Computer Networks and Internets, 5e
Chapter 6
Information Sources and Signals

Modified from the lecture slides of Lami Kaya
(LKaya@ieee.org) for use CECS 474, Fall 2008.

1

Introduction

- This chapter
    - Begins exploration of data communications in more detail
    - Examines the topics of information sources
    - Studies the characteristics of the signals that carry information

- Successive chapters
    - continue the exploration of data communications by explaining additional aspects

2

Information Sources

- A communication system accepts input from one or more sources and delivers it to a specified destination
- For the Internet, the source and destination of information are a pair of application programs
  - that generate and consume data
- Data communications theory concentrates on low-level communications systems
  - it applies to arbitrary sources of information
  - conventional computer peripherals such as keyboards and mice
  - information sources can include microphones, sensors, and measuring devices, such as thermometers and scales
  - destinations can include audio output devices such as earphones and loud speakers as well as devices such as LEDs that emit light

3

Analog and Digital Signals

- Data communications deals with two types of information:
  - analog
  - digital
- An analog signal is characterized by a continuous mathematical function
  - when the input changes from one value to the next, it does so by moving through all possible intermediate values
- A digital signal has a fixed set of valid levels
  - each change consists of an instantaneous move from one valid level to another

4

6.3 Analog and Digital Signals
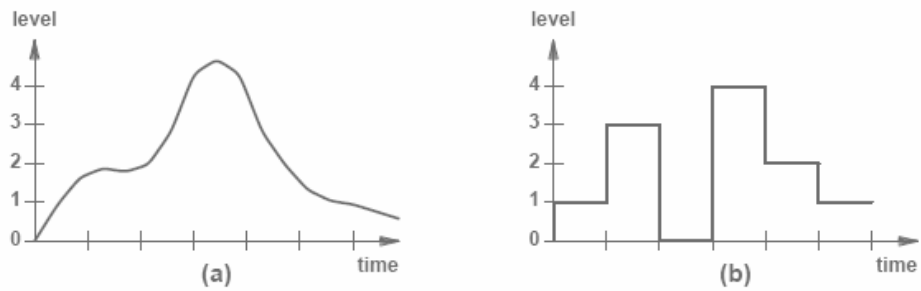
Figure 6.1 Illustration of (a) an analog signal, and (b) a digital signal.

5

Periodic and Aperiodic Signals

- Signals are broadly classified as
  - periodic
  - aperiodic (sometimes called nonperiodic)

  The classification depends on whether they repeat

- For example:
  - Figure 6.1a is aperiodic over the time interval shown because the signal does not repeat
  - Figure 6.2 illustrates a signal that is periodic (i.e., repeating) - periodic signal repeats

6

Sine Waves and Signal Characteristics

- Much of the analysis in data communications involves the use of sinusoidal trigonometric functions
  - especially sine, which is usually abbreviated sin
- Sine waves are especially important in information sources
  - because natural phenomena produce sine waves
  - when a microphone picks up an audible tone, the output is a sine
  - electromagnetic radiation can be represented as a sine wave
- We are interested in sine waves that correspond to a signal that oscillates in time

Sine Waves and Signal Characteristics
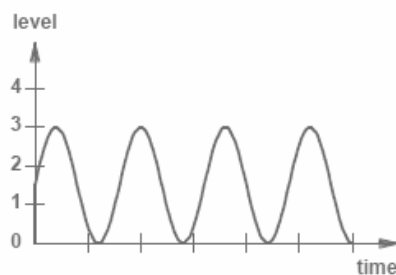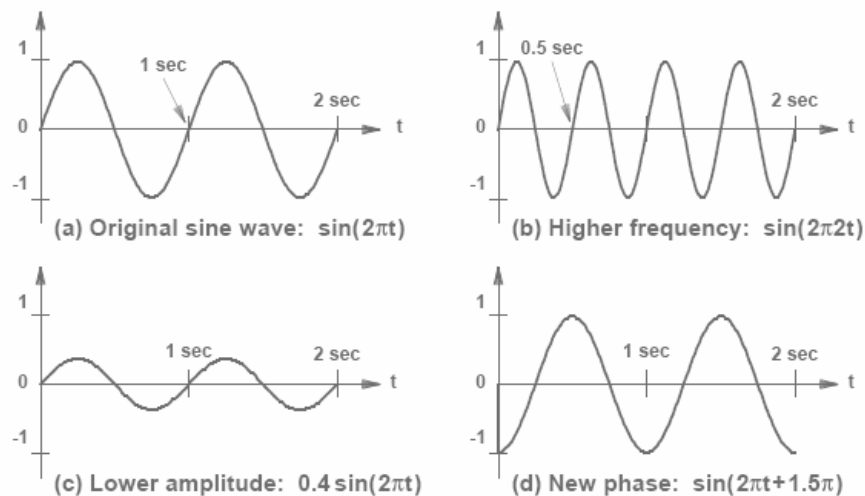

Figure 6.2 A periodic signal repeats.

Sine Waves and Signal Characteristics

- Important characteristics of signals that relate to sine waves:
- Frequency:
  - the number of oscillations per unit time (usually seconds)
- Amplitude:
  - the difference between the maximum and minimum signal heights
- Phase:
  - how far the start of the sine wave is shifted from a reference time
- Wavelength:
  - the length of a cycle as a signal propagates across a medium
  - is determined by the speed with which a signal propagates
- These characteristics can be expressed mathematically
  - Figure 6.3 illustrates the frequency, amplitude, and phase characteristics

9

Sine Waves and Signal Characteristics

(a) Original sine wave: $\sin(2\pi t)$

(b) Higher frequency: $\sin(2\pi 2t)$

(c) Lower amplitude: $0.4\sin(2\pi t)$

(d) New phase: $\sin(2\pi t + 1.5\pi)$

Figure 6.3 Illustration of frequency, amplitude, and phase characteristics.

10

Sine Waves and Signal Characteristics

- The frequency can be calculated as the inverse of the time required for one cycle, which is known as the period
- The example sine wave in Figure 6.3a has
  - a period $T = 1$ seconds
  - and a frequency of $1 / T$ or 1 Hertz
- The example in Figure 6.3b has
  - a period of $T = 0.5$ seconds
  - its frequency is 2 Hertz
- Both examples are considered extremely low frequencies
- Typical communication systems use high frequencies
  - often measured in millions of cycles per second
- To clarify high frequencies, engineers express time in fractions of a second or express frequency in units such as megahertz (MHz)

11

Sine Waves and Signal Characteristics

Time Unit	Value	Frequency Unit	Value
Seconds (s)	10^0 seconds	Hertz (Hz)	10^0 Hz
Milliseconds (ms)	10^3 seconds	Kilohertz (KHz)	10^3 Hz
Microseconds (μs)	10^6 seconds	Megahertz (MHz)	10^6 Hz
Nanoseconds (ns)	10^9 seconds	Gigahertz (GHz)	10^9 Hz
Picoseconds (ps)	10^{12} seconds	Terahertz (THz)	10^{12} Hz

Figure 6.4 Prefixes and abbreviations for units of time and frequency.

Figure 6.4 lists time scales and common prefixes used with frequency

12

Composite Signals

- Signals like the ones illustrated in Figure 6.3 are classified as simple
    - because they consist of a single sine wave that cannot be decomposed further
- Most signals are classified as composite
    - the signal can be decomposed into a set of simple sine waves
- Figure 6.5 illustrates a composite signal
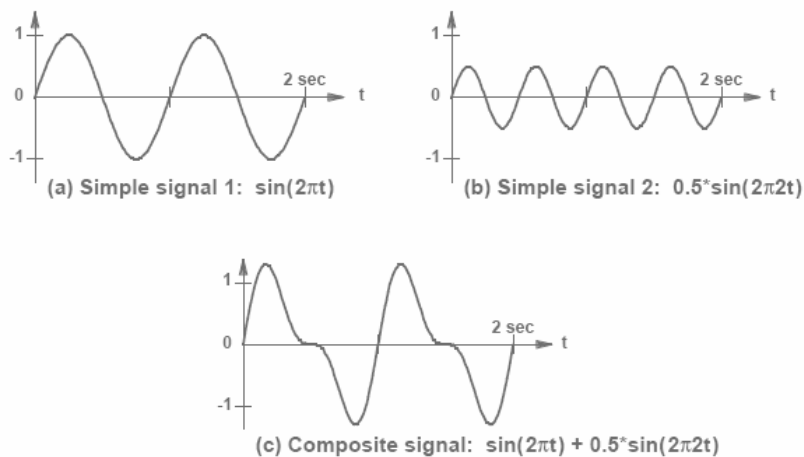   - formed by adding two simple sine waves

Composite Signals

(a) Simple signal 1: $\sin(2\pi t)$

(b) Simple signal 2: $0.5*\sin(2\pi 2t)$

(c) Composite signal: $\sin(2\pi t) + 0.5*\sin(2\pi 2t)$

Figure 6.5 Illustration of a composite signal formed from two simple signals.

# The Importance of Composite Signals and Sine Functions

- Why does data communications seem obsessed with sine functions and composite signals?
- When we discuss modulation and demodulation, we will understand one of the primary reasons:
  - the signals that result from modulation are usually composite signals
- A mathematician named Fourier discovered that
  - it is possible to decompose a composite signal into its constituent parts
    - a set of sine functions, each with a frequency, amplitude, and phase
- The analysis by Fourier shows that if the composite signal is periodic, the constituent parts will also be periodic
  - most systems use composite signals to carry information
  - a composite signal is created at the sending end
  - and the receiver decomposes the signal into the simple components

15

Time and Frequency Domain Representations

- Several methods have been invented to represent composite signals
- A graph of a signal as a function of time is known as time domain representation
- The other representation is known as a frequency domain
  - A frequency domain graph shows a set of simple sine waves that constitute a composite function
  - The y-axis gives the amplitude, and the x-axis gives the frequency
- The function **Asin(2pt)** is represented by a single line of height **A** that is positioned at **x = t**
- The frequency domain graph in Figure 6.6 represents a composite from Figure 6.5c
  - The figure shows a set of simple periodic signals

16

Time and Frequency Domain Representations
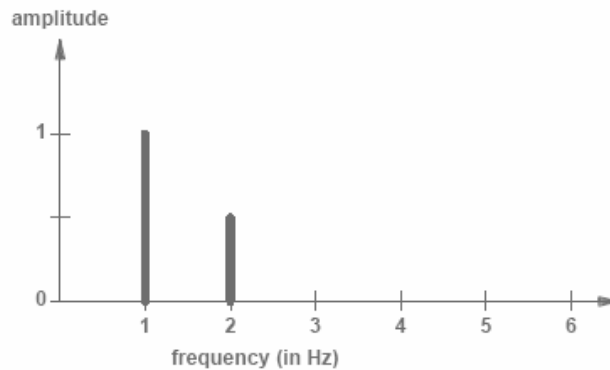
**Figure 6.6** Representation of $\sin(2\pi t)$ and $0.5\sin(2\pi 2t)$ in the frequency domain.

17

Time and Frequency Domain Representations

- A frequency domain representation can also be used with nonperiodic signals
  - but aperiodic representation is not essential to an understanding of the subject
- One of the advantages of the frequency domain representation arises from its compactness
  - a frequency domain representation is both small and easy to read because each sine wave occupies a single point along the x-axis
  - the advantage becomes clear when a composite signal contains many simple signals

18

Bandwidth of an Analog Signal

- What is network bandwidth?
- In networking and communication, the definition of bandwidth varies; here we describe analog bandwidth
  - We define the bandwidth of an analog signal to be the difference between the highest and lowest frequencies of the constituent parts (the highest and the lowest frequencies obtained by Fourier analysis)
- Figure 6.7 shows a frequency domain plot with frequencies measured in Kilohertz (KHz)
  - Such frequencies are in the range audible to a human ear
  - In the figure, the bandwidth is the difference between the highest and lowest frequency (5 KHz - 1 KHz = 4 KHz)


Bandwidth of an Analog Signal
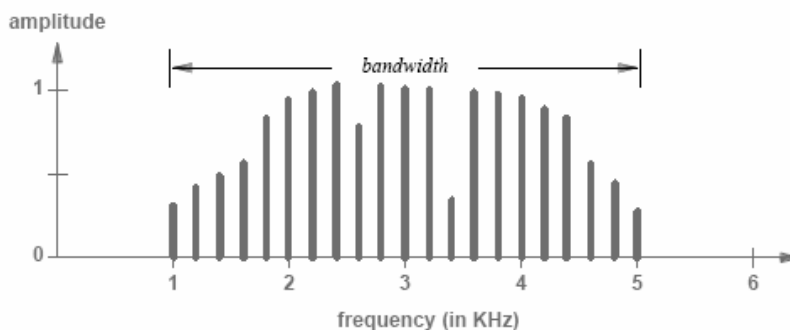
Figure 6.7 A frequency domain plot of an analog signal with a bandwidth of 4 KHz.

Digital Signals and Signal Levels

- Some systems use voltage to represent digital values
    - by making a positive voltage correspond to a logical one
    - and zero voltage correspond to a logical zero
- For example, **+5** volts can be used for a logical **one** and **0** volts for a logical **zero**
- If only two levels of voltage are used
    - each level corresponds to one data bit (0 or 1).
- Some physical transmission mechanisms can support more than two signal levels
- When multiple digital levels are available
    - each level can represent multiple bits
- For example, consider a system that uses four levels of voltage:
    - -5 volts, -2 volts, +2 volts, and +5 volts
    - Each level can correspond to two bits of data as Figure 6.8 illustrates

---

Digital Signals and Signal Levels
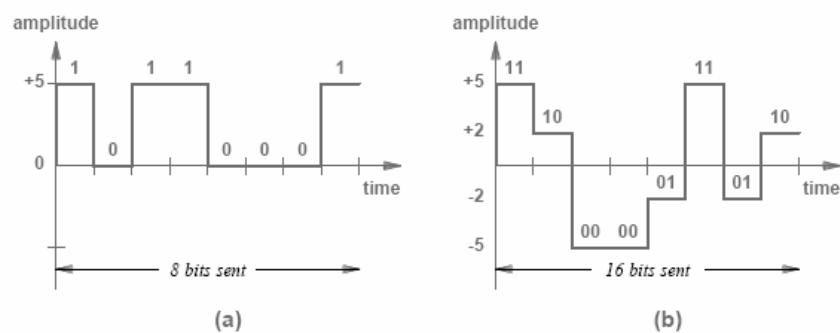
Figure 6.8 (a) A digital signal using two levels, and (b) a digital signal using four levels.

Digital Signals and Signal Levels

- The relationship between the number of levels required and the number of bits to be sent is straightforward
- There must be a signal level for each possible combination of bits
- There are 2^n combinations possible with n bits
  - a communication system must use 2^n levels to represent n bits
- One could achieve arbitrary numbers of levels by dividing voltage into arbitrarily small increments
  - Mathematically, one could create a million levels between 0 and 1 volts merely by using 0.0000001 volts for one level, 0.0000002 for the next level, and so on
- Practical electronic systems cannot distinguish between signals that differ by arbitrarily small amounts
  - Thus, practical systems are restricted to a few signal levels

Baud and Bits Per Second

- How much data can be sent in a given time?
  - The answer depends on two aspects of the communication system.
- The rate at which data can be sent depends on
  - the number of signal levels
  - the amount of time the system remains at a given level before moving to the next
- Figure 6.8a shows time along the x-axis, and the time is divided into eight segments
  - one bit being sent during each segment
- If the communication system is modified to use half as much time for a given bit
  - twice as many bits will be sent in the same amount of time

Baud and Bits Per Second

- As with signal levels, the hardware in a practical system places limits on how short the time can be
  - if the signal does not remain at a given level long enough, the receiving hardware will fail to detect it
- The accepted measure of a communication system does not specify a length of time
  - how many times the signal can change per second, which is defined as the baud
  - for example, if a system requires the signal to remain at a given level for .001 seconds, we say that the system operates at 1000 baud
- Both baud and number of signal levels control bit rate

Baud and Bits Per Second

- If a system with two signal levels operates at 1000 baud
  - the system can transfer exactly 1000 bits per second
- If a system that operates at 1000 baud has four signal levels
  - the system can transfer 2000 bits per second (because four signal levels can represent two bits)
- Equation 6.1 expresses the relationship between baud, signal levels, and bit rate

$$bits\ per\ second\ =\ baud\ \times\ \lfloor \log_2(\ levels\ ) \rfloor$$

Converting a Digital Signal to Analog

- How can a digital signal be converted into an equivalent analog signal?
- According to Fourier, an arbitrary curve can be represented as a composite of sine waves
  - where each sine wave in the set has a specific amplitude, frequency, and phase
- Fourier's theorem also applies to a digital signal
  - however, accurate representation of a digital signal requires an infinite set of sine waves
- Engineers adopt a compromise:
  - conversion of a signal from digital to analog is approximate
  - generate analog waves that closely approximate the digital signal
  - approximation involves building a composite signal from only a few sine waves
  - Figure 6.9 illustrates the approximation

Converting a Digital Signal to Analog

(a) digital signal

(b) sin(2πt/2)

(c) sin(2πt/2)+α sin(2π3t/2)   (d) sin(2πt/2)+α sin(2π3t/2)+β sin(2π5t/2)
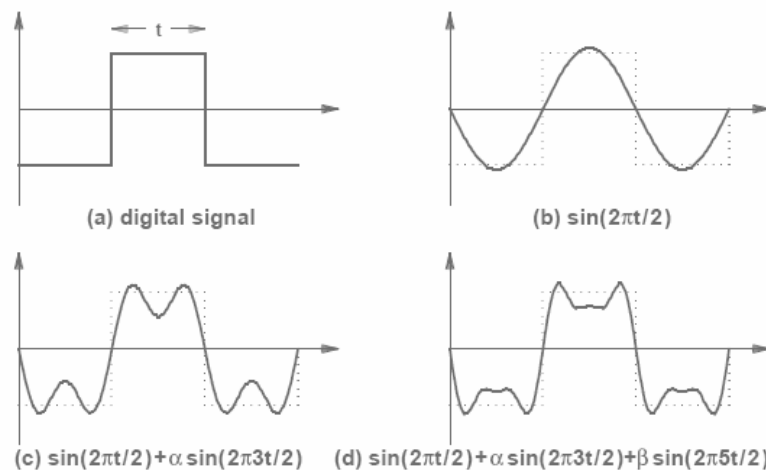
**Figure 6.9** Approximation of a digital signal with sine waves.

The Bandwidth of a Digital Signal

- What is the bandwidth of a digital signal?
- If Fourier analysis is applied to a square wave?
  - such as the digital signal illustrated in Figure 6.9a
- A digital signal has infinite bandwidth
  - because Fourier analysis of a digital signal produces an infinite set of sine waves with frequencies that grow to infinity

Synchronization and Agreement About Signals

- The electronics at both ends of a physical medium must have circuitry to measure time precisely
  - if one end transmits a signal with 10 elements per second, the other end must expect exactly 10 elements per second
- At slow speeds, making both ends agree is trivial
- Building electronic systems that agree at the high speeds used in modern networks is extremely difficult
- A fundamental problem that arises from the way data is represented concerns synchronization of sender/receiver
- Suppose a receiver misses the first bit that arrives, and starts interpreting data starting at the second bit
  - Figure 6.10 illustrates how a mismatch in interpretation can produce errors

Synchronization and Agreement About Signals
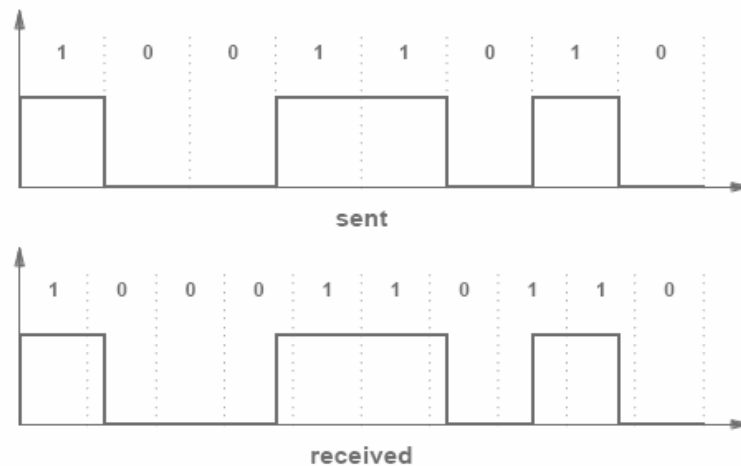
Figure 6.10 Illustration of a synchronization error in which the receiver allows slightly less time per bit than the sender.

31

Line Coding

- Several techniques have been invented that can help avoid synchronization errors
- In general, there are two broad approaches:
  1. Before transmitting, the sender transmits a known pattern of bits, typically a set of alternating 0s and 1s, for receiver to synchronize
  2. Data is represented by the signal in such a way that there can be no confusion about the meaning
- We use the term line coding to describe the way data is encoded in a signal
- As an example of line coding that eliminates ambiguity
- How one can use a transmission mechanism that supports three discrete signal levels?
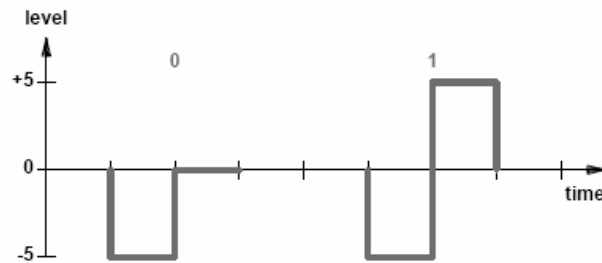 - Consider Figure 6.11

32

Line Coding

Figure 6.11 Example of two signal elements used to represent each bit.

Line Coding

- Using multiple signal elements to represent a single bit means fewer bits can be transmitted per unit time
- Designers prefer schemes that transmit multiple bits per signal element, such as the one that Figure 6.8b illustrates
- Figure 6.12 lists the names of line coding techniques in common use and groups them into related categories
  - choice depends on the specific needs of a given system
- A variety of line coding techniques are available that differ in
  - how they handle synchronization
 - as well as other properties such as the bandwidth used

Line Coding

Category	Scheme	Synchronization
Unipolar	NRZ	No, if many 0s or 1s are repeated
	NRZ-L	No, if many 0s or 1s are repeated
	NRZ-I	No, if many 0s or 1s are repeated
	Biphase	Yes
Bipolar	AMI	No, if many 0s are repeated
Multilevel	2B1Q	No, if many double bits are repeated
	8B6T	Yes
	4D-PAM5	Yes
Multiline	MLT-3	No, if many 0s are repeated

Figure 6.12 Names of line coding techniques in common use.

Manchester Encoding Used in Computer Networks

- In addition to the list in Figure 6.12, one particular standard for line coding is especially important for networks:
  - Manchester Encoding used with Ethernet
- Detecting a transition in signal level is easier than measuring the signal level
- This fact explains why Manchester Encoding uses transitions rather than levels to define bits
- In Manchester Encoding, a 1 corresponds to a transition from negative voltage level to a positive voltage level
  - Correspondingly, a 0 corresponds to a transition from a positive voltage level to a negative level
  - The transitions occur in the "middle" of the time slot of a bit
- Figure 6.13a illustrates the concept

Manchester Encoding Used in Computer Networks
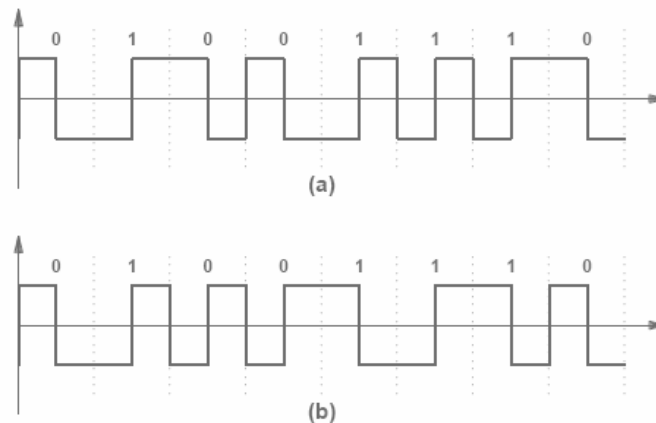
Figure 6.13 (a) Manchester and (b) Differential Manchester Encodings; each assumes the previous bit ended with a low signal level.

37

Manchester Encoding Used in Computer Networks

- A variation known as a Differential Manchester Encoding (also called a Conditional DePhase Encoding) uses relative transitions rather than absolute
  - the representation of a bit depends on the previous bit
- Each bit time slot contains one or two transitions
- A transition always occurs in the middle of the bit time
- The logical value of the bit is represented by the presence or absence of a transition at the beginning of a bit time:
  - logical 0 is represented by a transition
  - and logical 1 is represented by no transition
- Figure 6.13b illustrates Differential Manchester Encoding
  - most important property of differential encoding arises from a practical consideration:
    - the encoding works correctly even if the two wires carrying the signal are accidentally reversed

38

Converting an Analog Signal to Digital

- Many sources of information are analog
  - which means they must be converted to digital form for further processing (e.g., before they can be encrypted)
- There are two basic approaches:
  - pulse code modulation (PCM)
  - delta modulation (DM)
- In PCM, the level of an analog signal is measured repeatedly at fixed time intervals and converted to digital form
- The acronym PCM is ambiguous
  - because it can refer to the general idea or to a specific form of pulse code modulation used by the telephone system
- Figure 6.14 illustrates the steps
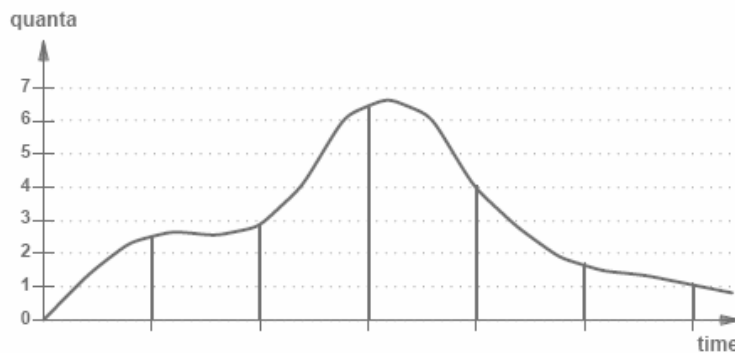

Converting an Analog Signal to Digital
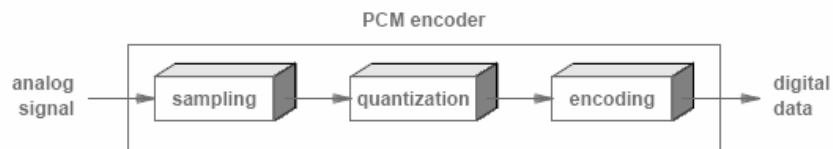
Figure 6.14 The three steps used in pulse code modulation.

Converting an Analog Signal to Digital

- The three steps used PCM
- First stage is known as sampling
  - each measurement is known as a sample
- A sample is quantized
  - by converting it into a small integer value
  - the quantized value is not a measure of voltage or any other property of the signal
  - the range of the signal from the minimum to maximum levels is divided into a set of slots, typically a power of 2
- Then encoded into a specific format
- Figure 6.15 illustrates the concept by showing a signal quantized into eight slots

---

Converting an Analog Signal to Digital

Figure 6.15 An illustration of the sampling and quantization used in pulse code modulation.

Converting an Analog Signal to Digital

- In the Figure 6.15
  - the six samples are represented by vertical gray lines
  - each sample is quantized by choosing the closest quantum interval
  - for example, the third sample, taken near the peak of the curve, is assigned a quantized value of 6
- In practice, slight variations in sampling have been invented
- For example, to avoid inaccuracy caused by a brief spike or a dip in the signal, averaging can be used
  - instead of relying on a single measurement for each sample, three measurements can be taken close together and an arithmetic mean can be computed

43

Converting an Analog Signal to Digital

- Delta modulation (DM) also takes samples like PCM
  - however, does not do a quantization for each sample
  - DM sends one quantization value followed by a string of values that give the difference between the previous value and the current value
- Transmitting differences requires fewer bits than sending full values, especially if the signal does not vary rapidly
- The main tradeoff with DM arises from the effect of an error
  - if any item in the sequence is lost or damaged, all successive values will be misinterpreted
  - communication systems that expect data values to be lost or changed during transmission usually use PCM

44

The Nyquist Theorem and Sampling Rate

- An analog signal must be sampled in PCM or DM
- How frequently should an analog signal be sampled?
  - Taking too few samples (undersampling) means that the digital values only give a crude approximation of the original signal
  - Taking too many samples (oversampling) means that more digital data will be generated, which uses extra bandwidth
- A mathematician named Nyquist discovered the answer to the question of how much sampling is required:

$$sampling\ rate\ =\ 2 \times f_{max}$$

  - where f_{max} is the highest frequency in the composite signal
- Nyquist Theorem provides a practical solution to the problem:
  - sample a signal at least twice as fast as the highest frequency that must be preserved

45

Nyquist Theorem and Telephone System Transmission

- Consider the telephone system that was originally designed to transfer voice as an example
  - Measurements of human speech have shown that preserving frequencies between 0 - 4000 Hz provides acceptable audio quality
- When converting a voice signal from analog to digital
  - the signal should be sampled at a rate of 8000 samples per second
- The PCM standard used by the phone system quantifies each sample into an 8 bit value for quality
  - the range of input is divided into 256 possible levels so that each sample has a value between 0 and 255
- The rate generated for a single telephone call is:

$$digitized\ voice\ call\ =\ 8000\frac{samples}{second}\ \times\ 8\frac{bits}{sample}\ =\ 64,000\frac{bits}{second}$$

46

Encoding and Data Compression

- Data compression refers to a technique that reduces the number of bits required to represent data
- Data compression is relevant to a communication system
  - because reducing the number of bits used to represent data reduces the time required for transmission
  - a communication system can be optimized by compressing data
- Chapter 29 discusses compression in multimedia applications
- There are two types of compression:
  - Lossy - some information is lost during compression
  - Lossless - all information is retained in the compressed version

Encoding and Data Compression

- Lossy compression is generally used with data that a human consumes, such as an image, video/audio
- The key idea is that the compression only needs to preserve details to the level of human perception
  - a change is acceptable if humans cannot detect the change
  - JPEG (used for images) compression or MPEG-3 (abbreviated MP3 and used for audio recordings) employ lossy compression
- Lossless compression preserves the original data without any change
  - lossless compression can be used for documents or in any situation where data must be preserved exactly
  - when used for communication, a sender compresses the data before transmission and the receiver decompresses the result
  - arbitrary data can be compressed by a sender and decompressed by a receiver to recover an exact copy of the original