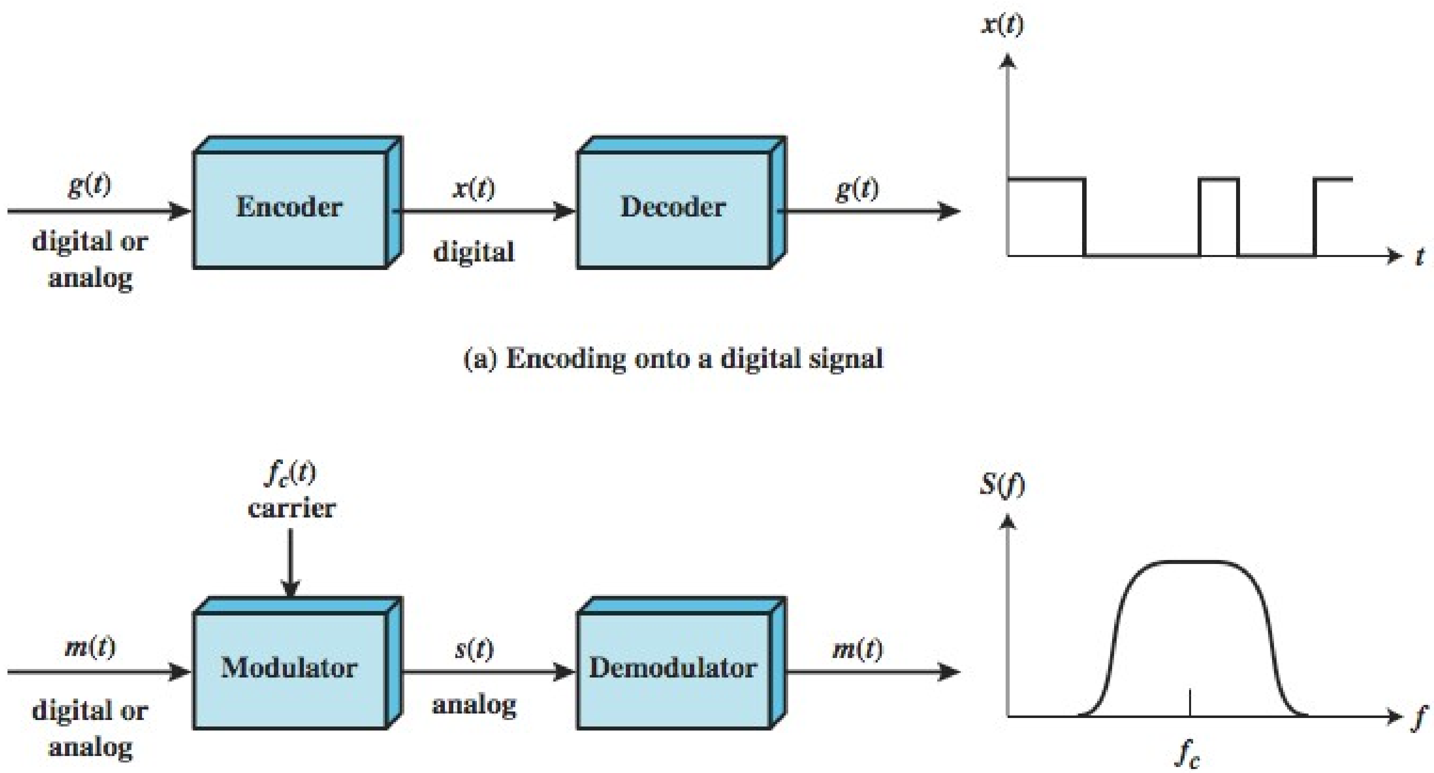
Signal Encoding Techniques

Signal Encoding Techniques



(a) Encoding onto a digital signal

(b) Modulation onto an analog signal

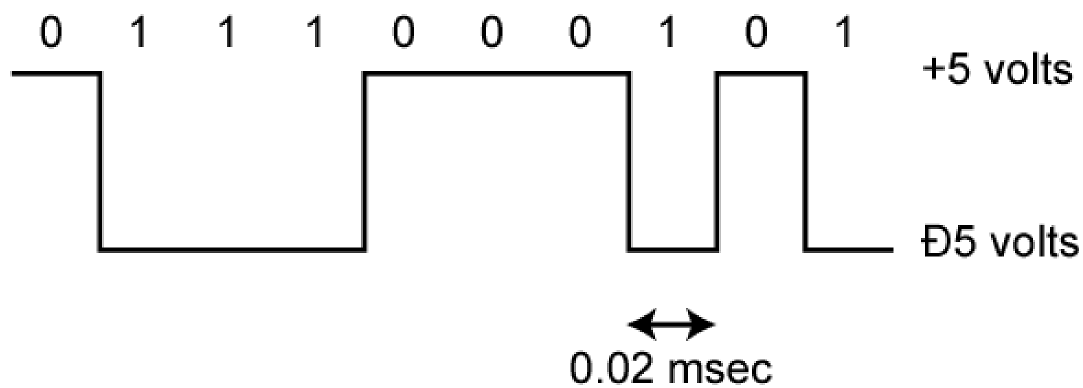
Figure 5.1 Encoding and Modulation Techniques

- **Digital data, digital signals:** simplest form of digital encoding of digital data
- **Digital data, analog signal:** A modem converts digital data to an analog signal so that it can be transmitted over an analog
- **Analog data, digital signals:** Analog data, such as voice and video, are often digitized to be able to use digital transmission facilities
- **Analog data, analog signals:** Analog data are modulated by a carrier frequency to produce an analog signal in a different frequency band, which can be utilized on an analog transmission system

Stallings DCC8e Fig emphasizes the process involved in this. For **digital signaling**, a data source *g*(*t*), which may be either digital or analog, is encoded into a digital signal *x*(*t*). The basis for **analog signaling** is a continuous constant-frequency f_c signal known as the **carrier signal**. Data may be transmitted using a carrier signal by modulation, which is the process of encoding source data onto the carrier signal. All modulation techniques involve operation on one or more of the three fundamental frequency domain parameters: amplitude, frequency, and phase. The input signal *m*(*t*) may be analog or digital and is called the modulating signal, and the result of modulating the carrier signal is called the modulated signal *s*(*t*).

Digital Data, Digital Signal

- Digital signal
  - discrete, discontinuous voltage pulses
  - each pulse is a signal element
  - binary data encoded into signal elements



- **Encoding - Digital data to digital signals:** A digital signal is a sequence of discrete, discontinuous voltage pulses, as illustrated in Stallings DCC8e. Each pulse is a signal element. Binary data are transmitted by encoding each data bit into signal elements. In the simplest case, there is a one-to-one correspondence between bits and signal elements. More complex encoding schemes are used to improve performance, by altering the spectrum of the signal and providing synchronization capability. In general, the equipment for encoding digital data into a digital signal is less complex and less expensive than digital-to-analog modulation equipment

Some Terms

- unipolar
- polar
- data rate
- duration or length of a bit
- modulation rate
- mark and space

- Before discussing this further, we need to define some terms:
- Unipolar - All signal elements have the same sign
- Polar - One logic state represented by positive voltage the other by negative voltage
- Data rate - Rate of data (R) transmission in bits per second
- Duration or length of a bit - Time taken for transmitter to emit the bit (1/R)
- Modulation rate -Rate at which the signal level changes, measured in baud = signal elements per second. Depends on type of digital encoding used.
- Mark and Space - Binary 1 and Binary 0 respectively

Interpreting Signals

- need to know
  - timing of bits - when they start and end
  - signal levels
- factors affecting signal interpretation
  - signal to noise ratio
  - data rate
  - bandwidth
  - encoding scheme

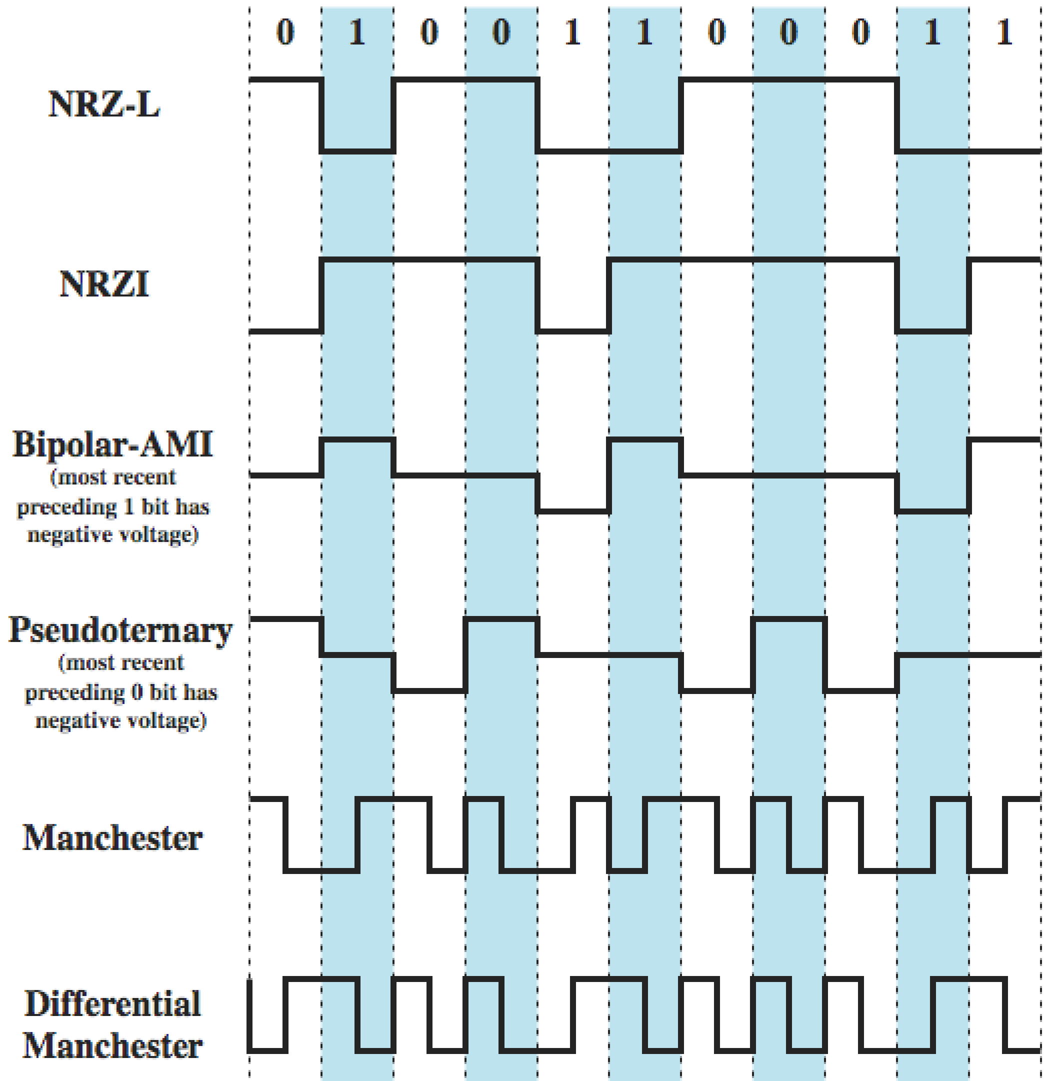
- The tasks involved in interpreting digital signals at the receiver can be summarized as follows. First, the receiver must know the timing of each bit, knowing with some accuracy when a bit begins and ends. Second, the receiver must determine whether the signal level for each bit position is high (0) or low (1). These tasks can be performed by sampling each bit position in the middle of the interval and comparing the value to a threshold. Because of noise and other impairments, there will be errors. As was shown in Chapter 3, three factors are important: the signal-to-noise ratio, the data rate, and the bandwidth. With other factors held constant, the following statements are true:
  - • An increase in data rate increases bit error rate (BER).
  - • An increase in SNR decreases bit error rate.
  - • An increase in bandwidth allows an increase in data rate.
- There is another factor that can be used to improve performance, and that is the encoding scheme. The encoding scheme is simply the mapping from data bits to signal elements. A variety of approaches have been tried. In what follows, we describe some of the more common ones.

Comparison of Encoding Schemes

- signal spectrum
- clocking
- error detection
- signal interference and noise immunity
- cost and complexity

- Before describing the various encoding techniques, consider the following ways of evaluating or comparing them:
- Signal Spectrum - Lack of high frequencies reduces required bandwidth, lack of dc component allows ac coupling via transformer, providing isolation, should concentrate power in the middle of the bandwidth
- Clocking - need for synchronizing transmitter and receiver either with an external clock or with a sync mechanism based on signal
- Error detection - useful if can be built in to signal encoding
- Signal interference and noise immunity - some codes are better than others
- Cost and complexity - Higher signal rate (& thus data rate) lead to higher costs, some codes require signal rate greater than data rate

Encoding Schemes



They include:

- Nonreturn to Zero-Level (NRZ-L)
- Nonreturn to Zero Inverted (NRZI)
- Bipolar -AMI
- Pseudoternary
- Manchester
- Differential Manchester
- B8ZS
- HDB3

Nonreturn to Zero-Level (NRZ-L)

- two different voltages for 0 and 1 bits
- voltage constant during bit interval
  - no transition I.e. no return to zero voltage
  - such as absence of voltage for zero, constant positive voltage for one
  - more often, negative voltage for one value and positive for the other

- The most common, and easiest, way to transmit digital signals is to use two different voltage levels for the two binary digits. Codes that follow this strategy share the property that the voltage level is constant during a bit interval; there is no transition (no return to a zero voltage level). Can have absence of voltage used to represent binary 0, with a constant positive voltage used to represent binary 1. More commonly a negative voltage represents one binary value and a positive voltage represents the other. This is known as **Nonreturn to Zero-Level (**NRZ-L). NRZ-L is typically the code used to generate or interpret digital data by terminals and other devices.

Nonreturn to Zero Inverted

- nonreturn to zero inverted on ones
- constant voltage pulse for duration of bit
- data encoded as presence or absence of signal transition at beginning of bit time
  - transition (low to high or high to low) denotes binary 1
  - no transition denotes binary 0
- example of differential encoding since have
  - data represented by changes rather than levels
  - more reliable detection of transition rather than level
  - easy to lose sense of polarity

- A variation of NRZ is known as **NRZI** (Nonreturn to Zero, invert on ones). As with NRZ-L, NRZI maintains a constant voltage pulse for the duration of a bit time. The data bits are encoded as the presence or absence of a signal transition at the beginning of the bit time. A transition (low to high or high to low) at the beginning of a bit time denotes a binary 1 for that bit time; no transition indicates a binary 0.
- NRZI is an example of **differential encoding**. In differential encoding, the information to be transmitted is represented in terms of the changes between successive signal elements rather than the signal elements themselves. The encoding of the current bit is determined as follows: if the current bit is a binary 0, then the current bit is encoded with the same signal as the preceding bit; if the current bit is a binary 1, then the current bit is encoded with a different signal than the preceding bit. One benefit of differential encoding is that it may be more reliable to detect a transition in the presence of noise than to compare a value to a threshold. Another benefit is that with a complex transmission layout, it is easy to lose the sense of the polarity of the signal.

NRZ Pros & Cons

- Pros
  - easy to engineer
  - make good use of bandwidth
- Cons
  - dc component
  - lack of synchronization capability
- used for magnetic recording
- not often used for signal transmission

- The NRZ codes are the easiest to engineer and, in addition, make efficient use of bandwidth. Most of the energy in NRZ and NRZI signals is between dc and half the bit rate. The main limitations of NRZ signals are the presence of a dc component and the lack of synchronization capability. Consider that with a long string of 1s or 0s for NRZ-L or a long string of 0s for NRZI, the output is a constant voltage over a long period of time. Under these circumstances, any drift between the clocks of transmitter and receiver will result in loss of synchronization between the two.
- Because of their simplicity and relatively low frequency response characteristics, NRZ codes are commonly used for digital magnetic recording. However, their limitations make these codes unattractive for signal transmission applications.

Multilevel Binary
Bipolar-AMI

- Use more than two levels
- Bipolar-AMI
  - zero represented by no line signal
  - one represented by positive or negative pulse
  - one pulses alternate in polarity
  - no loss of sync if a long string of ones
  - long runs of zeros still a problem
  - no net dc component
  - lower bandwidth
  - easy error detection

- A category of encoding techniques known as multilevel binary addresses some of the deficiencies of the NRZ codes. These codes use more than two signal levels. Two examples of this scheme was illustrated in Figure.
- In the **bipolar-AMI** scheme, a binary 0 is represented by no line signal, and a binary 1 is represented by a positive or negative pulse. The binary 1 pulses must alternate in polarity. There are several advantages to this approach. First, there will be no loss of synchronization if a long string of 1s occurs. Each 1 introduces a transition, and the receiver can resynchronize on that transition. A long string of 0s would still be a problem. Second, because the 1 signals alternate in voltage from positive to negative, there is no net dc component. Also, the bandwidth of the resulting signal is considerably less than the bandwidth for NRZ. Finally, the pulse alternation property provides a simple means of error detection. Any isolated error, whether it deletes a pulse or adds a pulse, causes a violation of this property.

Multilevel Binary Pseudoternary

- one represented by absence of line signal
- zero represented by alternating positive and negative
- no advantage or disadvantage over bipolar-AMI
- each used in some applications

- The comments on **bipolar-AMI** also apply to **pseudoternary**. In this case, it is the binary 1 that is represented by the absence of a line signal, and the binary 0 by alternating positive and negative pulses. There is no particular advantage of one technique versus the other, and each is the basis of some applications.

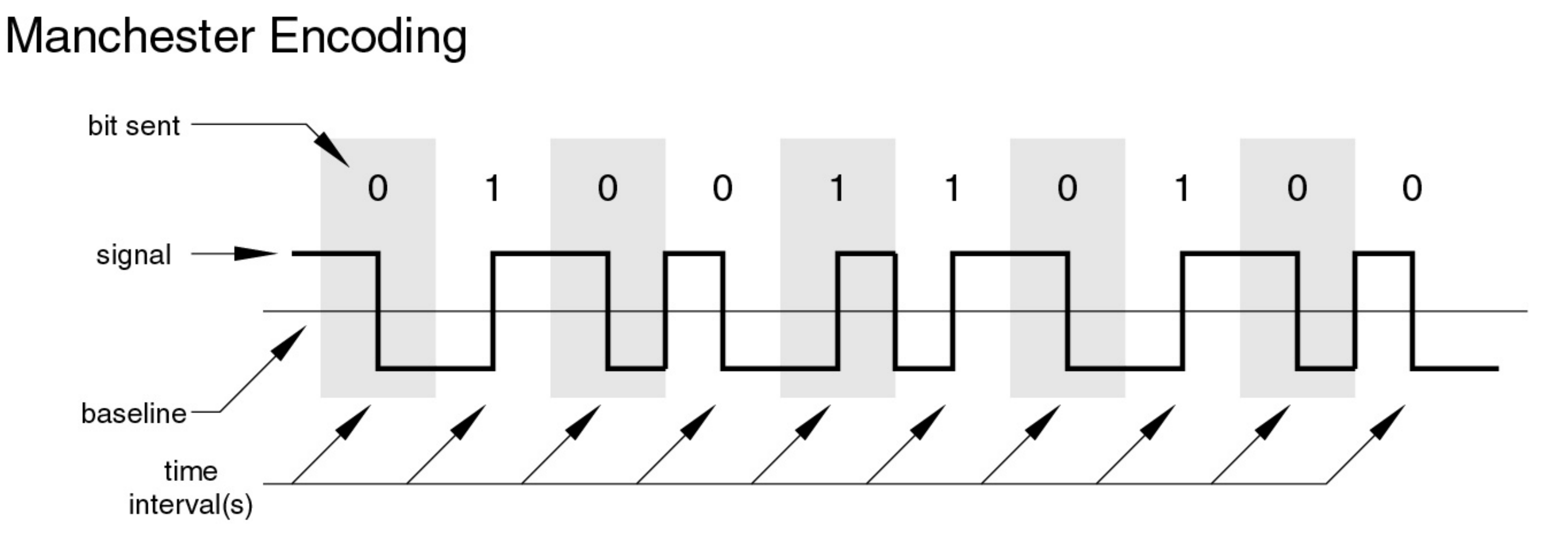
Multilevel Binary Issues

- synchronization with long runs of 0's or 1's
  - can insert additional bits, cf ISDN
  - scramble data (later)
- not as efficient as NRZ
  - each signal element only represents one bit
    - receiver distinguishes between three levels: +A, -A, 0
  - a 3 level system could represent $\log_2 3 = 1.58$ bits
  - requires approx. 3dB more signal power for same probability of bit error

- Although a degree of synchronization is provided with these codes, a long string of 0s in the case of AMI or 1s in the case of pseudoternary still presents a problem. Several techniques have been used to address this deficiency. One approach is to insert additional bits that force transitions. This technique is used in ISDN (integrated services digital network) for relatively low data rate transmission. Of course, at a high data rate, this scheme is expensive, because it results in an increase in an already high signal transmission rate. To deal with this problem at high data rates, a technique that involves scrambling the data is used.
- Thus, with suitable modification, multilevel binary schemes overcome the problems of NRZ codes. Of course, as with any engineering design decision, there is a tradeoff. With multilevel binary coding, the line signal may take on one of three levels, but each signal element, which could represent $\log_2 3 = 1.58$ bits of information, bears only one bit of information, since the receiver of multilevel binary signals has to distinguish between three levels ($+A$, $-A$, 0) instead of just two levels in the signaling formats previously discussed. Because of this, the multilevel binary signal requires approximately 3 dB more signal power than a two-valued signal for the same probability of bit error. Put another way, the bit error rate for NRZ codes, at a given signal-to-noise ratio, is significantly less than that for multilevel binary.

Manchester Encoding

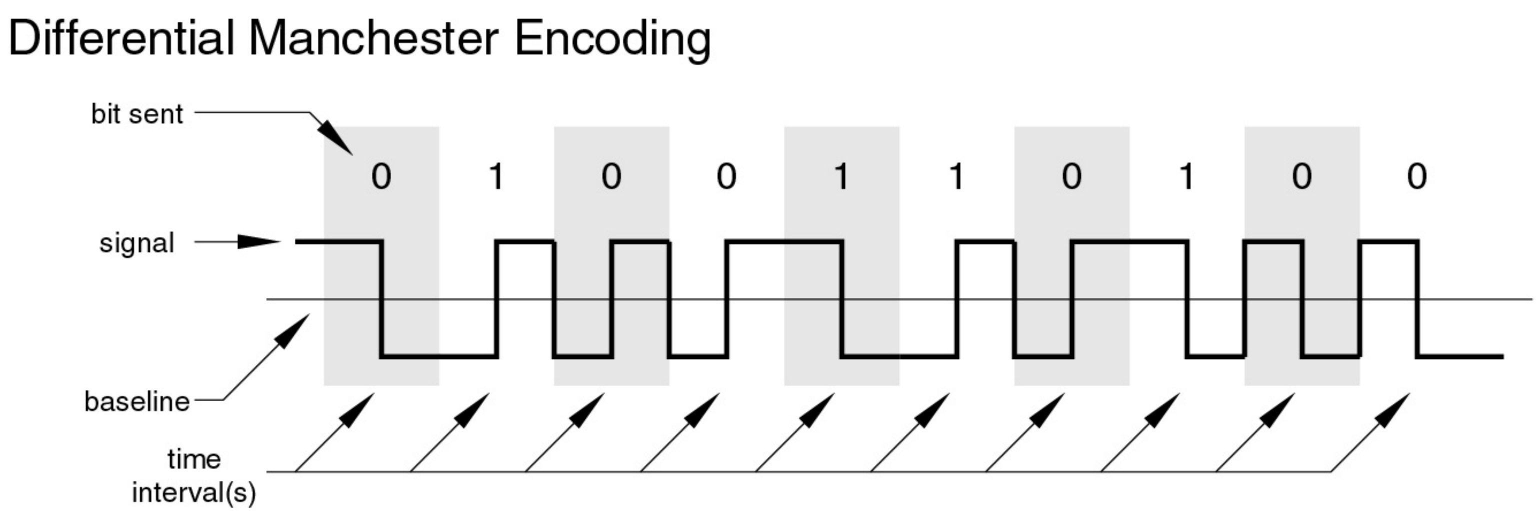
- has transition in middle of each bit period
- transition serves as clock and data
- low to high represents one
- high to low represents zero
- used by IEEE 802.



- There is another set of coding techniques, grouped under the term *biphase*, that overcomes the limitations of NRZ codes. Two of these techniques, Manchester and differential Manchester, are in common use.
- In the **Manchester** code, there is a transition at the middle of each bit period. The midbit transition serves as a clocking mechanism and also as data: a low-to-high transition represents a 1, and a high-to-low transition represents a 0. Biphase codes are popular techniques for data transmission. The more common Manchester code has been specified for the IEEE 802.3 (Ethernet) standard for baseband coaxial cable and twisted-pair bus LANs.

Differential Manchester Encoding

- midbit transition is clocking only
- transition at start of bit period representing 0
- no transition at start of bit period representing 1
  - this is a differential encoding scheme
- used by IEEE 802.5



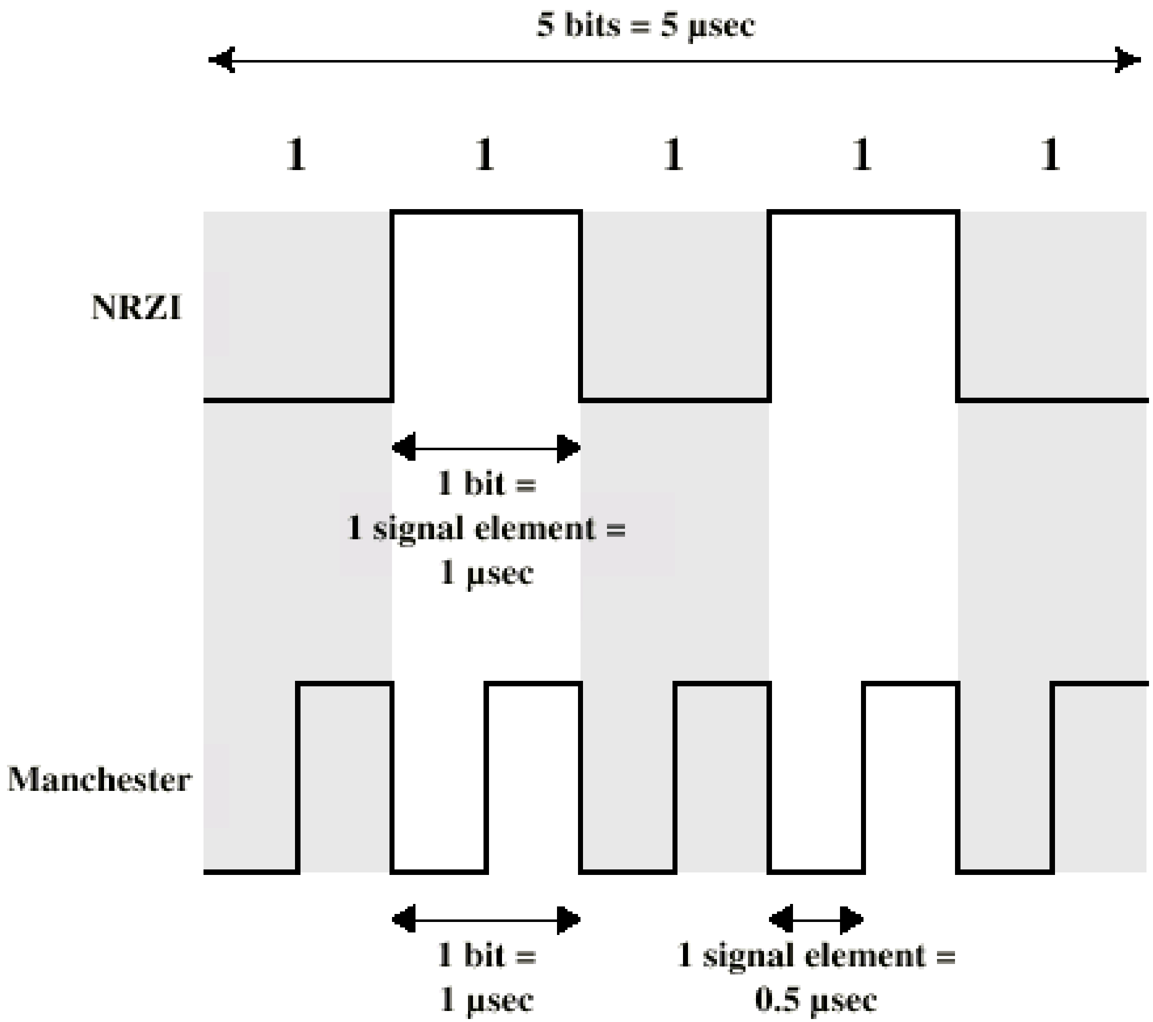
- In **differential Manchester**, the midbit transition is used only to provide clocking. The encoding of a 0 is represented by the presence of a transition at the beginning of a bit period, and a 1 is represented by the absence of a transition at the beginning of a bit period. Differential Manchester has the added advantage of employing differential encoding.
- Differential Manchester has been specified for the IEEE 802.5 token ring LAN, using shielded twisted pair.

Biphase Pros and Cons

- Con
  - at least one transition per bit time and possibly two
  - maximum modulation rate is twice NRZ
  - requires more bandwidth
- Pros
  - synchronization on mid bit transition (self clocking)
  - has no dc component
  - has error detection

- All of the biphase techniques require at least one transition per bit time and may have as many as two transitions. Thus, the maximum modulation rate is twice that for NRZ; this means that the bandwidth required is correspondingly greater. The bandwidth for biphase codes is reasonably narrow and contains no dc component. However, it is wider than the bandwidth for the multilevel binary codes.
- On the other hand, the biphase schemes have several advantages:
  - **Synchronization:** Because there is a predictable transition during each bit time, the receiver can synchronize on that transition, known as self-clocking codes.
  - **No dc component:** Biphase codes have no dc component
  - **Error detection:** The absence of an expected transition can be used to detect errors. Noise on the line would have to invert both the signal before and after the expected transition to cause an undetected error.

Modulation Rate



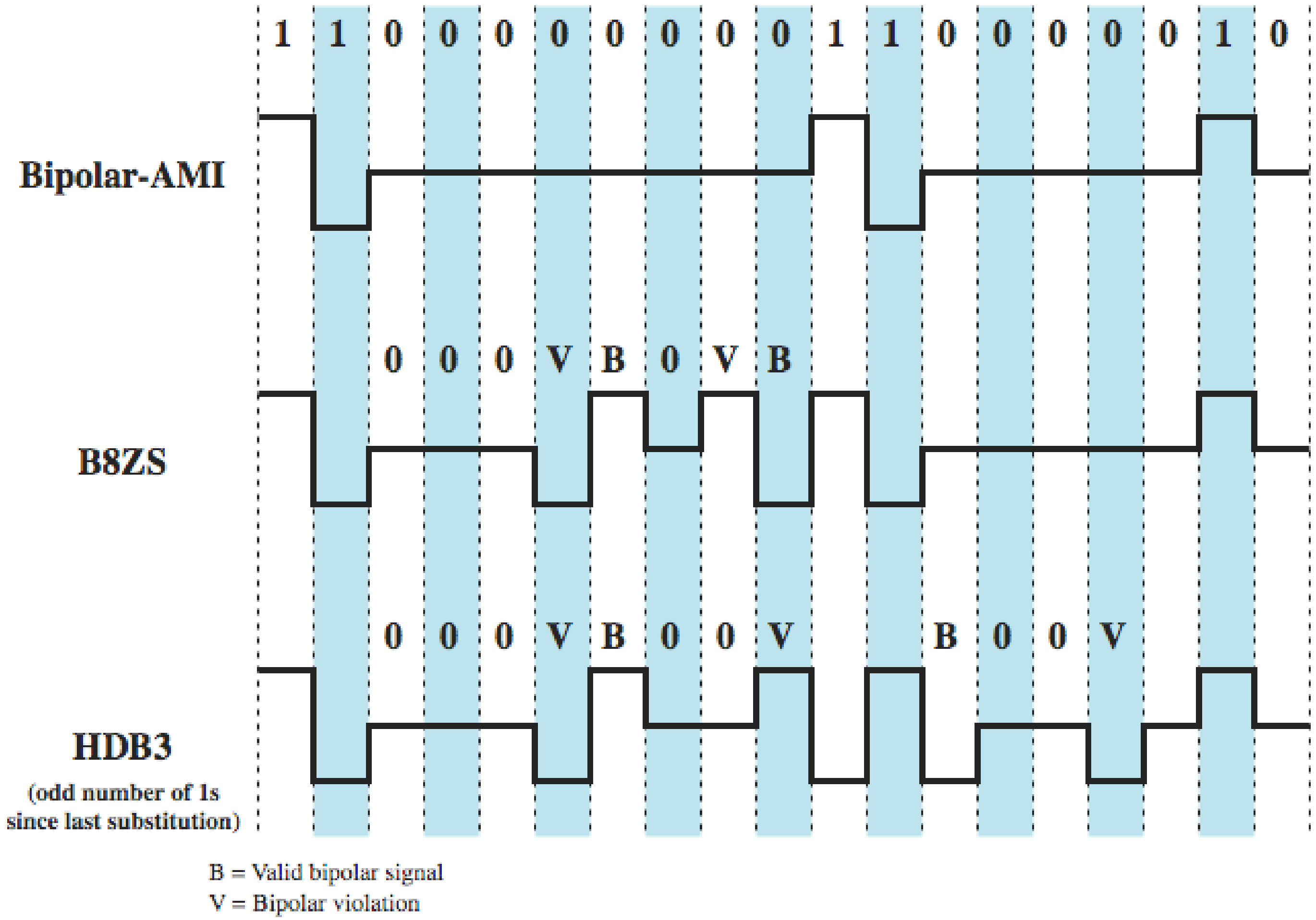
- When signal-encoding techniques are used, a distinction needs to be made between data rate (expressed in bits per second) and modulation rate (expressed in baud). The data rate, or bit rate, is $1/T_b$, where T_b = bit duration. The modulation rate is the rate at which signal elements are generated. Consider, for example, Manchester encoding. The minimum size signal element is a pulse of one-half the duration of a bit interval. For a string of all binary zeroes or all binary ones, a continuous stream of such pulses is generated. Hence the maximum modulation rate for Manchester is $2/T_b$. This situation is illustrated in Stallings DCC8e Figure, which shows the transmission of a stream of binary 1s at a data rate of 1 Mbps using NRZI and Manchester.
- One way of characterizing the modulation rate is to determine the average number of transitions that occur per bit time. In general, this will depend on the exact sequence of bits being transmitted. Stallings DCC8e Table 5.3 compares transition rates for various techniques.

Scrambling

- use scrambling to replace sequences that would produce constant voltage
- these filling sequences must
  - produce enough transitions to sync
  - be recognized by receiver & replaced with original
  - be same length as original
- design goals
  - have no dc component
  - have no long sequences of zero level line signal
  - have no reduction in data rate
  - give error detection capability

- Although the biphase techniques have achieved widespread use in local area network applications at relatively high data rates (up to 10 Mbps), they have not been widely used in long-distance applications. The principal reason for this is that they require a high signaling rate relative to the data rate. This sort of inefficiency is more costly in a long-distance application.
- Another approach is to make use of some sort of scrambling scheme. The idea behind this approach is simple: sequences that would result in a constant voltage level on the line are replaced by filling sequences that will provide sufficient transitions for the receiver's clock to maintain synchronization. The filling sequence must be recognized by the receiver and replaced with the original data sequence. The filling sequence is the same length as the original sequence, so there is no data rate penalty. The design goals for this approach can be summarized as follows:
- • No dc component
- • No long sequences of zero-level line signals
- • No reduction in data rate
- • Error-detection capability

B8ZS and HDB3



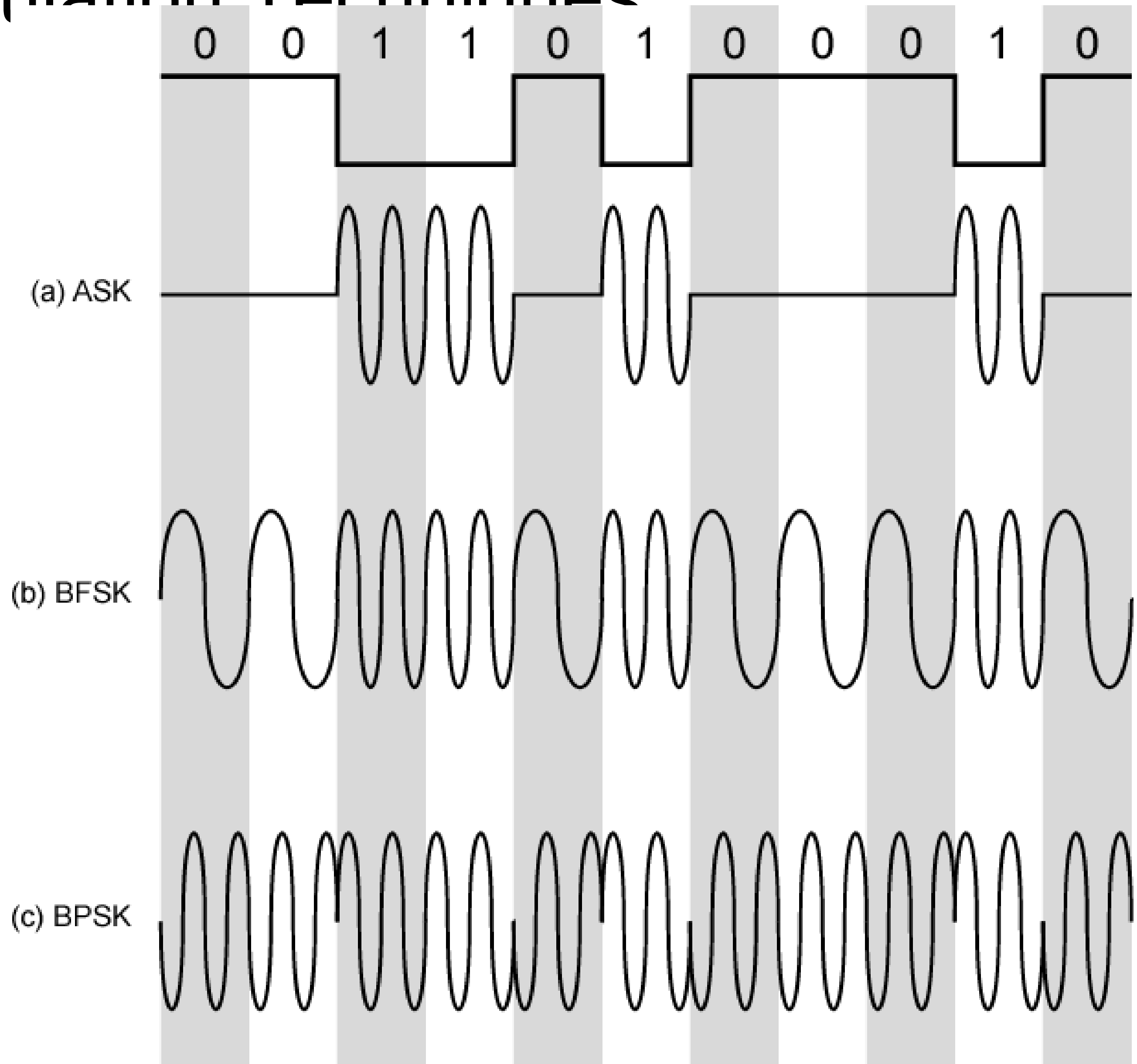
- Two techniques are commonly used in long-distance transmission services; these are illustrated in Stallings DCC8e Figure.
- A coding scheme that is commonly used in North America, based on a bipolar-AMI, is known as **bipolar with 8-zeros substitution (B8ZS).** To overcome the drawback of the AMI code that a long string of zeros may result in loss of synchronization, the encoding is amended with the following rules:
- • If an octet of all zeros occurs and the last voltage pulse preceding this octet was positive, then the eight zeros of the octet are encoded as 000+–0–+.
- • If an octet of all zeros occurs and the last voltage pulse preceding this octet was negative, then the eight zeros of the octet are encoded as 000–+0+–.
- This technique forces two code violations (signal patterns not allowed in AMI) of the AMI code, an event unlikely to be caused by noise or other transmission impairment. The receiver recognizes the pattern and interprets the octet as consisting of all zeros.
- A coding scheme that is commonly used in Europe and Japan is known as the **high-density bipolar-3 zeros (HDB3)** code. It is also based on the use of AMI encoding. In this case, the scheme replaces strings of four zeros with sequences containing one or two pulses. In each case, the fourth zero is replaced with a code violation. In addition, a rule is needed to ensure that successive violations are of alternate polarity so that no dc component is introduced. Thus, if the last violation was positive, this violation must be negative and vice versa.
- Neither of these codes has a dc component. Most of the energy is concentrated in a relatively sharp spectrum around a frequency equal to one-half the data rate. Thus, these codes are well suited to high data rate transmission.

Digital Data, Analog Signal

- main use is public telephone system
  - has freq range of 300Hz to 3400Hz
  - use modem (modulator-demodulator)
- encoding techniques
  - Amplitude shift keying (ASK)
  - Frequency shift keying (FSK)
  - Phase shift keying (PK)

- We turn now to the case of transmitting digital data using analog signals. The most familiar use of this transformation is for transmitting digital data through the public telephone network. The telephone network was designed to receive, switch, and transmit analog signals in the voice-frequency range of about 300 to 3400 Hz. It is not at present suitable for handling digital signals from the subscriber locations (although this is beginning to change). Thus digital devices are attached to the network via a modem (modulator-demodulator), which converts digital data to analog signals, and vice versa.
- Have stated that modulation involves operation on one or more of the three characteristics of a carrier signal: amplitude, frequency, and phase. Accordingly, there are three basic encoding or modulation techniques for transforming digital data into analog signals, as illustrated in Stallings DCC8e Figure (next slide): amplitude shift keying (ASK), frequency shift keying (FSK), and phase shift keying (PSK). In all these cases, the resulting signal occupies a bandwidth centered on the carrier frequency.

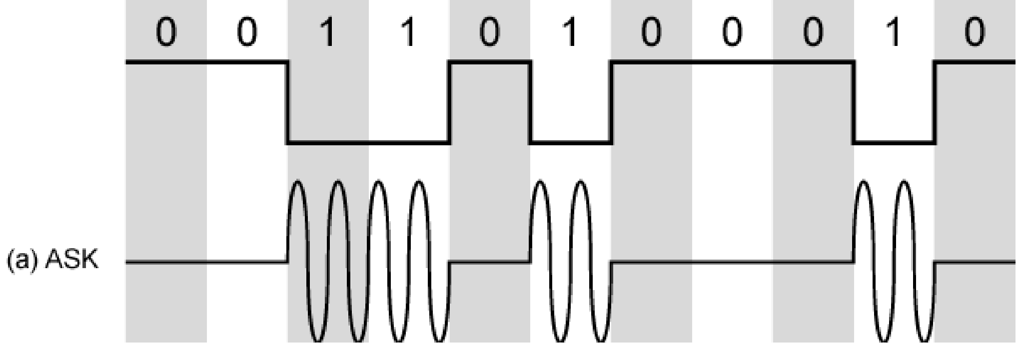
Modulation Techniques



- Have stated that modulation involves operation on one or more of the three characteristics of a carrier signal: amplitude, frequency, and phase. Accordingly, there are three basic encoding or modulation techniques for transforming digital data into analog signals, as illustrated in Stallings DCC8e Figure (above): amplitude shift keying (ASK), frequency shift keying (FSK), and phase shift keying (PSK). In all these cases, the resulting signal occupies a bandwidth centered on the carrier frequency.

Amplitude Shift Keying

- encode 0/1 by different carrier amplitudes
  - usually have one amplitude zero
- susceptible to sudden gain changes
- inefficient
- used for
  - up to 1200bps on voice grade lines
  - very high speeds over optical fiber

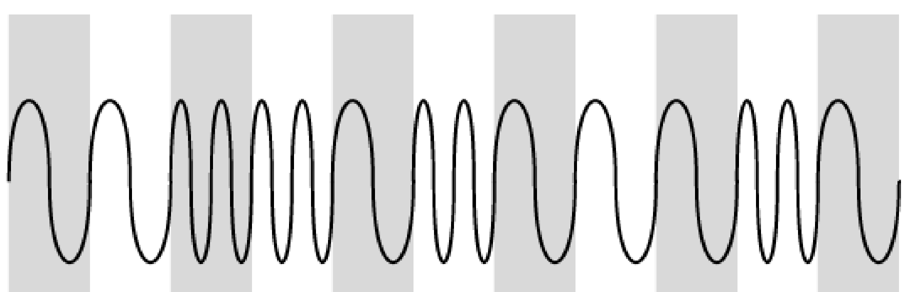


- In ASK, the two binary values are represented by two different amplitudes of the carrier frequency. Commonly, one of the amplitudes is zero; that is, one binary digit is represented by the presence, at constant amplitude, of the carrier, the other by the absence of the carrier, as shown in Stallings DCC8e Figure.
- ASK is susceptible to sudden gain changes and is a rather inefficient modulation technique. On voice-grade lines, it is typically used only up to 1200 bps.
- The ASK technique is used to transmit digital data over optical fiber, where one signal element is represented by a light pulse while the other signal element is represented by the absence of light.

Binary Frequency Shift Keying

- most common is binary FSK (BFSK)
- two binary values represented by two different frequencies (near carrier)
- less susceptible to error than ASK
- used for
  - up to 1200bps on voice grade lines
  - high frequency radio
  - even higher frequency on LANs using co-ax

(b) BFSK



- The most common form of FSK is binary FSK (BFSK), in which the two binary values are represented by two different frequencies near the carrier frequency, as shown in Stallings DCC8e Figure.
- BFSK is less susceptible to error than ASK. On voice-grade lines, it is typically used up to 1200 bps. It is also commonly used for high-frequency (3 to 30 MHz) radio transmission. It can also be used at even higher frequencies on local area networks that use coaxial cable.

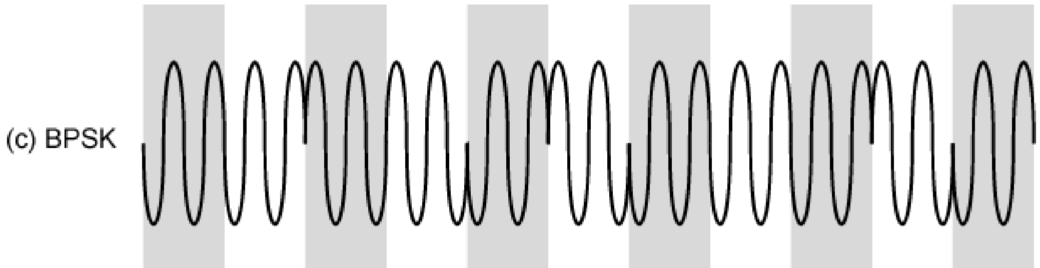
Multiple FSK

- each signalling element represents more than one bit
- more than two frequencies used
- more bandwidth efficient
- more prone to error

- A signal that is more bandwidth efficient, but also more susceptible to error, is multiple FSK (MFSK), in which more than two frequencies are used. In this case each signaling element represents more than one bit. To match the data rate of the input bit stream, each output signal element is held for a period of $T_s = LT$ seconds, where T is the bit period (data rate = $1/T$). Thus, one signal element, which is a constant-frequency tone, encodes L bits. The total bandwidth required is $2Mf_d$. It can be shown that the minimum frequency separation required is $2f_d = 1/T_s$. Therefore, the modulator requires a bandwidth of $W_d = 2Mf_d = M/T_s$.

Phase Shift Keying

- phase of carrier signal is shifted to represent data
- binary PSK
  - two phases represent two binary digits
- differential PSK
  - phase shifted relative to previous transmission rather than some reference signal



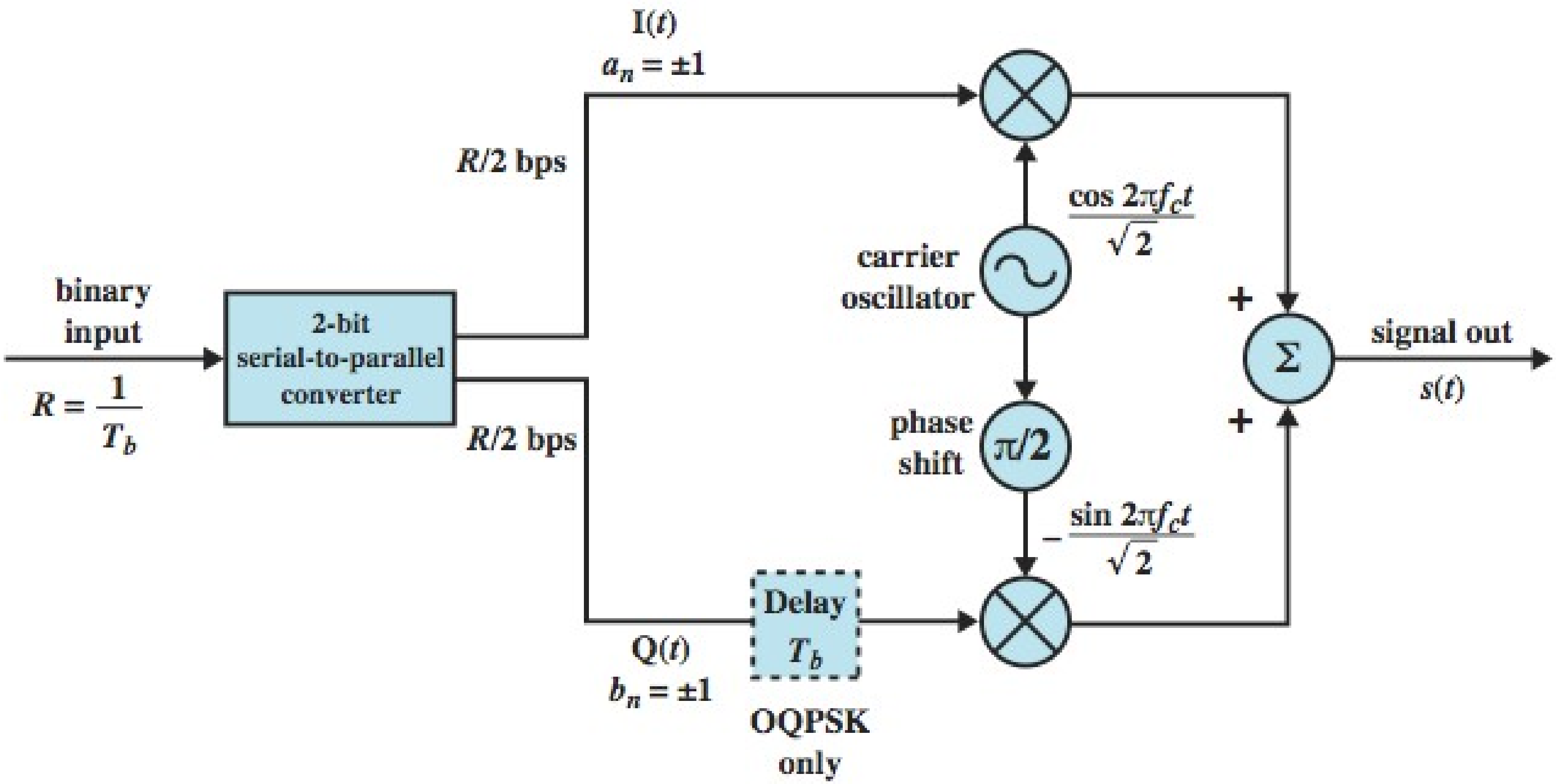
- In PSK, the phase of the carrier signal is shifted to represent data. The simplest scheme uses two phases to represent the two binary digits (Figure) and is known as binary phase shift keying.
- An alternative form of two-level PSK is differential PSK (DPSK). In this scheme, a binary 0 is represented by sending a signal burst of the same phase as the previous signal burst sent. A binary 1 is represented by sending a signal burst of opposite phase to the preceding one. This term *differential* refers to the fact that the phase shift is with reference to the previous bit transmitted rather than to some constant reference signal. In differential encoding, the information to be transmitted is represented in terms of the changes between successive data symbols rather than the signal elements themselves. DPSK avoids the requirement for an accurate local oscillator phase at the receiver that is matched with the transmitter. As long as the preceding phase is received correctly, the phase reference is accurate.

Quadrature PSK

- get more efficient use if each signal element represents more than one bit
  - eg. shifts of $\pi/2$ (90^0)
  - each element represents two bits
  - split input data stream in two & modulate onto carrier & phase shifted carrier
- can use 8 phase angles & more than one amplitude
  - 9600bps modem uses 12 angles, four of which have two amplitudes

- More efficient use of bandwidth can be achieved if each signaling element represents more than one bit. For example, instead of a phase shift of 180˚, as allowed in BPSK, a common encoding technique, known as quadrature phase shift keying (QPSK), uses phase shifts separated by multiples of $\pi/2$ (90˚). Thus each signal element represents two bits rather than one. The input is a stream of binary digits with a data rate of $R = 1/T_b$, where T_b is the width of each bit. This stream is converted into two separate bit streams of $R/2$ bps each, by taking alternate bits for the two streams. The two data streams are referred to as the I (in-phase) and Q (quadrature phase) streams. The streams are modulated on a carrier of frequency f_c by multiplying the bit stream by the carrier, and the carrier shifted by 90˚. The two modulated signals are then added together and transmitted. Thus, the combined signals have a symbol rate that is half the input bit rate.
- The use of multiple levels can be extended beyond taking bits two at a time. It is possible to transmit bits three at a time using eight different phase angles. Further, each angle can have more than one amplitude. For example, a standard 9600 bps modem uses 12 phase angles, four of which have two amplitude values, for a total of 16 different signal elements.

QPSK and OQPSK Modulators



- Stallings DCC8e Figure shows the QPSK modulation scheme in general terms. The input is a stream of binary digits with a data rate of $R = 1/T_b$, where T_b is the width of each bit. This stream is converted into two separate bit streams of $R/2$ bps each, by taking alternate bits for the two streams. The two data streams are referred to as the I (in-phase) and Q (quadrature phase) streams. The streams are modulated on a carrier of frequency f_c by multiplying the bit stream by the carrier, and the carrier shifted by 90˚. The two modulated signals are then added together and transmitted. Thus, the combined signals have a symbol rate that is half the input bit rate.
- This figure also shows the variation of QPSK known as offset QPSK (OQPSK), or orthogonal QPSK. The difference is that a delay of one bit time is introduced in the Q stream. Because OQPSK differs from QPSK only by the delay in the Q stream, its spectral characteristics and bit error performance are the same as that of QPSK.

Performance of Digital to Analog Modulation Schemes

- bandwidth
  - ASK/PSK bandwidth directly relates to bit rate
  - multilevel PSK gives significant improvements
- in presence of noise:
  - bit error rate of PSK and QPSK are about 3dB superior to ASK and FSK
  - for MFSK & MPSK have tradeoff between bandwidth efficiency and error performance

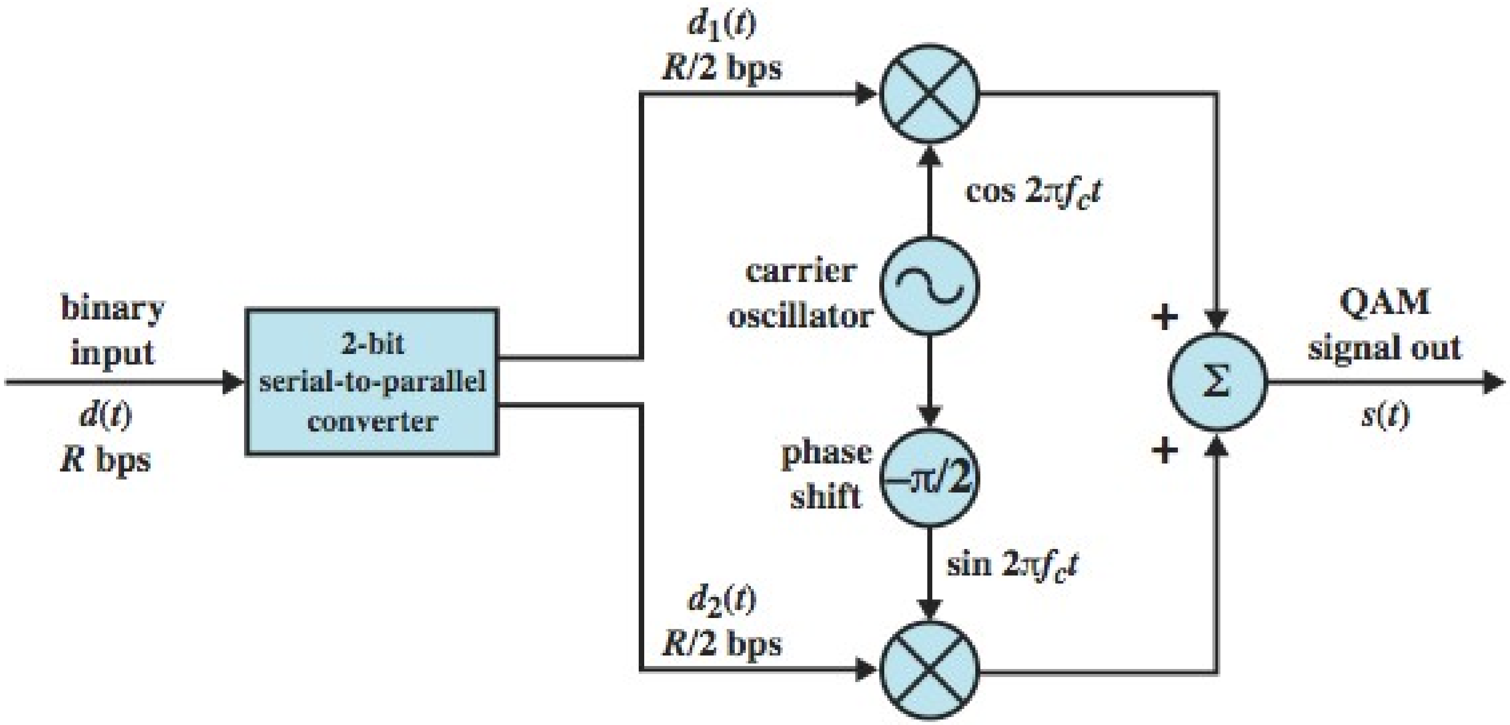
- In looking at the performance of various digital-to-analog modulation schemes, the first parameter of interest is the bandwidth of the modulated signal. This depends on a variety of factors, including the definition of bandwidth used and the filtering technique used to create the bandpass signal. For ASK & PSK the bandwidth is directly related to the bit rate. With multilevel PSK (MPSK), significant improvements in bandwidth can be achieved.
- Nothing has yet been said of performance in the presence of noise. Stallings DCC8e Figure shows the bit error rate plotted as a function of the ratio E_b/N_0. As that ratio increases, the bit error rate drops. Further, DPSK and BPSK are about 3 dB superior to ASK and BFSK. Stallings DCC8e Figure shows the same information for various levels of *M* for MFSK and MPSK. There is an important difference. For MFSK, the error probability for a given value E_b/N_0 of decreases as *M* increases, while the opposite is true for MPSK. On the other hand, comparing Equations (5.10) and (5.11), the bandwidth efficiency of MFSK decrease as M increases, while the opposite is true of MPSK. Thus, in both cases, there is a tradeoff between bandwidth efficiency and error performance: an increase in bandwidth efficiency results in an increase in error probability.

Quadrature Amplitude Modulation

- QAM used on asymmetric digital subscriber line (ADSL) and some wireless
- combination of ASK and PSK
- logical extension of QPSK
- send two different signals simultaneously on same carrier frequency
  - use two copies of carrier, one shifted 90°
  - each carrier is ASK modulated
  - two independent signals over same medium
  - demodulate and combine for original binary output

- Quadrature amplitude modulation (QAM) is a popular analog signaling technique that is used in the asymmetric digital subscriber line (ADSL), described in Chapter 8, and in some wireless standards. This modulation technique is a combination of ASK and PSK. QAM can also be considered a logical extension of QPSK. QAM takes advantage of the fact that it is possible to send two different signals simultaneously on the same carrier frequency, by using two copies of the carrier frequency, one shifted by 90° with respect to the other. For QAM, each carrier is ASK modulated. The two independent signals are simultaneously transmitted over the same medium. At the receiver, the two signals are demodulated and the results combined to produce the original binary input.

QAM Modulator



- Stallings DCC8e Figure shows the QAM modulation scheme in general terms. The input is a stream of binary digits arriving at a rate of *R* bps. This stream is converted into two separate bit streams of *R*/2 bps each, by taking alternate bits for the two streams. In the diagram, the upper stream is ASK modulated on a carrier of frequency f_c by multiplying the bit stream by the carrier. Thus, a binary zero is represented by the absence of the carrier wave and a binary one is represented by the presence of the carrier wave at a constant amplitude. This same carrier wave is shifted by 90˚ and used for ASK modulation of the lower binary stream. The two modulated signals are then added together and transmitted.

QAM Variants

- two level ASK
  - each of two streams in one of two states
  - four state system
  - essentially QPSK
- four level ASK
  - combined stream in one of 16 states
- have 64 and 256 state systems
- improved data rate for given bandwidth
  - but increased potential error rate

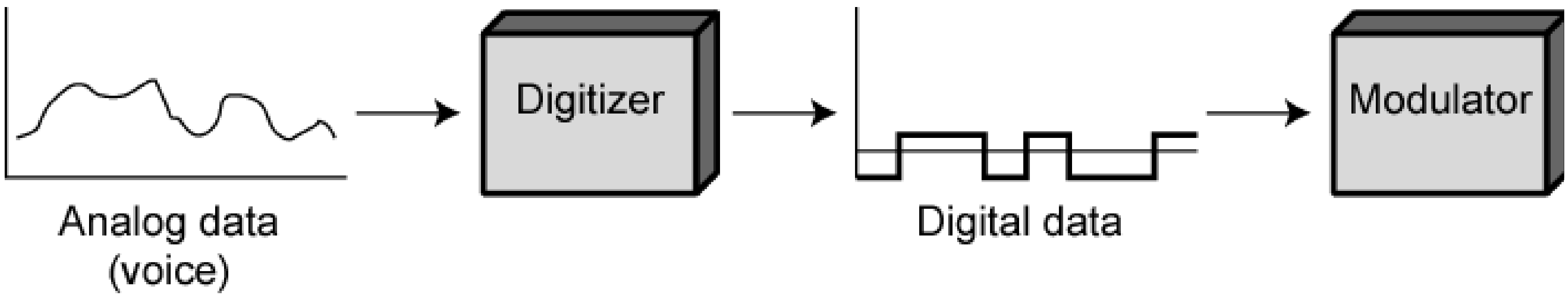
- If two-level ASK is used, then each of the two streams can be in one of two states and the combined stream can be in one of $4 = 2 \times 2$ states. This is essentially QPSK. If four-level ASK is used (i.e., four different amplitude levels), then the combined stream can be in one of $16 = 4 \times 4$ states. Systems using 64 and even 256 states have been implemented. The greater the number of states, the higher the data rate that is possible within a given bandwidth. Of course, as discussed previously, the greater the number of states, the higher the potential error rate due to noise and attenuation.

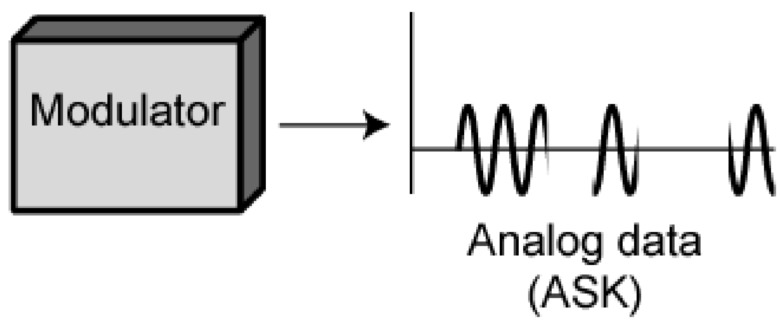
Analog Data, Digital Signal

- digitization is conversion of analog data into digital data which can then:
  - be transmitted using NRZ-L
  - be transmitted using code other than NRZ-L
  - be converted to analog signal
- analog to digital conversion done using a codec
  - pulse code modulation
  - delta modulation

- In this section we examine the process of transforming analog data into digital signals. Analog data, such as voice and video, is often digitized to be able to use digital transmission facilities. Strictly speaking, it might be more correct to refer to this as a process of converting analog data into digital data; this process is known as digitization. Once analog data have been converted into digital data, a number of things can happen. The three most common are:
- **1.** The digital data can be transmitted using NRZ-L. In this case, we have in fact gone directly from analog data to a digital signal.
- **2.** The digital data can be encoded as a digital signal using a code other than NRZ-L. Thus an extra step is required.
- **3.** The digital data can be converted into an analog signal, using one of the modulation techniques discussed in Section 5.2.
- The device used for converting analog data into digital form for transmission, and subsequently recovering the original analog data from the digital, is known as a **codec** (coder-decoder). In this section we examine the two principal techniques used in codecs, pulse code modulation and delta modulation.

Digitizing Analog Data





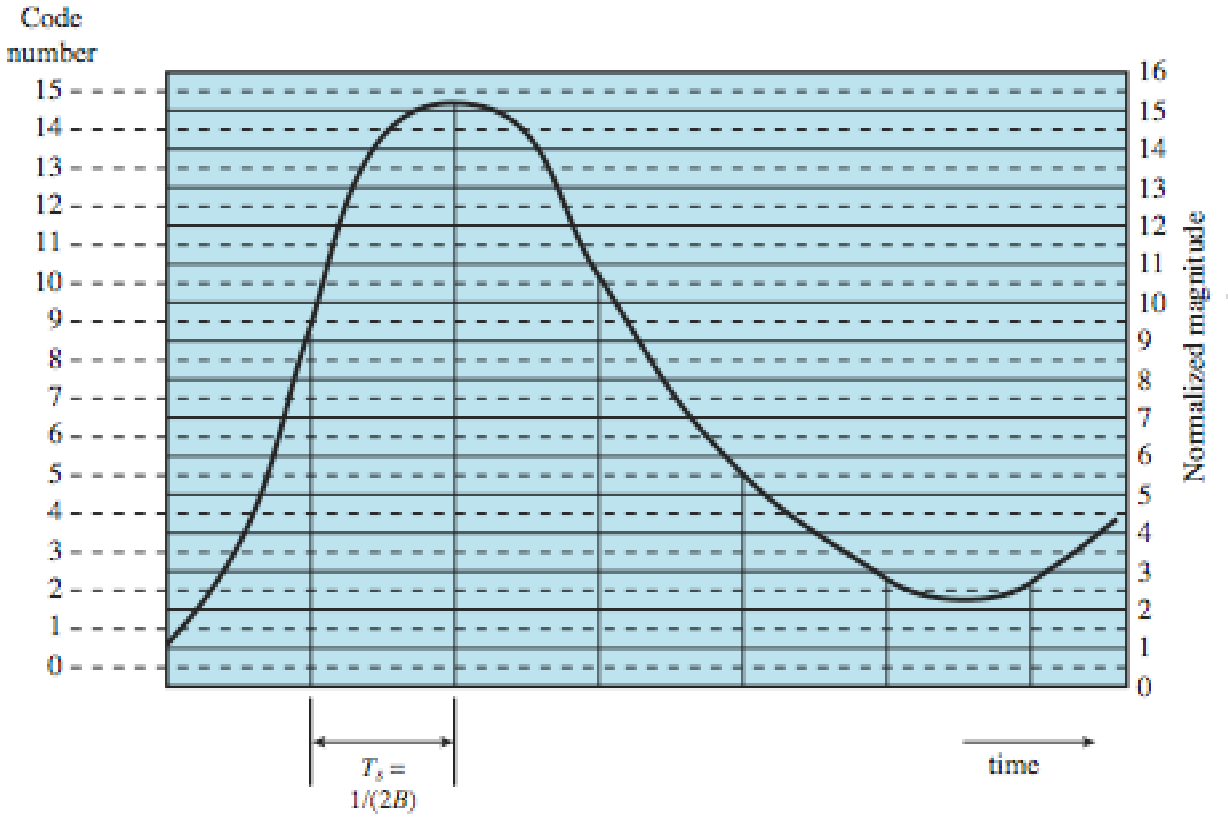
- Stallings DCC8e Figure illustrates the 3rd alternative, which shows voice data that are digitized and then converted to an analog ASK signal. This allows digital transmission in the sense defined in Chapter 3. The voice data, because they have been digitized, can be treated as digital data, even though transmission requirements (eg. use of microwave) dictate that an analog signal be used.

Pulse Code Modulation (PCM)

- sampling theorem:
  - “If a signal is sampled at regular intervals at a rate higher than twice the highest signal frequency, the samples contain all information in original signal”
  - eg. 4000Hz voice data, requires 8000 sample per sec
- strictly have analog samples
  - Pulse Amplitude Modulation (PAM)
- so assign each a digital value

- The simplest technique for transforming analog data into digital signals is pulse code modulation (PCM), which involves sampling the analog data periodically and quantizing the samples. Pulse code modulation (PCM) is based on the sampling theorem (quoted above). Hence if voice data is limited to frequencies below 4000 Hz (a conservative procedure for intelligibility), 8000 samples per second would be sufficient to characterize the voice signal completely. Note, however, that these are analog samples, called **pulse amplitude modulation (PAM)** samples. To convert to digital, each of these analog samples must be assigned a binary code.

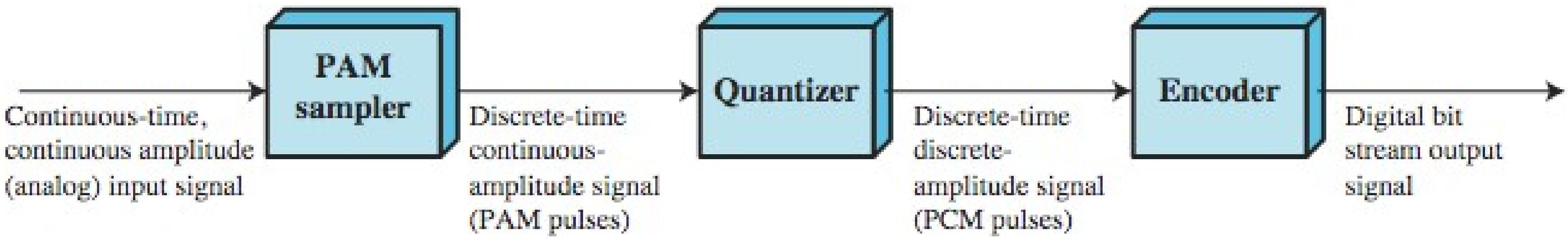
PCM Example



PAM value	1.1	9.2	15.2	10.8	5.6	2.8	2.7
quantized code number	1	9	15	10	5	2	2
PCM code	0001	1001	1111	1010	0101	0010	0010

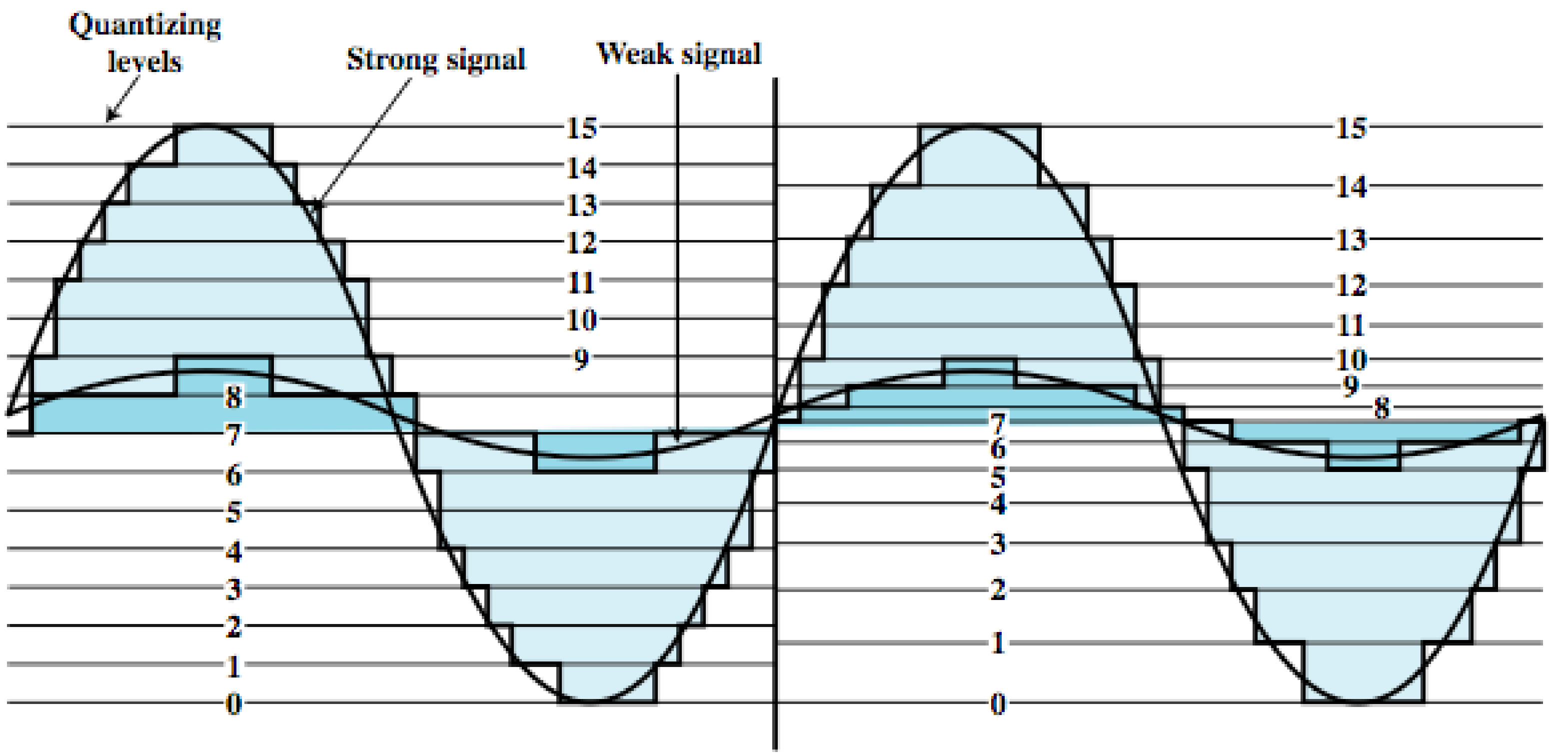
- Stallings DCC8e Figure shows an example in which the original signal is assumed to be bandlimited with a bandwidth of *B*. PAM samples are taken at a rate of 2*B*, or once every $T_s = 1/2B$ seconds. Each PAM sample is approximated by being *quantized* into one of 16 different levels. Each sample can then be represented by 4 bits. But because the quantized values are only approximations, it is impossible to recover the original signal exactly. By using an 8-bit sample, which allows 256 quantizing levels, the quality of the recovered voice signal is comparable with that achieved via analog transmission. Note that this implies that a data rate of 8000 samples per second $\times$ 8 bits per sample = 64 kbps is needed for a single voice signal.

PCM Block Diagram



- Thus, PCM starts with a continuous-time, continuous-amplitude (analog) signal, from which a digital signal is produced, as shown in Stallings DCC8e Figure. The digital signal consists of blocks of *n* bits, where each *n*-bit number is the amplitude of a PCM pulse. On reception, the process is reversed to reproduce the analog signal. Notice, however, that this process violates the terms of the sampling theorem. By quantizing the PAM pulse, the original signal is now only approximated and cannot be recovered exactly. This effect is known as **quantizing error** or **quantizing noise**. Each additional bit used for quantizing increases SNR by about 6 dB, which is a factor of 4.

Non-Linear Coding

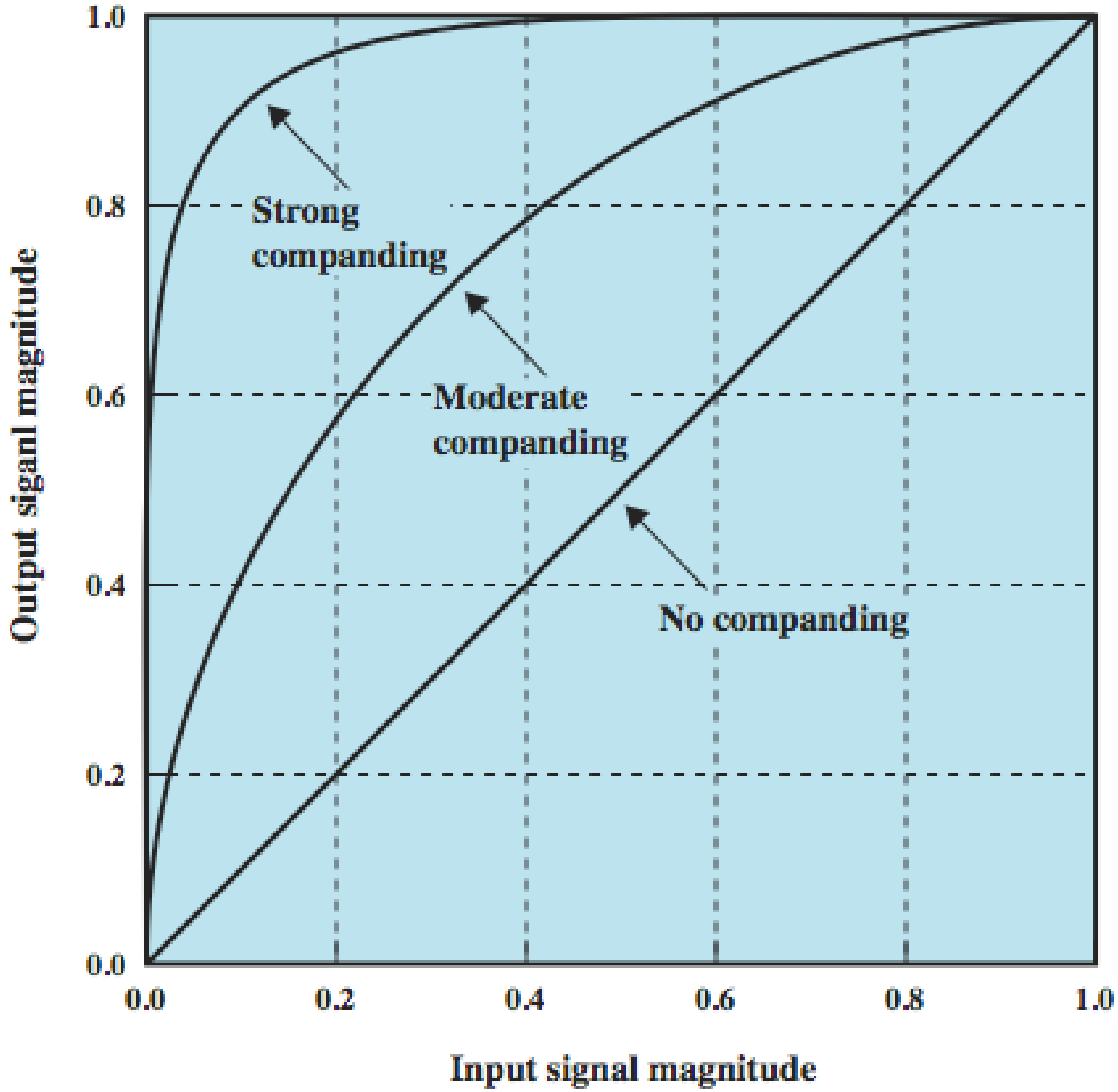


(a) Without nonlinear encoding

(b) With nonlinear encoding

- Typically, the PCM scheme is refined using a technique known as nonlinear encoding, which means, in effect, that the quantization levels are not equally spaced. The problem with equal spacing is that the mean absolute error for each sample is the same, regardless of signal level. Consequently, lower amplitude values are relatively more distorted. By using a greater number of quantizing steps for signals of low amplitude, and a smaller number of quantizing steps for signals of large amplitude, a marked reduction in overall signal distortion is achieved, as shown in Stallings DCC8e Figure.
- Nonlinear encoding can significantly improve the PCM SNR ratio. For voice signals, improvements of 24 to 30 dB have been achieved.

Companding



1.0
0.8
0.6
0.4
0.2
0.0
Output siganl magnitude
Strong
companding
Moderate
companding
No companding
0.0
0.2
0.4
0.6
0.8
1.0
Input signal magnitude

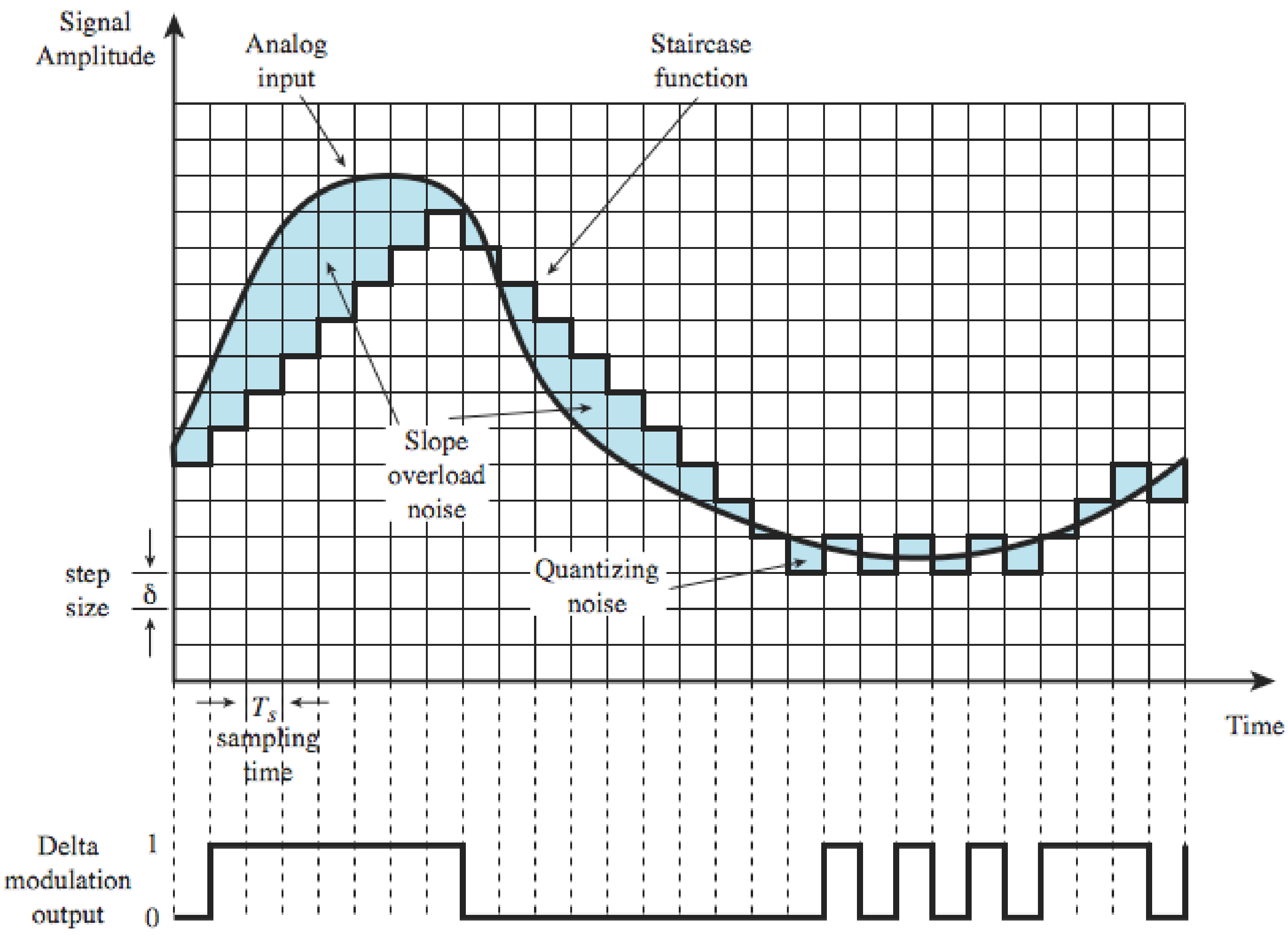
- The same effect can be achieved by using uniform quantizing but companding (compressing-expanding) the input analog signal. Companding is a process that compresses the intensity range of a signal by imparting more gain to weak signals than to strong signals on input. At output, the reverse operation is performed. Stallings DCC8e Figure shows typical companding functions. Note that the effect on the input side is to compress the sample so that the higher values are reduced with respect to the lower values. Thus, with a fixed number of quantizing levels, more levels are available for lower-level signals. On the output side, the compander expands the samples so the compressed values are restored to their original values.

Delta Modulation

- analog input is approximated by a staircase function
  - can move up or down one level (δ) at each sample interval
- has binary behavior
  - since function only moves up or down at each sample interval
  - hence can encode each sample as single bit
  - 1 for up or 0 for down

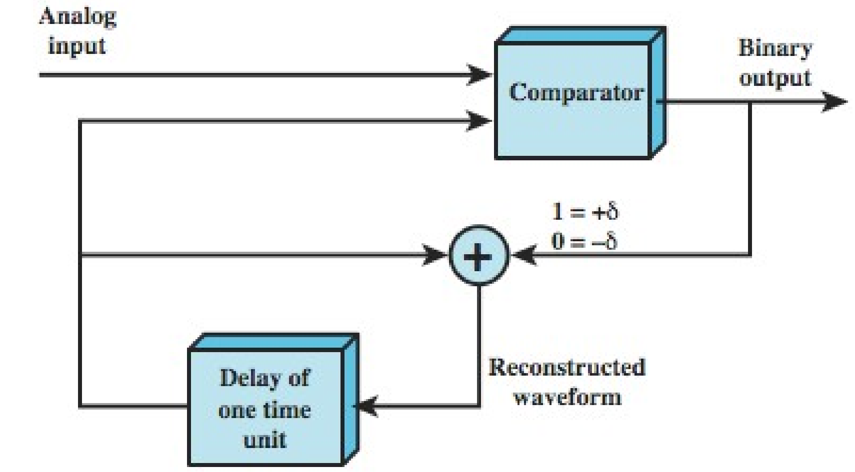
- A variety of techniques have been used to improve the performance of PCM or to reduce its complexity. One of the most popular alternatives to PCM is delta modulation (DM). With delta modulation, an analog input is approximated by a staircase function that moves up or down by one quantization level (δ) at each sampling interval (T_s). The important characteristic of this staircase function is that its behavior is binary: At each sampling time, the function moves up or down a constant amount δ. Thus, the output of the delta modulation process can be represented as a single binary digit for each sample. In essence, a bit stream is produced by approximating the derivative of an analog signal rather than its amplitude: A 1 is generated if the staircase function is to go up during the next interval; a 0 is generated otherwise.

Delta Modulation Example

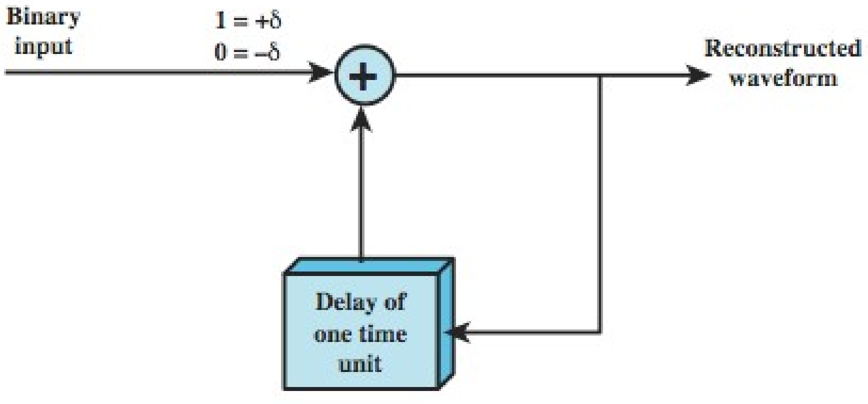


- Stallings DCC8e Figure shows an example where the staircase function is overlaid on the original analog waveform. A 1 is generated if the staircase function is to go up during the next interval; a 0 is generated otherwise. The transition (up or down) that occurs at each sampling interval is chosen so that the staircase function tracks the original analog waveform as closely as possible.
- There are two important parameters in a DM scheme: the size of the step assigned to each binary digit, δ, and the sampling rate. As the above figure illustrates, δ must be chosen to produce a balance between two types of errors or noise. When the analog waveform is changing very slowly, there will be quantizing noise. This noise increases as δ is increased. On the other hand, when the analog waveform is changing more rapidly than the staircase can follow, there is slope overload noise. This noise increases as δ is decreased. It should be clear that the accuracy of the scheme can be improved by increasing the sampling rate. However, this increases the data rate of the output signal.

Delta Modulation Operation



(a) Transmission



(b) Reception

- Stallings DCC8e Figure illustrates the logic of the process, which is essentially a feedback mechanism. For transmission, the following occurs: At each sampling time, the analog input is compared to the most recent value of the approximating staircase function. If the value of the sampled waveform exceeds that of the staircase function, a 1 is generated; otherwise, a 0 is generated. Thus, the staircase is always changed in the direction of the input signal. The output of the DM process is therefore a binary sequence that can be used at the receiver to reconstruct the staircase function. The staircase function can then be smoothed by some type of integration process or by passing it through a lowpass filter to produce an analog approximation of the analog input signal.

PCM verses Delta Modulation

- DM has simplicity compared to PCM
- but has worse SNR
- issue of bandwidth used
  - eg. for good voice reproduction with PCM
    - want 128 levels (7 bit) & voice bandwidth 4khz
    - need 8000 x 7 = 56kbps
- data compression can improve on this
- still growing demand for digital signals
  - use of repeaters, TDM, efficient switching
- PCM preferred to DM for analog signals

- The principal advantage of DM over PCM is the simplicity of its implementation. In general, PCM exhibits better SNR characteristics at the same data rate. Good voice reproduction via PCM can be achieved with 128 quantization levels, or 7-bit coding ($2^7 = 128$). A voice signal, conservatively, occupies a bandwidth of 4 kHz. Thus, according to the sampling theorem, samples should be taken at a rate of 8000 samples per second. This implies a data rate of $8000 \times 7 = 56$ kbps for the PCM-encoded digital data. But using the Nyquist criterion from Chapter 3, this digital signal could require on the order of 28 kHz of bandwidth. Even more severe differences are seen with higher bandwidth signals. For example, a common PCM scheme for color television uses 10-bit codes, which works out to 92 Mbps for a 4.6-MHz bandwidth signal. In spite of these numbers, digital techniques continue to grow in popularity for transmitting analog data. The principal reasons for this are
- • Because repeaters are used instead of amplifiers, there is no cumulative noise.
- • use time division multiplexing (TDM) for digital signals with no intermodulation noise, verses of the frequency division multiplexing (FDM) used for analog signals.
- • The conversion to digital signaling allows the use of the more efficient digital switching techniques.
- Furthermore, techniques have been developed to provide more efficient codes.
- Studies also show that PCM-related techniques are preferable to DM-related techniques for digitizing analog signals that represent digital data.

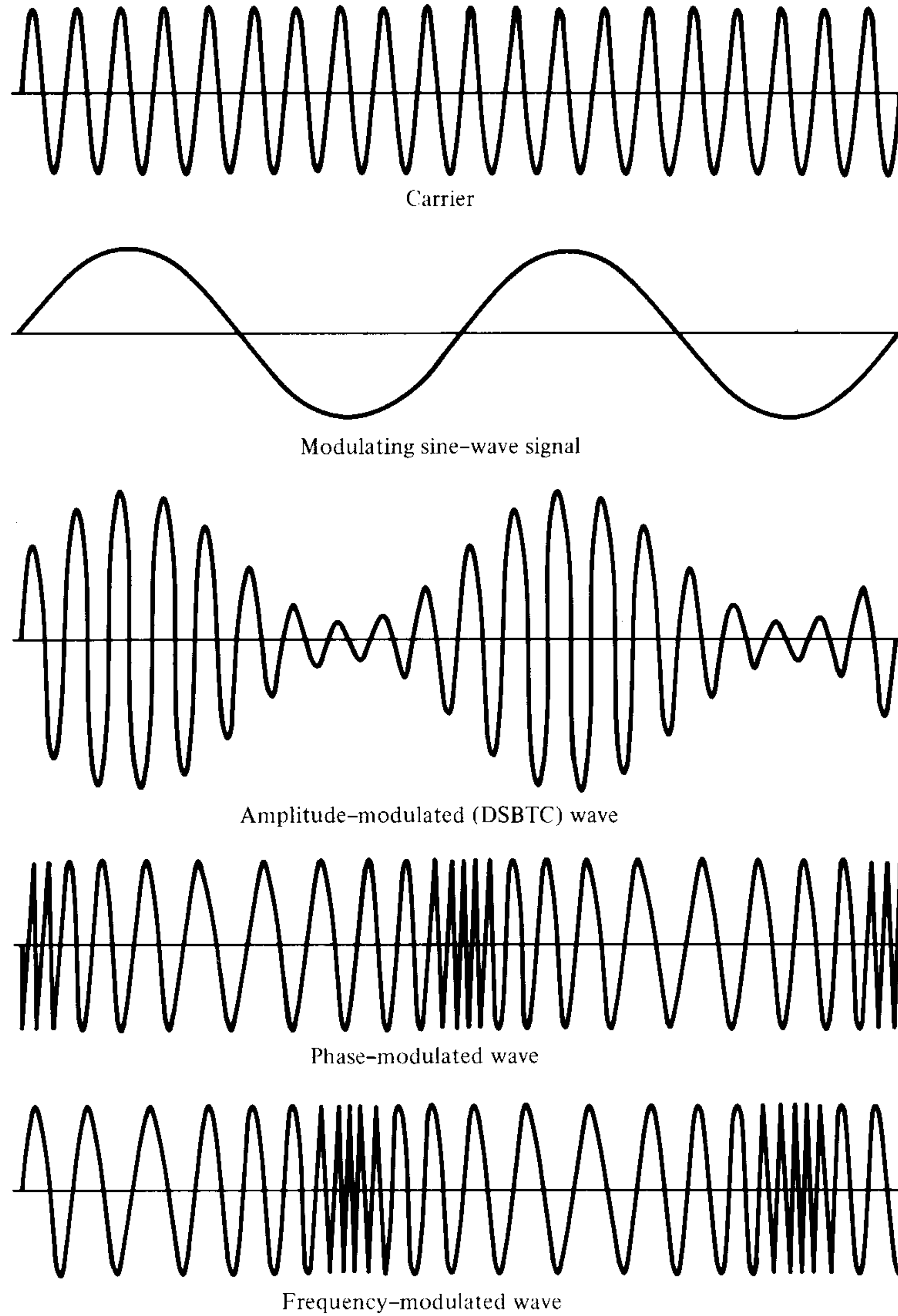
Analog Data, Analog Signals

- modulate carrier frequency with analog data
- why modulate analog signals?
  - higher frequency can give more efficient transmission
  - permits frequency division multiplexing (chapter 8)
- types of modulation
  - Amplitude
  - Frequency
  - Phase

- Analog data can be modulated by a carrier frequency to produce an analog signal in a different frequency band, which can be utilized on an analog transmission system. The basic techniques are amplitude modulation (AM), frequency modulation (FM), and phase modulation (PM).
- Modulation has been defined as the process of combining an input signal $m(t)$ and a carrier at frequency f_c to produce a signal $s(t)$ whose bandwidth is (usually) centered on f_c. For digital data, the motivation for modulation should be clear: When only analog transmission facilities are available, modulation is required to convert the digital data to analog form. The motivation when the data are already analog is less clear. After all, voice signals are transmitted over telephone lines at their original spectrum (referred to as baseband transmission). There are two principal reasons for analog modulation of analog signals:
  - A higher frequency may be needed for effective transmission, since for unguided transmission, it is virtually impossible to transmit baseband signals;
  - Modulation permits frequency division multiplexing, an important technique explored in Chapter 8.
- In this section we look at the principal techniques for modulation using analog data: amplitude modulation (AM), frequency modulation (FM), and phase modulation (PM). As before, the three basic characteristics of a signal are used for modulation.

Analog Modulation Techniques

- Amplitude Modulation
- Frequency Modulation
- Phase Modulation



- Amplitude modulation (AM) is the simplest form of modulation, and involves the multiplication of the input signal by the carrier f_c. The spectrum consists of the original carrier plus the spectrum of the input signal translated to f_c. The portion of the spectrum for $|f| > |f_c|$ is the *upper sideband*, and the portion of the spectrum for $|f| < |f_c|$ is *lower sideband*. Both the upper and lower sidebands are replicas of the original spectrum $M(f)$, with the lower sideband being frequency reversed. A popular variant of AM, known as single sideband (SSB), takes advantage of this fact by sending only one of the sidebands, eliminating the other sideband and the carrier.
- Frequency modulation (FM) and phase modulation (PM) are special cases of angle modulation. For phase modulation, the phase is proportional to the modulating signal. For frequency modulation, the derivative of the phase is proportional to the modulating signal. As with AM, both FM and PM result in a signal whose bandwidth is centered at f_c, but can show that the magnitude of that bandwidth is very different, hence both FM and PM require greater bandwidth than AM.
- Stallings DCC8e Figure illustrates these various techniques showing amplitude, phase, and frequency modulation by a sine wave. The shapes of the FM and PM signals are very similar. Indeed, it is impossible to tell them apart without knowledge of the modulation function.