Chapter 3

Introduction to Physical Layer

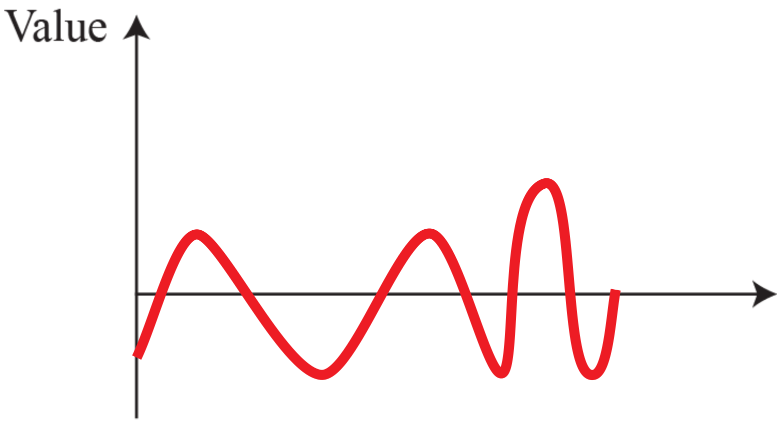
Analog and Digital Data

- Data can be analog or digital.
- The term analog data refers to information that is continuous
- Digital data refers to information that has discrete states.
- Analog data, such as the sounds made by a human voice, take on continuous values. When someone speaks, an analog wave is created in the air. This can be captured by a microphone and converted to an analog signal or sampled and converted to a digital signal.
- Digital data take on discrete values. For example, data are stored in computer memory in the form of 0s and 1s. They can be converted to a digital signal or modulated into an analog signal for transmission across a medium.

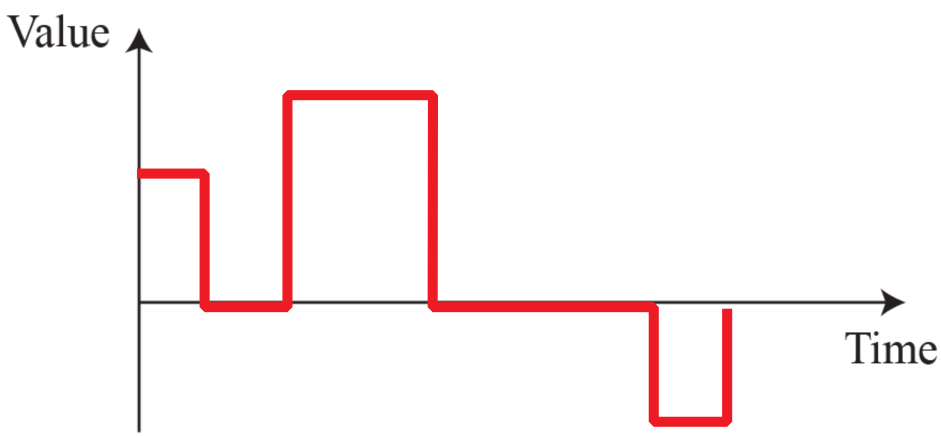
Analog and Digital Signals

- Like the data they represent, signals can be either analog or digital.
- An analog signal has infinitely many levels of intensity over a period of time.
- As the wave moves from value A to value B, it passes through and includes an infinite number of values along its path.
- A digital signal, on the other hand, can have only a limited number of defined values.
- Although each value can be any number, it is often as simple as 1 and 0.

Figure 3.2: Comparison of analog and digital signals



a. Analog signal



b. Digital signal

Periodic and Nonperiodic

- Both analog and digital signals can take one of two forms: periodic or nonperiodic
- A periodic signal completes a pattern within a measurable time frame, called a period, and repeats that pattern over subsequent identical periods.
- The completion of one full pattern is called a cycle.
- A nonperiodic signal changes without exhibiting a pattern or cycle that repeats over time.
- Both analog and digital signals can be periodic or nonperiodic
- In data communications, we commonly use periodic analog signals and nonperiodic digital signals

Periodic Analog Signal

- Periodic analog signals can be classified as simple or composite.
- A simple periodic analog signal, a sine wave, cannot be decomposed into simpler signals.
- A composite periodic analog signal is composed of multiple sine waves.

Sine Wave

- The sine wave is the most fundamental form of a periodic analog signal.
- When we visualize it as a simple oscillating curve, its change over the course of a cycle is smooth and consistent, a continuous, rolling flow.
- Each cycle consists of a single arc above the time axis followed by a single arc below it.
- A sine wave can be represented by three parameters: the peak amplitude, the frequency, and the phase.
- These three parameters fully describe a sine wave

Figure 3.3: ***A sine wave***

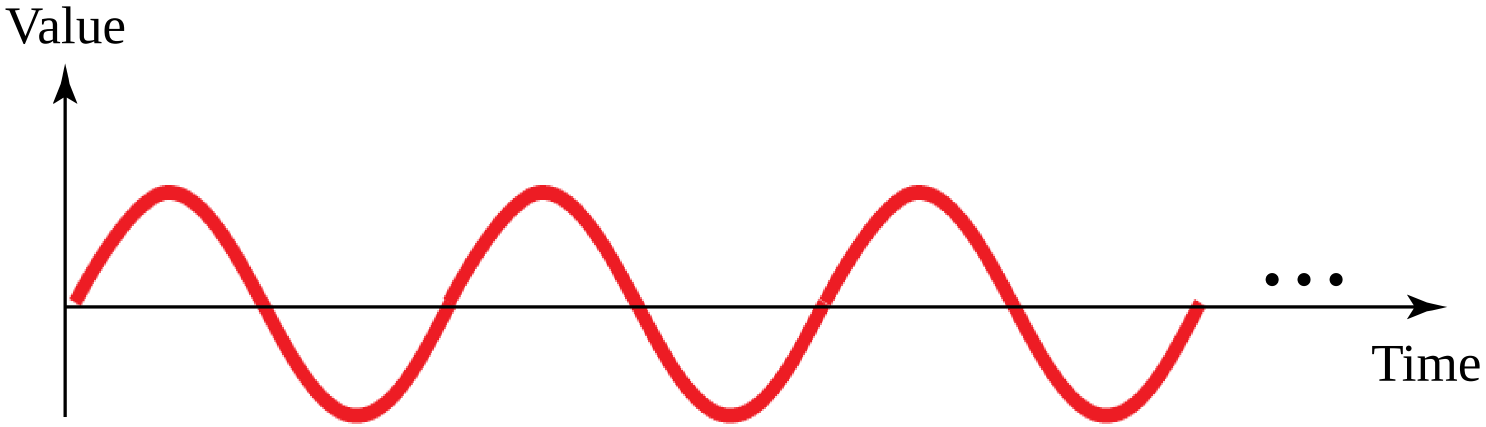
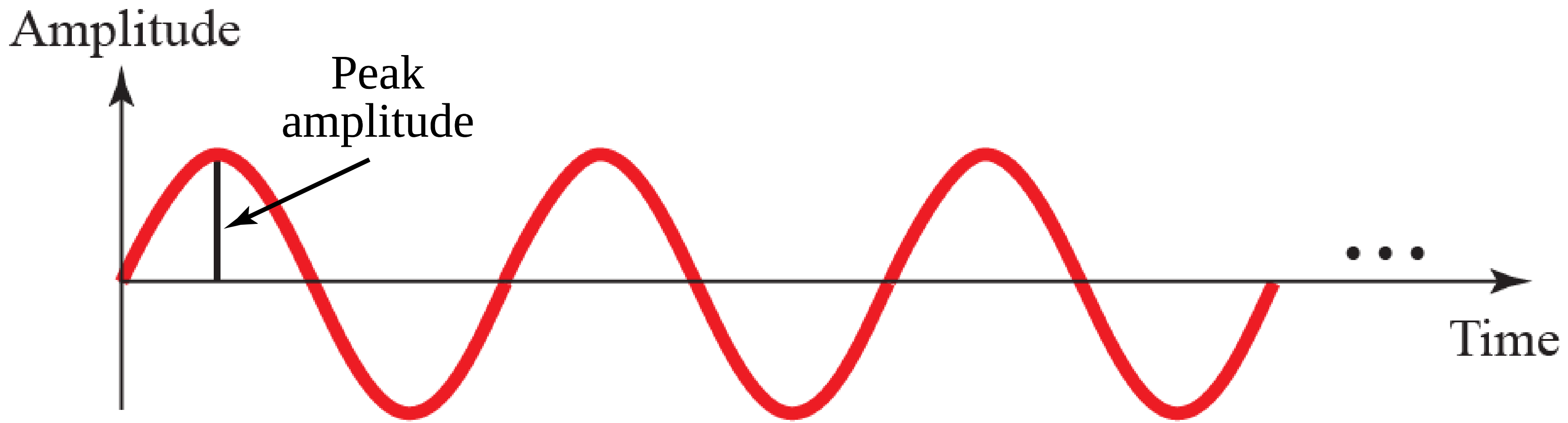
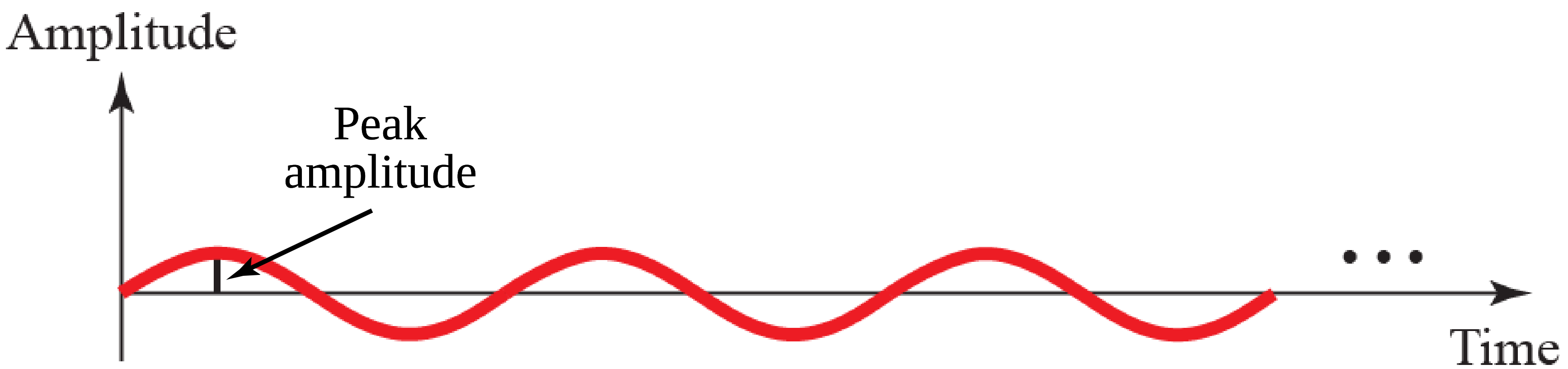


Figure 3.4: Peak Amplitude



a. A signal with high peak amplitude



b. A signal with low peak amplitude

Example 3.1

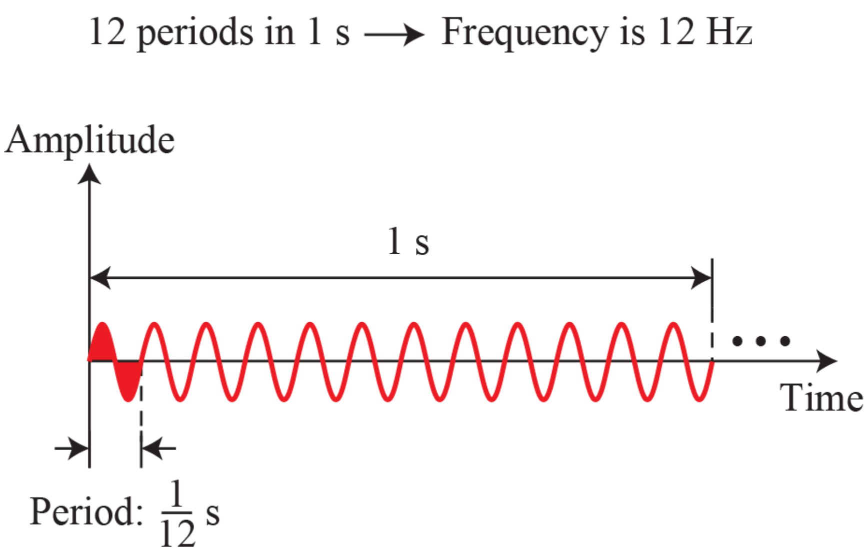
- The power in your house can be represented by a sine wave with a peak amplitude of 310 V.
- However, the voltage of the power in Bangladesh homes is 220V.
- This discrepancy is due to the fact that these are root mean square (rms) values.
- The signal is squared and then the average amplitude is calculated.
- The peak value = $\sqrt{2} \times$ rms value.

Period and Frequency

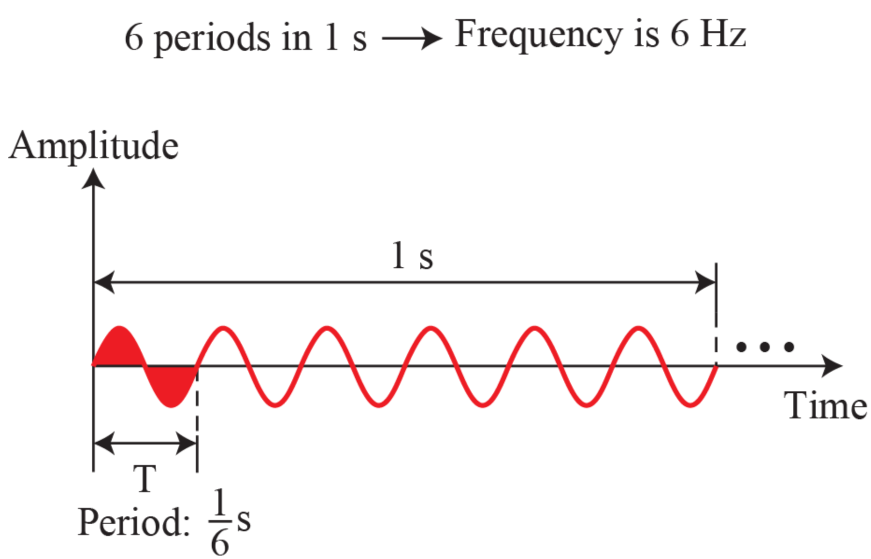
- Period refers to the amount of time, in seconds, a signal needs to complete 1 cycle.
- Frequency refers to the number of periods in 1s.
- Note that period and frequency are just one characteristic defined in two ways.
- Period is the inverse of frequency, and frequency is the inverse of period

$$f = \frac{1}{T} \qquad \text{and} \qquad T = \frac{1}{f}$$

Figure 3.5: ***Two signals with same phase, different amplitudes and frequency***



a. A signal with a frequency of 12 Hz



b. A signal with a frequency of 6 Hz

Table 3.1: Units of period and frequency

Period		*Frequency*	
Unit	*Equivalent*	*Unit*	*Equivalent*
Seconds (s)	1 s	Hertz (Hz)	1 Hz
Milliseconds (ms)	10^{-3} s	Kilohertz (kHz)	10^{3} Hz
Microseconds (μs)	10^{-6} s	Megahertz (MHz)	10^{6} Hz
Nanoseconds (ns)	10^{-9} s	Gigahertz (GHz)	10^{9} Hz
Picoseconds (ps)	10^{-12} s	Terahertz (THz)	10^{12} Hz

Example 3.4

Express a period of 100 ms in microseconds.

Solution

From Table 3.1 we find the equivalents of 1 ms (1 ms is 10–3 s) and 1 s (1 s is 106 μs). We make the following substitutions:

$$100 \text{ ms} = 100 \times 10^{-3} \text{ s} = 100 \times 10^{-3} \times 10^{6} \ \mu\text{s} = 10^{2} \times 10^{-3} \times 10^{6} \ \mu\text{s} = 10^{5} \ \mu\text{s}$$

Example 3.3

The power we use in USA has a frequency of 60 Hz (50 Hz in Europe/Asia). The period of this sine wave can be determined as follows:

$$T = \frac{1}{f} = \frac{1}{60} = 0.0166 \text{ s} = 0.0166 \times 10^3 \text{ ms} = 16.6 \text{ ms}$$

This means that the period of the power for lights at US home is 0.0116 s, or 16.6 ms. Our eyes are not sensitive enough to distinguish these rapid changes in amplitude.

Example 3.5

The period of a signal is 100 ms. What is its frequency in kilohertz?.

Solution

First we change 100 ms to seconds, and then we calculate the frequency from the period (1 Hz = 10^{-3} kHz).

$$100 \text{ ms} = 100 \times 10^{-3} \text{ s} = 10^{-1} \text{ s}$$

$$f = \frac{1}{T} = \frac{1}{10^{-1}} \text{Hz} = 10 \text{ Hz} = 10 \times 10^{-3} \text{ kHz} = 10^{-2} \text{ kHz}$$

More About Frequency

Frequency is the rate of change with respect to time. Change in a short span of time means high frequency. Change over a long span of time means low frequency.

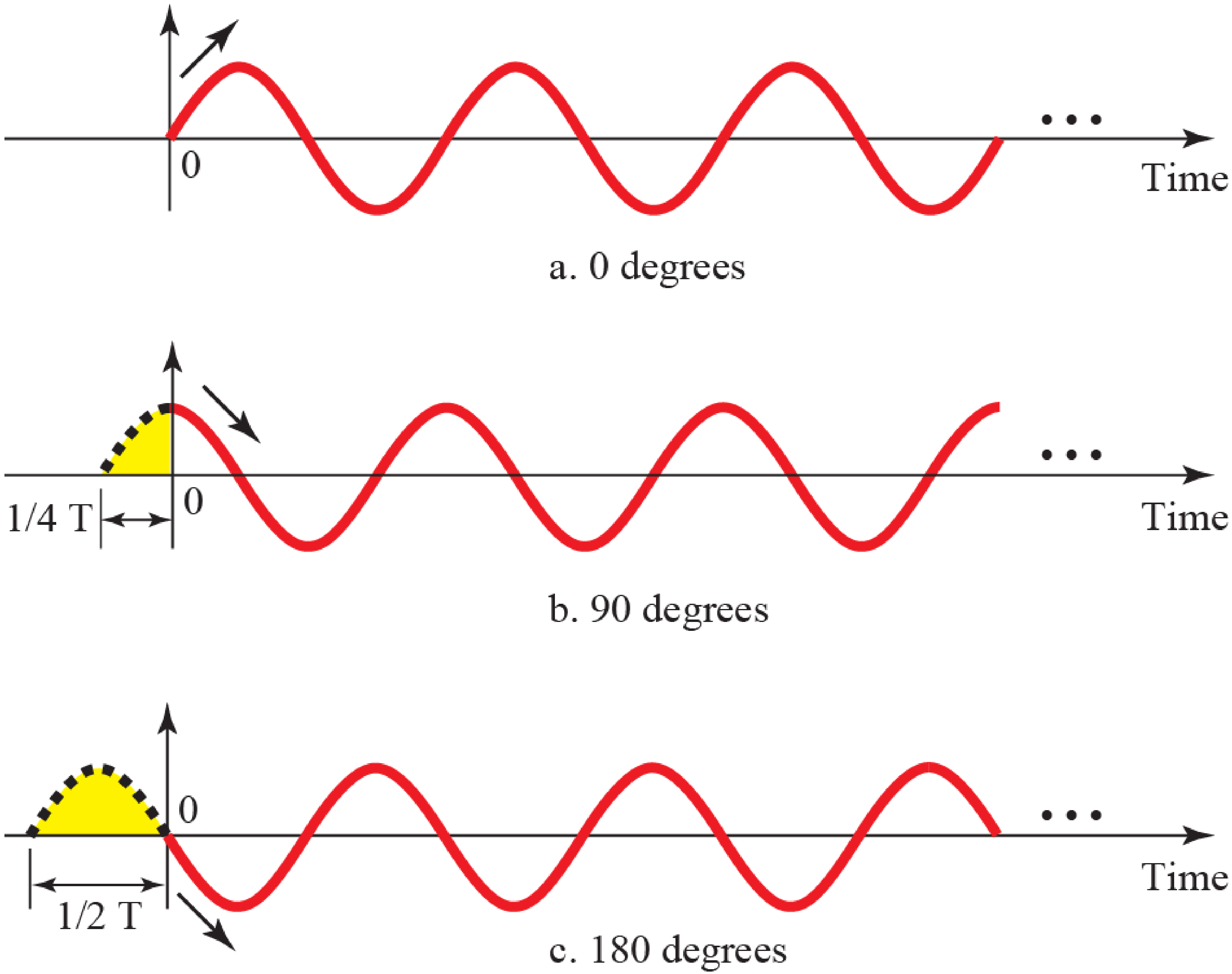
Two Extremes

If a signal does not change at all, its frequency is zero.
If a signal changes instantaneously, its frequency is infinite.

Phase

- The term phase, or phase shift, describes the position of the waveform relative to time 0.
- If we think of the wave as something that can be shifted backward or forward along the time axis, phase describes the amount of that shift.
- It indicates the status of the first cycle.
- Phase is measured in degrees or radians [360^o is 2π rad; 1^o is $2\pi/360$ rad, and 1 rad is $360/(2\pi)$].
- A phase shift of 360^o corresponds to a shift of a complete period; a phase shift of $180°$ corresponds to a shift of one-half of a period; and a phase shift of 90^o corresponds to a shift of one-quarter of a period

Figure 3.6: ***Three sine waves with different phases***



Example 3.6

A sine wave is offset 1/6 cycle with respect to time 0. What is its phase in degrees and radians?

Solution

We know that 1 complete cycle is 360°. Therefore, 1/6 cycle is

$$\frac{1}{6} \times 360 = 60^\circ = 60 \times \frac{2\pi}{360} \text{ rad} = \frac{\pi}{3} \text{ rad} = 1.046 \text{ rad}$$

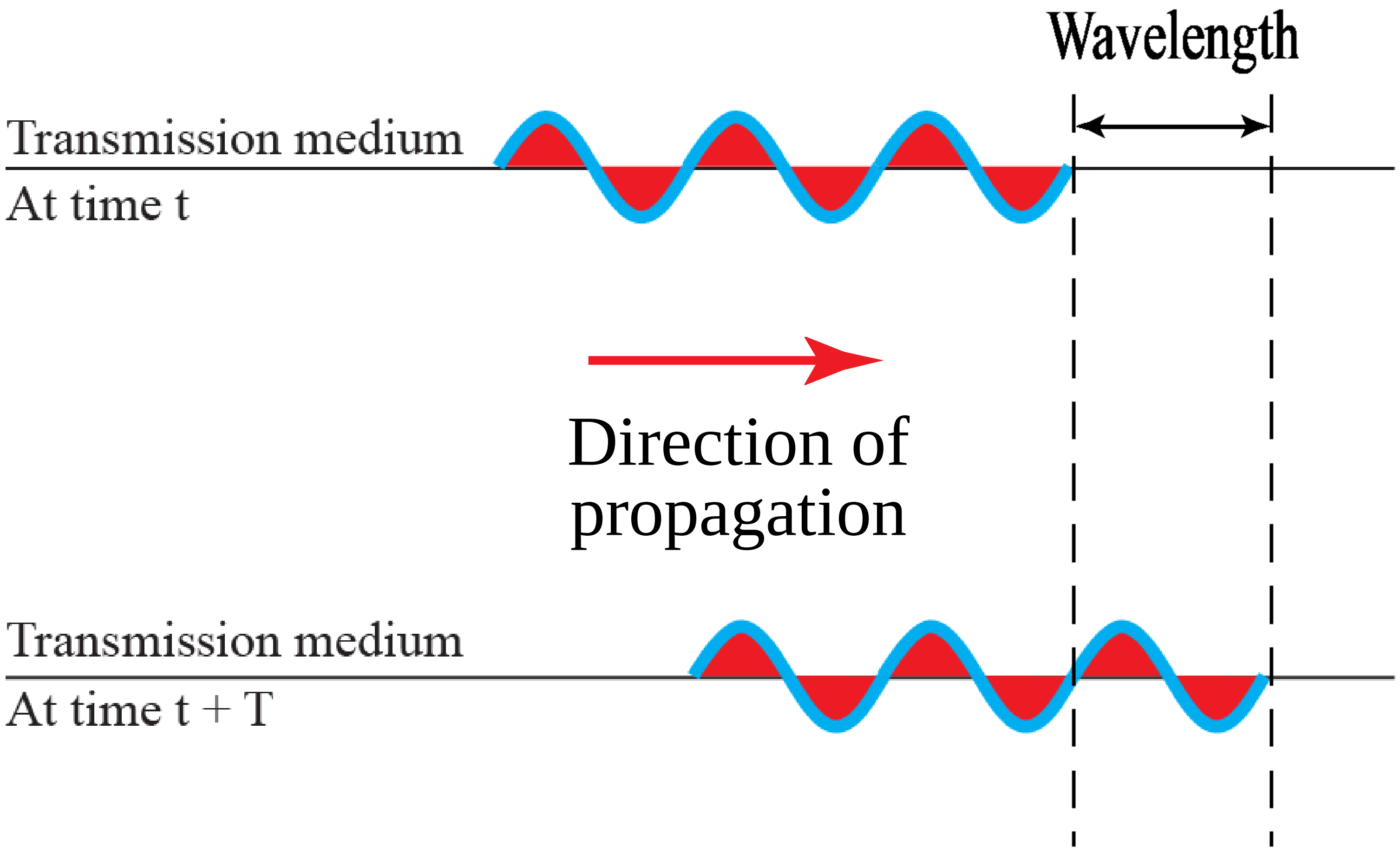
Wavelength

- Wavelength is another characteristic of a signal traveling through a transmission medium.
- Wavelength binds the period or the frequency of a simple sine wave to the propagation speed of the medium

$$\text{Wavelength} = (\text{propagation speed}) \times \text{period} = \frac{\text{propagation speed}}{\text{frequency}}$$

$$\lambda = \frac{c}{f}$$

Figure 3.7: Wavelength and period



Example

In a vacuum, light is propagated with a speed of 3×10^8 m/s. That speed is lower in air and even lower in cable.

The wavelength is normally measured in micrometers (microns) instead of meters.

For example, the wavelength of red light (frequency $= 4 \times 10^{14}$) in air is:

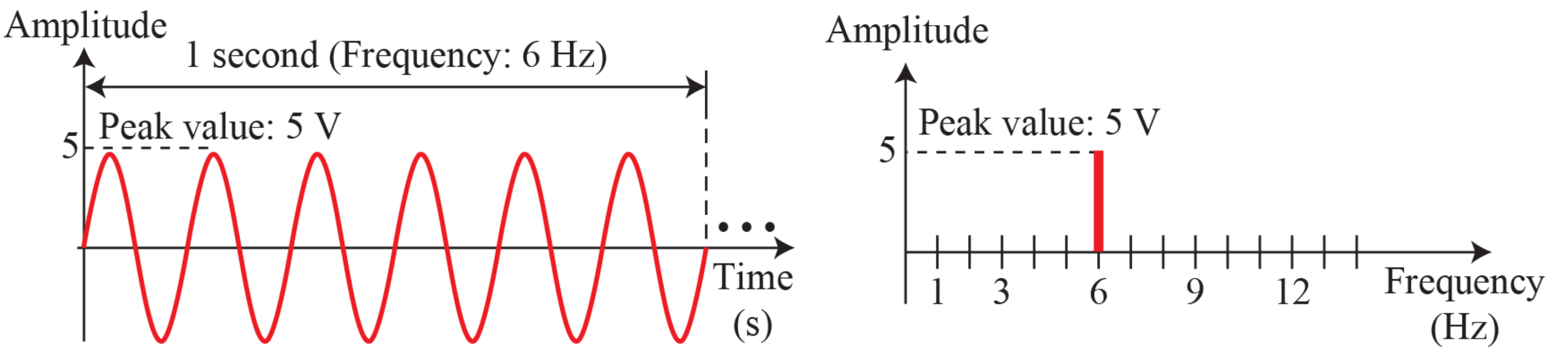
$$\lambda = \frac{c}{f} = \frac{3 \times 10^8}{4 \times 10^{14}} = 0.75 \times 10^{-6}\ \text{m} = 0.75\ \mu\text{m}$$

In a coaxial or fiber-optic cable, however, the wavelength is shorter (0.5 μm) because the propagation speed in the cable is decreased.

Time and Frequency Domains

- A sine wave is comprehensively defined by its amplitude, frequency, and phase.
- We have been showing a sine wave by using what is called a time domain plot.
- The time-domain plot shows changes in signal amplitude with respect to time (it is an amplitude-versus-time plot).
- Phase is not explicitly shown on a time-domain plot.
- To show the relationship between amplitude and frequency, we can use what is called a frequency-domain plot.
- A frequency-domain plot is concerned with only the peak value and the frequency.
- Changes of amplitude during one period are not shown.

Figure 3.8: ***The time and frequency-domain plots of a sine wave***



a. A sine wave in the time domain

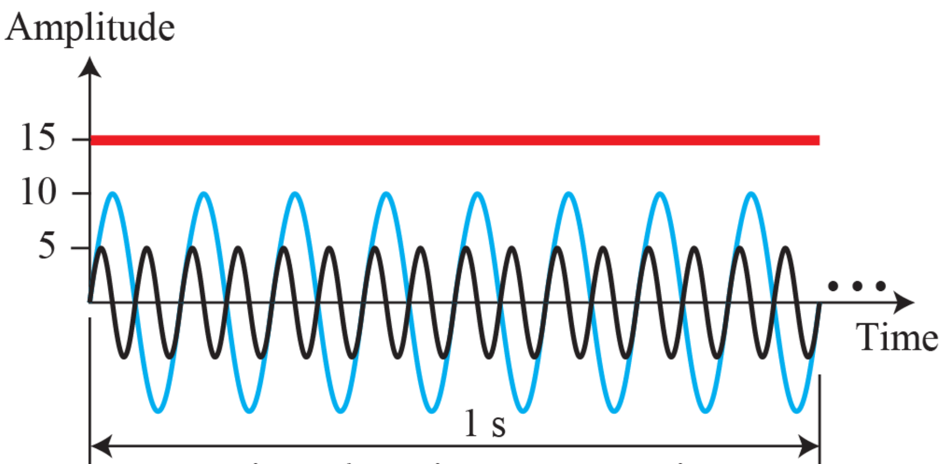
b. The same sine wave in the frequency domain

A complete sine wave in the time domain can be represented by one single spike in the frequency domain.

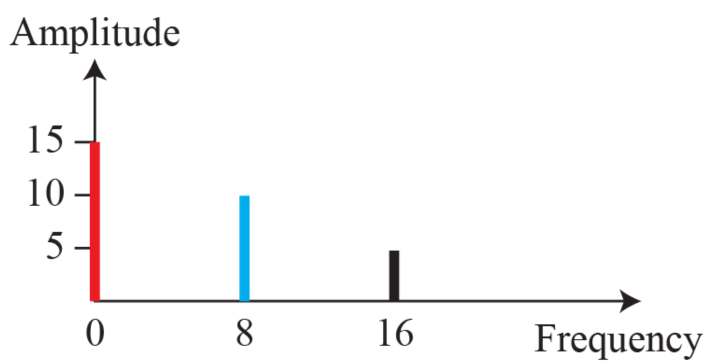
Example 3.7

The frequency domain is more compact and useful when we are dealing with more than one sine wave. For example, Figure 3.9 shows three sine waves, each with different amplitude and frequency. All can be represented by three spikes in the frequency domain.

Figure 3.9: The time and frequency domain of three sine waves



a. Time-domain representation



b. Frequency-domain representation

Composite Signals

- We can send a single sine wave to carry electric energy from one place to another. For example, the power company sends a single sine wave with a frequency of 60 Hz to distribute electric energy to houses and businesses.
- However, If we had only one single sine wave to convey a conversation over the phone, it would make no sense and carry no information. We would just hear a buzz.
- We need to send a composite signal to communicate data.
- A composite signal is made of many simple sine waves.
- According to Fourier analysis, any composite signal is a combination of simple sine waves with different frequencies, amplitudes, and phases.
- If the composite signal is periodic, the decomposition gives a series of signals with discrete frequencies; if the composite signal is nonperiodic, the decomposition gives a combination of sine waves with continuous frequencies.

Example 3.8

Figure 3.10 shows a periodic composite signal with frequency f. This type of signal is not typical of those found in data communications. We can consider it to be three alarm systems, each with a different frequency. The analysis of this signal can give us a good understanding of how to decompose signals. It is very difficult to manually decompose this signal into a series of simple sine waves. However, there are tools, both hardware and software, that can help us do the job. We are not concerned about how it is done; we are only interested in the result. Figure 3.11 shows the result of decomposing the above signal in both the time and frequency domains.

Figure 3.10: ***A composite periodic signal***

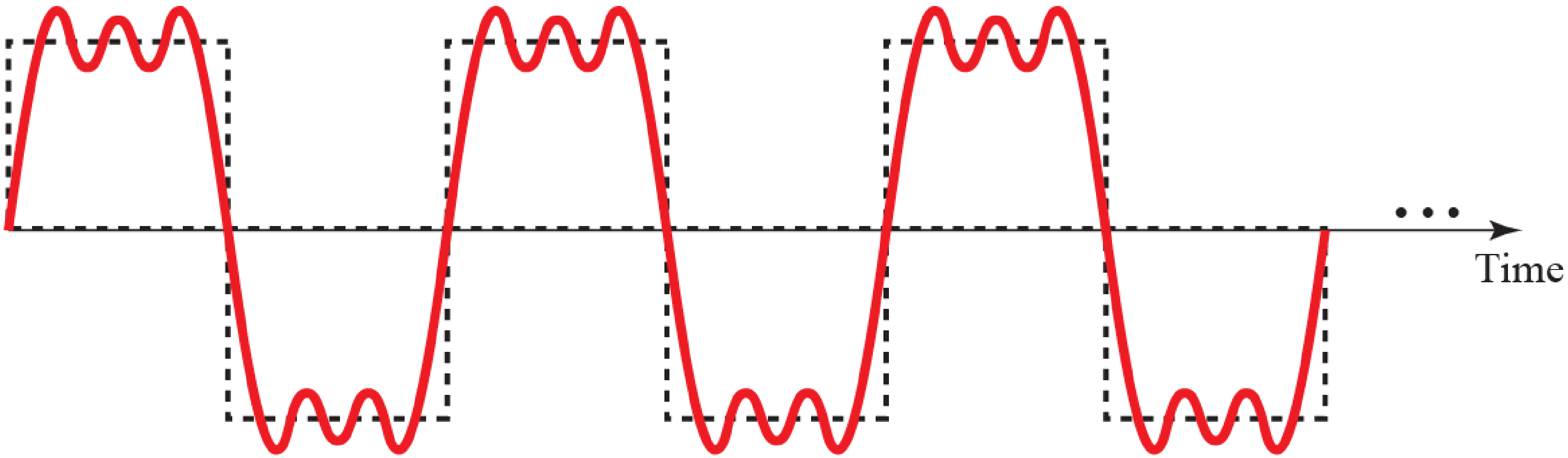
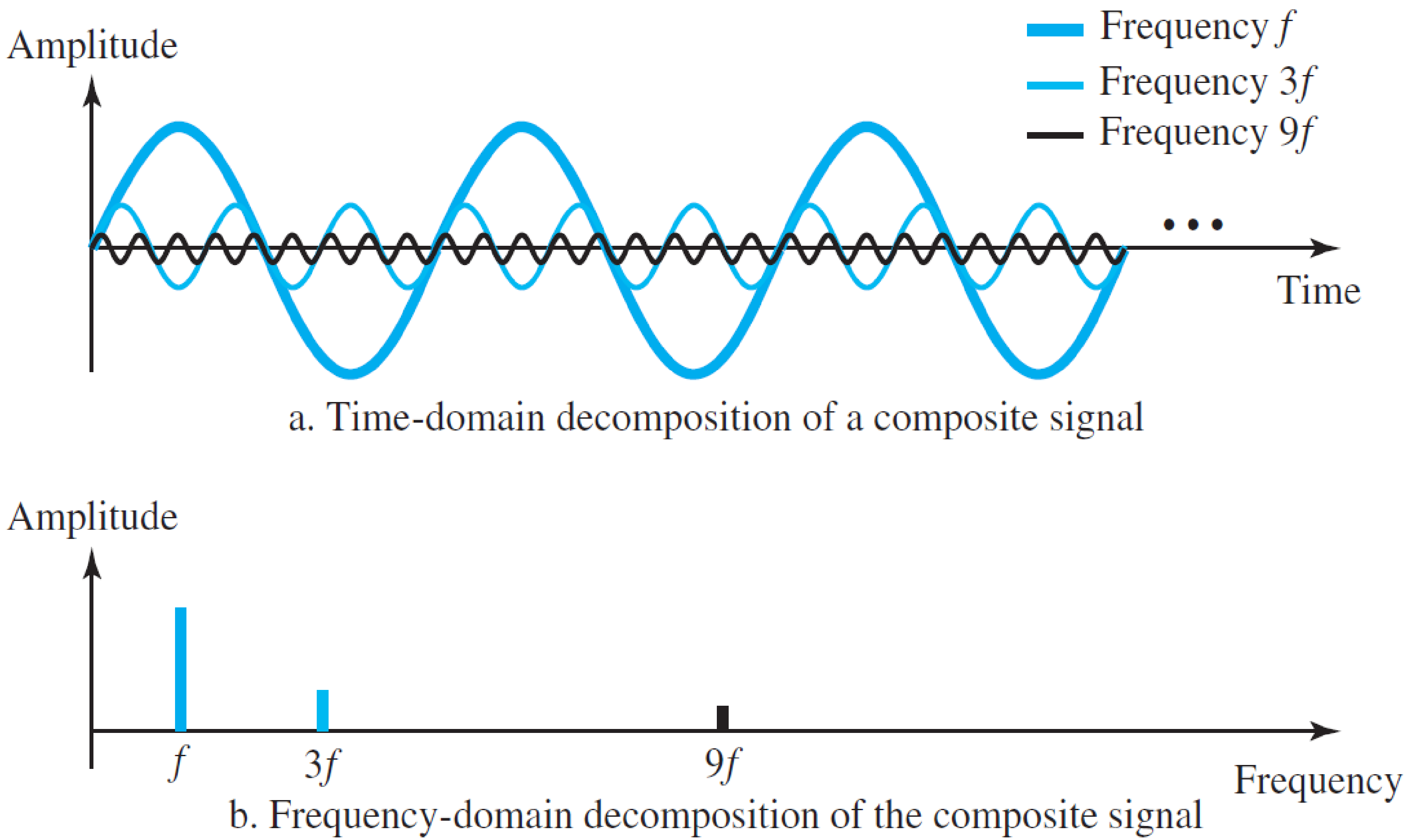


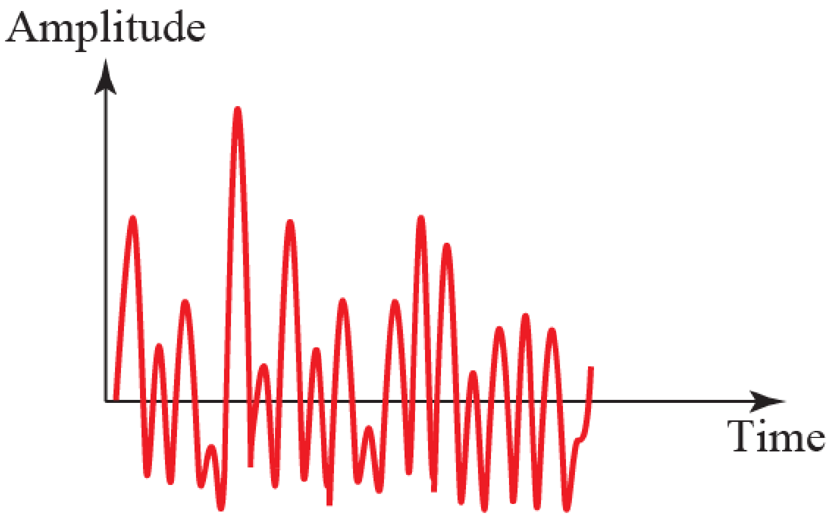
Figure 3.11: Decomposition of a composite periodic signal



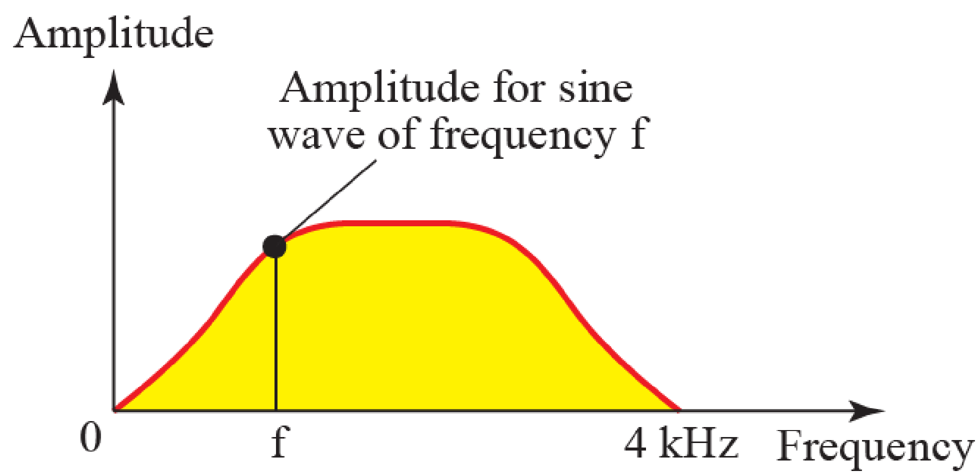
Example 3.9

Figure 3.12 shows a nonperiodic composite signal. It can be the signal created by a microphone or a telephone set when a word or two is pronounced. In this case, the composite signal cannot be periodic, because that implies that we are repeating the same word or words with exactly the same tone.

Figure 3.12: ***Time and frequency domain of a non-periodic signal***



a. Time domain



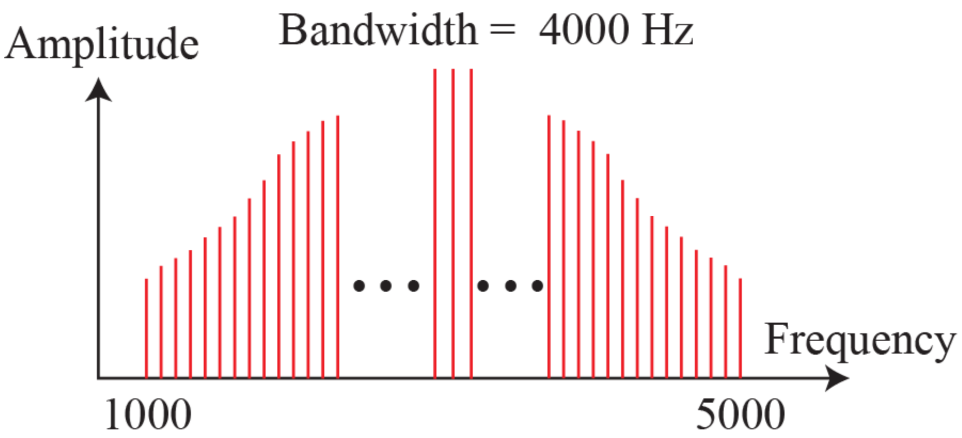
b. Frequency domain

Bandwidth

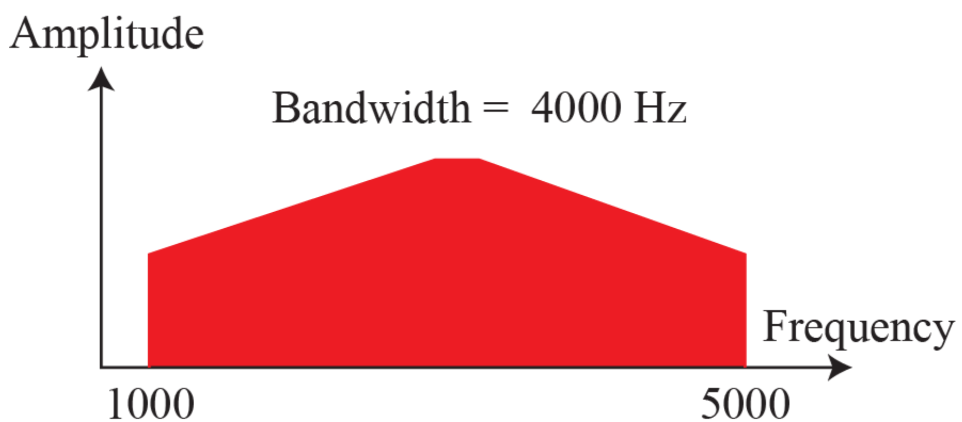
- The range of frequencies contained in a composite signal is its bandwidth.
- The bandwidth is normally a difference between two numbers. For example, if a composite signal contains frequencies between 1000 and 5000, its bandwidth is 5000 − 1000, or 4000.

The bandwidth of a composite signal is the difference between the highest and the lowest frequencies contained in that signal.

Figure 3.13: ***The bandwidth of periodic and nonperiodic composite signals***



a. Bandwidth of a periodic signal



b. Bandwidth of a nonperiodic signal

Example 3.10

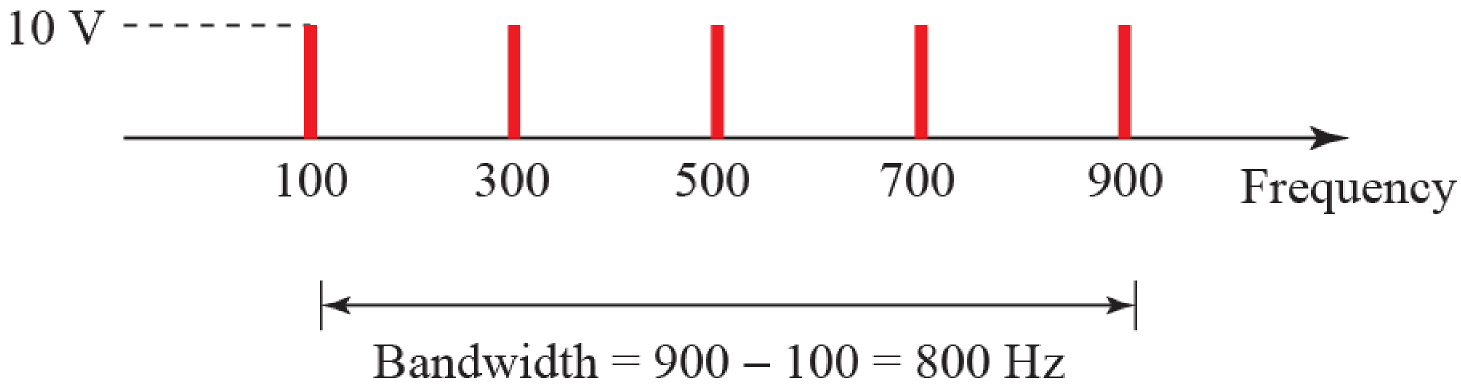
If a periodic signal is decomposed into five sine waves with frequencies of 100, 300, 500, 700, and 900 Hz, what is its bandwidth? Draw the spectrum, assuming all components have a maximum amplitude of 10 V.

Solution

Let f_h be the highest frequency, f_l the lowest frequency, and B the bandwidth. Then

$$B = f_h - f_l = 900 - 100 = 800 \text{ Hz}$$

Figure 3.14: The bandwidth for example 3.10



Example 3.11

A periodic signal has a bandwidth of 20 Hz. The highest frequency is 60 Hz. What is the lowest frequency? Draw the spectrum if the signal contains all frequencies of the same amplitude.

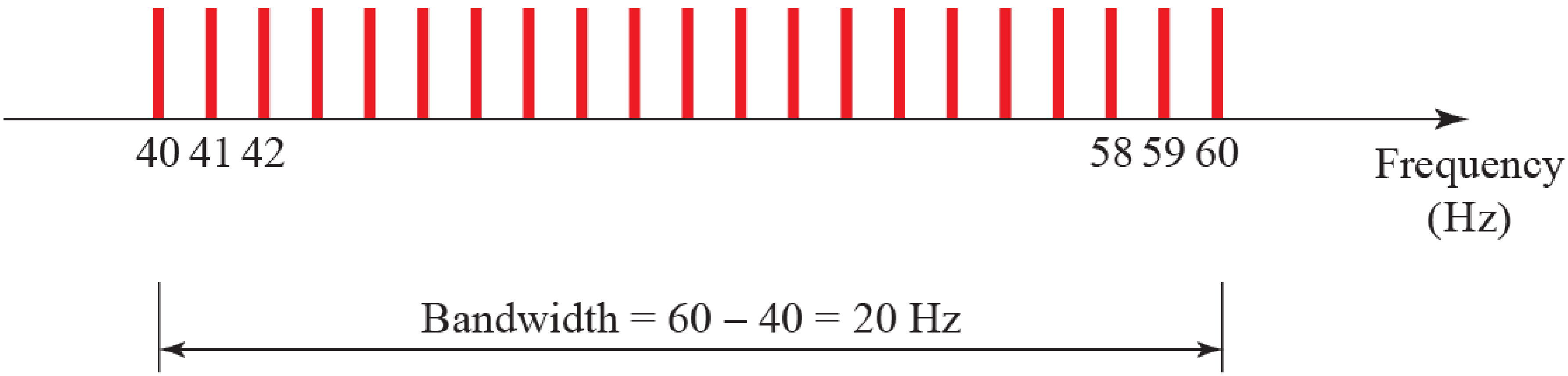
Solution

Let f_h be the highest frequency, f_l the lowest frequency, and B the bandwidth. Then

$$B = f_h - f_l \longrightarrow 20 = 60 - f_l \longrightarrow f_l = 60 - 20 = 40 \text{ Hz}$$

The spectrum contains all integer frequencies. We show this by a series of spikes (see Figure 3.15).

Figure 3.15: The bandwidth for example 3.11



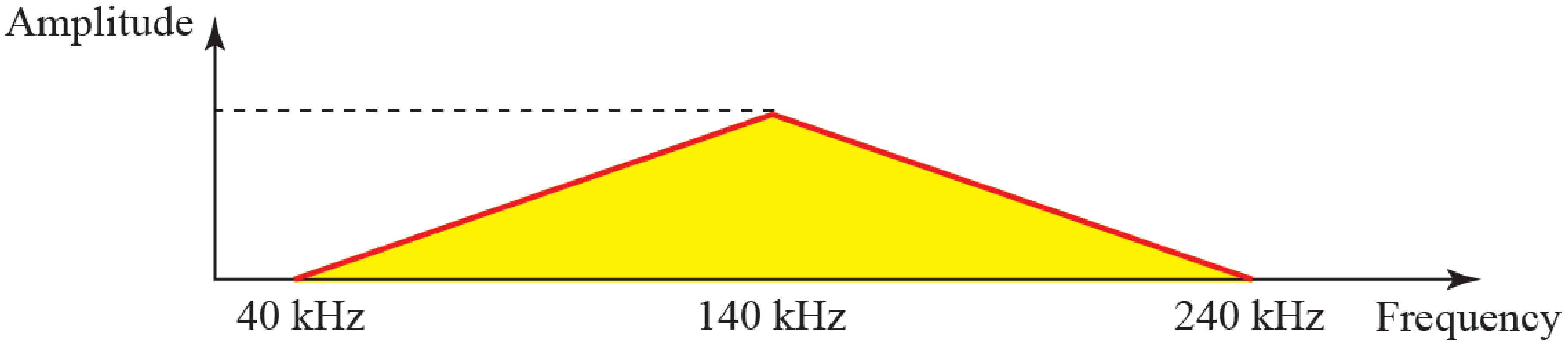
Example 3.12

A nonperiodic composite signal has a bandwidth of 200 kHz, with a middle frequency of 140 kHz and peak amplitude of 20 V. The two extreme frequencies have an amplitude of 0. Draw the frequency domain of the signal.

Solution

The lowest frequency must be at 40 kHz and the highest at 240 kHz. Figure 3.16 shows the frequency domain and the bandwidth.

Figure 3.16: ***The bandwidth for example 3.12***



Example 3.15

Another example of a nonperiodic composite signal is the signal received by an old-fashioned analog black-and-white TV. A TV screen is made up of pixels (picture elements) with each pixel being either white or black. The screen is scanned 30 times per second. If we assume a resolution of 525 × 700 (525 vertical lines and 700 horizontal lines), which is a ratio of 3:4, we have 367,500 pixels per screen. If we scan the screen 30 times per second, this is 367,500 × 30 = 11,025,000 pixels per second. The worst-case scenario is alternating black and white pixels. In this case, we need to represent one color by the minimum amplitude and the other color by the maximum amplitude. We can send 2 pixels per cycle.

Example 3.15 (continued)

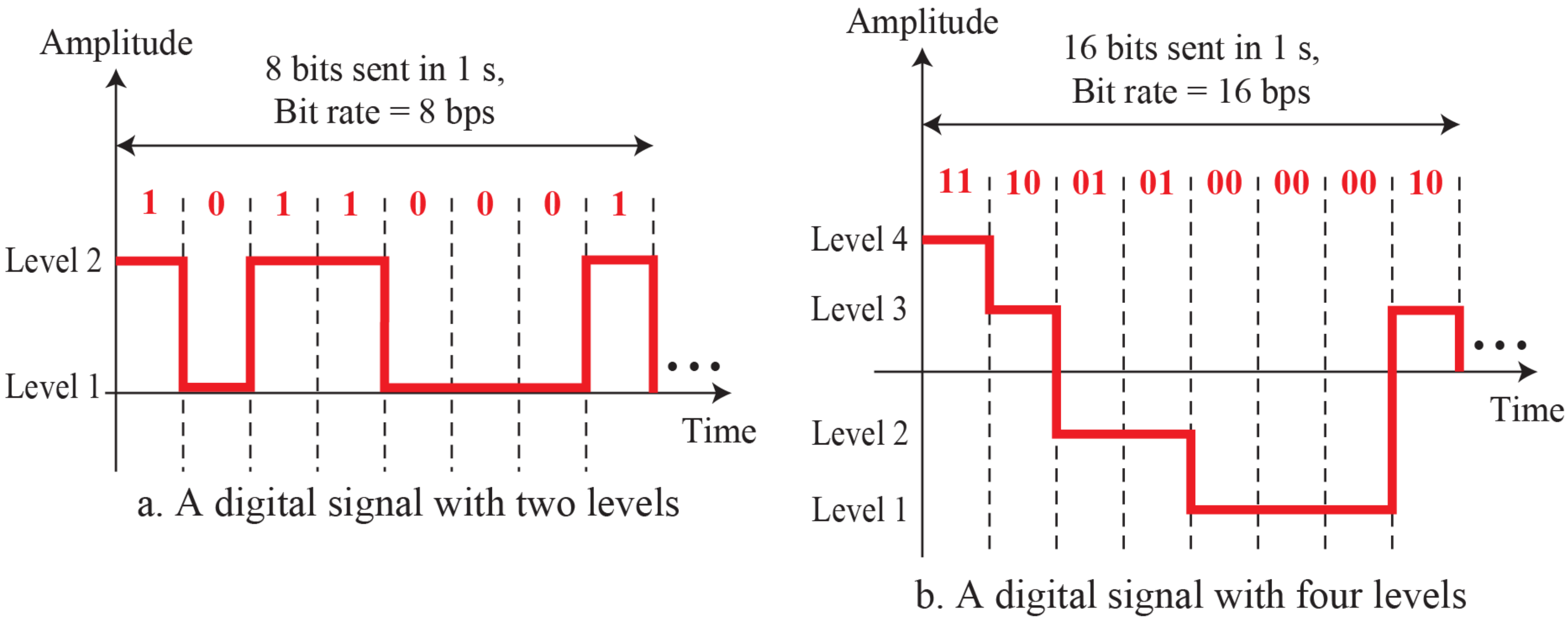
Therefore, we need 11,025,000 / 2 = 5,512,500 cycles per second, or Hz. The bandwidth needed is 5.5124 MHz. This worst-case scenario has such a low probability of occurrence that the assumption is that we need only 70 percent of this bandwidth, which is 3.85 MHz. Since audio and synchronization signals are also needed, a 4-MHz bandwidth has been set aside for each black and white TV channel. An analog color TV channel has a 6-MHz bandwidth.

DIGITAL SIGNALS

In addition to being represented by an analog signal, information can also be represented by a digital signal.

For example, a 1 can be encoded as a positive voltage and a 0 as zero voltage. A digital signal can have more than two levels. In this case, we can send more than 1 bit for each level. Figure 3.17 shows two signals, one with two levels
and the other with four.

Figure 3.17: ***Two digital signals: one with two signal levels and the other with four signal levels***



In general, if a signal has L levels, each level needs $\log_2 L$ bits.

Example 3.16

A digital signal has eight levels. How many bits are needed per level? We calculate the number of bits from the following formula. Each signal level is represented by 3 bits.

$$\text{Number of bits per level} = \log_2 8 = 3$$

Example 3.17

A digital signal has nine levels. How many bits are needed per level? We calculate the number of bits by using the formula. Each signal level is represented by 3.17 bits. However, this answer is not realistic. The number of bits sent per level needs to be an integer as well as a power of 2. For this example, 4 bits can represent one level.

Bit Rate

- Most digital signals are nonperiodic, and thus period and frequency are not appropriate characteristics.
- Another term—bit rate (instead of frequency)—is used to describe digital signals.
- The bit rate is the number of bits sent in 1s, expressed in bits per second (bps).

Example 3.18

Assume we need to download text documents at the rate of 100 pages per second. What is the required bit rate of the channel?

Solution

A page is an average of 24 lines with 80 characters in each line. If we assume that one character requires 8 bits, the bit rate is

$$100 \times 24 \times 80 \times 8 = 1{,}536{,}000 \text{ bps} = 1.536 \text{ Mbps}$$

Example 3.19

A digitized voice channel is made by digitizing a 4-kHz bandwidth analog voice signal. We need to sample the signal at twice the highest frequency (two samples per hertz). We assume that each sample requires 8 bits. What is the required bit rate?

Solution

The bit rate can be calculated as:

$$2 \times 4000 \times 8 = 64{,}000 \text{ bps} = 64 \text{ kbps}$$

Example 3.20

What is the bit rate for high-definition TV (HDTV)?

Solution

HDTV uses digital signals to broadcast high quality video signals. The HDTV screen is normally a ratio of 16 : 9 (in contrast to 4 : 3 for regular TV), which means the screen is wider. There are 1920 by 1080 pixels per screen, and the screen is renewed 30 times per second. Twenty-four bits represents one color pixel. We can calculate the bit rate as

$$1920 \times 1080 \times 30 \times 24 = 1{,}492{,}992{,}000 \approx 1.5 \text{ Gbps}$$

The TV stations reduce this rate to 20 to 40 Mbps through compression.

Bit Length

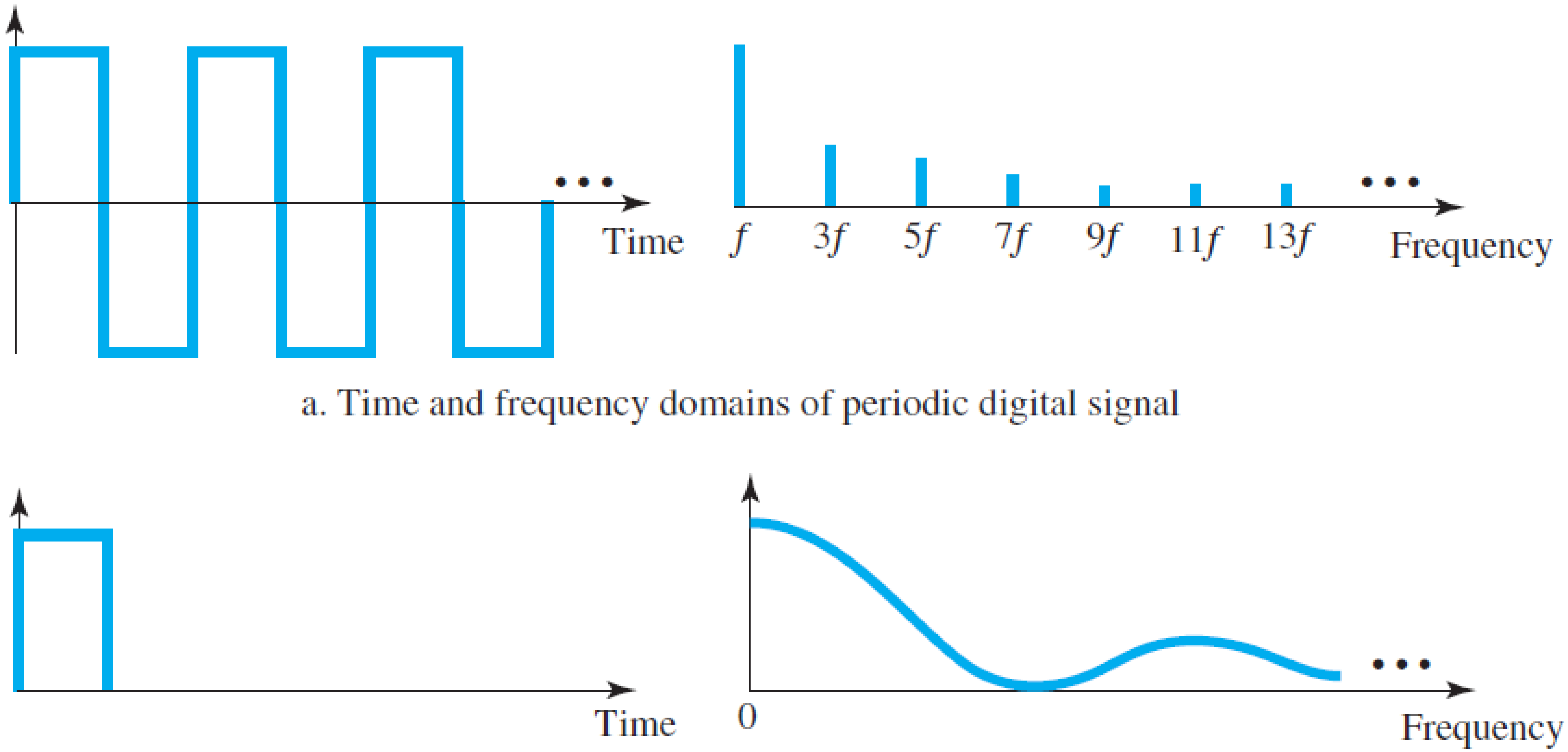
- We discussed the concept of the wavelength for an analog signal: the distance one cycle occupies on the transmission medium.
- We can define something similar for a digital signal: the bit length.
- The bit length is the distance one bit occupies on the transmission medium.

Bit length = propagation speed X **bit duration**

Digital Signal as a Composite Analog Signal

- Based on Fourier analysis, a digital signal is a composite analog signal.
- A digital signal, in the time domain, comprises connected vertical and horizontal line segments.
- A vertical line in the time domain means a frequency of infinity (sudden change in time)
- A horizontal line in the time domain means a frequency of zero (no change in time).
- Going from a frequency of zero to a frequency of infinity (and vice versa) implies all frequencies in between are part of the domain.
- Fourier analysis can be used to decompose a digital signal.
- If the digital signal is periodic, which is rare in data communications, the decomposed signal has a frequency domain representation with an infinite bandwidth and discrete frequencies.
- If the digital signal is non-periodic, the decomposed signal still has an infinite bandwidth, but the frequencies are continuous.

Figure 3.18 ***The time and frequency domains of periodic and nonperiodic digital signals***



a. Time and frequency domains of periodic digital signal

b. Time and frequency domains of nonperiodic digital signal

Transmission of Digital Signals

- A digital signal, periodic or non-periodic, is a composite analog signal with frequencies between zero and infinity.
- For the remainder of the discussion, let us consider the case of a non-periodic digital signal, similar to the ones we encounter in data communications.
- The fundamental question is, how can we send a digital signal from point A to point B?
- We can transmit a digital signal by using one of two different approaches:
  - **Baseband transmission** or
  - **Broadband transmission (using modulation).**

Baseband Transmission

- Baseband transmission means sending a digital signal over a channel without changing the digital signal to an analog signal
- Baseband transmission requires that we have a low-pass channel, a channel with a bandwidth that starts from zero.
- This is the case if we have a dedicated medium with a bandwidth constituting only one channel.
- For example, the entire bandwidth of a cable connecting two computers is one single channel.
- As another example, we may connect several computers to a bus, but not allow more than two stations to communicate at a time.

Figure 3.19 Baseband transmission

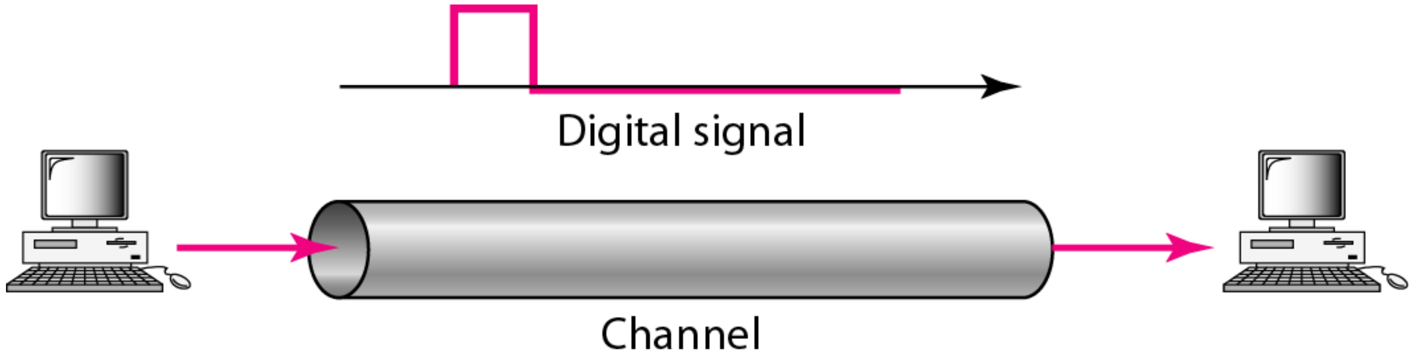
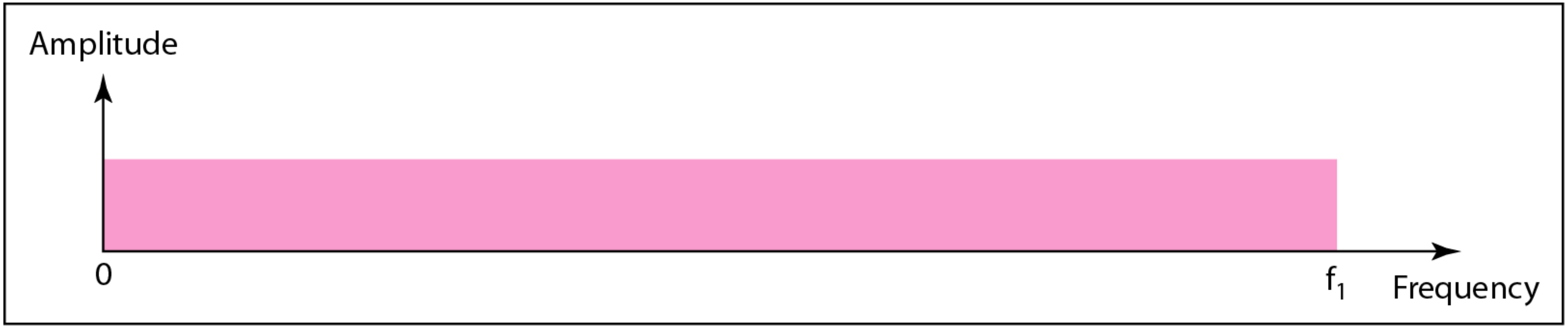


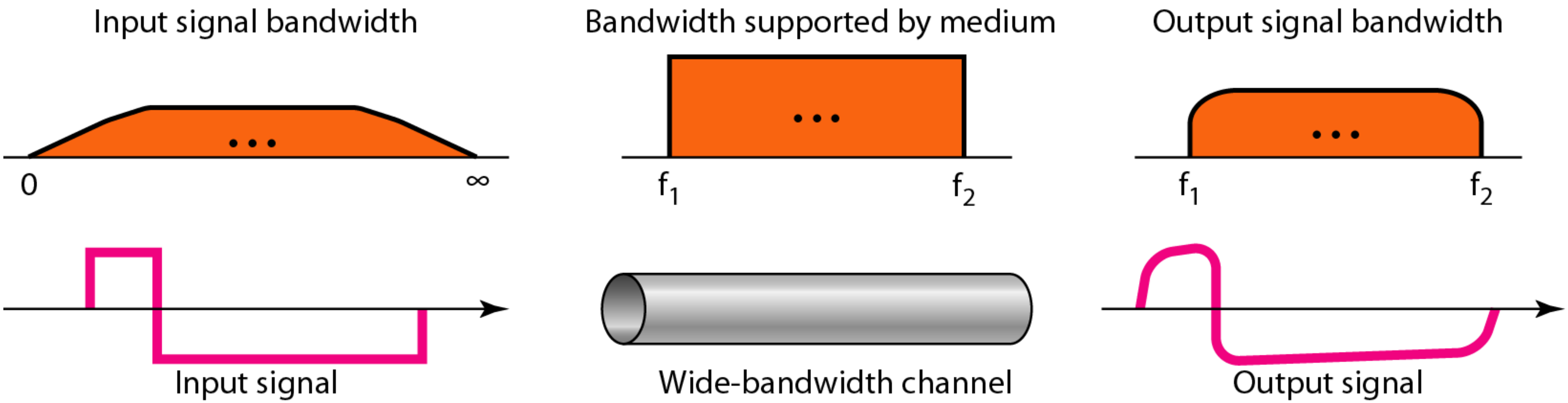
Figure 3.20 **Bandwidths of two low-pass channels**



a. Low-pass channel, wide bandwidth

b. Low-pass channel, narrow bandwidth

Figure 3.21 Baseband transmission using a dedicated medium



Note

Baseband transmission of a digital signal that preserves the shape of the digital signal is possible only if we have a low-pass channel with an infinite or very wide bandwidth.

Example 3.21

An example of a dedicated channel where the entire bandwidth of the medium is used as one single channel is a LAN. Almost every wired LAN today uses a dedicated channel for two stations communicating with each other. In a bus topology LAN with multipoint connections, only two stations can communicate with each other at each moment in time (timesharing); the other stations need to refrain from sending data. In a star topology LAN, the entire channel between each station and the hub is used for communication between these two entities.

Figure 3.22 Rough approximation of a digital signal using the first harmonic for worst case

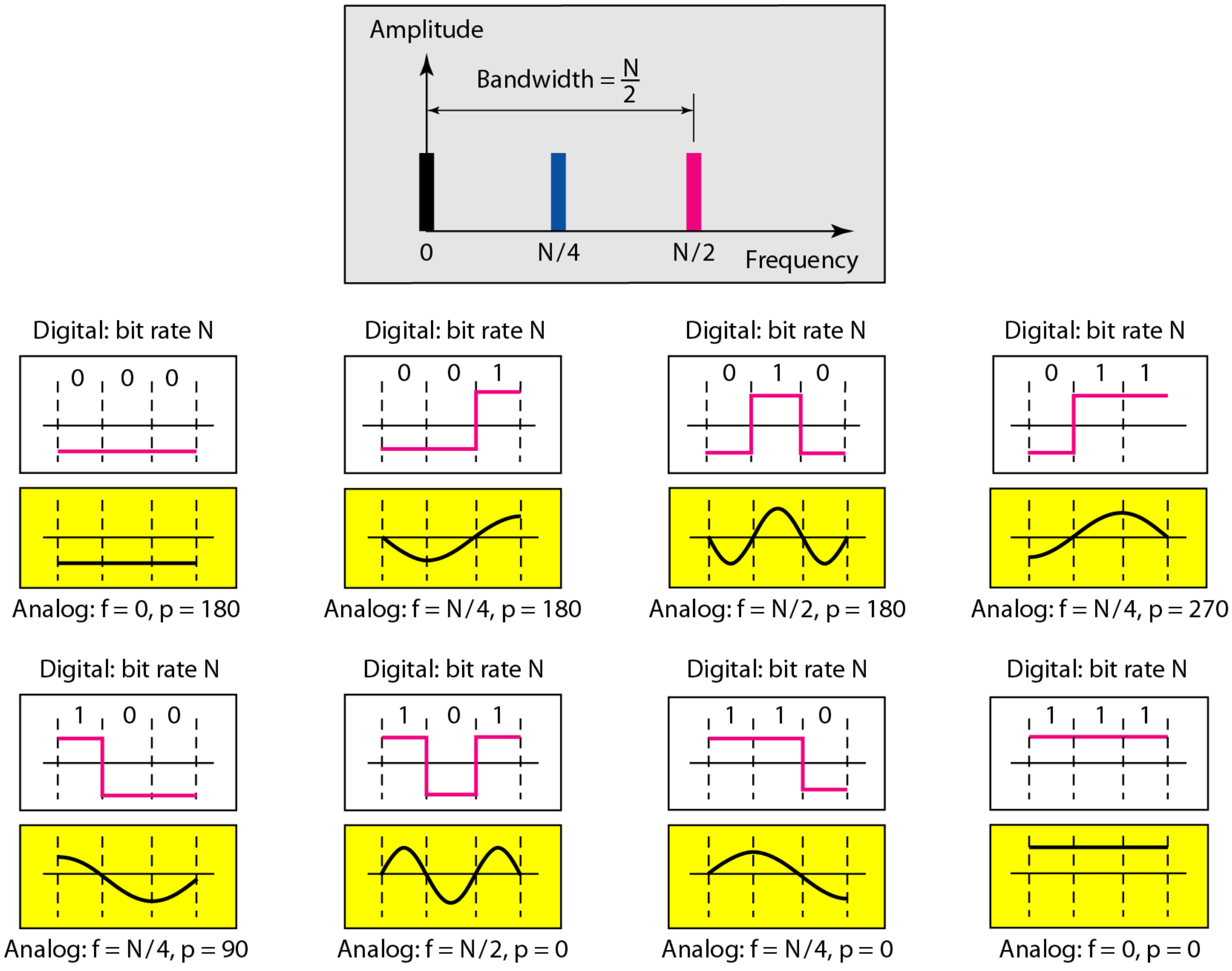
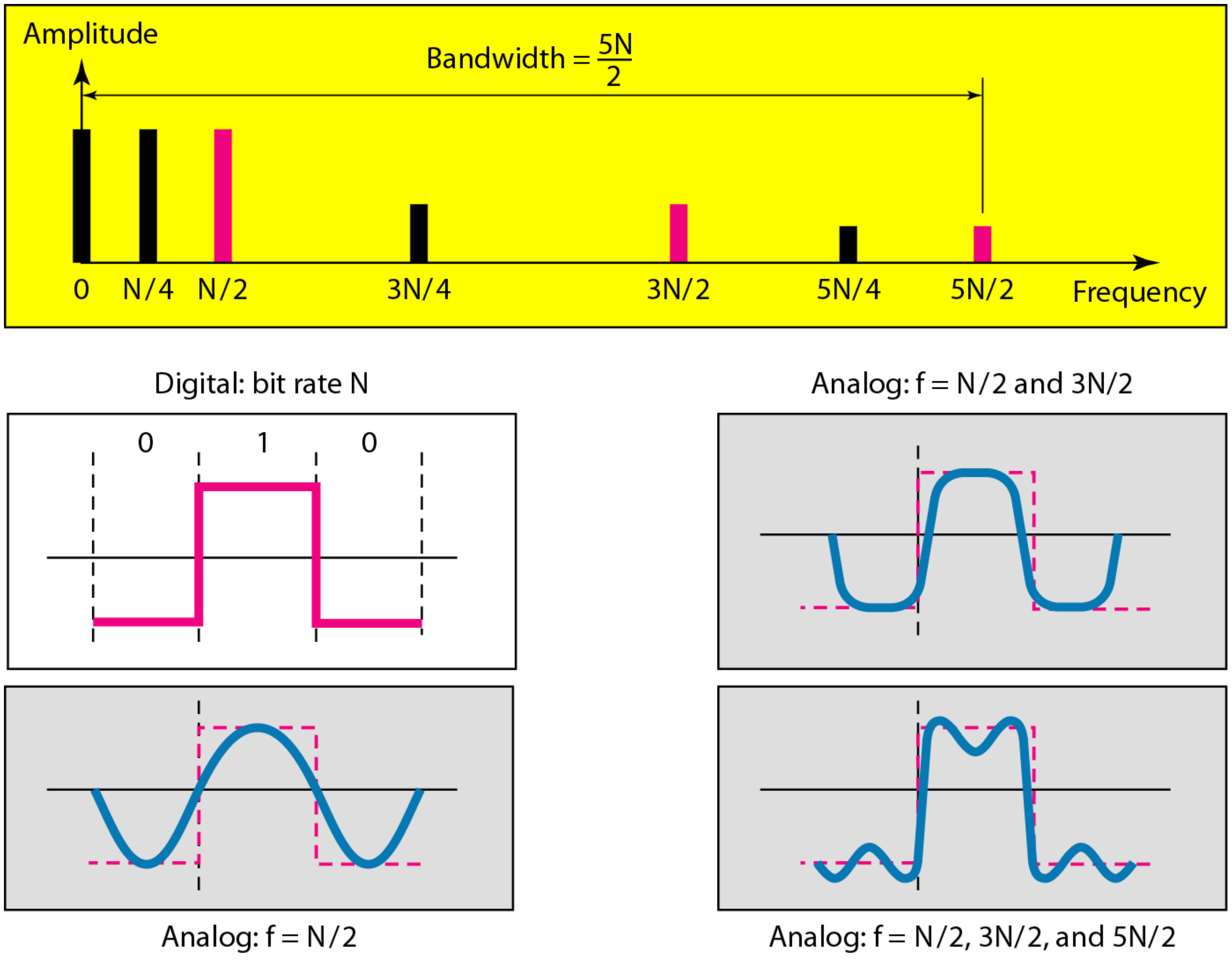


Figure 3.23 Simulating a digital signal with first three harmonics



Note

In baseband transmission, the required bandwidth is proportional to the bit rate; if we need to send bits faster, we need more bandwidth.

Table 3.2 ***Bandwidth requirements***

Bit Rate	*Harmonic 1*	*Harmonics 1, 3*	*Harmonics 1, 3, 5*
$n = 1$ kbps	$B = 500$ Hz	$B = 1.5$ kHz	$B = 2.5$ kHz
$n = 10$ kbps	$B = 5$ kHz	$B = 15$ kHz	$B = 25$ kHz
$n = 100$ kbps	$B = 50$ kHz	$B = 150$ kHz	$B = 250$ kHz

Example 3.22

What is the required bandwidth of a low-pass channel if we need to send 1 Mbps by using baseband transmission?

Solution

The answer depends on the accuracy desired.

a. The minimum bandwidth, is B = bit rate /2, or 500 kHz.

b. A better solution is to use the first and the third harmonics with $B = 3 \times 500$ kHz $= 1.5$ MHz.

c. Still a better solution is to use the first, third, and fifth harmonics with $B = 5 \times 500$ kHz $= 2.5$ MHz.

Example 3.23

We have a low-pass channel with bandwidth 100 kHz. What is the maximum bit rate of this channel?

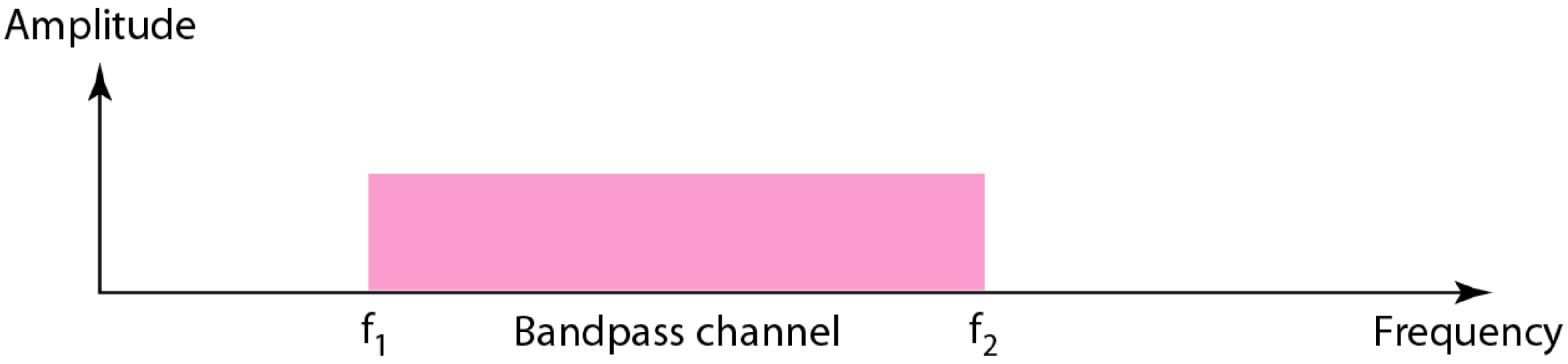
Solution

The maximum bit rate can be achieved if we use the first harmonic. The bit rate is 2 times the available bandwidth, or 200 kbps.

Broadband Transmission (Using Modulation)

- Broadband transmission or modulation means changing the digital signal to an analog signal for transmission.
- Modulation allows us to use a bandpass channel—a channel with a bandwidth that does not start from zero.
- This type of channel is more available than a low-pass channel.

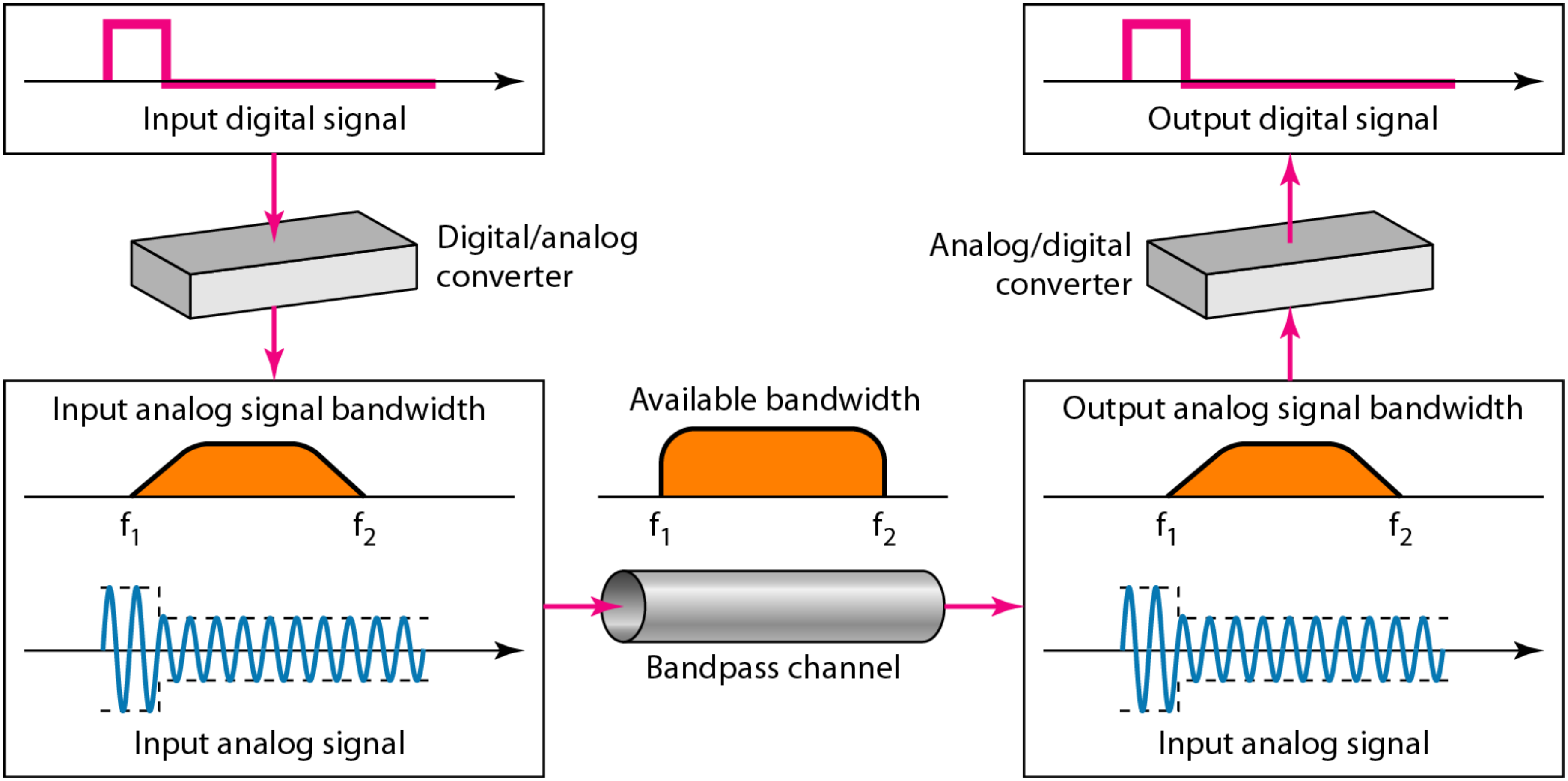
Figure 3.24 ***Bandwidth of a bandpass channel***



Note

If the available channel is a bandpass channel, we cannot send the digital signal directly to the channel;
we need to convert the digital signal to an analog signal before transmission.

Figure 3.25 ***Modulation of a digital signal for transmission on a bandpass channel***



Example 3.24

An example of broadband transmission using modulation is the sending of computer data through a telephone subscriber line, the line connecting a resident to the central telephone office. These lines are designed to carry voice with a limited bandwidth. The channel is considered a bandpass channel. We convert the digital signal from the computer to an analog signal, and send the analog signal. We can install two converters to change the digital signal to analog and vice versa at the receiving end. The converter, in this case, is called a modem.

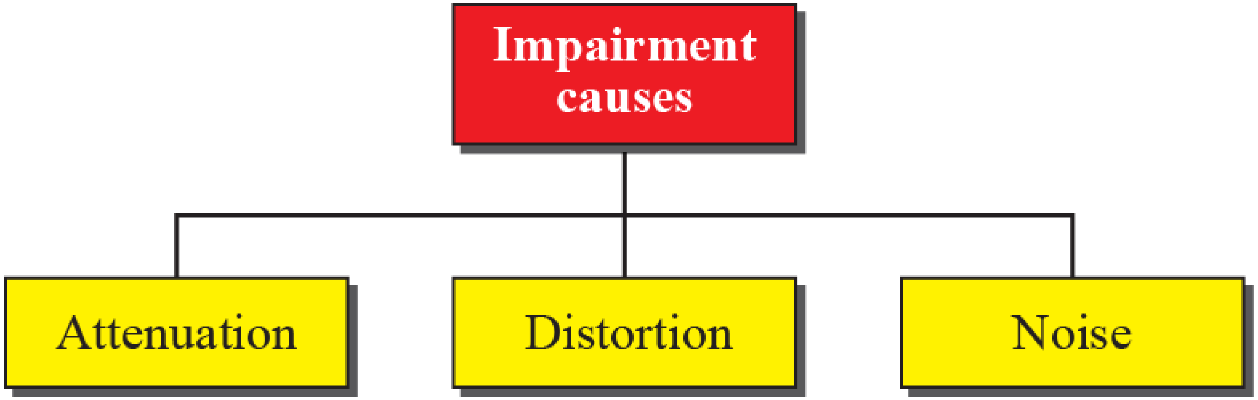
Example 3.25

A second example is the digital cellular telephone. For better reception, digital cellular phones convert the analog voice signal to a digital signal. Although the bandwidth allocated to a company providing digital cellular phone service is very wide, we still cannot send the digital signal without conversion. The reason is that we only have a bandpass channel available between caller and callee. We need to convert the digitized voice to a composite analog signal before sending.

TRANSMISSION IMPAIRMENT

Signals travel through transmission media, which are not perfect. The imperfection causes signal impairment. This means that the signal at the beginning of the medium is not the same as the signal at the end of the medium. What is sent is not what is received. Three causes of impairment are attenuation, distortion, and noise

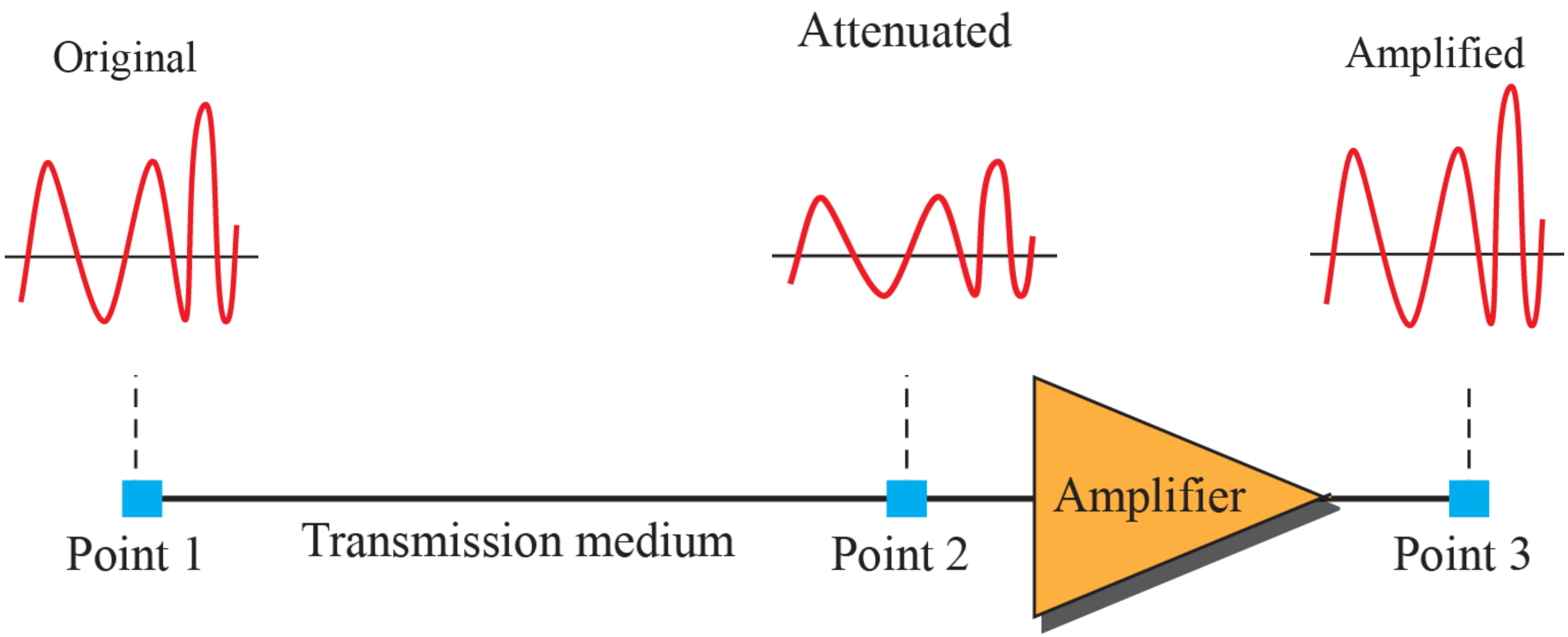
Figure 3.26: *Causes of impairment*



Attenuation

- Attenuation means a loss of energy.
- When a signal, simple or composite, travels through a medium, it loses some of its energy in overcoming the resistance of the medium.
- That is why a wire carrying electric signals gets warm, if not hot, after a while.
- Some of the electrical energy in the signal is converted to heat.
- To compensate for this loss, amplifiers are used to amplify the signal.

Figure 3.27: Attenuation and amplification



Example 3.26

Suppose a signal travels through a transmission medium and its power is reduced to one half. This means that P2 = 0.5 P1. In this case, the attenuation (loss of power) can be calculated as

$$\mathbf{10 \log_{10} P_2/P_1 = 10 \log_{10} (0.5\, P_1)/P_1 = 10 \log_{10} 0.5 = 10 \times (-0.3) = -3\ dB.}$$

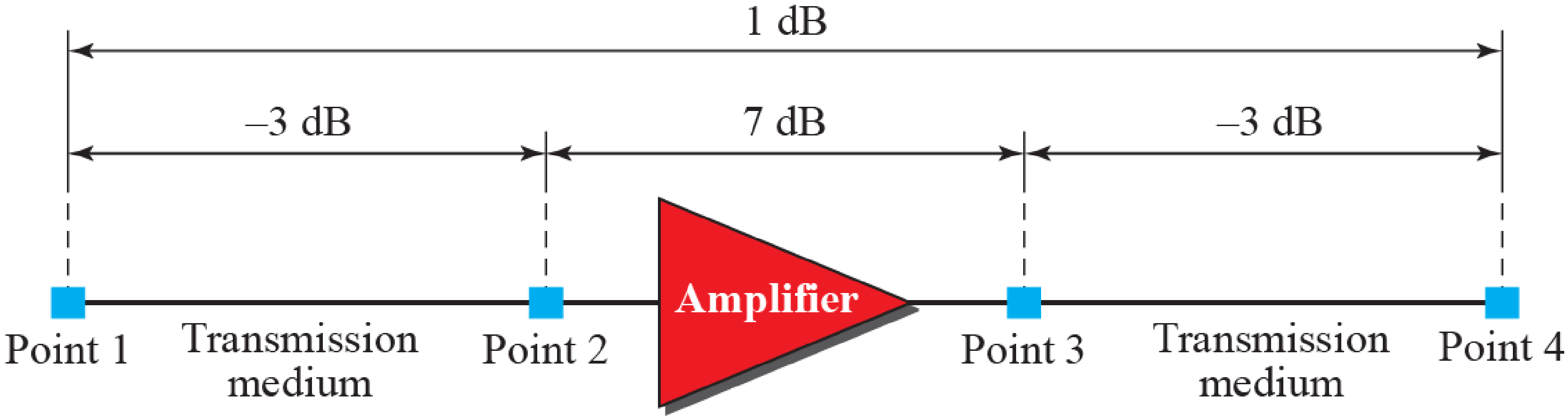
A loss of 3 dB (–3 dB) is equivalent to losing one-half the power.

Example 3.27

A signal travels through an amplifier, and its power is increased 10 times. This means that $P_2 = 10P_1$. In this case, the amplification (gain of power) can be calculated as

$$10 \log_{10} \frac{P_2}{P_1} = 10 \log_{10} \frac{10P_1}{P_1} = 10 \log_{10} 10 = 10(1) = 10 \text{ dB}$$

Figure 3.28: Decibels for Example 3.28



Example 3.28

One reason that engineers use the decibel to measure the changes in the strength of a signal is that decibel numbers can be added (or subtracted) when we are measuring several points (cascading) instead of just two. In Figure 3.28 a signal travels from point 1 to point 4. The signal is attenuated by the time it reaches point 2. Between points 2 and 3, the signal is amplified. Again, between points 3 and 4, the signal is attenuated. We can find the resultant decibel value for the signal just by adding the decibel measurements between each set of points. In this case, the decibel value can be calculated as

$$\mathrm{dB} = -3 + 7 - 3 = +1$$

Example 3.29

Sometimes the decibel is used to measure signal power in milliwatts. In this case, it is referred to as dB_m and is calculated as dB_m = 10 log10 Pm, where Pm is the power in milliwatts. Calculate the power of a signal if its $dB_m = -30$.

Solution

We can calculate the power in the signal as

$$\mathrm{dB_m} = 10 \log_{10} \longrightarrow dB_m = -30 \longrightarrow \log_{10} P_m = -3 \longrightarrow P_m = 10^{-3}\,\mathrm{mW}$$

Example 3.30

The loss in a cable is usually defined in decibels per kilometer (dB/km). If the signal at the beginning of a cable with −0.3 dB/km has a power of 2 mW, what is the power of the signal at 5 km?

Solution

The loss in the cable in decibels is $5 \times (-0.3) = -1.5$ dB. We can calculate the power as

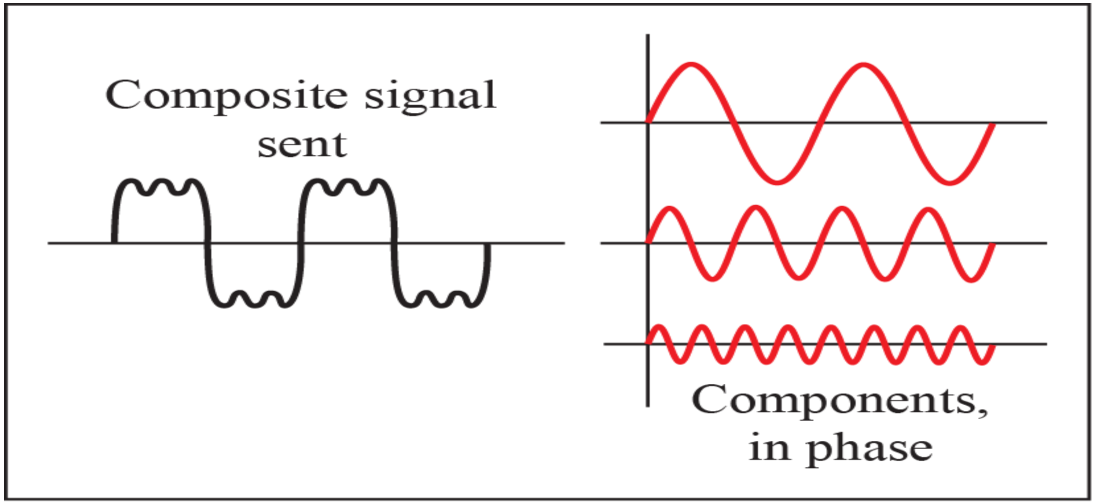
$$\text{dB} = 10 \log_{10} (P_2 / P_1) = -1.5 \quad \longrightarrow \quad (P_2 / P_1) = 10^{-0.15} = 0.71$$

$$P_2 = 0.71 P_1 = 0.7 \times 2 \text{ mW} = 1.4 \text{ mW}$$

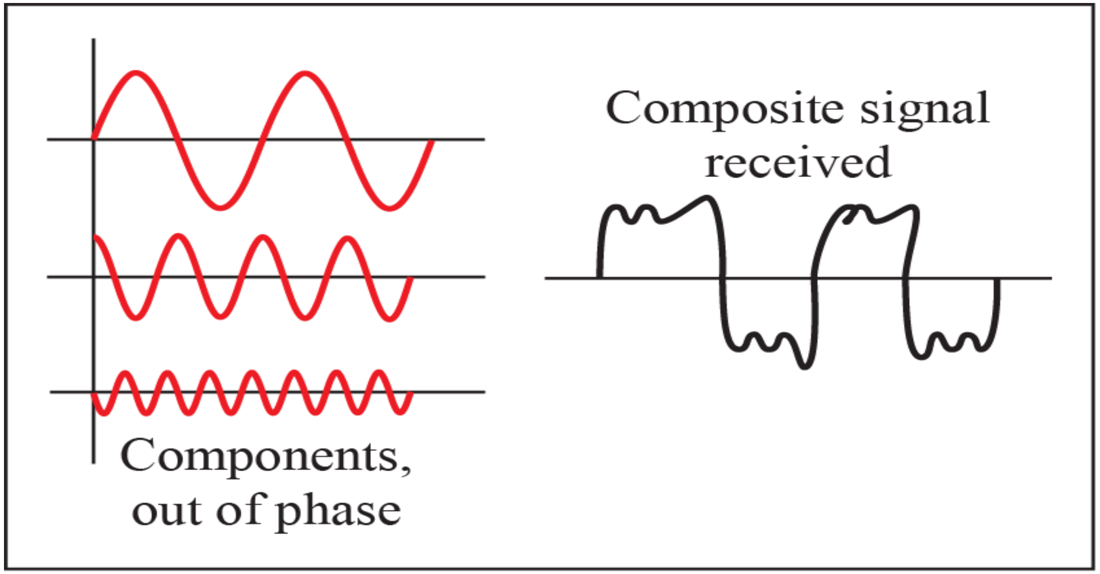
Distortion

- Distortion means that the signal changes its form or shape.
- Distortion can occur in a composite signal made of different frequencies.
- Each signal component has its own propagation speed through a medium and, therefore, its own delay in arriving at the final destination.
- Differences in delay may create a difference in phase if the delay is not exactly the same as the period duration.

Figure 3.29: *Distortion*



At the sender

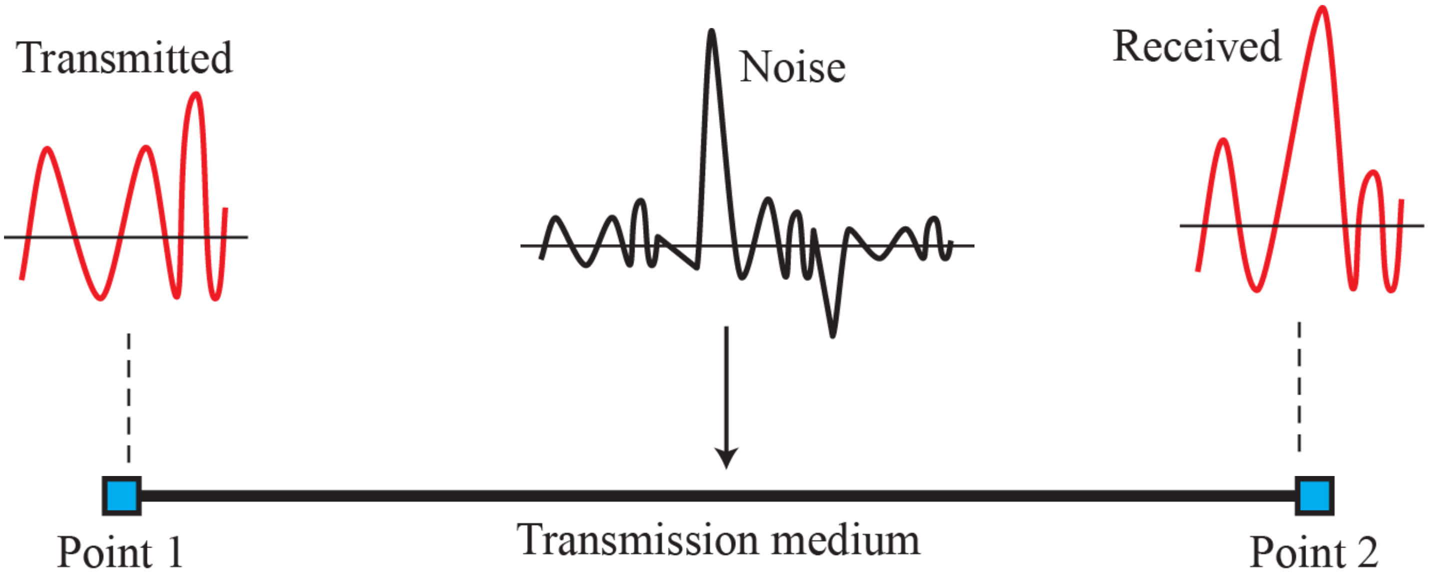


At the receiver

Noise

- Noise is another cause of impairment.
- Several types of noise, such as thermal noise, induced noise, crosstalk, and impulse noise, may corrupt the signal.
- Thermal noise is the random motion of electrons in a wire, which creates an extra signal not originally sent by the transmitter.
- Induced noise comes from sources such as motors.
- Crosstalk is the effect of one wire on the other.

Figure 3.30: Noise



Signal-to-Noise Ratio (SNR)

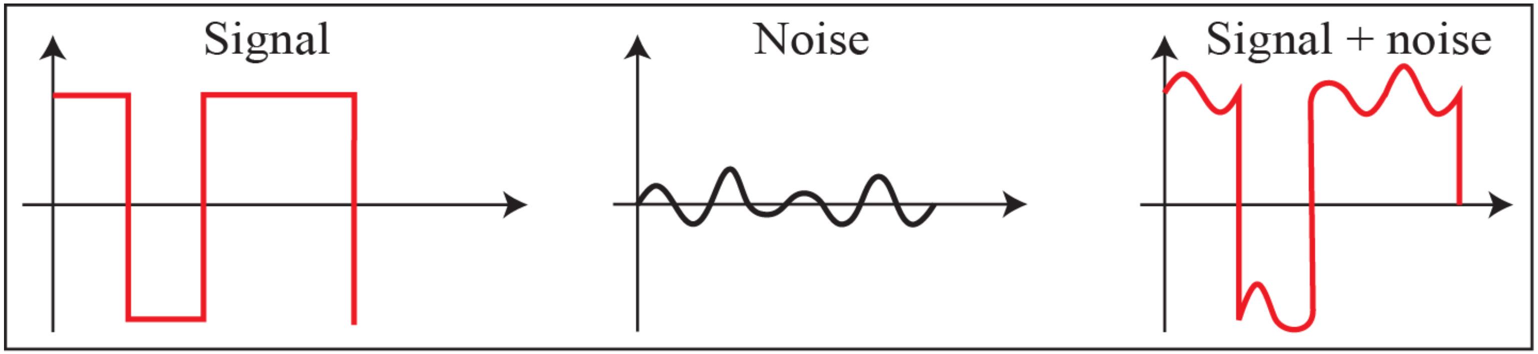
- To find the theoretical bit rate limit, we need to know the ratio of the signal power to the noise power.
- The signal-to-noise ratio is defined as:

$$\text{SNR} = \frac{\text{average signal power}}{\text{average noise power}}$$

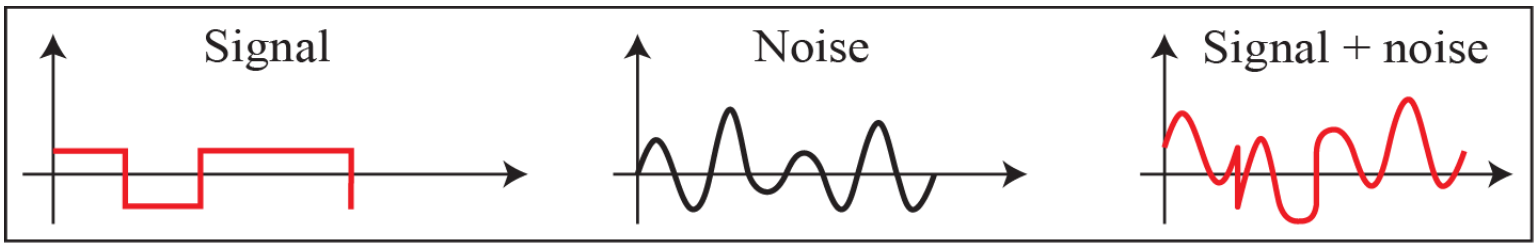
- We need to consider the average signal power and the average noise power because these may change with time.
- SNR is actually the ratio of what is wanted (signal) to what is not wanted (noise).
- A high SNR means the signal is less corrupted by noise; a low SNR means the signal is more corrupted by noise.
- Because SNR is the ratio of two powers, it is often described in decibel units, SNR_{dB}, defined as

$$\text{SNR}_{\text{dB}} = 10 \log_{10} \text{SNR}$$

Figure 3.31: ***Two cases of SNR: a high SNR and a low SNR***



a. High SNR



b. Low SNR

Example 3.31

The power of a signal is 10 mW and the power of the noise is 1 μW; what are the values of SNR and SNR_{dB}?

Solution

The values of SNR and SNR_{dB} can be calculated as follows:

$$SNR = (10{,}000\ \mu w) / (1\ \mu w) = 10{,}000 \qquad SNR_{dB} = 10 \log_{10} 10{,}000 = 10 \log_{10} 10^4 = 40$$

Example 3.32

The values of SNR and SNR_{dB} for a noiseless channel are

Solution

The values of SNR and SNRdB for a noiseless channel are

$$\mathbf{SNR = (signal\ power) / 0 = \infty \quad \longrightarrow \quad SNR_{dB} = 10 \log_{10} \infty = \infty}$$

We can never achieve this ratio in real life; it is an ideal.

DATA RATE LIMITS

A very important consideration in data communications is how fast we can send data, in bits per second, over a channel. Two theoretical formulas were developed to calculate the data rate: one by Nyquist for a noiseless channel, another by Shannon for a noisy channel.

Noiseless Channel: Nyquist Rate

- For a noiseless channel, the Nyquist bit rate formula defines the theoretical maximum bit rate.

$$\text{BitRate} = 2 \times \text{bandwidth} \times \log_2 L$$

- In this formula, bandwidth is the bandwidth of the channel, L is the number of signal levels used to represent data, and BitRate is the bit rate in bits per second.
- According to the formula, we might think that, given a specific bandwidth, we can have any bit rate we want by increasing the number of signal levels.
- Although the idea is theoretically correct, practically there is a limit.
- Increasing the levels of a signal reduces the reliability of the system.

Example 3.33

Does the Nyquist theorem bit rate agree with the intuitive bit rate described in baseband transmission?

Solution

They match when we have only two levels. We said, in baseband transmission, the bit rate is 2 times the bandwidth if we use only the first harmonic in the worst case. However, the Nyquist formula is more general than what we derived intuitively; it can be applied to baseband transmission and modulation. Also, it can be applied when we have two or more levels of signals.

Example 3.34

Consider a noiseless channel with a bandwidth of 3000 Hz transmitting a signal with two signal levels. The maximum bit rate can be calculated as

$$\mathbf{BitRate = 2 \times 3000 \times \log_2 2 = 6000\ bps}$$

Example 3.35

Consider the same noiseless channel transmitting a signal with four signal levels (for each level, we send 2 bits). The maximum bit rate can be calculated as

$$\text{BitRate} = 2 \times 3000 \times \log_2 4 = 12{,}000 \text{ bps}$$

Example 3.36

We need to send 265 kbps over a noiseless channel with a bandwidth of 20 kHz. How many signal levels do we need?

Solution

We can use the Nyquist formula as shown:

$$265{,}000 = 2 \times 20{,}000 \times \log_2 L \longrightarrow \log_2 L = 6.625 \longrightarrow L = 2^{6.625} = 98.7 \text{ levels}$$

Since this result is not a power of 2, we need to either increase the number of levels or reduce the bit rate. If we have 128 levels, the bit rate is 280 kbps. If we have 64 levels, the bit rate is 240 kbps.

Noisy Channel: Shannon Capacity

- In reality, we cannot have a noiseless channel; the channel is always noisy.
- In 1944, Claude Shannon introduced a formula, called the Shannon capacity, to determine the theoretical highest data rate for a noisy channel:

$$\text{Capacity} = \text{bandwidth} \times \log_2(1 + \text{SNR})$$

- In this formula, bandwidth is the bandwidth of the channel, SNR is the signal-to-noise ratio, and capacity is the capacity of the channel in bits per second.
- Note that in the Shannon formula there is no indication of the signal level, which means that no matter how many levels we have, we cannot achieve a data rate higher than the capacity of the channel.
- In other words, the formula defines a characteristic of the channel, not the method of transmission.

Example 3.37

Consider an extremely noisy channel in which the value of the signal-to-noise ratio is almost zero. In other words, the noise is so strong that the signal is faint. For this channel the capacity C is calculated as

$$C = B \log_2 (1 + \text{SNR}) = B \log_2(1 + 0) = B \log_2 1 = B \times 0 = 0$$

This means that the capacity of this channel is zero regardless of the bandwidth. In other
words, we cannot receive any data through this channel.

Example 3.38

We can calculate the theoretical highest bit rate of a regular telephone line. A telephone line normally has a bandwidth of 3000 Hz (300 to 3300 Hz) assigned for data communications. The signal-to-noise ratio is usually 3162. For this channel the capacity is calculated as

$$C = B \log_2 (1 + \text{SNR}) = 3000 \log_2(1 + 3162) = 3000 \times 11.62 = 34{,}860 \text{ bps}$$

This means that the highest bit rate for a telephone line is 34.860 kbps. If we want to send data faster than this, we can either increase the bandwidth of the line or improve the signal-to-noise ratio.

Example 3.39

The signal-to-noise ratio is often given in decibels. Assume that SNR_{dB} = 36 and the channel bandwidth is 2 MHz. The theoretical channel capacity can be calculated as

$$SNR_{dB} = 10 \log_{10} SNR \longrightarrow SNR = 10^{SNR_{dB}/10} \longrightarrow SNR = 10^{3.6} = 3981$$

$$C = B \log_2(1 + SNR) = 2 \times 10^6 \times \log_2 3982 = 24 \text{ Mbps}$$

Example 3.40

When the SNR is very high, we can assume that SNR + 1 is almost the same as SNR. In these cases, the theoretical channel capacity can be simplified to $C = B \times SNR_{dB}/3$. For example, we can calculate the theoretical capacity of the previous example as

$$C = 2\ \text{MHz} \times (36 / 3) = 24\ \text{Mbps}$$

Using Both Limits

In practice, we need to use both methods to find the limits and signal levels. Let us show this with an example.

Example 3.41

We have a channel with a 1-MHz bandwidth. The SNR for this channel is 63. What are the appropriate bit rate and signal level?

Solution

First, we use the Shannon formula to find the upper limit.

$$C = B \log_2(1 + \text{SNR}) = 10^6 \log_2(1 + 63) = 10^6 \log_2 64 = 6 \text{ Mbps}$$

The Shannon formula gives us 6 Mbps, the upper limit. For better performance we choose something lower, 4 Mbps. Then we use the Nyquist formula to find the number of signal levels.

$$4 \text{ Mbps} = 2 \times 1 \text{ MHz} \times \log_2 L \quad \longrightarrow \quad L = 4$$

PERFORMANCE

Up to now, we have discussed the tools of transmitting data (signals) over a network and how the data behave. One important issue in networking is the performance of the network—how good is it? In this section, we introduce terms that we need for future chapters.

Bandwidth

- One characteristic that measures network performance is bandwidth.
- However, the term can be used in two different contexts with two different measuring values: bandwidth in hertz and bandwidth in bits per second
- **Bandwidth in Hertz:**. Bandwidth in hertz is the range of frequencies contained in a composite signal or the range of frequencies a channel can pass.
- **Bandwidth in Bits per Seconds:** The term bandwidth can also refer to the number of bits per second that a channel, a link, or even a network can transmit.
- **Relationship:** There is an explicit relationship between the bandwidth in hertz and bandwidth in bits per second. Basically, an increase in bandwidth in hertz means an increase in bandwidth in bits per second. The relationship depends on whether we have baseband transmission or transmission with modulation.

Example 3.42

The bandwidth of a subscriber line is 4 kHz for voice or data. The bandwidth of this line for data transmission can be up to 56,000 bps using a sophisticated modem to change the digital signal to analog.

Example 3.43

If the telephone company improves the quality of the line and increases the bandwidth to 8 kHz, we can send 112,000 bps by using the same technology as mentioned in Example 3.42.

Throughput

- The throughput is a measure of how fast we can actually send data through a network.
- Although, at first glance, bandwidth in bits per second and throughput seem the same, they are different.
- A link may have a bandwidth of B bps, but we can only send T bps through this link with T always less than B.
- The bandwidth is a potential measurement of a link; the throughput is an actual measurement of how fast we can send data.
- For example, we may have a link with a bandwidth of 1 Mbps, but the devices connected to the end of the link may handle only 200 kbps. This means that we cannot send more than 200 kbps through this link.

Example 3.44

A network with bandwidth of 10 Mbps can pass only an average of 12,000 frames per minute with each frame carrying an average of 10,000 bits. What is the throughput of this network?

Solution

We can calculate the throughput as

$$\text{Throughput} = (12{,}000 \times 10{,}000) / 60 = 2 \text{ Mbps}$$

The throughput is almost one-fifth of the bandwidth in this case.

Latency (Delay)

- The latency or delay defines how long it takes for an entire message to completely arrive at the destination from the time the first bit is sent out from the source.
- We can say that latency is made of four components: **propagation time**, **transmission time**, **queuing time** and **processing delay.**

Latency = propagation time + transmission time + queuing time + processing delay

Example 3.45

What is the propagation time if the distance between the two points is 12,000 km? Assume the propagation speed to be 2.4×10^8 m/s in cable.

Solution

We can calculate the propagation time as

$$\textbf{Propagation time} = (12{,}000 \times 10{,}000) / (2.4 \times 2^8) = 50 \text{ ms}$$

The example shows that a bit can go over the Atlantic Ocean in only 50 ms if there is a direct cable between the source and the destination.

Example 3.46

What are the propagation time and the transmission time for a 2.5-KB (kilobyte) message if the bandwidth of the network is 1 Gbps? Assume that the distance between the sender and the receiver is 12,000 km and that light travels at 2.4×10^8 m/s.

Solution

We can calculate the propagation and transmission time as

$$\text{Propagation time} = (12{,}000 \times 1000) / (2.4 \times 10^8) = 50 \text{ ms}$$

$$\text{Transmission time} = (2500 \times 8) / 10^9 = 0.020 \text{ ms}$$

Note that in this case, because the message is short and the bandwidth is high, the dominant factor is the propagation time, not the transmission time.

Example 3.47

What are the propagation time and the transmission time for a 5-MB (megabyte) message (an image) if the bandwidth of the network is 1 Mbps? Assume that the distance between the sender and the receiver is 12,000 km and that light travels at 2.4×10^8 m/s.

Solution

We can calculate the propagation and transmission times as

$$\text{Propagation time} = (12{,}000 \times 1000) / (2.4 \times 10^8) = 50 \text{ ms}$$

$$\text{Transmission time} = (5{,}000{,}000 \times 8) / 10^6 = 40 \text{ s}$$

We can calculate the propagation and transmission times as

Bandwidth-Delay Product

- Bandwidth and delay are two performance metrics of a link.
- However, as we will see in this chapter and future chapters, what is very important in data communications is the product of the two, the bandwidth-delay product.
- Let us elaborate on this issue, using two hypothetical cases as examples.

Figure 3.32: Filling the links with bits for Case 1

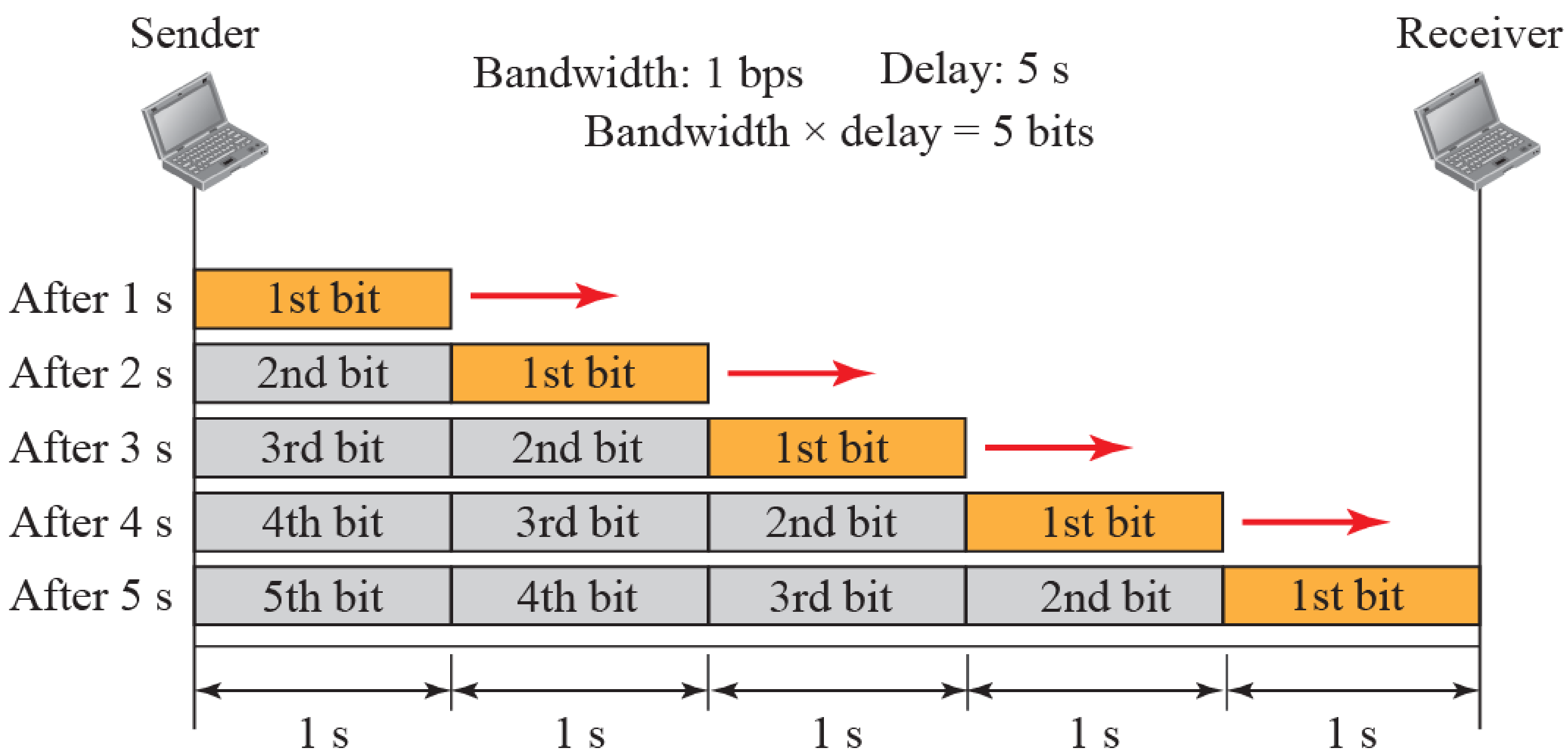
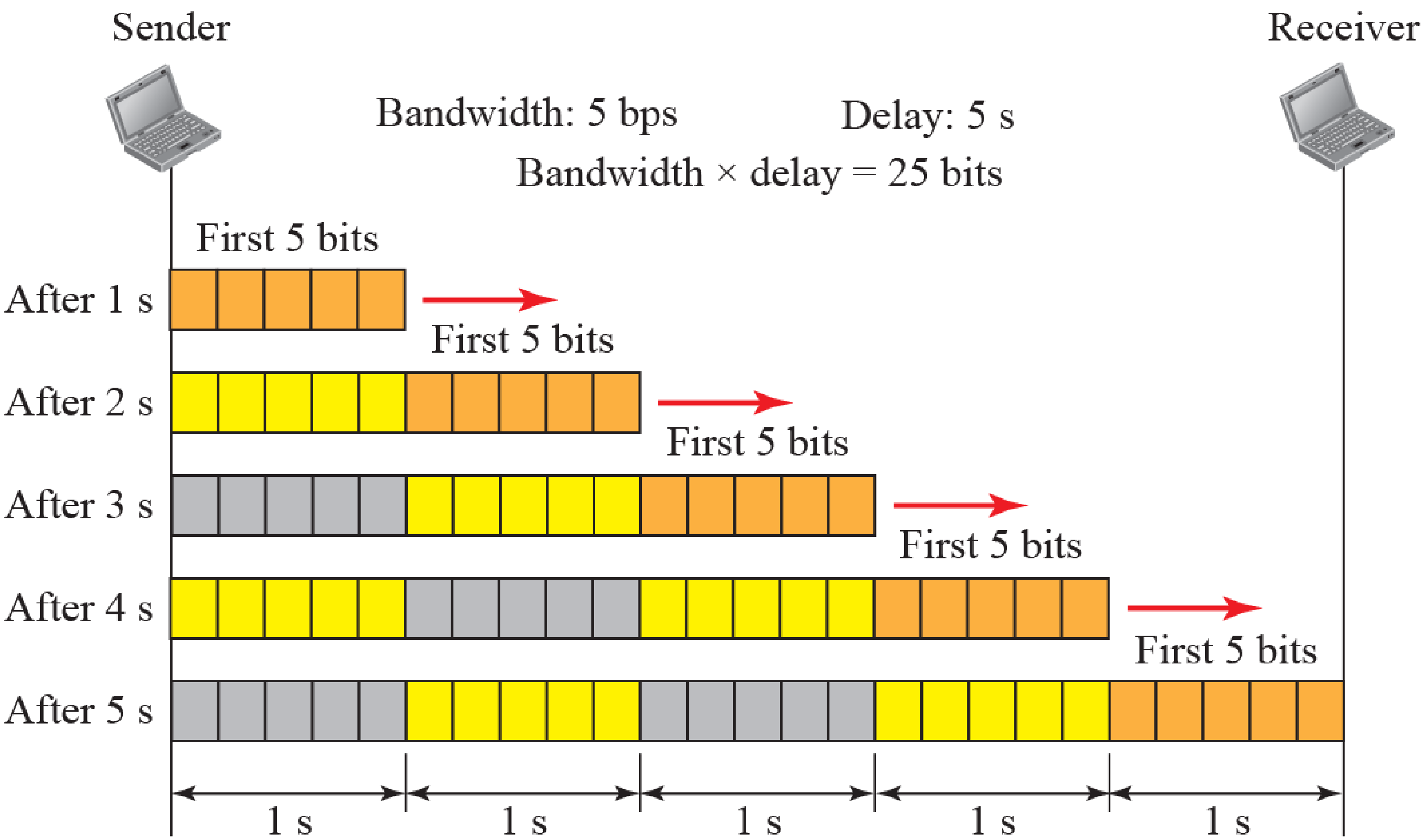


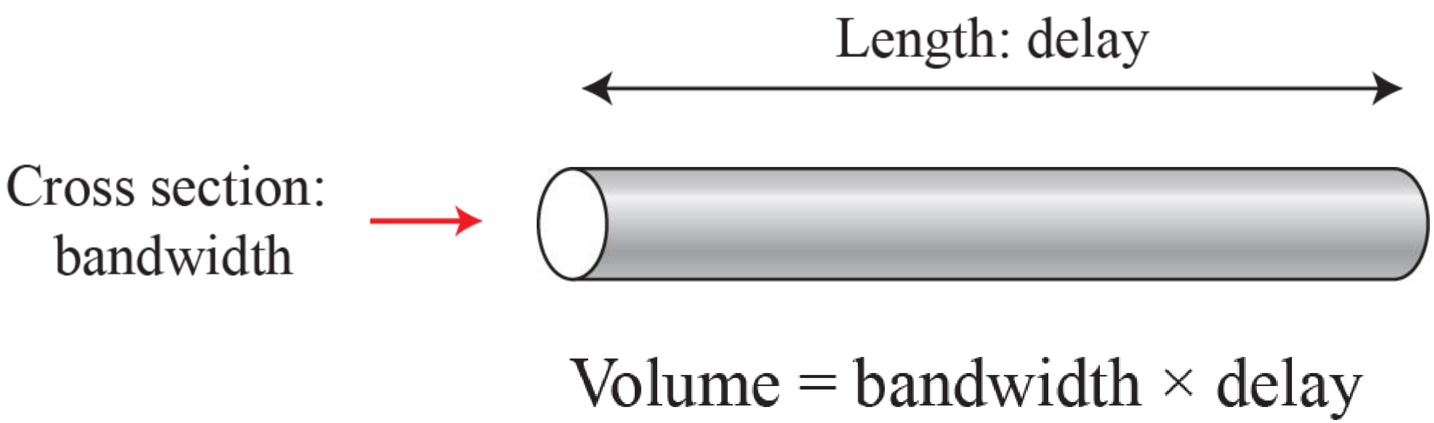
Figure 3.33: Filling the pipe with bits for Case 2



Example 3.48

We can think about the link between two points as a pipe. The cross section of the pipe represents the bandwidth, and the length of the pipe represents the delay. We can say the volume of the pipe defines the bandwidth-delay product, as shown in Figure 3.34.

Figure 3.34: ***Concept of bandwidth-delay product***



Jitter

- Another performance issue that is related to delay is jitter.
- We can roughly say that jitter is a problem if different packets of data encounter different delays and the application using the data at the receiver site is time-sensitive (audio and video data, for example).
- If the delay for the first packet is 20 ms, for the second is 45 ms, and for the third is 40 ms, then the real-time application that uses the packets endures jitter.