is electromagnetic radiation ranging in the hundreds and thousands of billions of cycles per second, it offers extremely large signal capacity. Its high frequency means that it can be turned on and off quickly, making it capable of carrying a great deal of digital data in a short amount of time over a single fiber. Thus, a single glass strand of optical fiber has more than 600 times the capacity of a coaxial cable; in short, fiber-optic cable has huge bandwidth, with all of the attendant implications for economic competition, even against satellite

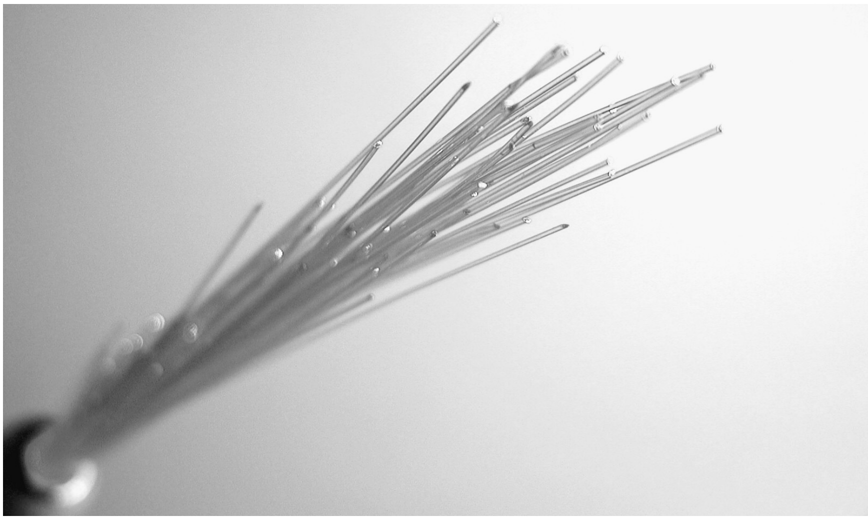

Photo 3.4 Fiber-optic cable with its light-carrying capability makes it possible to carry tremendous bandwidth compared to other types of cable.

technology. Third, electrical interference is nonexistent in optical fiber; this means clearer signals for both audio and video. Finally, when properly installed, fiber optics cable exhibits very little signal loss (*attenuation*), so fewer amplifiers (called regenerative repeaters) are needed to maintain signal strength. These benefits are attractive to cable television companies, which long for technologies that offer virtually unlimited capacity.

Of course, optical fiber technology has its downside. First, conversion from metal to glass is labor intensive and therefore expensive. In addition, connecting and switching fiber-optic transmissions can be tricky, as the glass used in the lines must not include any impurities. Furthermore, since the frequencies of natural light are too incoherent to travel effectively through fiber, lasers or LEDs must be used. Another problem is that light waves traveling through the pipe may take slightly different paths—some travel straight down the axis of the pipe, whereas others reflect and bounce their way through. Because all these paths have different lengths, signals arrive at their destinations at slightly different times, resulting in what has come to be called *smearing* or *pulse spreading*. This factor limits the rate at which data can be sent over a line without being corrupted.

Nevertheless, the move is on to make the switch. As of 1994, the regional Bell operating companies, or RBOCs (rhymes with "car locks"), had already installed more than 2 million miles of fiber-optic cable, while the major cable television companies, all of which now use fiber in some parts of their systems, had laid over 100,000 miles (Dominick, Sherman, & Copeland, 1996, p. 126; Dominick, Sherman, & Messere, 2000, p. 68).

AT&T began installing submarine fiber-optic cable in the 1980s to supplement undersea copper telephone cables. Such deployments offer a cost-effective option for intercontinental delivery of television programs instead of satellite systems.

The service capacity of fiber-optic lines has been truly impressive. According to some experts, the upper limit of a single fiber-optic channel is about 200 Gbps, or 200 billion bits per second, enough to carry 4,000 television signals. At this rate, if all of the light spectrum that could be harnessed for optical fiber were used, the theoretical capacity would be 50,000 Gbps or 50 Tbps (terabits or trillions of bits per second), enough to supply carriage for a million television programs to a single household (imagine the reruns).

One difficulty with fiber-optic systems as they have developed over the years is their incompatibility with one another. Interconnecting fiber systems built to different specifications means that special interfaces must be designed for handing off signals from one to another.

Terrestrial Microwave Transmission. Radio transmission in the form of microwaves (in current terms, roughly from 1 to 50 GHz, practically speaking) is used to carry broadcast signals and other services in the United States.

In terrestrial microwave transmission, radio waves are sent between antennas using microwave dish antennas and/or horn-shaped metal pipes called *waveguides* (see Figure 3.19). Some of the uses of microwave bands are for long-distance

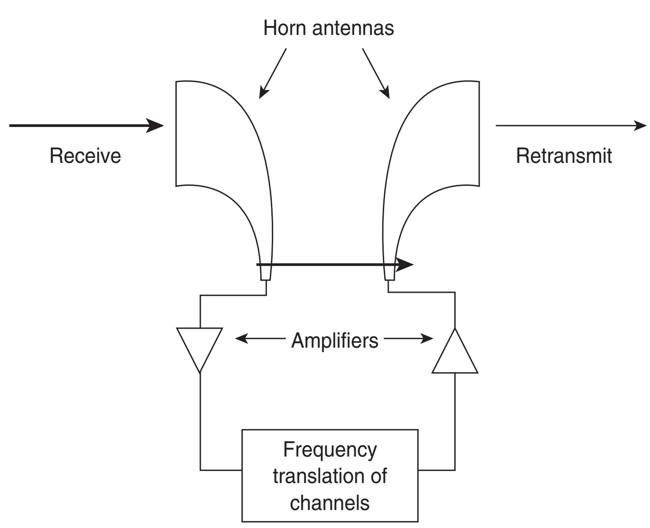

Figure 3.19 Diagram of a terrestrial microwave transmitter.

SOURCE: Adapted from A. M. Noll (1998), *Introduction to Telephones and Telephones Systems, 3/e.* Reprinted with permission from Archtech House.

NOTE: Microwave relays receive radio signals, change their frequencies, and retransmit them.

telecommunication, cable broadcasts, and cable point-to-point transmissions. As is well known in broadcasting, the nature of microwave radio energy is such that it must follow a line-of-sight path. Therefore, communication between antennas is possible only if there is an unobstructed view between them. For a microwave radio signal to cross the country, therefore, terrestrial towers must be spaced at no more than about 26 miles from one another to avoid having their signals blocked by the curvature of the Earth.

Microwave signals also dissipate relatively quickly. Therefore, to maintain signal strength, repeaters are used to amplify signals along their path. The usual practice is to beam a signal from a waveguide to a target antenna, where its frequency is shifted to a different channel to avoid jamming. Then it is amplified and sent on its way through another waveguide.

Satellites. In some situations, it is either too costly or infeasible to build a series of terrestrial microwave towers for delivering broadcasting products and services to audiences and subscribers. For example, sparsely populated rural areas are not cost-effective investments for such infrastructure, and stringing towers across oceans is simply impractical. In these cases, placing a microwave transmitter up in the sky aboard a satellite is the answer.

Satellite technology enables microwave signals to cross the ocean in one hop (see Figure 3.20). Satellites placed over the equator in **geosynchronous** orbits (meaning that they stay above the same spot on Earth at all times) must be at an altitude of 22,300 miles. When deployed this way, satellites take 24 hours to make one revolution of the Earth. Radio transmission between a satellite and a ground station is constant and reliable when the satellite maintains the same position relative to the Earth while in orbit. This is a great advantage because ground stations do not have to be moved to maintain contact and signal strength with the satellite. Satellites can then receive reliable signals, amplify them, shift them to a new

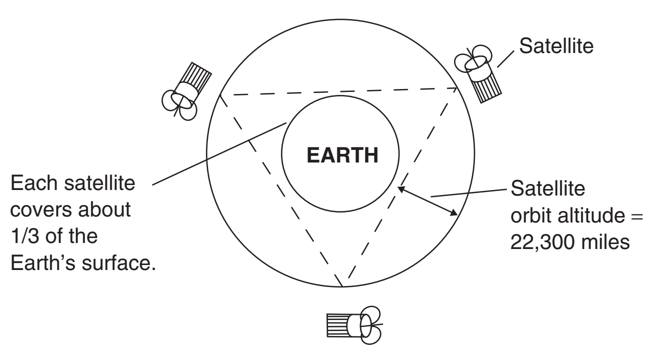

Figure 3.20 Diagram of satellites in orbit in relation to the Earth.

frequency (to avoid interference), and then retransmit them to ground stations thousands of miles away. Circuits that do this are called *transponders.*

The effective coverage area of a signal beamed to Earth from a satellite is called a *footprint* (see Figure 3.21). Access to the satellite's signal varies as a function of the size of the footprint, which is in large part determined by the satellite's transmitting power.

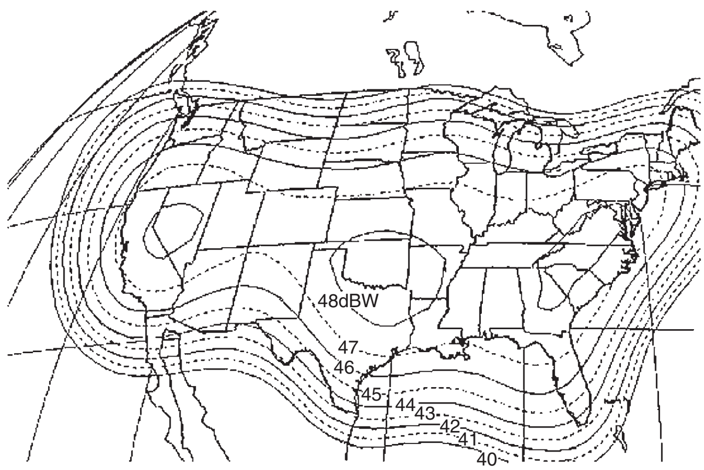

Figure 3.21 Satellite footprint showing area of strongest signal strength in the central portion of the coverage area.

One fundamental advantage of satellites over older terrestrial relay technology is that whereas repeaters on land can link one location with only two others, a satellite can link a group of program or service suppliers (e.g., television networks) to an unlimited number of ground stations in multiple locations at no extra cost. For this reason, satellites are called *distance insensitive.*

A geosynchronous satellite signal takes about ¼ second for it to travel up to and ¼ second to travel down from the satellite. Such delays, while insignificant for bursty data, can be annoying for an interactive telephone conversation. For this reason, satellites are not the best choice for voice transmission. However, they are excellent for one-way distribution of television signals (called TVRO for TV receive-only signals) among networks, cable companies, and affiliate stations. In addition, there is growth in direct-broadcast satellite (DBS) programming from program suppliers directly to homes.

Unfortunately, in reality, satellites deployed in geosynchronous orbit are not perfectly fixed in relation to the Earth; they require adjustment. To do this, noncommunication (*telemetry*) signals are traded between satellites and Earth stations to make corrections using fuel and rockets aboard the satellites. Eventually, the fuel runs out. When it does, satellites must be replaced.

Cellular Telephony

Without doubt, the success in recent years of wireless cellular telephone service has been among the most impressive developments in digital communication, with deep implications for video production and distribution.

Photo 3.5 Cell phones are no longer just for phone calls. These versatile digital devices now enable users to download and watch videos wirelessly from the Internet, providing video and audio capability just like the MP-3 player (or video i-Pod) next to it.

In 1990, 6 years after cellular service began in the United States, there were 5.3 million U.S. subscribers. By 1996, that number had grown to more than 44 million. In 2006, the number of U.S. subscribers has reached over 200 million. The ability to watch streaming video wirelessly is largely the result of the successful development and deployment of cellular telephone technology.

Access to the World Wide Web via cell phones is also on the rise. It is estimated that by the end of 2001, about 100 million subscribers were accessing text-based content via their cell phones. By 2004, according to some estimates, the worldwide number of cellular wireless Web surfers had grown to more than 750 million. In America alone, some analysts have predicted that by 2006, the number of Americans with Web-enabled phones have topped 50 million.

03-Shyles-45161.qxd 1/5/2007 7:30 PM Page 83

Cell phone technology connects users to regular wire and portable (wireless) telephones using two-way radio transmission. In the early (precellular) days of mobile telephony, a few dozen channels in an area provided telephone service with a single transmitter covering a 25-mile radius. Access was limited because if a user engaged a frequency, it was unavailable for anyone else until the user completed a call. This approach quickly led to congestion problems. The introduction of the basic principles of cellular service, developed by Bell Labs in the 1940s, solved the congestion problem. By the 1970s, cellular technology became viable for widespread service.

Technical Principles. In cell service, the big improvement over early mobile telephone technology is the reuse of channels to increase access. The ability to reuse channels is due to a scheme that places a low-power transmitter or base station into a small area (called a cell) inside the larger service area (see Figure 3.22).

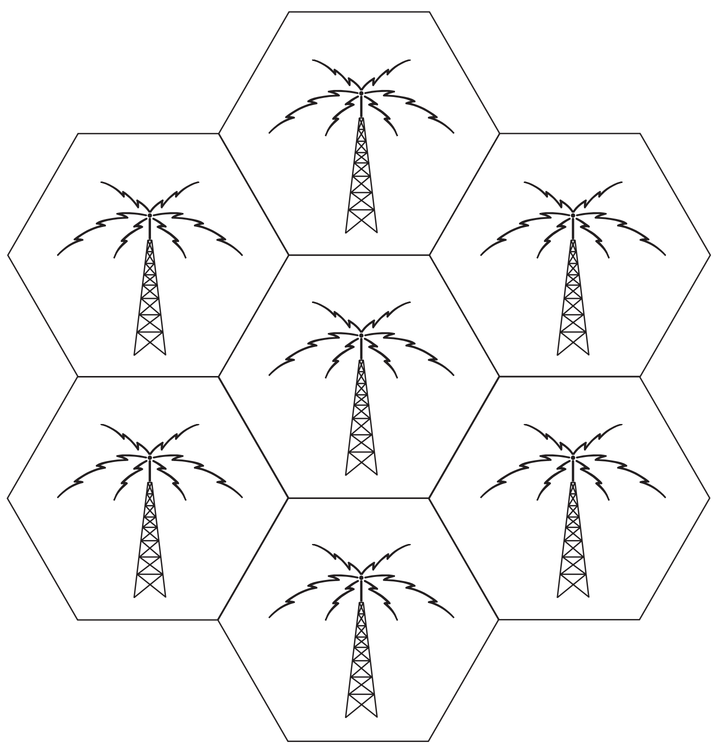

Figure 3.22 Diagram of a scheme for placing low-power transmitters into small areas for carrying phone traffic.

SOURCE: Adapted from A. M. Noll (1998), *Introduction to Telephones and Telephones Systems, 3/e.* Reprinted with permission from Archtech House.

NOTE: Cellular telephone areas are ideally shaped like hexagons. Six are placed on a circular arrangement with a seventh in the center; each has its own low-power transmitter.

Mobile Telephone Switching Office. Cells are typically 6 to 12 miles in radius. Each base station is equipped with a low-power (less than 100-watt) transmitter and is receiver-controlled by the service provider's mobile telephone switching office (MTSO). The low power of each transmitter means that the same frequency can be reused in other parts of the service area without causing interference.

The MTSO directs all call activity (see Figure 3.23). When a cell phone initiates a call, a control channel receives a signal to assign a channel pair for service; if available, the MTSO assigns one. As the cell phone moves from one cell to another, the MTSO tracks its movement through adjacent base stations, all of which can sense the signal strength of the cell phone's transmitter. The base station receiving the strongest signal assumes service. The assumption of service by an adjacent base station is called a *handoff.*

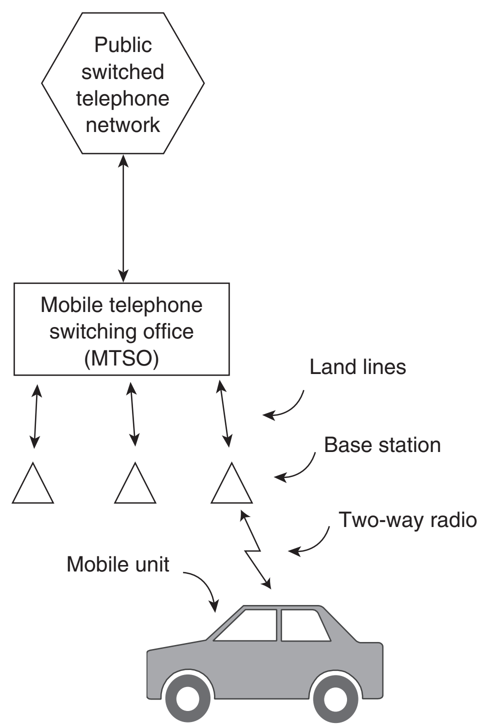

Figure 3.23 Diagram illustrating the relationship of the mobile telephone switching office (MTSO) to the public switched telephone network (PSTN).

SOURCE: Adapted from A. M. Noll (1998), *Introduction to Telephones and Telephones Systems, 3/e.* Reprinted with permission from Archtech House

NOTE: Each cell phone communicates via two-way radio with a base station. Base stations communicate over landlines with the MTSO, which maintains communication with the PSTN. Systems such as this permit wireless downloads of MP-3s and videos in addition to handling phone traffic.

To have a full understanding of how technology enables us to create, store, transmit, receive, consume, and manipulate television and other content through both wire and wireless means, it is essential to know something about broadcasting, computers, and distribution networks. All three infrastructures are critical to commerce in both mass media products and services; without all three, digital video would not be what it is today.

KEY TERMS

frequency 35
period 35
amplitude 35
wavelength 35
phase 35
electromagnetic waves 36
high fidelity 37
transducer 37
amplitude modulation (AM) 40
frequency modulation (FM) 40
demodulation 41
resolution 42
photoconductivity 44
photoemissive effect 44
pickup tube 45
charge-coupled device (CCD) 45
camera control unit (CCU) 45
viewfinder 45
pixels 46
persistence of vision 48
field 48
frame 48
aspect ratio 49
chrominance 49
luminance 49
saturation 49
digital television (DTV) 52
high-definition television (HDTV or HD) 52
progressive scanning 54
letterboxing 56
sampling 57
quantizing 57
broadband 57
narrowband 57
binary code 58
artificial intelligence (AI) 66
semiconductors 66
bandwidth 69
packet switching 72
throughput 74
lossy compression 75
lossless compression 75
geosynchronous 80

QUESTIONS FOR REVIEW

How does signal modulation make communication at a distance possible?

Describe the process of transduction in a microphone.

How do AM and FM differ? Why are FM signals less susceptible than AM to static and interference?

Why did video engineers choose a scanning method for video signal transmission?

What components of the video signal must be sent to reproduce a coherent program at the receiver?

What are some production implications of current television technology?

What are some differences between the NTSC standard and the new DTV standard?

Describe several implications that result from integrating television technology with computers and telecommunications networks.

NOTES

1. With digital television (DTV) systems, three scanning standards have emerged, including both interlaced and progressive formats, described in more detail later.
2. Actual picture resolution and frame rates of a received video image may differ from the resolution and frame rate of the sent signal, depending on the user's needs. For example, a DTV receiver may receive an interlaced image at one level of resolution but display it differently; that is, it may play a program back at 24-p (24 frames per second, progressively scanned), even though the program may have originally been sent at 60-i (60 frames per second, interlaced scanned), performing such feats through the use of digital signal processors. Both the 480- and 1,080-line formats offer interlaced or progressive scanning.
3. Silicon-based chips for computers are the semiconductor industry standard, but industry researchers are planning for a transition away from conventional silicon transistors to nanotechnology using organic molecules and carbon-based materials as a way to increase computing power beyond what is believed possible with silicon. The transition to nanotechnology is now scheduled for 2015 (see Markoff, 2005d).
4. A thorough review of the history of computers and binary code is beyond the scope of this book. Interested readers wanting more information about these topics should read "Computers in Communication" in Shyles (2003, pp. 45–93). Excellent additional sources abound in print and on the Internet. In print, see Augarten (1984) and Petzold. On-line sources include http://goldenink.com/computersandnetworks.shtml (retrieved December 26, 2005) and http://inventors.about.com/library/blcoindex.htm (retrieved December 26, 2005). Individuals who deserve further study include Blaise Pascal, Gottfried Leibnitz, Joseph-Marie Jacquard, Charles Babbage, George Boole, Claude Shannon, John Vincent Atanosoff, George Stibitz, Alan Turing, John Mauchly, J. Presper Eckert, John von Neumann, Jack Kilby, and John Noyce. While this list is not exhaustive, it is a great start.
5. At this writing, the Motion Picture Experts Group has MPEG versions up to MPEG-7, but it appears that the MPEG-2 standard is still among the leaders.

04-Shyles-45161.qxd 1/5/2007 7:06 PM Page 87

PART II

Elements and Techniques of Video Production