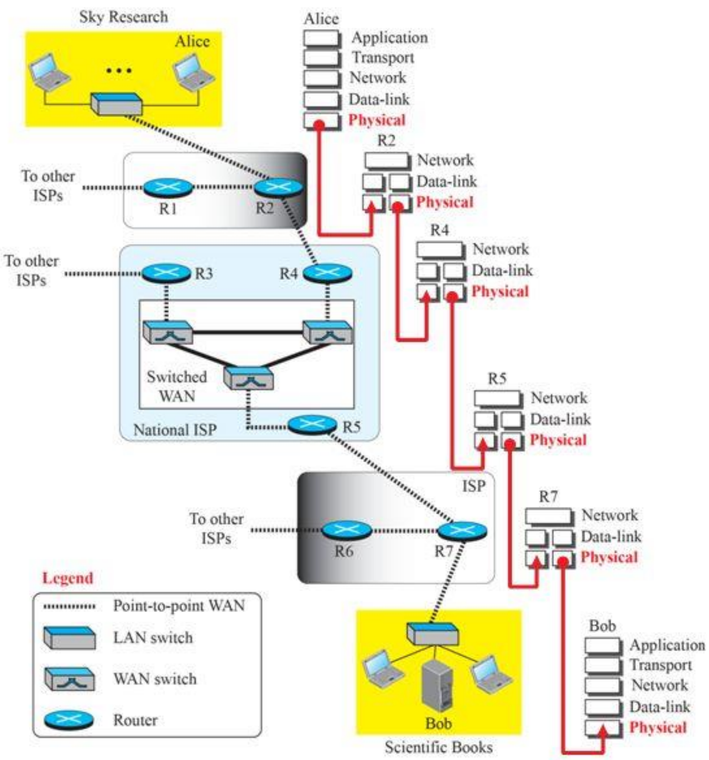
3.1 DATA AND SIGNALS

Figure 3.1: Communication at the physical layer



3.6

Communication at application, transport, network, or data- link is logical; communication at the physical layer is physical. we have shown only ; host- to- router, router-to- router, and router- to- host, but the switches are also involved in the physical communication.

Although Alice and Bob need to exchange data, communication at the physical

layer means exchanging signals. Data need to be transmitted and received, but the media have to change data to signals. Both data and the signals that represent them can be either analog or digital in form.

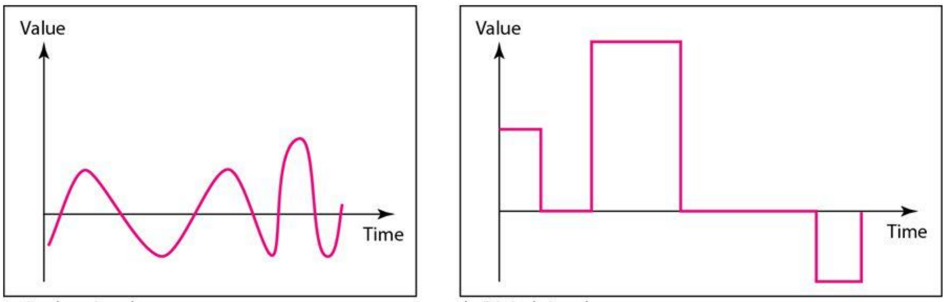
3.1.1 Analog and Digital Data

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 Os and ls.

3.1.2 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. The curve representing the analog signal passes through an infinite number of points. The vertical lines of the digital signal, however, demonstrate the sudden jump that the signal makes from value to value.

Figure 3.1 *Comparison of analog and digital signals*



a. Analog signal b. Digital signal

3.3

3.1.3 Periodic and Nonperiodic

Both analog and digital signals can take one of two forms: periodic or nonperiodic sometimes referred to as aperiodic; 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.

3.2 PERIODIC ANALOG SIGNALS

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.

3.2.1 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.

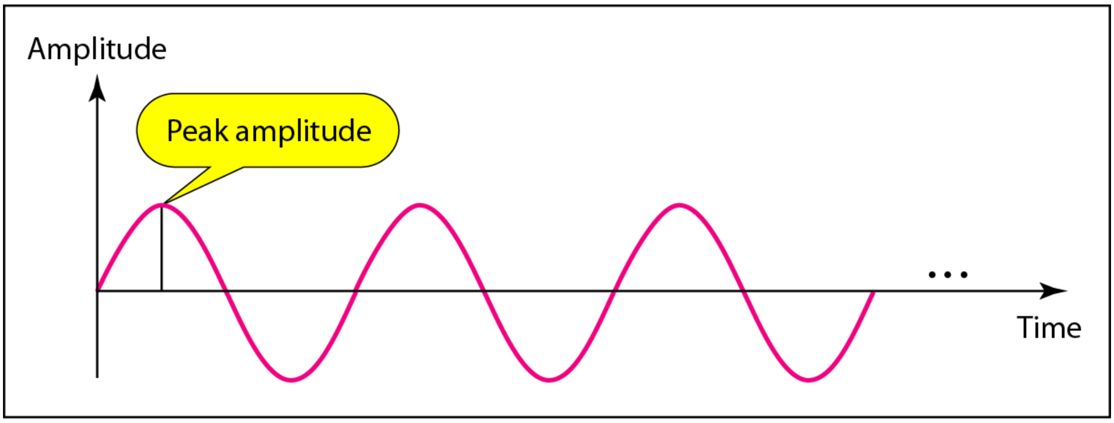
A sine wave can be represented by three parameters: the peak amplitude, the frequency, and the phase. These three parameters fully describe a sine wave.

Figure 3.2 *A sine wave*

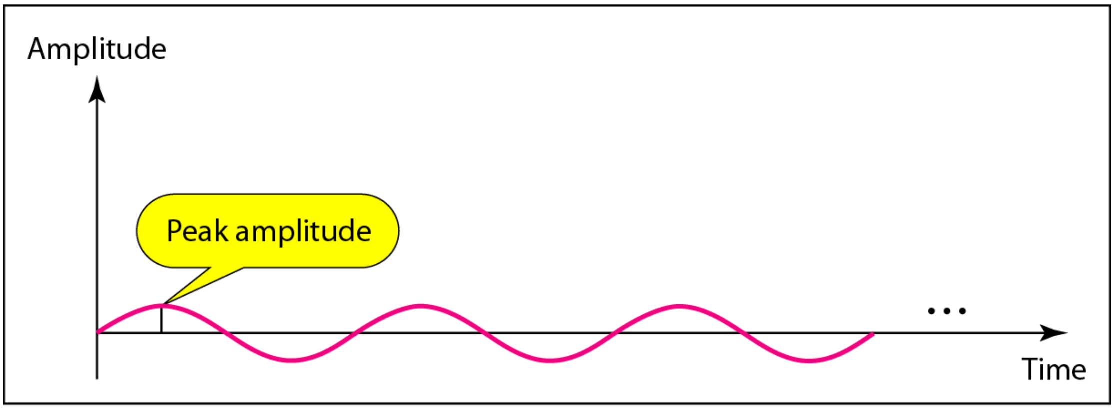
3.9

Peak Amplitude

The peak amplitude of a signal is the absolute value of its highest intensity, proportional to the energy it carries. For electric signals, peak amplitude is normally measured in volts.



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 --------- V. However, it is common knowledge that the voltage of the power in Iraq homes is 220 V. 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 is equal to $2^{1/2}$ * rms value.

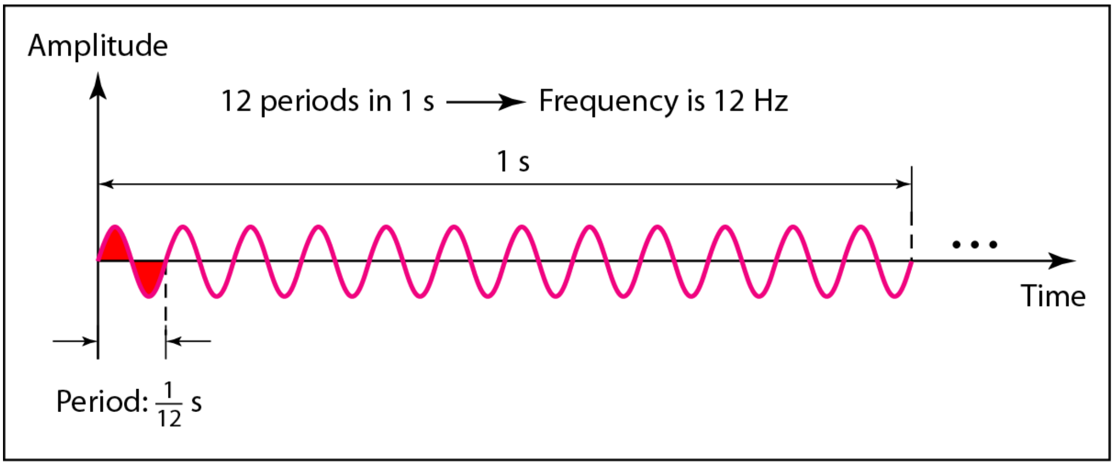
Example 3.2

The voltage of a battery is a constant. For example, the peak value of an AA battery is normally 1. 5 V.

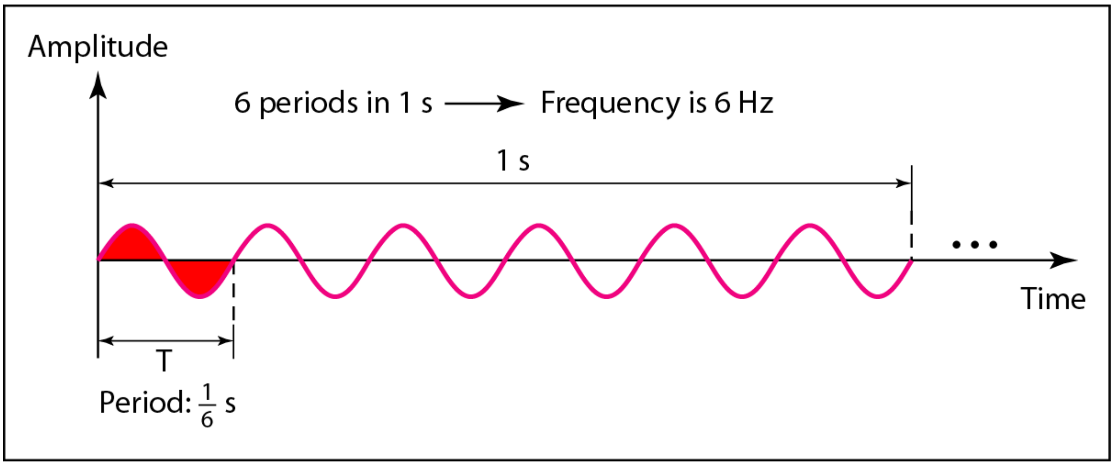
Period and Frequency

Period refers to the amount of time, in seconds, a signal needs to complete 1 cycle. Frequency refers to the number of periods in 1 s. 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, as the following formulas show.

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



a. A signal with a frequency of 12 Hz



b. A signal with a frequency of 6 Hz

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 3.3

The power we use at home has a frequency of 60 Hz. The period of this sine wave can be determined as follows:

This means that the period of the power for our lights at home is 0.0116 s, or 16. 6 ms.

Example 3.4

Express a period of 100 ms in microseconds.

Solution :

Example 3. 5

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

Solution:

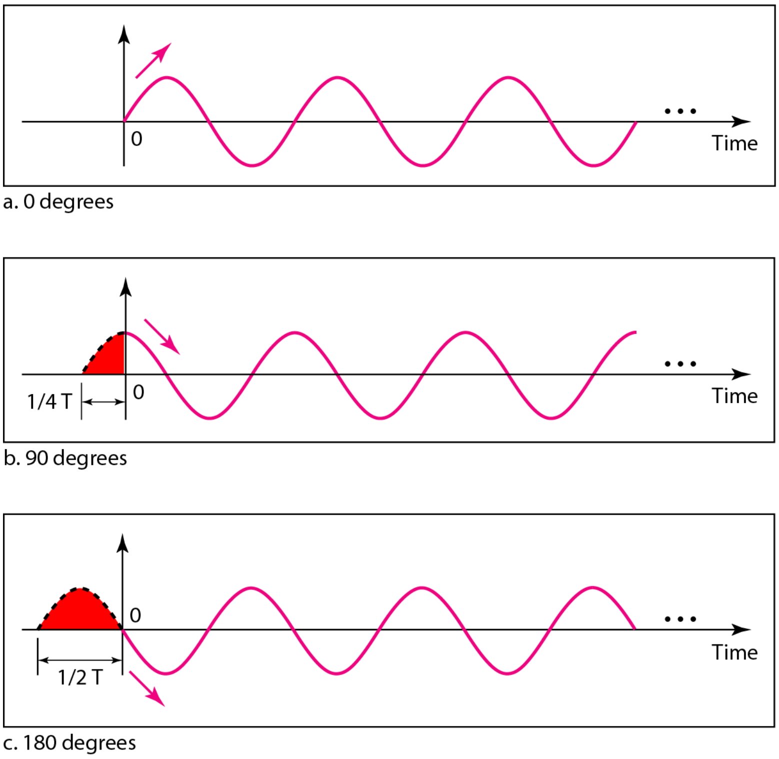
-Frequency is the rate of change with respect to time. Change in a short span of time means high frequency. Change over a long span of time means low frequency.

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

3.2.2 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° is 2pi rad; 1° is 2pi/360 rad, and 1 rad is 360/(2pi)]. A phase shift of 360° 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° corresponds to a shift of one-quarter of a period.



Example 3.6

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

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

3.23 Wavelength

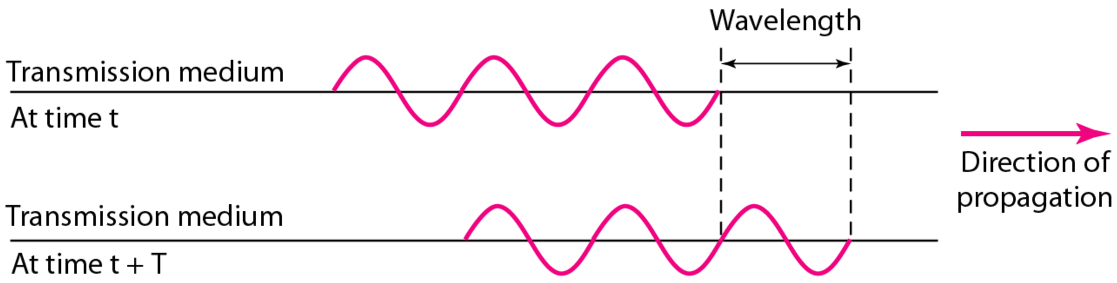
Wavelength is another characteristic of a signal traveling through a transmission medium. Wavelength binds the period or the frequency of a simple sine wave to the propagation speed of the medium

While the frequency of a signal is independent of the medium, the wavelength depends on both the frequency and the medium. Wavelength is a property of any type of signal. In data communications, we often use wavelength to describe the transmission of light in an optical fiber. The wavelength is the distance a simple signal can travel in one period.

Wavelength can be calculated if one is given the propagation speed (the speed of light) and the period of the signal.

Wavelength = (propagation speed) * period= propagation speed / frequency

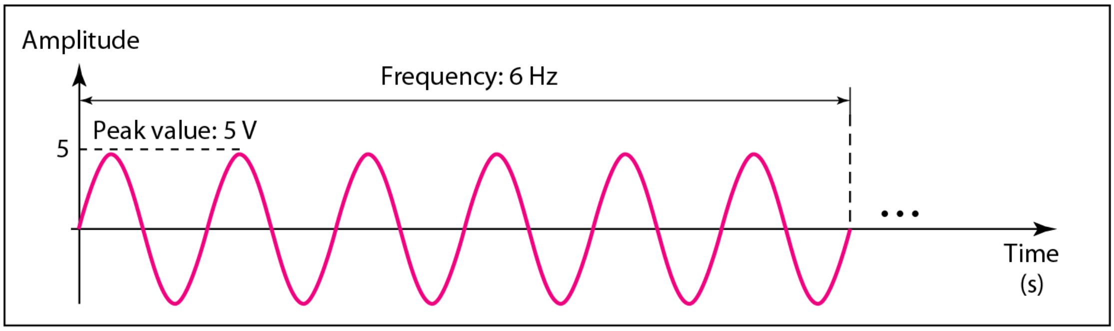
The propagation speed of electromagnetic signals depends on the medium and on the frequency of the signal. For example, in a vacuum, light is propagated with a speed of 3 x 10^8 m/ s. That speed is lower in air and even lower in cable. The wavelength is normally measured in micrometers (microns) instead of meters. For example, the wavelength of red light (frequency= 4×10^{14}) in air is



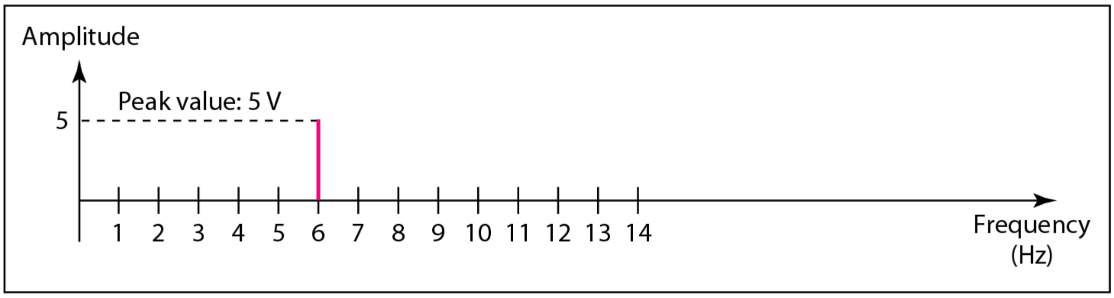
3. 2.4 Time and Frequency Domains

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.

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.



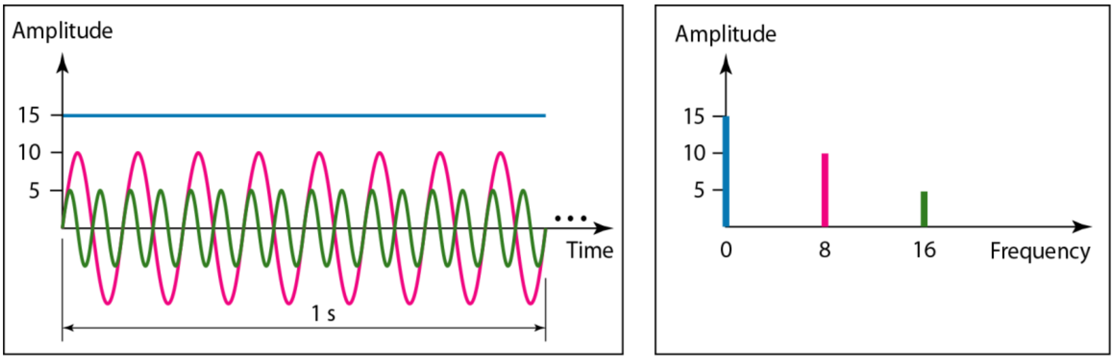
a. A sine wave in the time domain (peak value: 5 V, frequency: 6 Hz)



b. The same sine wave in the frequency domain (peak value: 5 V, frequency: 6 Hz)

The advantage of the frequency domain is that we can immediately see the values of the frequency and peak amplitude. A complete sine wave is represented by one spike. The position of the spike shows the frequency; its height shows the peak amplitude.

Example 3. 7



a. Time-domain representation of three sine waves with frequencies 0, 8, and 16

b. Frequency-domain representation of the same three signals

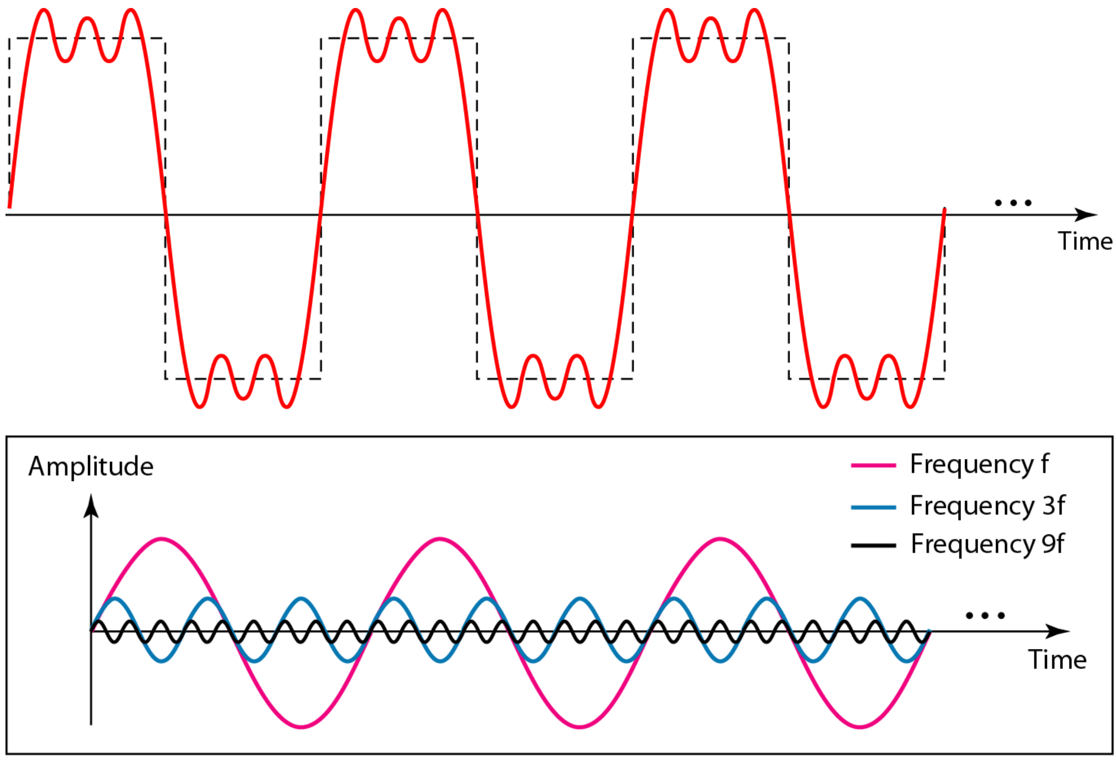
3.2.5 Composite Signals

Simple sine waves have many applications in daily life. 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. As another example, we can use a single sine wave to send an alarm to a security center when a burglar opens a door or window in the house. In the first case, the sine wave is carrying energy; in the second, the sine wave is a signal of danger. 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. Fourier showed that any composite signal is actually a combination of simple sine waves with different frequencies, amplitudes, and phases.

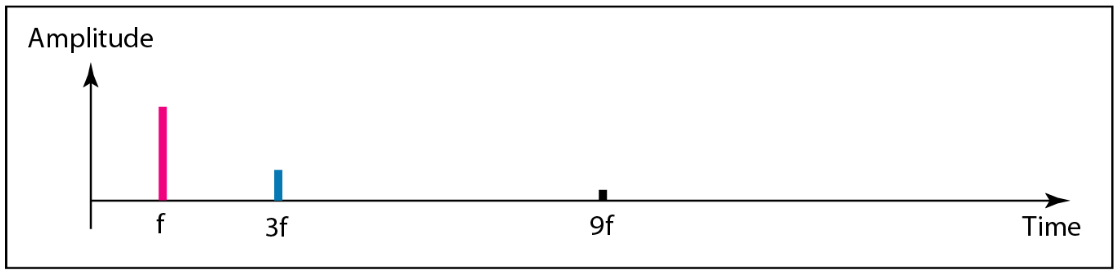
A composite signal can be periodic or nonperiodic. A periodic composite signal can be decomposed into a series of simple sine waves with discrete frequencies that have integer values (1, 2, 3, and so on). A nonperiodic composite signal can be decomposed into a combination of an infinite number of simple sine waves with continuous frequencies, frequencies that have real values.

Example 3.8

Figure 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. The amplitude of the sine wave with frequency *f* is almost the same as the peak amplitude of the composite signal. The amplitude of the sine wave with frequency *3f* is one third of that of the first, and the amplitude of the sine wave with frequency *9f* is one- ninth of the first. The frequency of the sine wave with frequency *f* is the same as the frequency of the composite signal; it is called the fundamental frequency, or first harmonic. The sine wave with frequency *3f* has a frequency of 3 times the fundamental frequency; it is called the third harmonic. The third sine wave with frequency *9f* has a frequency of 9 times the fundamental frequency; it is called the ninth harmonic. Note that the frequency decomposition of the signal is discrete; it has frequencies *f*,*3f*, and *9f*. Because *f* is an integral number, *3f* and *9f* are also integral numbers. There are no frequencies such as *1.2f or 2.6f*. The frequency domain of a periodic composite signal is always made of discrete spikes.



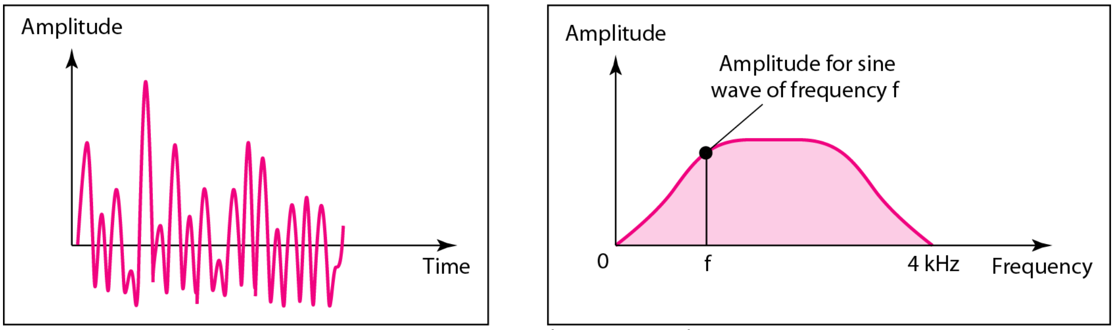
a. Time-domain decomposition of a composite signal



b. Frequency-domain decomposition of the composite signal

Example 3.9

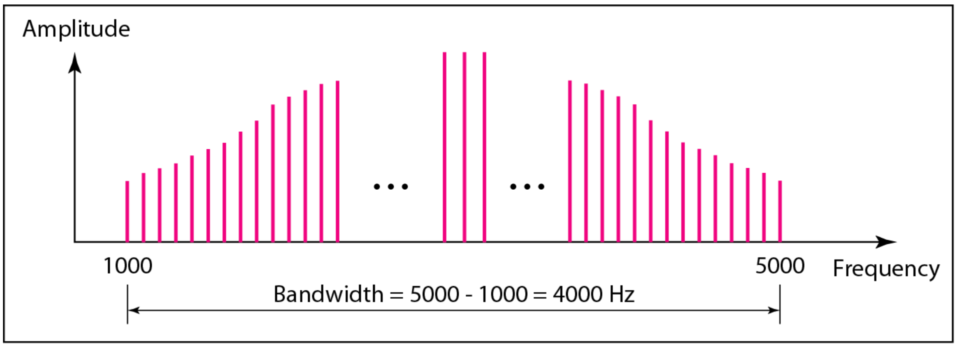
Figure shows a nonperiodic composite signal. It can be the signal created by a microphone or a telephone set when a word or two is pronounced. In a time-domain representation of this composite signal, there are an infinite number of simple sine frequencies. Although the number of frequencies in a human voice is infinite, the range is limited. A normal human being can create a continuous range of frequencies between 0 and 4 kHz. Note that the frequency decomposition of the signal yields a continuous curve. There are an infinite number of frequencies between 0.0 and 4000.0 (real values). To find the amplitude related to frequency *f*, we draw a vertical line at *f* to intersect the envelope curve. The height of the vertical line is the amplitude of the corresponding frequency.



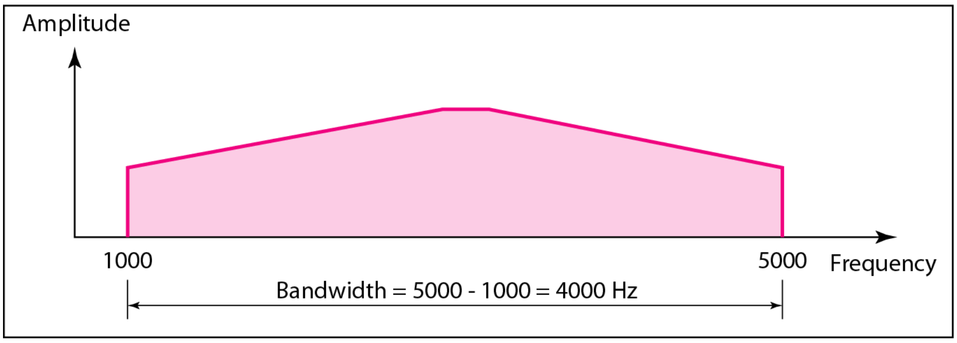
a. Time domain b. Frequency domain

3.2.6 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 figure depicts two composite signals, one periodic and the other nonperiodic. The bandwidth of the periodic signal contains all integer frequencies between 1000 and 5000(1000, 1001, 1002,..). The bandwidth of the nonperiodic signals has the same range, but the frequencies are continuous



a. Bandwidth of a periodic signal



b. Bandwidth of a nonperiodic signal

Example 3.10

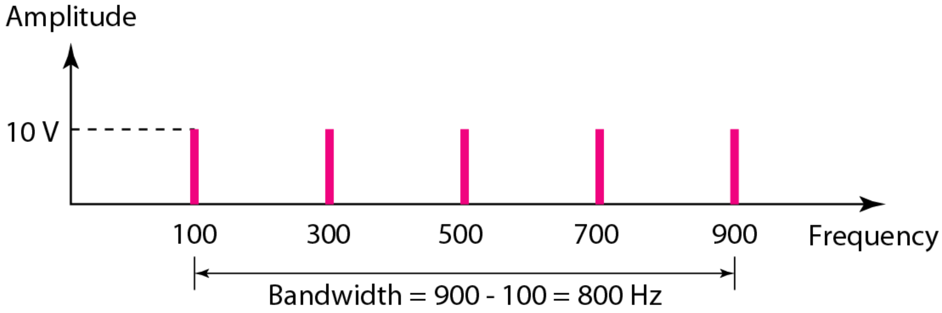
If a periodic signal is decomposed into five sine waves with frequencies of 100, 300, 500, 700, and 900 Hz, what is its bandwidth? Draw the spectrum, assuming all components have a maximum amplitude of 10 V.

Solution

Let *fh* be the highest frequency, *fl* the lowest frequency, and *B* the bandwidth. Then

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

The spectrum has only five spikes, at 100, 300, 500, 700, and 900 Hz (see the Figure)



Example 3.11

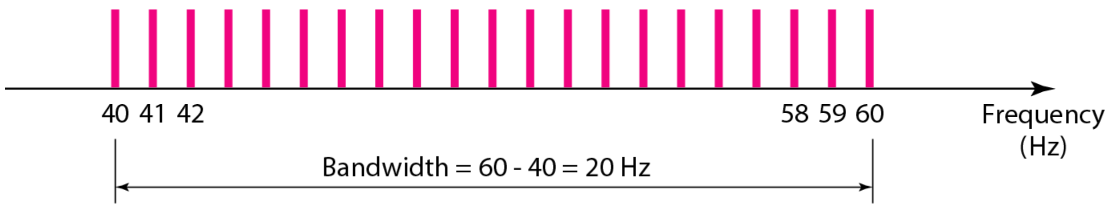
A periodic signal has a bandwidth of 20 Hz. The highest frequency is 60 Hz. What is the lowest frequency? Draw the spectrum if the signal contains all frequencies of the same amplitude.

Solution

Let *fh* be the highest frequency, *fl* the lowest frequency, and *B* the bandwidth. Then

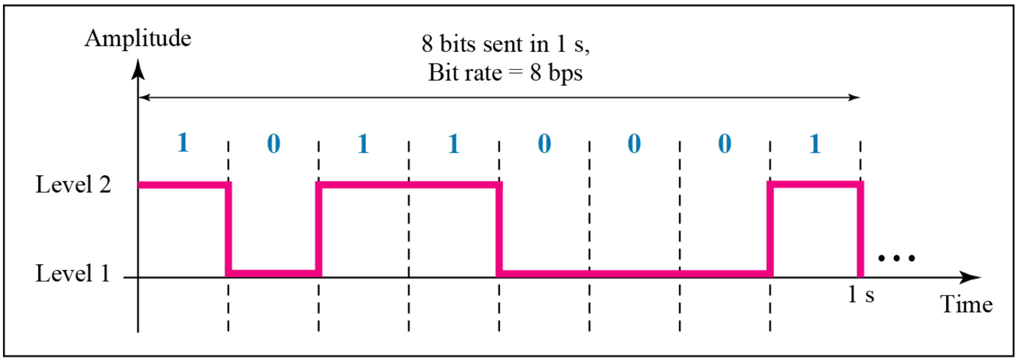
$$\boxed{B = f_h - f_l \quad \Rightarrow \quad 20 = 60 - f_l \quad \Rightarrow \quad f_l = 60 - 20 = 40 \text{ Hz}}$$

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

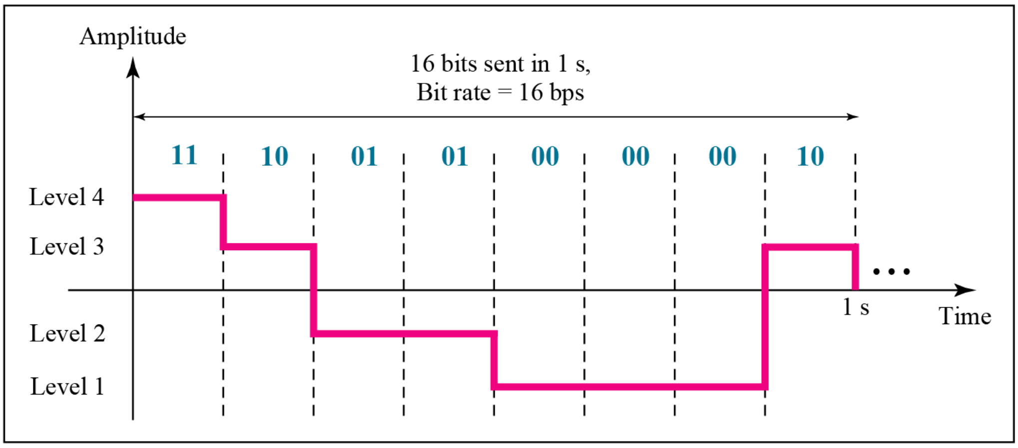


3.3 DIGITAL SIGNALS

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.



a. A digital signal with two levels



b. A digital signal with four levels

Example 3.16

A digital signal has eight levels. How many bits are needed per level? We calculate the number of bits from the formula

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

3.3.1 Bit Rate

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

Example 3.18

Assume we need to download text documents at the rate of 100 pages per minute. 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

$$\boxed{100 \times 24 \times 80 \times 8 = 1{,}636{,}000 \text{ bps} = 1.636 \text{ Mbps}}$$

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

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

3.3.2 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.

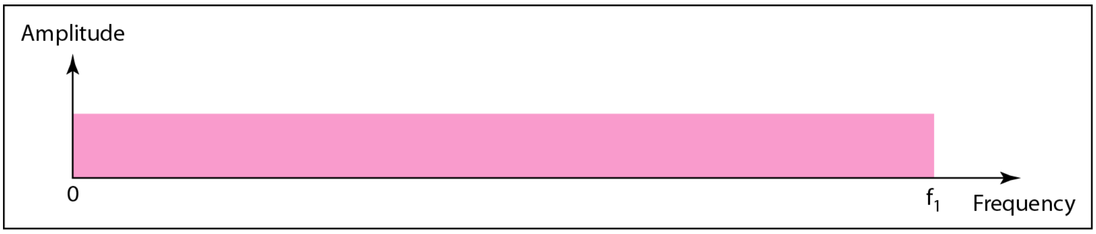
Bit length =propagation speed * bit duration

3.3.4 Transmission of Digital Signals

A digital signal, periodic or nonperiodic, is a composite analog signal with frequencies between zero and infinity. Let us consider the case of a nonperiodic digital signal, similar to the ones we encounter in data communications. 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. 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. The figure shows two low-pass channels: one with a narrow bandwidth and the other with a wide bandwidth.

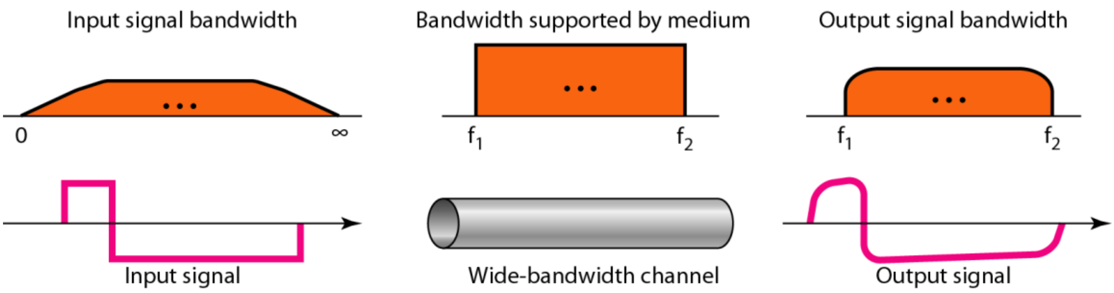


a. Low-pass channel, wide bandwidth

b. Low-pass channel, narrow bandwidth

Case 1: Low-Pass Channel with Wide Bandwidth

If we want to preserve the exact form of a nonperiodic digital signal with vertical segments vertical and horizontal segments horizontal, we need to send the entire spectrum, the continuous range of frequencies between zero and infinity. This is possible if we have a dedicated medium with an infinite bandwidth between the sender and receiver that preserves the exact amplitude of each component of the composite signal. Although this may be possible inside a computer (e.g., between CPU and memory), it is not possible between two devices. Fortunately, the amplitudes of the frequencies at the border of the bandwidth are so small that they can be ignored. This means that if we have a medium, such as a coaxial cable or fiber optic, with a very wide bandwidth, two stations can communicate by using digital signals with very good accuracy.



Example 3.21

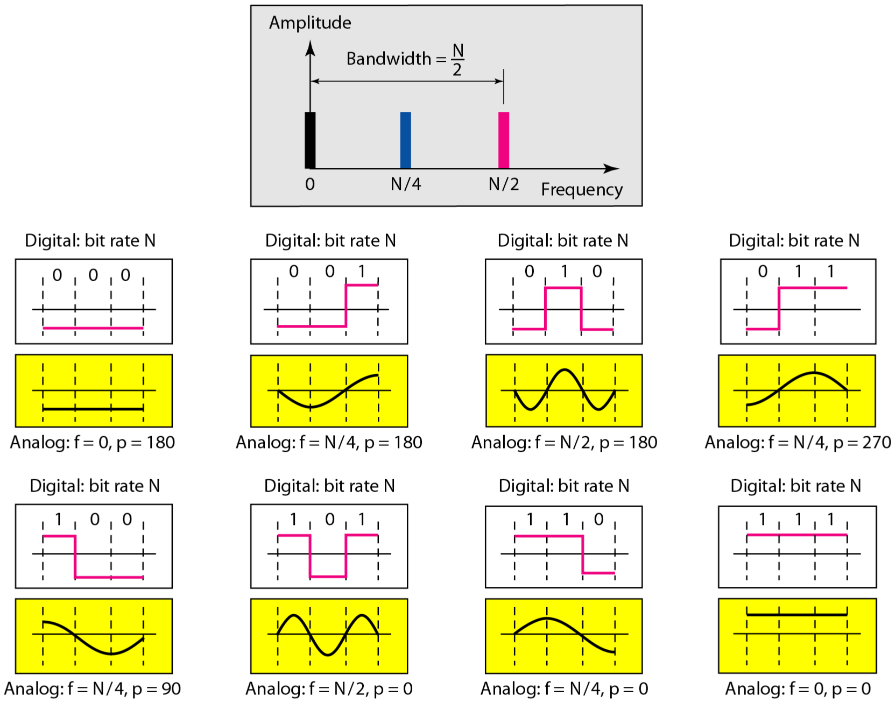
An example of a dedicated channel where the entire bandwidth of the medium is used as one single channel is a LAN. Almost every wired LAN today uses a dedicated channel for two stations communicating with each other. In a bus topology LAN with multipoint connections, only two stations can communicate 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.

Case 2: Low-Pass Channel with Limited Bandwidth

In a low-pass channel with limited bandwidth, we approximate the digital signal with an analog signal. The level of approximation depends on the bandwidth available.

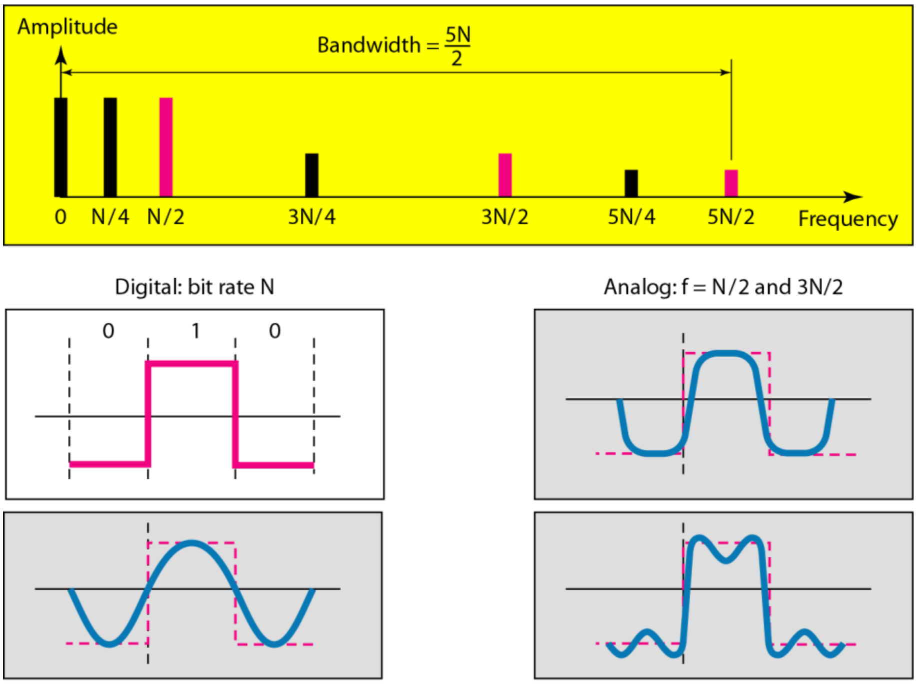
Rough Approximation

Let us assume that we have a digital signal of bit rate *N*. If we want to send analog signals to roughly simulate this signal, we need to consider the worst case, a maximum number of changes in the digital signal. This happens when the signal carries the sequence 01010101 ... or the sequence 10101010….. . To simulate these two cases, we need an analog signal of frequency *f* = *N/2*. Let 1 be the positive peak value and 0 be the negative peak value. We send 2 bits in each cycle; the frequency of the analog signal is one-half of the bit rate, or *N/2*. However, just this one frequency cannot make all patterns; we need more components. The maximum frequency is *N/2*. As an example of this concept, let us see how a digital signal with a 3-bit pattern can be simulated by using analog signals. The Figure shows the idea. The two similar cases (000 and 111) are simulated with a signal with frequency f =0 and a phase of 180° for 000 and a phase of 0° for 111. The two worst cases (010 and 101) are simulated with an analog signal with frequency =*N/2* and phases of 180° and 0°. The other four cases can only be simulated with an analog signal with f = N/4 and phases of 180°, 270°, 90°, and 0°. In other words, we need a channel that can handle frequencies 0, N/4, and N/2. This rough approximation is referred to as using the first harmonic (N/2) frequency. The required bandwidth is



Better Approximation

To make the shape of the analog signal look more like that of a digital signal, we need to add more harmonics of the frequencies. We need to increase the bandwidth. We can increase the bandwidth to 3N/2, 5N/2, 7N/2, and so on. The Figure shows the effect of this increase for one of the worst cases, the pattern 010.



Note that we have shown only the highest frequency for each hannonic. We use the first, third, and fifth hannonics. The required bandwidth is now 5N/2, the difference between the lowest frequency 0 and the highest frequency 5N/2. As we emphasized before, we need to remember that the required bandwidth is proportional to the bit rate.

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 3.23

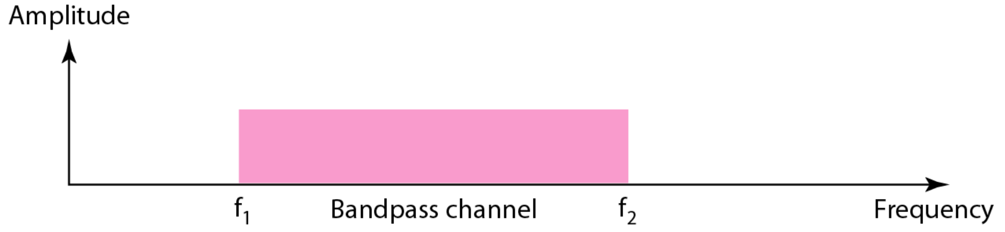
We have a low-pass channel with bandwidth 100 kHz. What is the maximum bit rate of this channel?

Solution

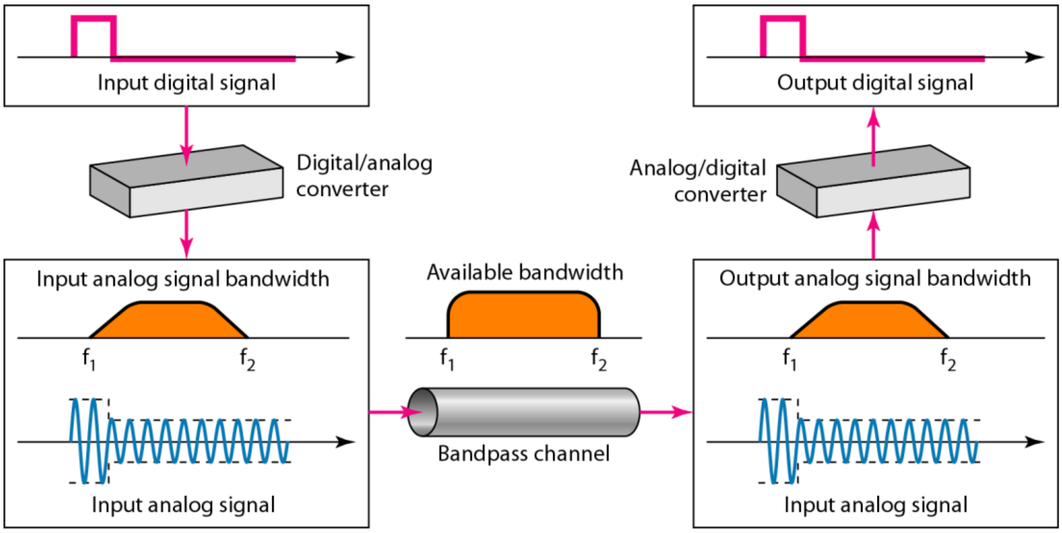
The maximum bit rate can be achieved if we use the first harmonic. The bit rate is 2 times the available bandwidth, or 200 kbps.

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. The Figure shows a bandpass channel.



The next Figure shows the modulation of a digital signal. A digital signal is converted to a composite analog signal. We have used a single-frequency analog signal (called a carrier); the amplitude of the carrier has been changed to look like the digital signal. The result, however, is not a single-frequency signal; it is a composite signal. At the receiver, the received analog signal is converted to digital, and the result is a replica of what has been sent.

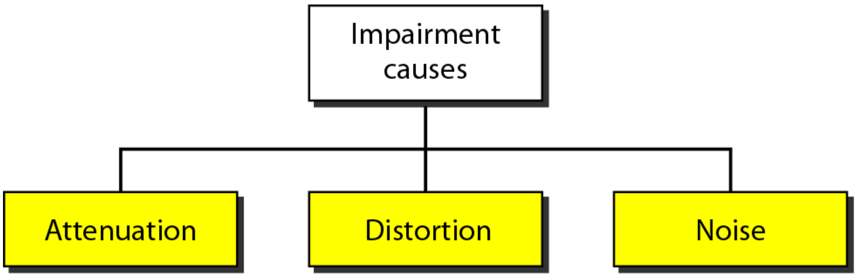


Example 3.24

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, installed many years ago, are designed to carry voice (analog signal) with a limited bandwidth (frequencies between 0 and 4 kHz). Although this channel can be used as a low-pass channel, it is normally considered a bandpass channel. One reason is that the bandwidth is so narrow (4 kHz) that if we treat the channel as low-pass and use it for baseband transmission, the maximum bit rate can be only 8 kbps. The solution is to consider the channel a bandpass channel, 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 (modulator/demodulator)

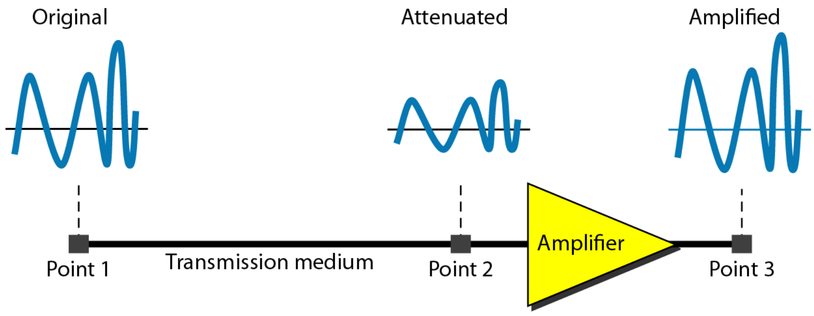
3.4 TRANSMISSION IMPAIRMENT

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 (see Figure)



3.4.1 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. The Figure shows the effect of attenuation and amplification.



Decibel

To show that a signal has lost or gained strength, engineers use the unit of the decibel. The decibel (dB) measures the relative strengths of two signals or one signal at two different points. Note that the decibel is negative if a signal is attenuated and positive if a signal is amplified.

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

Example 3.26

Suppose a signal travels through a transmission medium and its power is reduced to one-half. This means that *P2* is *(1/2)P1*. In this case, the attenuation (loss of power) can be calculated as

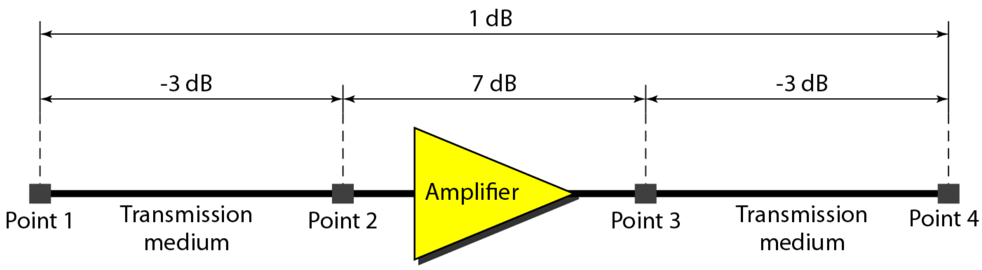
$$10 \log_{10} \frac{P_2}{P_1} = 10 \log_{10} \frac{0.5P_1}{P_1} = 10 \log_{10} 0.5 = 10(-0.3) = -3 \text{ dB}$$

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

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 the Figure a signal travels from point 1 to point 4. In this case, the decibel value can be calculated as

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



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 $dB_m = 10 \log10 P_m$, where P_m is the power in milliwatts. Calculate the power of a signal with $dB_m = -30$.

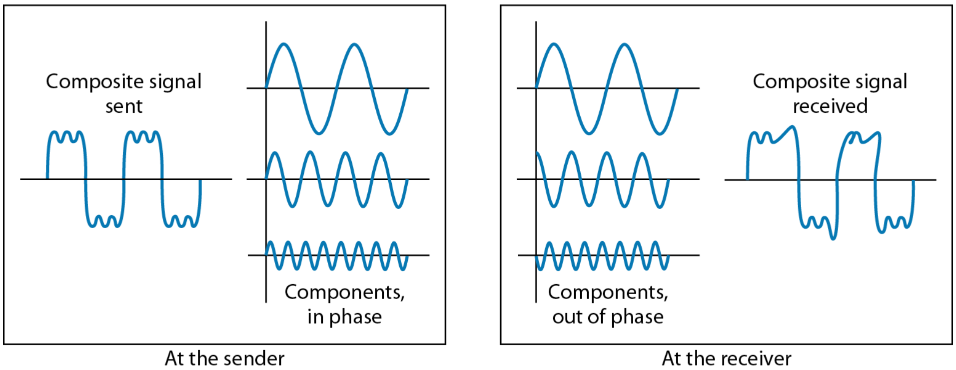
Solution

We can calculate the power in the signal as

$$\text{dB}_m = 10 \log_{10} P_m = -30$$
$$\log_{10} P_m = -3 \qquad P_m = 10^{-3} \text{ mW}$$

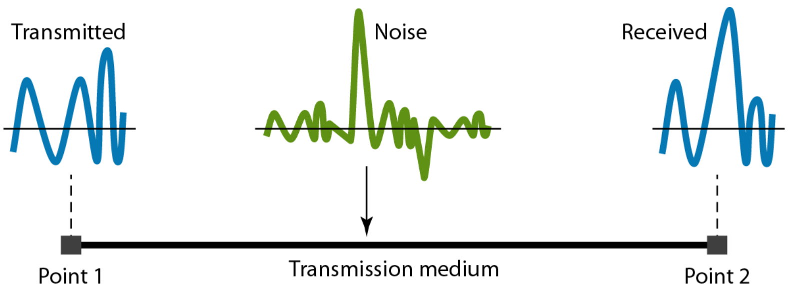
3.4.2 Distortion

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. In other words, signal components at the receiver have phases different from what they had at the sender. The shape of the composite signal is therefore not the same. The Figure shows the effect of distortion on a composite signal.



3.4.3 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 and appliances. These devices act as a sending antenna, and the transmission medium acts as the receiving antenna. Crosstalk is the effect of one wire on the other. One wire acts as a sending antenna and the other as the receiving antenna. Impulse noise is a spike (a signal with high energy in a very short time) that comes from power lines, lightning, and so on. The Figure shows the effect of noise on a signal.



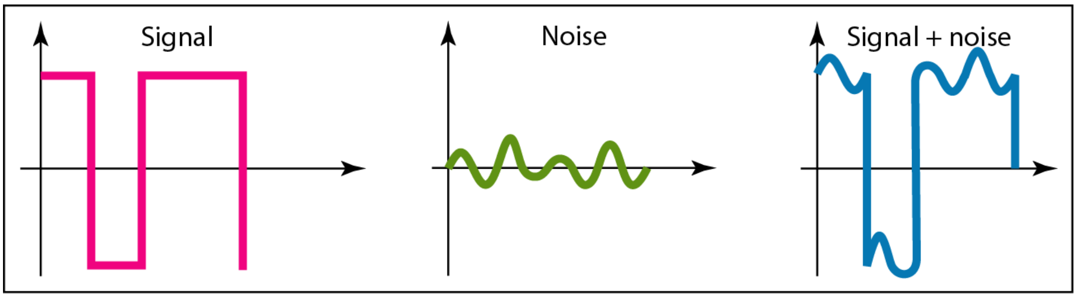
Signal-to-Noise Ratio (SNR)

As we will see later, 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

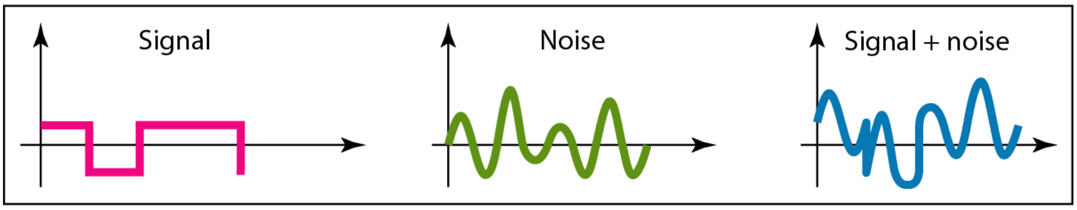
SNR=average signal power / average noise power

We need to consider the average signal power and the average noise power because these may change with time. The Figure shows the idea of SNR. 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

$$\mathbf{SNR_{dB} = 10 \log10 \ SNR}$$



a. Large SNR



b. Small 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:

$$SNR = \frac{10{,}000\ \mu W}{1\ mW} = 10{,}000$$
$$SNR_{dB} = 10 \log_{10} 10{,}000 = 10 \log_{10} 10^4 = 40$$

3.5 DATA RATE LIMITS

A very important consideration in data communications is how fast we can send data, in bits per second. Over a channel. Data rate depends on three factors:

1. The bandwidth available

2. The level of the signals we use

3. The quality of the channel (the level of noise)

Two theoretical formulas were developed to calculate the data rate: one by Nyquist for a noiseless channel. Another by Shannon for a noisy channel.

3.5.1 Noiseless Channel: Nyquist Bit Rate

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

$$\textbf{Bit Rate} = 2 * \textbf{bandwidth} * \log_2 L$$

bandwidth is the bandwidth of the channel, L is the number of signal levels used to represent data, and Bit Rate 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 signa11eve1s. Although the idea is theoretically correct, practically there is a limit. When we increase the number of signal leve1s, we impose a burden on the receiver. If the number of levels in a signal is just 2, the receiver can easily distinguish between a 0 and a 1. If the level of a signal is 64, the receiver must be very sophisticated to distinguish between 64 different levels. In other words, 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

$$\textbf{BitRate} = 2 * 3000 * \log_2 2 = 6000 \textbf{ bps}$$

3.5.2 Noisy Channel: Shannon Capacity

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

$$\textbf{Capacity} = \textbf{bandwidth} * \log2 (1 + \textbf{SNR})$$

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

$$C = B \log_2 (1 + \text{SNR}) = 3000 \log_2 (1 + 3162) = 3000 \log_2 3163$$
$$= 3000 \times 11.62 = 34{,}860 \text{ bps}$$

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.

3.5.3 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?

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

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