Slide 1

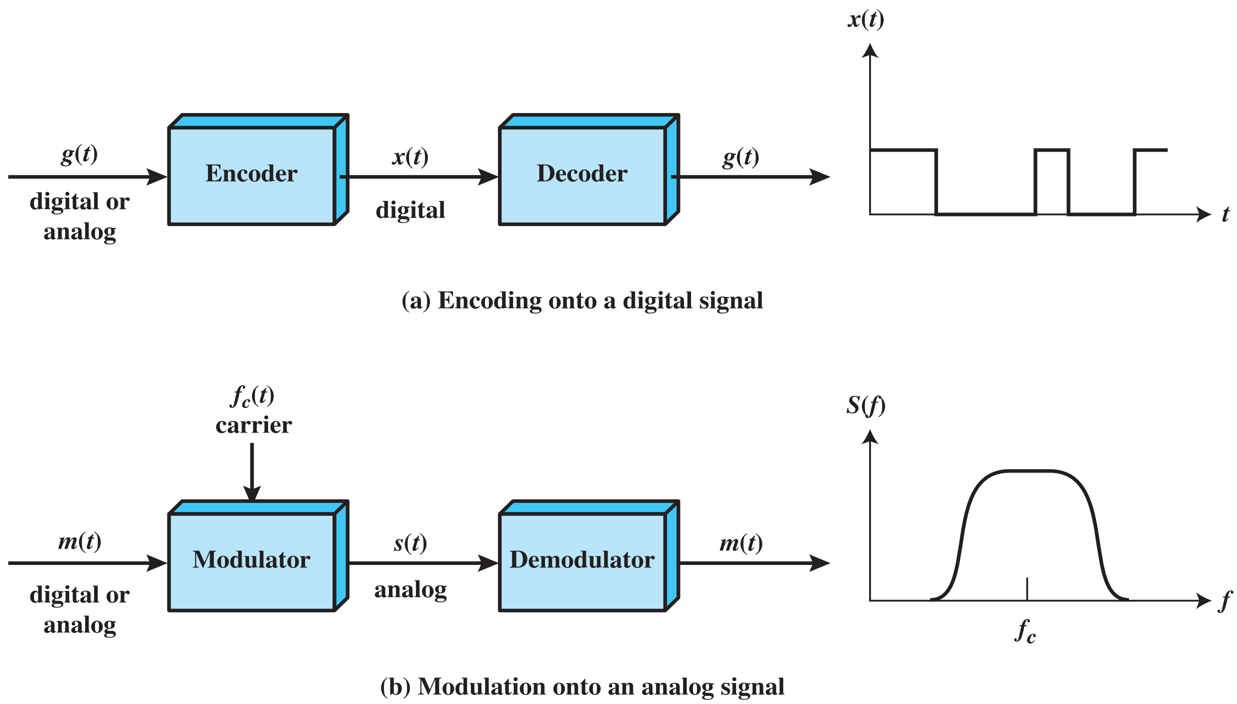


Figure 5.1 Encoding and Modulation Techniques

In Chapter 3 a distinction was made between analog and digital data and analog and digital signals. Figure 3.14 suggested that either form of data could be encoded into either form of signal.

Figure 5.1 is another depiction that emphasizes the process involved. For digital signaling, a data source $g(t)$, which may be either digital or analog, is encoded into a digital signal $x(t)$. The actual form of $x(t)$ depends on the encoding technique and is chosen to optimize use of the transmission medium. For example, the encoding may be chosen to conserve bandwidth or to minimize errors.

The basis for analog signaling is a continuous constant-frequency signal known as the carrier signal. The frequency of the carrier signal is chosen to be compatible with the transmission medium being used. Data may be transmitted using a carrier signal by modulation. Modulation is the process of encoding source data onto a carrier signal with frequency f_c. 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 or baseband signal. The result of modulating the carrier signal is called the modulated signal $s(t)$. As Figure 5.1b indicates, $s(t)$ is a bandlimited (bandpass) signal. The location of the bandwidth on the spectrum is related to f_c and is often centered on f_c. Again, the actual form of the encoding is chosen to optimize some characteristic of the transmission.

Each of the four possible combinations depicted in Figure 5.1 is in widespread use. The reasons for choosing a particular combination for any given communication task vary. We list here some representative reasons:

Digital data, digital signal: In general, the equipment for encoding digital data into a digital signal is less complex and less expensive than digital-to-analog modulation equipment.

Analog data, digital signal: Conversion of analog data to digital form permits the use of modern digital transmission and switching equipment. The advantages of the digital approach were outlined in Section 3.2.

Digital data, analog signal: Some transmission media, such as optical fiber and unguided media, will only propagate analog signals.

Analog data, analog signal: Analog data in electrical form can be transmitted as baseband signals easily and cheaply. This is done with voice transmission over voice-grade lines. One common use of modulation is to shift the bandwidth of a baseband signal to another portion of the spectrum. In this way multiple signals, each at a different position on the spectrum, can share the same transmission medium. This is known as frequency division multiplexing.

Digital Data, Digital Signal

- Digital signal
  - Sequence of discrete, discontinuous voltage pulses
  - Each pulse is a signal element
  - Binary data are transmitted by encoding each data bit into signal elements

A digital signal is a sequence of discrete, discontinuous voltage pulses. 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. An example is shown in Figure 3.16, in which binary 1 is represented by a lower voltage level and binary 0 by a higher voltage level. We show in this section that a variety of other encoding schemes are also used.

Terminology

- **Unipolar** – all signal elements have the same sign
- **Polar** – one logic state represented by positive voltage and the other by negative voltage
- **Data rate** – rate, in bits per second that data are transmitted
- **Duration or length of a bit** – time taken for transmitter to emit the bit = 1/Data rate.
- **Modulation rate** – rate at which the signal level is changed; the rate is expressed in baud, which means signal elements per second
- **Mark and space** – refer to the binary digits 1 and 0
  (1) (0)

First, we define some terms. If the signal elements all have the same algebraic sign, that is, all positive or negative, then the signal is unipolar. In polar signaling, one logic state is represented by a positive voltage level, and the other by a negative voltage level. The data signaling rate, or just data rate, of a signal is the rate, in bits per second, that data are transmitted. The duration or length of a bit is the amount of time it takes for the transmitter to emit the bit; for a data rate R, the bit duration is $1/R$. The modulation rate, in contrast, is the rate at which the signal level is changed. This will depend on the nature of the digital encoding, as explained later. The modulation rate is expressed in baud, which means signal elements per second. Finally, the terms mark and space, for historical reasons, refer to the binary digits 1 and 0, respectively.

Slide 4

Term	Units	Definition
Data element	Bits	A single binary one or zero
Data rate	Bits per second (bps)	The rate at which data elements are transmitted
Signal element	Digital: a voltage pulse of constant amplitude Analog: a pulse of constant frequency, phase, and amplitude	That part of a signal that occupies the shortest interval of a signaling code
Signaling rate or modulation rate	Signal elements per second (baud)	The rate at which signal elements are transmitted

Table 5.1 summarizes key terms; these should be clearer when we see an example later in this section.

Slide 5

Interpreting Signals

Tasks involved in interpreting digital signal at the receiver:

- Timing of bits - when they start and end
- Signal levels

Factors affecting signal interpretation:

- Signal to noise ratio
- Data rate
- Bandwidth

The tasks involved in interpreting digital signals at the receiver can be summarized by again referring to Figure 3.16. First, the receiver must know the timing of each bit. That is, the receiver must know 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). In Figure 3.16, these tasks are 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 shown.

What factors determine how successful the receiver will be in interpreting the incoming signal? We saw in Chapter 3 that 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.

Slide 6

Nonreturn to Zero-Level (NRZ-L) 0 = high level 1 = low level **Nonreturn to Zero Inverted (NRZI)** 0 = no transition at beginning of interval (one bit time) 1 = transition at beginning of interval **Bipolar-AMI** 0 = no line signal 1 = positive or negative level, alternating for successive ones **Pseudoternary** 0 = positive or negative level, alternating for successive zeros 1 = no line signal **Manchester** 0 = transition from high to low in middle of interval 1 = transition from low to high in middle of interval **Differential Manchester** Always a transition in middle of interval 0 = transition at beginning of interval 1 = no transition at beginning of interval **B8ZS** Same as bipolar AMI, except that any string of eight zeros is replaced by a string with two code violations **HDB3** Same as bipolar AMI, except that any string of four zeros is replaced by a string with one code violation	Table 5.2 Definition of Digital Signal Encoding Formats (This table can be found on page 153 in the textbook)

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; they are defined in Table 5.2

Q: What is the strenght & weakness?

Slide 7

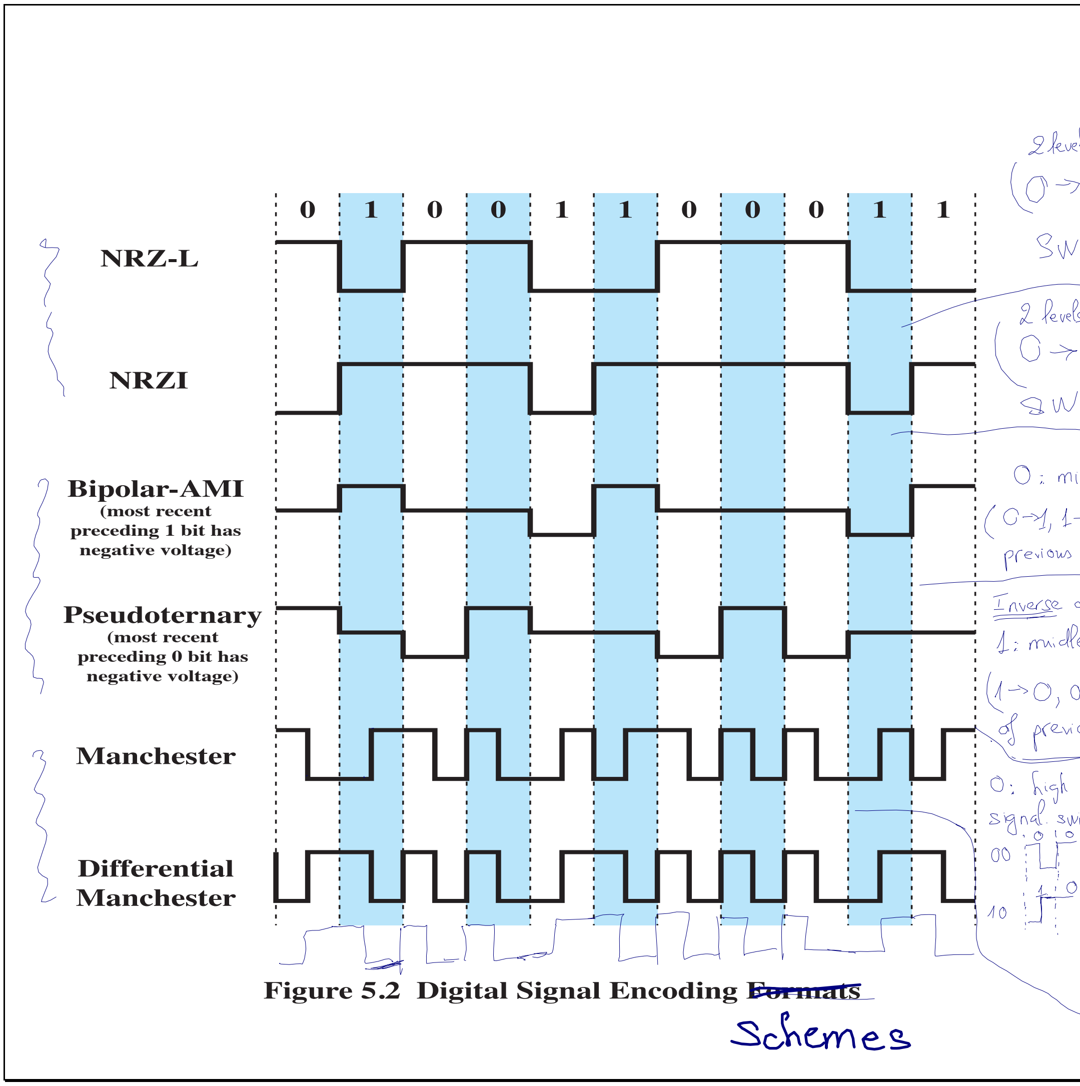


Figure 5.2 Digital Signal Encoding ~~Formats~~ Schemes

2 levels: 0 1
(0→1, 1→0)
switch.

2 levels
(0→1, 1→1)
switch.

0: middle level; 1: posi/neg
(0→1, 1→1): switch & inverse of previous switch.

Inverse of Bipolar AMI
1: middle level; 0: posi/neg.
(1→0, 0→0): switch & inverse of previous switch.

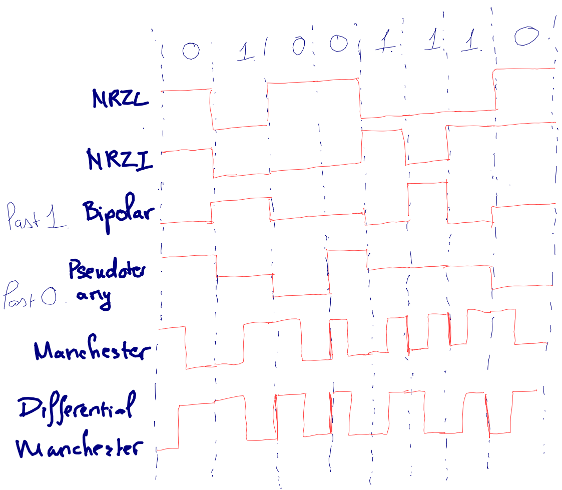
0: high 1: low
signal. switch level at the middle,
00 01 and base on its current position (after switch) v/s.
10 11 next value & do the next switch.
(The 1st half of the pulse indicates current signal value)

Figure 5.2 shows the signal encoding for the binary sequence 01001100011 using six different signal encoding schemes.

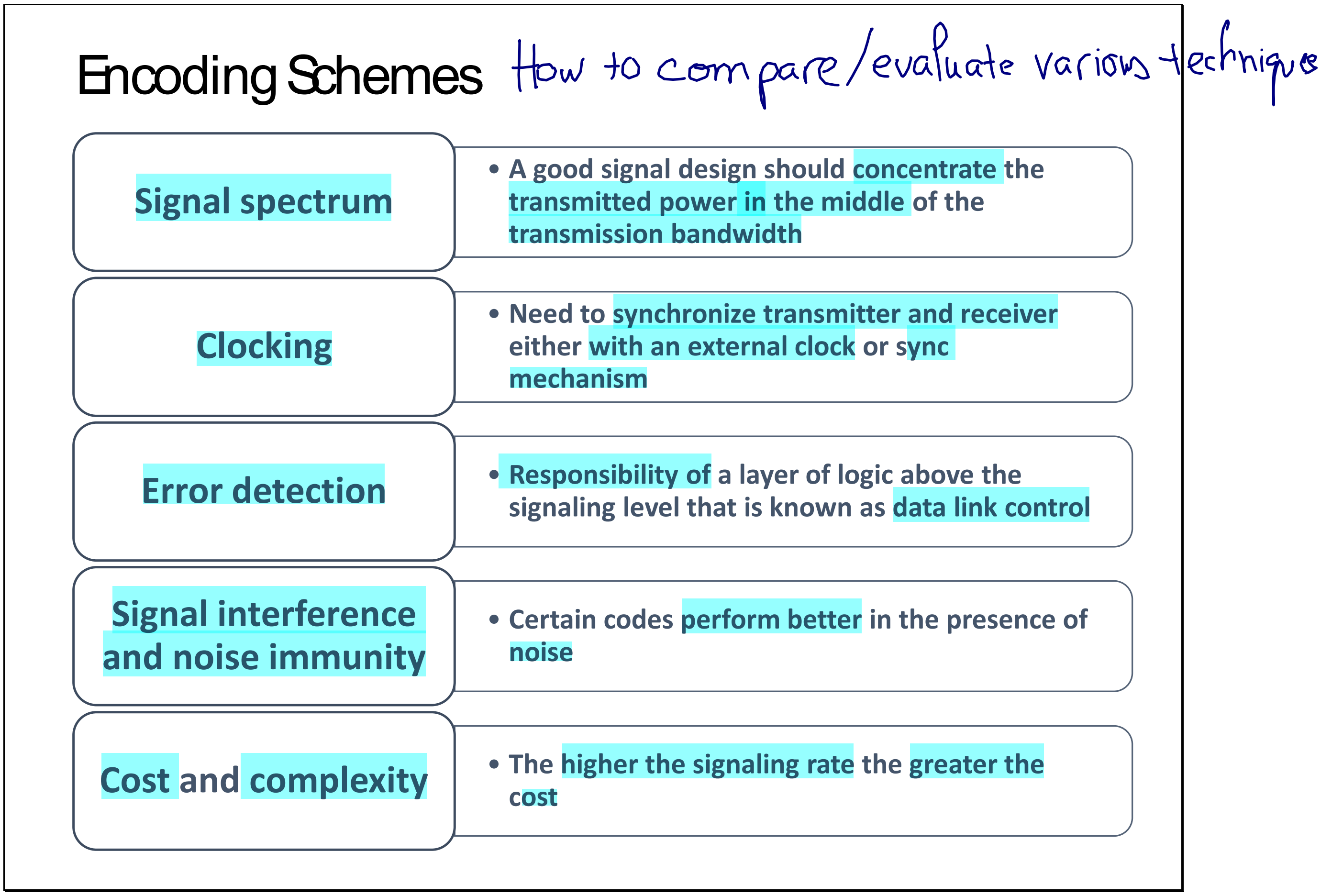
NRZ-L
NRZI: got 1 then change

3 levels
Bipolar AMI: 0 (middle), got 1 then change positive/negative v/s the last 1
Pseudoternary: 1 (middle), got 0 then change positive/negative v/s last 0

switch at the middle
Manchester:
Differential Manchester: 0: start with / follow last 1. 1: inverse last 1 (or 0).



Slide 8



Before describing these techniques, let us consider the following ways of evaluating or comparing the various techniques.

Signal spectrum: Several aspects of the signal spectrum are important. A lack of high-frequency components means that less bandwidth is required for transmission. In addition, lack of a direct-current (dc) component is also desirable. With a dc component to the signal, there must be direct physical attachment of transmission components. With no dc component, ac coupling via transformer is possible; this provides excellent electrical isolation, reducing interference. Finally, the magnitude of the effects of signal distortion and interference depend on the spectral properties of the transmitted signal. In practice, it usually happens that the transmission characteristics of a channel are worse near the band edges. Therefore, a good signal design should concentrate the transmitted power in the middle of the transmission bandwidth. In such a case, a smaller distortion should be present in the received signal. To meet this objective, codes can be designed with the aim of shaping the spectrum of the transmitted signal.

Clocking: We mentioned the need to determine the beginning and end of each bit position. This is no easy task. One rather expensive approach is to provide a separate clock lead to synchronize the transmitter and receiver. The alternative is to provide some synchronization mechanism that is based on the transmitted signal. This can be achieved with suitable encoding, as explained subsequently.

Error detection: We will discuss various error-detection techniques in Chapter 6 and show that these are the responsibility of a layer of logic above the signaling level that is known as data link control. However, it is useful to have some error detection capability built into the physical signaling encoding scheme. This permits errors to be detected more quickly.

Signal interference and noise immunity: Certain codes exhibit superior performance in the presence of noise. Performance is usually expressed in terms of a BER.

Cost and complexity: Although digital logic continues to drop in price, this factor should not be ignored. In particular, the higher the signaling rate to achieve a given data rate, the greater the cost. We shall see that some codes require a signaling rate that is greater than the actual data rate.

Nonreturn to Zero

- Easiest way to transmit digital signals is to use two different voltages for 0 and 1 bits
- Voltage level is constant during a bit interval (unlike Manchester*)
  - No transition (no return to a zero voltage level)
  - Absence of voltage for 0, constant positive voltage for 1
  - More often, a negative voltage represents one value and a positive voltage represents the other (NRZ-L)

$\left(\pm \frac{V}{2}\right)$ is better than $(0, V)$ in terms of power consumption

$$E = P T_b = \frac{V^2}{R} T_b$$

$$\frac{E_1}{E_2} = \frac{\frac{V^2}{R} T}{\frac{2\left(\frac{V}{2}\right)^2 T}{R}} = \frac{V^2}{\frac{2V^2}{4}} = 2 \Rightarrow E_1 = 2E_2$$

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). For example, the absence of voltage can be 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 latter code, known as Nonreturn to Zero-Level (NRZ-L), is illustrated in Figure 5.2. NRZ-L is typically the code used to generate or interpret digital data by terminals and other devices. If a different code is to be used for transmission, it is generated from an NRZ-L signal by the transmission system [in terms of Figure 5.1, NRZ-L is $g(t)$ and the encoded signal is $x(t)$].

Non-return to Zero Inverted (NRZI)

- Non-return to zero, invert on ones
- Maintains a constant voltage pulse for duration of a bit time
- Data are encoded as presence or absence of signal transition at the beginning of the bit time
  - Transition (low to high, high to low) denotes binary 1
  - No transition denotes binary 0

Is an example of differential encoding
• Data are represented by changes rather than levels • More reliable to detect a transition in the presence of noise than to compare a value to a threshold • 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 themselves 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. For example, on a multidrop twisted-pair line, if the leads from an attached device to the twisted pair are accidentally inverted, all 1s and 0s for NRZ-L will be inverted. This does not happen with differential encoding.

Slide 11

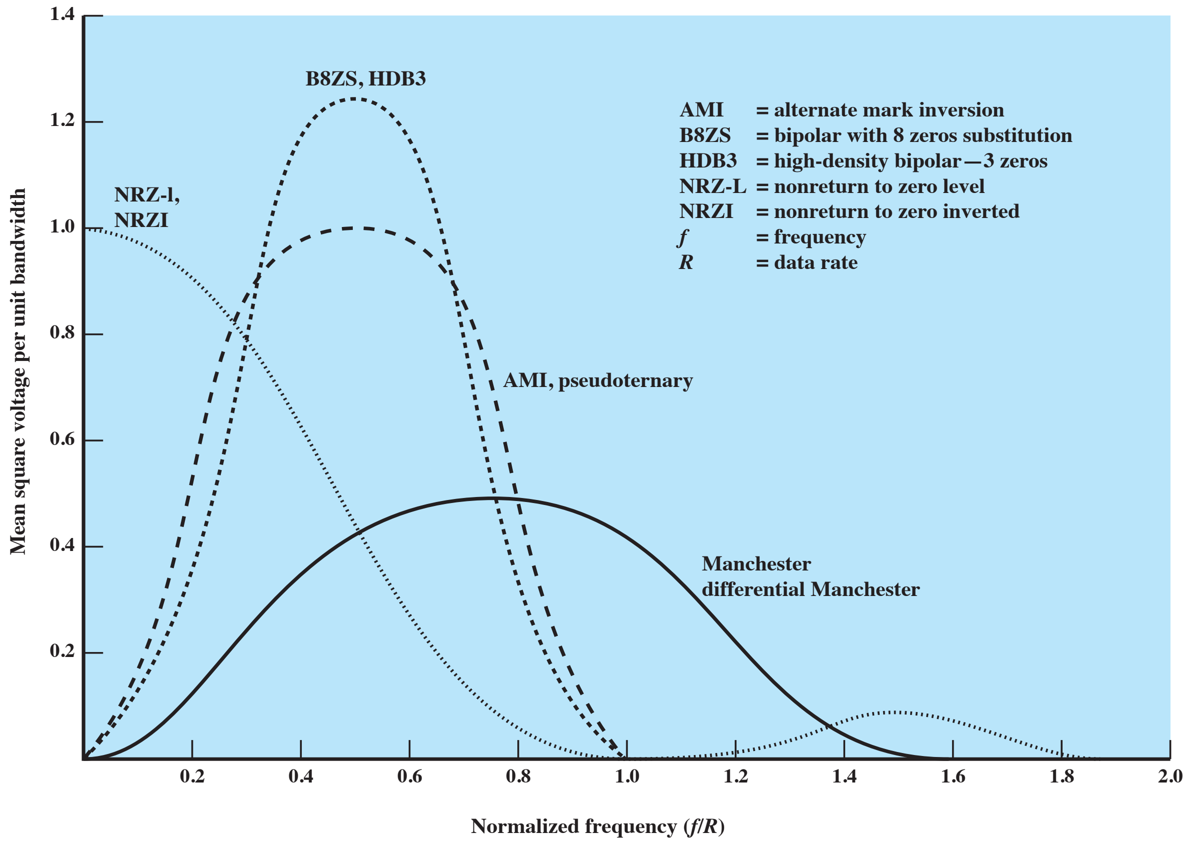


Figure 5.3 Spectral Density of Various Signal Encoding Schemes

The NRZ codes are the easiest to engineer and, in addition, make efficient use of bandwidth. This latter property is illustrated in Figure 5.3, which compares the spectral density of various encoding schemes. In the figure, frequency is normalized to the data rate. Most of the energy in NRZ and NRZI signals is between dc and half the bit rate. For example, if an NRZ code is used to generate a signal with data rate of 9600 bps, most of the energy in the signal is concentrated between dc and 4800 Hz.

The main limitations of NRZ signals are the presence of a dc component and the lack of synchronization capability. To picture the latter problem, 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 signal levels
- Bipolar-AMI
  - Binary 0 represented by no line signal
  - Binary 1 represented by positive or negative pulse
  - Binary 1 pulses alternate in polarity
  - No loss of sync if a long string of 1s occurs
  - No net dc component
  - Lower bandwidth
  - Easy error detection

Multilevel binary encoding techniques address some of the deficiencies of the NRZ codes. These codes use more than two signal levels. Two examples of this scheme are illustrated in Figure 5.2, bipolar-AMI (alternate mark inversion) and pseudoternary.

In the case of 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 (Figure 5.3). 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

- Binary 1 represented by absence of line signal
- Binary 0 represented by alternating positive and negative pulses
- No advantage or disadvantage over bipolar-AMI and each is the basis of some applications

The comments of the previous paragraph 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.

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. We examine two examples of this technique later in this section.

Slide 14

Multilevel Binary Issues

- Synchronization with long runs of 0's or 1's
  - Can insert additional bits that force transitions
  - Scramble data
- 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 approximately 3dB more signal power for same probability of bit error

Thus, with suitable modification, multilevel binary schemes overcome the problems of NRZ codes. Of course, as with any engineering design decision, there is a trade-off. With multilevel binary coding, the line signal may take on one of three levels, but each signal element, which could represent log2 3 = 1.58 bits of information, bears only one bit of information. Thus multilevel binary is not as efficient as NRZ coding. Another way to state this is that 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. This is illustrated in Figure 5.4. 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.

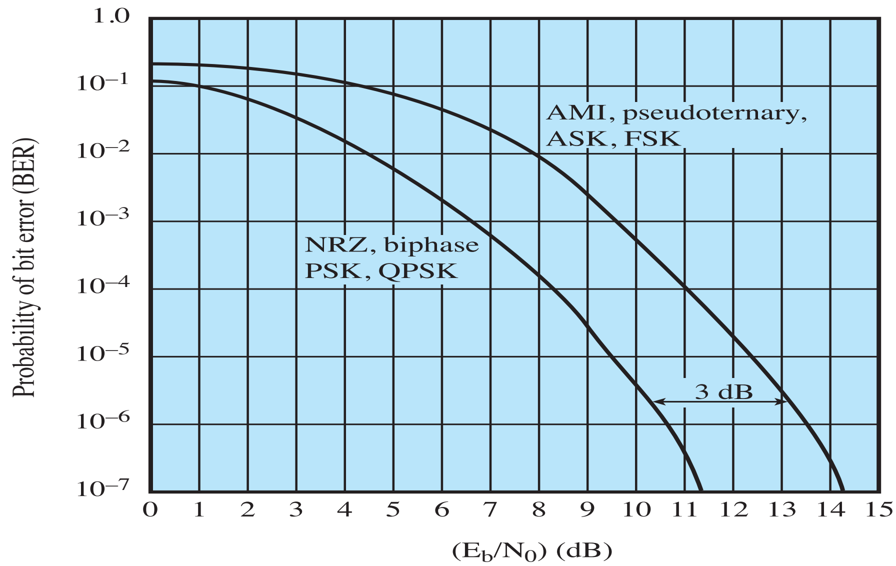


Figure 5.4 Theoretical Bit Error Rate for Various Encoding Schemes

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 log2 3 = 1.58 bits of information, bears only one bit of information. Thus multilevel binary is not as efficient as NRZ coding. Another way to state this is that 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. This is illustrated in Figure 5.4. 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

- There is a transition at the middle of each bit period
- Midbit transition serves as a clocking mechanism and also as data
- Low to high transition represents a 1
- High to low transition represents a 0

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.

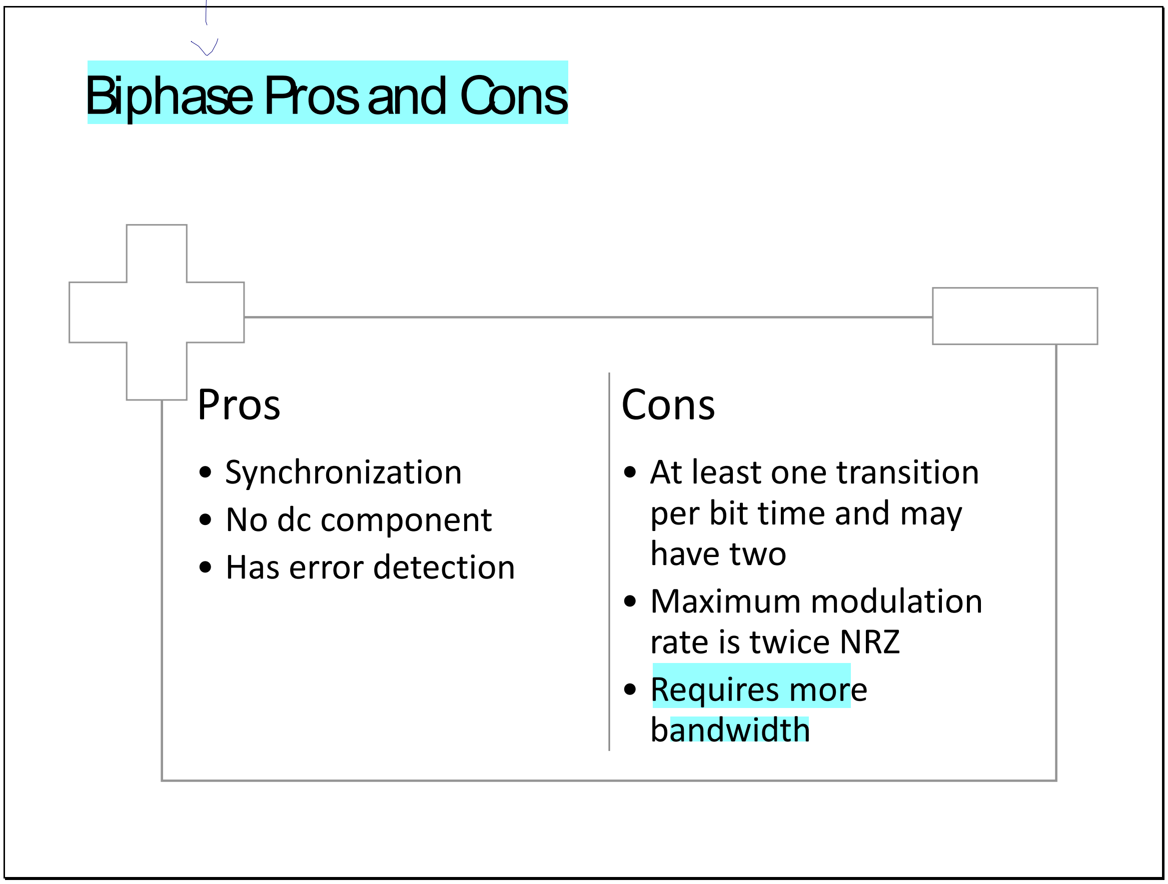
Differential Manchester Encoding

- Midbit transition is only used for clocking
- The encoding of a 0 is represented by the presence of a transition at the beginning of a bit period
- A 1 is represented by the absence of a transition at the beginning of a bit period
- Has the added advantage of employing differential encoding

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.

Slide 18

transition mid bit.



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. 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. For this reason, the biphase codes are known as self-clocking codes.

No dc component: Biphase codes have no dc component, yielding the benefits described earlier.

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.

As can be seen from Figure 5.3, 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.

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 has been specified for the IEEE 802.5 token ring LAN, using shielded twisted pair.

Slide 19

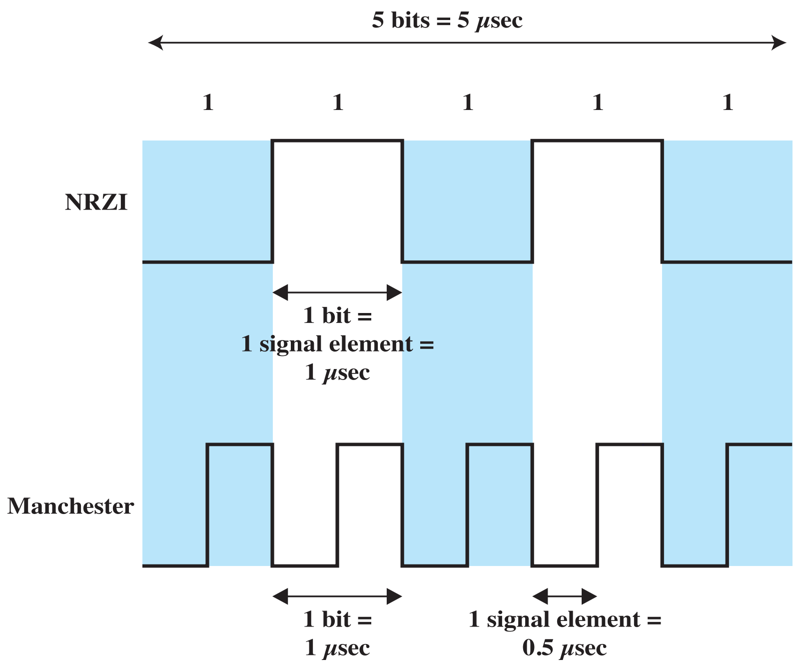


Figure 5.5 A Stream of Binary Ones at 1 Mbps

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 $1/Tb$, where Tb = 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 $2/Tb$. This situation is illustrated in Figure 5.5, which shows the transmission of a stream of binary 1s at a data rate of 1 Mbps using NRZI and Manchester.

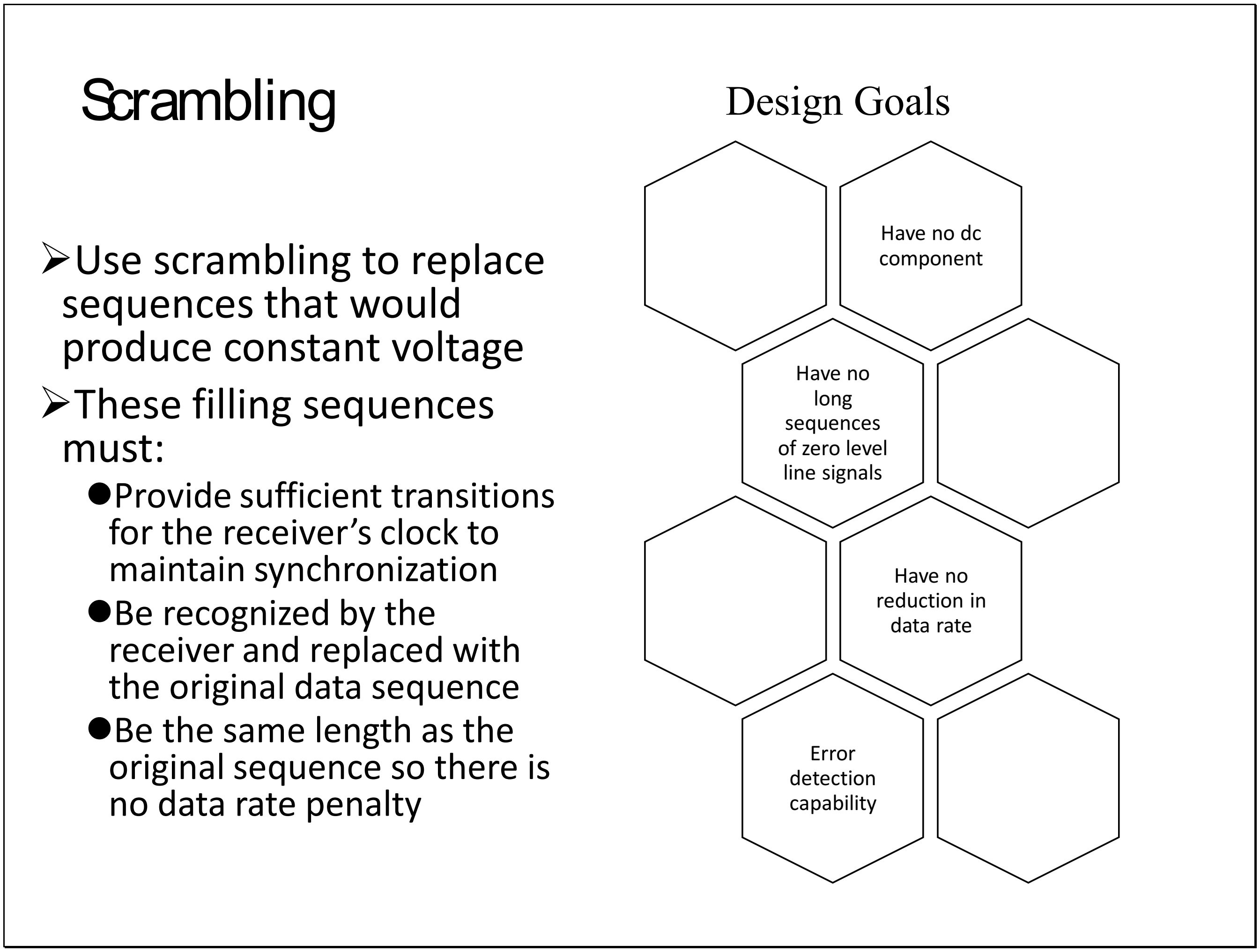
Table 5.3

Normalized Signal Transition Rate of Various Digital Signal Encoding Schemes

	Minimum	101010. . .	Maximum
NRZ-L	0 (all 0s or 1s)	1.0	1.0
NRZI	0 (all 0s)	0.5	1.0 (all 1s)
Bipolar-AMI	0 (all 0s)	1.0	1.0
Pseudoternary	0 (all 1s)	1.0	1.0
Manchester	1.0 (1010 . . .)	1.0	2.0 (all 0s or 1s)
Differential Manchester	1.0 (all 1s)	1.5	2.0 (all 0s)

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. Table 5.3 compares transition rates for various techniques. It indicates the signal transition rate in the case of a data stream of alternating 1s and 0s, and for the data stream that produces the minimum and maximum modulation rate.

Slide 21



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

- Bipolar with 8-zeros substitution
- Coding scheme commonly used in North America
- Based on a bipolar-AMI
  - 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+-

The bipolar with 8-zeros substitution (B8ZS) coding scheme is commonly used in North America. The coding scheme is based on a bipolar-AMI. We have seen that the drawback of the AMI code is that a long string of zeros may result in loss of synchronization. To overcome this problem, 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.

	Number of Bipolar Pulses (ones) since Last Substitution	
Polarity of Preceding Pulse	Odd	Even
-	000-	+00+
+	000+	-00-

A coding scheme that is commonly used in Europe and Japan is known as the high-density bipolar-3 zeros (HDB3) code (Table 5.4). As before, it is 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. Table 5.4
shows that this condition is tested for by determining (1) whether the number of pulses since the last violation is even or odd and (2) the polarity of the last pulse before the occurrence of the four zeros.

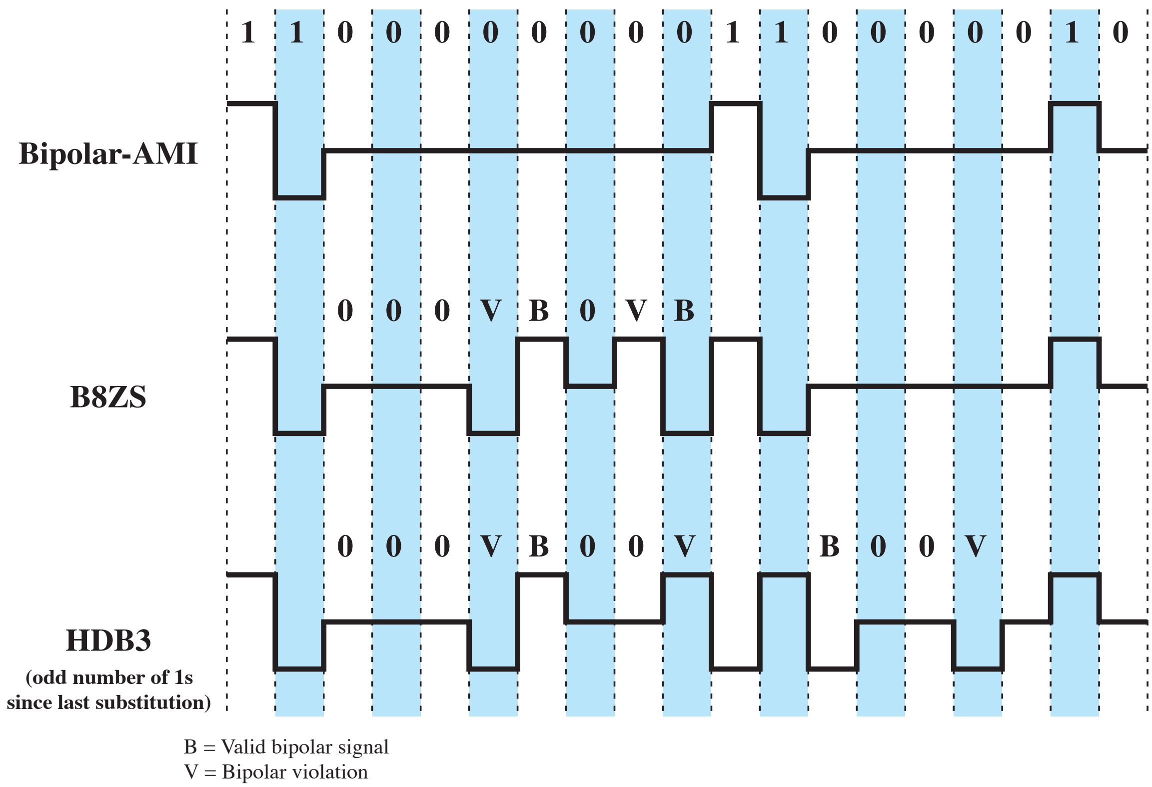


Figure 5.6 Encoding Rules for B8ZS and HDB3

Figure 5.6 shows the signal encoding for the binary sequence 1100000000110000010 using AMI, and then scrambled using B8ZS and HDB3. The original sequence includes a continuous strings of eight zeros and five zeros. B8ZS eliminates the string of eight zeros. HDB3 eliminates both strings. The total number of transitions for this sequence is 7 for Bipolar AMI, 12 for B8ZS, and 14 for HDB3.

Digital Data, Analog Signal

- Main use is public telephone system
  - Was designed to receive, switch, and transmit analog signals
  - Has a frequency range of 300Hz to 3400Hz
  - Is not at present suitable for handling digital signals from the subscriber locations
  - Uses modem (modulator-demodulator) to convert digital data to analog signals and vice versa

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.

For the telephone network, modems are used that produce signals in the voice-frequency range. The same basic techniques are used for modems that produce signals at higher frequencies (e.g., microwave). This section introduces these techniques and provides a brief discussion of the performance characteristics of the alternative approaches.

Slide 26

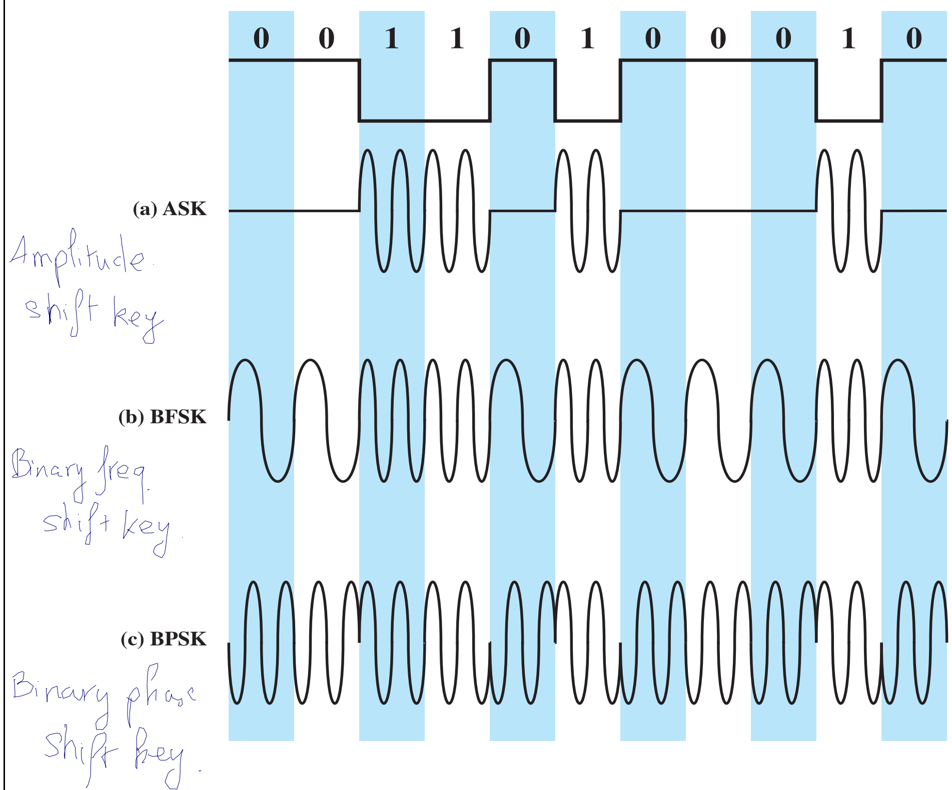


Figure 5.7 Modulation of Analog Signals for Digital Data

We mentioned 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 Figure 5.7: 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 (ASK)

- 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 (Figure 5.7a).

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. For LED (light-emitting diode) transmitters, Equation (5.2) is valid. That is, one signal element is represented by a light pulse while the other signal element is represented by the absence of light. Laser transmitters normally have a fixed "bias" current that causes the device to emit a low light level. This low level represents one signal element, while a higher-amplitude lightwave represents another signal element.

Binary Frequency Shift Keying (BFSK)

- Most common form of FSK
- Two binary values are 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 coaxial cable

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 (Figure 5.7b).

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.

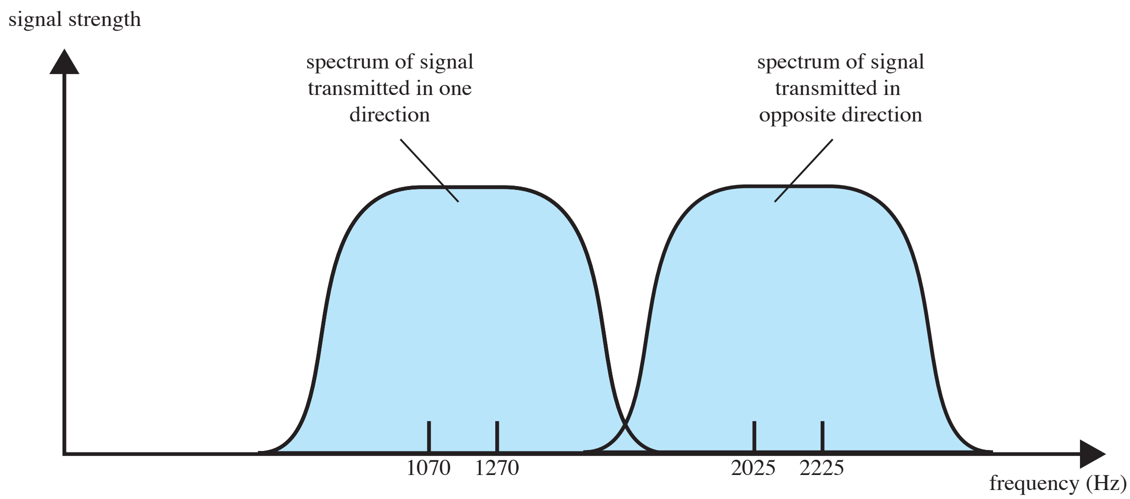


Figure 5.8 Full-Duplex FSK Transmission on a Voice-Grade Line

Figure 5.8 shows an example of the use of BFSK for full-duplex operation over a voice-grade line. The figure is a specification for the Bell System 108 series modems. Recall that a voice-grade line will pass frequencies in the approximate range 300 to 3400 Hz, and that *full duplex* means that signals are transmitted in both directions at the same time. To achieve full-duplex transmission, this bandwidth is split. In one direction (transmit or receive), the frequencies used to represent 1 and 0 are centered on 1170 Hz, with a shift of 100 Hz on either side. The effect of alternating between those two frequencies is to produce a signal whose spectrum is indicated as the shaded area on the left in Figure 5.8. Similarly, for the other direction (receive or transmit) the modem uses frequencies shifted 100 Hz to each side of a center frequency of 2125 Hz. This signal is indicated by the shaded area on the right in Figure 5.8. Note that there is little overlap and thus little interference.

Multiple FSK (MFSK)

- Each signaling element represents more than one bit
- More than two frequencies are used
- More bandwidth efficient
- More susceptible 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 $W_d = 2Mf_d = M/T_s$.

Q. calculation of bw?

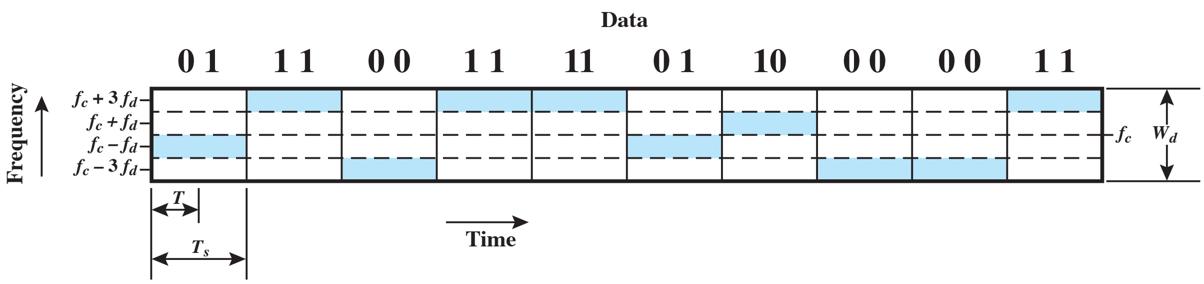


Figure 5.9 MFSK Frequency Use (*M* = 4)

Figure 5.9 shows an example of MFSK with M = 4. An input bit stream of 20 bits is encoded 2 bits at a time, with each of the four possible 2-bit combinations transmitted as a different frequency. The display in the figure shows the frequency transmitted (y -axis) as a function of time (x -axis). Each column represents a time unit T_s in which a single 2-bit signal element is transmitted. The shaded rectangle in the column indicates the frequency transmitted during that time unit.

Phase Shift Keying (PSK)

- The phase of the carrier signal is shifted to represent data
- Binary PSK
  - Two phases represent the 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 5.7c) 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.

Slide 33

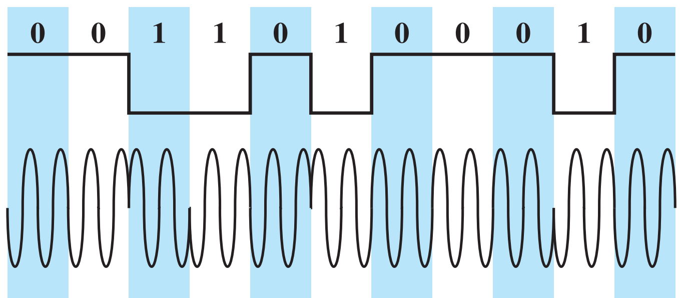


Figure 5.10 Differential Phase-Shift Keying (DPSK)

Figure 5.10 shows an example of DPSK.

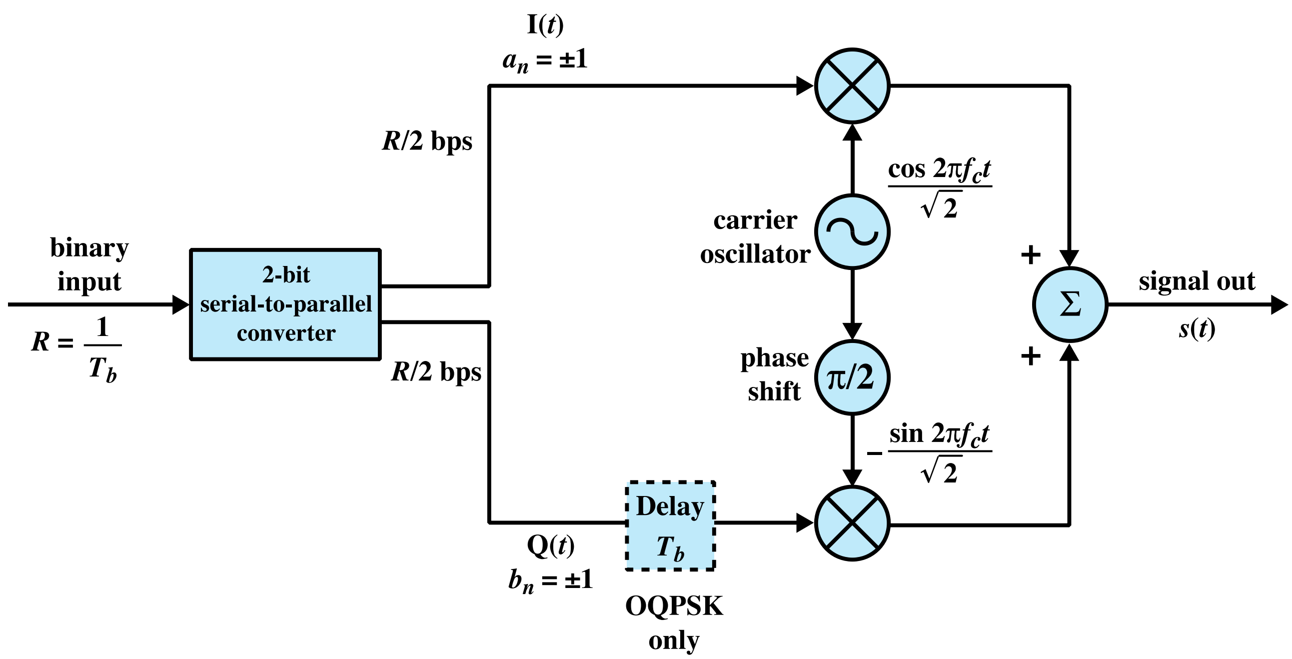


Figure 5.11 QPSK and OQPSK Modulators

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 π /2 (90°).

Thus each signal element represents two bits rather than one.

Figure 5.11 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.

Figure 5.11 also shows a 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.

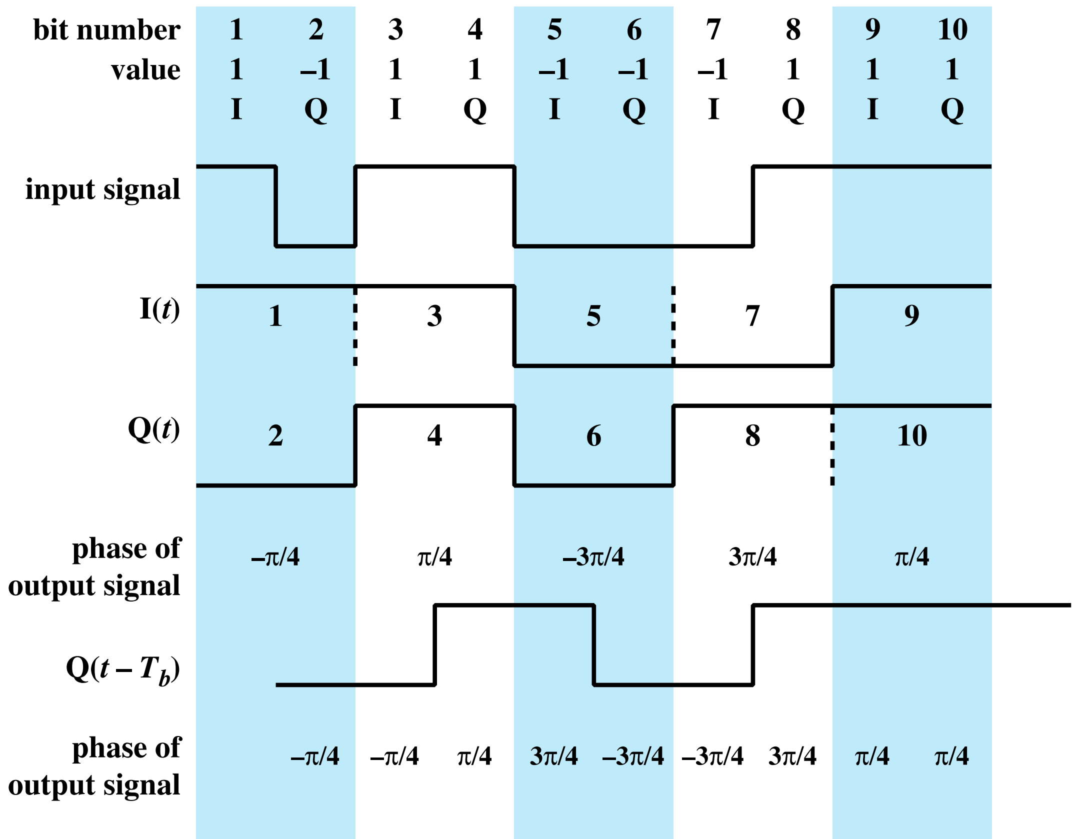


Figure 5.12 Example of QPSK and OQPSK Waveforms

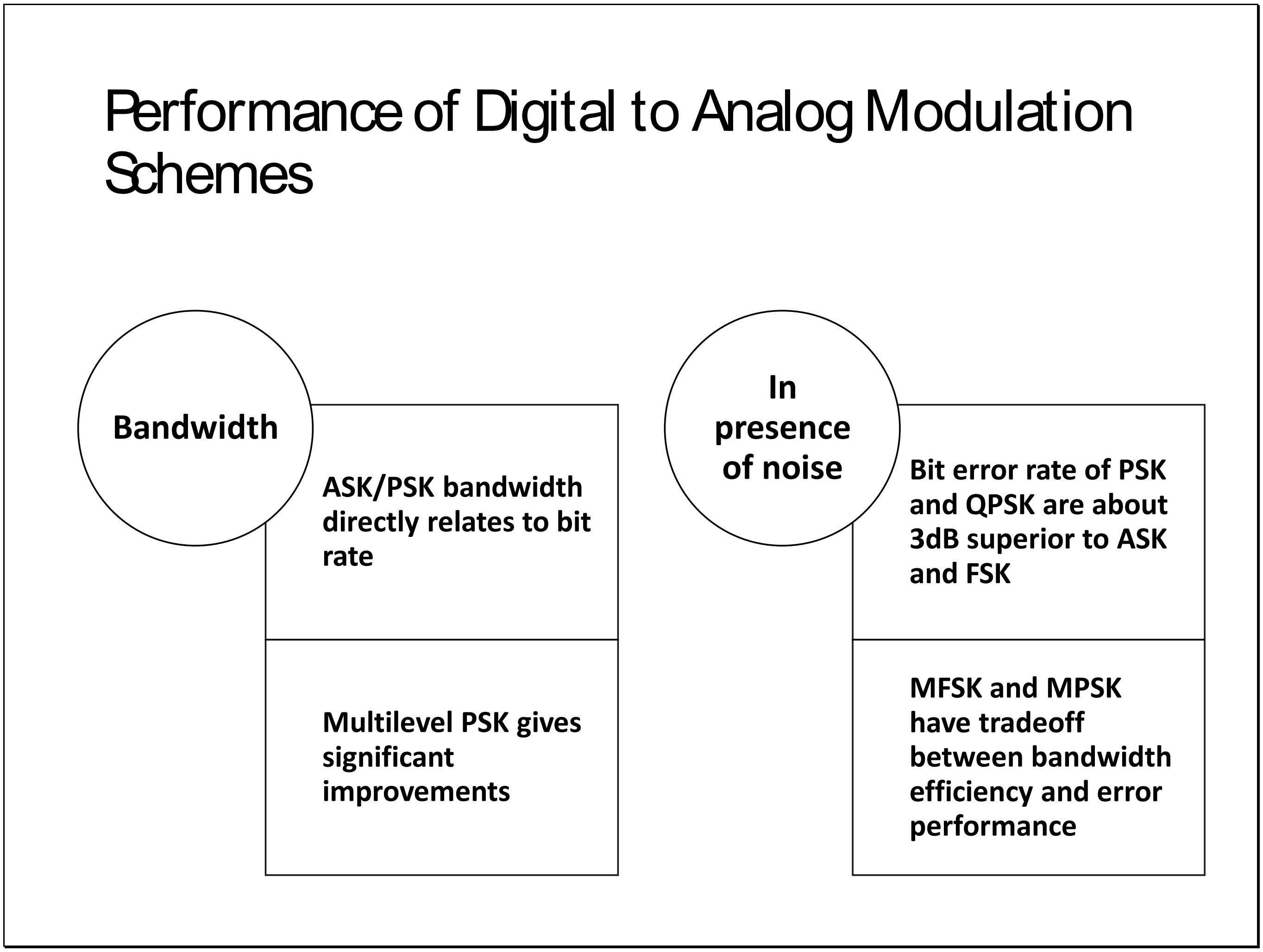
Figure 5.12 shows an example of QPSK coding. Each of the two modulated streams is a BPSK signal at half the data rate of the original bit stream. Thus, the combined signals have a symbol rate that is half the input bit rate. Note that from one symbol time to the next, a phase change of as much as 180° (π) is possible.

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. From Figure 5.12, we can observe that only one of two bits in the pair can change sign at any time and thus the phase change in the combined signal never exceeds 90° ($\pi/2$) . This can be an advantage because physical limitations on phase modulators make large phase shifts at high transition rates difficult to perform. OQPSK also provides superior performance when the transmission channel (including transmitter and receiver) has significant nonlinear components. The effect of nonlinearities is a spreading of the signal bandwidth, which may result in adjacent channel interference. It is easier to control this spreading if the phase changes are smaller, hence the advantage of OQPSK over QPSK.

	$r = 0$	$r = 0.5$	$r = 1$
ASK	1.0	0.67	0.5
Multilevel FSK			
$M = 4, L = 2$	0.5	0.33	0.25
$M = 8, L = 3$	0.375	0.25	0.1875
$M = 16, L = 4$	0.25	0.167	0.125
$M = 32, L = 5$	0.156	0.104	0.078
PSK	1.0	0.67	0.5
Multilevel PSK			
$M = 4, L = 2$	2.00	1.33	1.00
$M = 8, L = 3$	3.00	2.00	1.50
$M = 16, L = 4$	4.00	2.67	2.00
$M = 32, L = 5$	5.00	3.33	2.50

Table 5.5 shows the ratio of data rate, R, to transmission bandwidth for various schemes. This ratio is also referred to as the bandwidth efficiency. As the name suggests, this parameter measures the efficiency with which bandwidth can be used to transmit data. The advantage of multilevel signaling methods now becomes clear.

Slide 37



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. Figure 5.4 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.

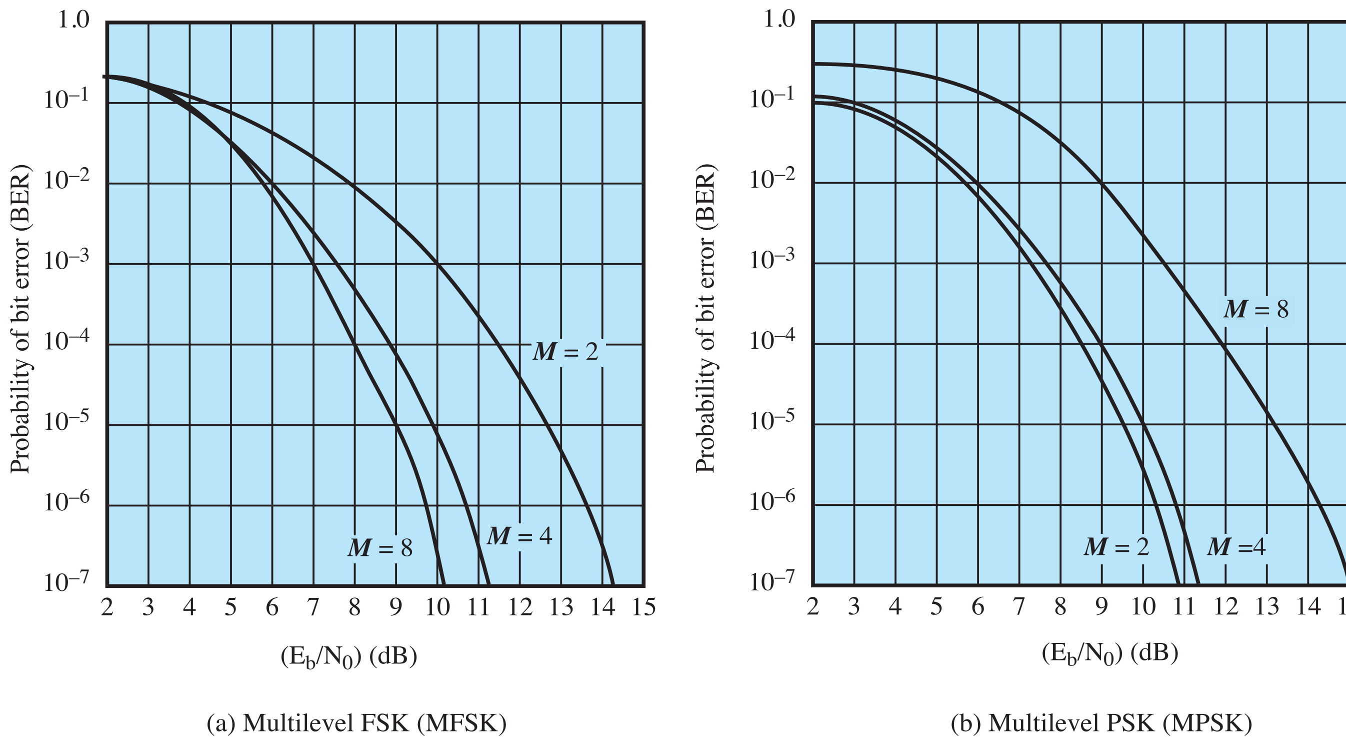


Figure 5.13 Theoretical Bit Error Rate for Multilevel FSK and PSK

Figure 5.13 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. The fact that these trade-offs move in opposite directions with respect to the number of levels M for MFSK and MPSK can be derived from the underlying equations. A discussion of the reasons for this difference is beyond the scope of this book. See [SKLA01] for a full treatment.

Quadrature Amplitude Modulation (QAM)

- QAM is used in the asymmetric digital subscriber line (ADSL), in cable modems, and in some wireless standards
- Is a combination of ASK and PSK
- Logical extension of QPSK
- Send two different signals simultaneously on the same carrier frequency
  - Use two copies of carrier, one shifted 90°
  - Each carrier is ASK modulated
  - Two independent signals simultaneously transmitted over the same medium
  - At the receiver, the two signals are demodulated and the results are combined to produce the original binary input

Quadrature amplitude modulation (QAM) is a popular analog signaling technique that is used in the asymmetric digital subscriber line (ADSL) and in cable modems, 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.

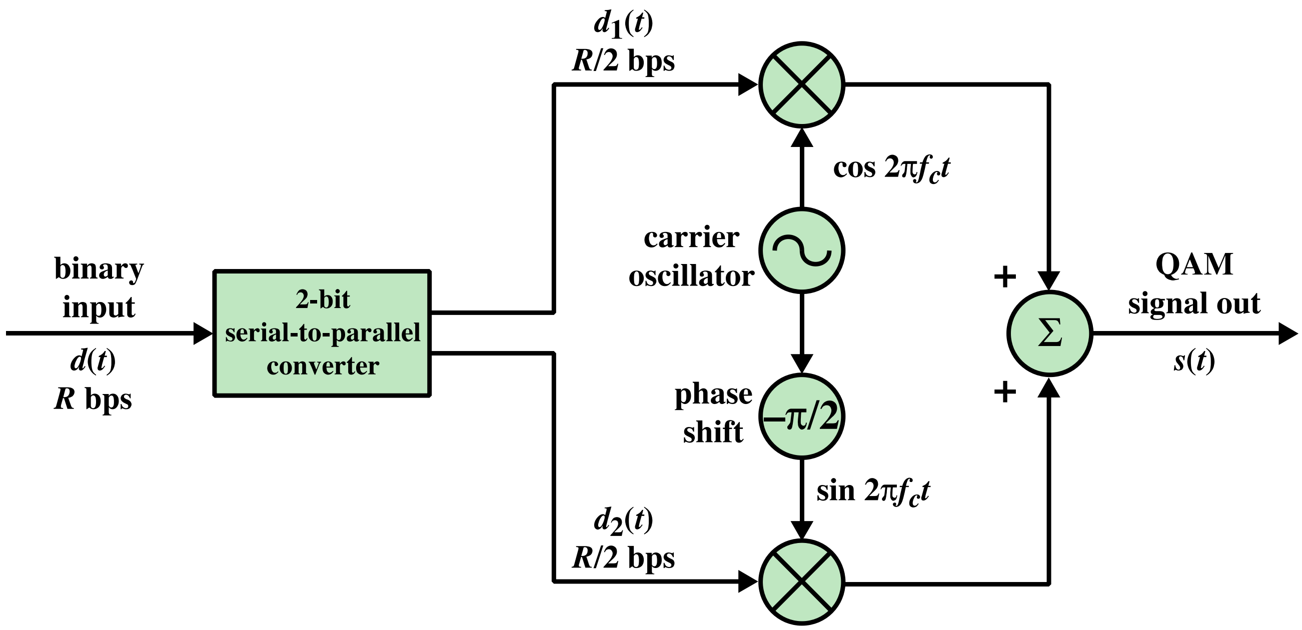


Figure 5.14 QAM Modulator

Figure 5.14 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.

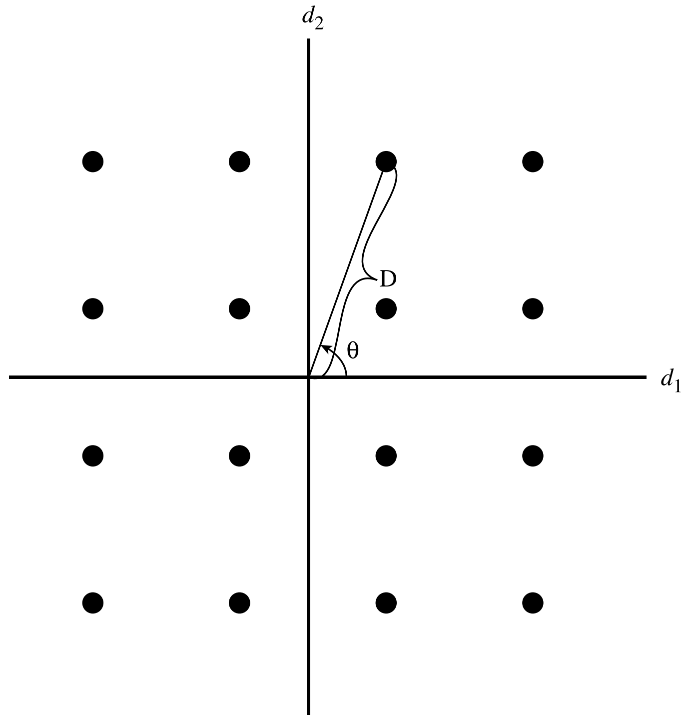


Figure 15.15 16-QAM Constellation

Figure 5.15 shows the possible combinations of instantaneous values of the digital signals d_1 (t) and d_2 (t). For 16-QAM, each digital signals encodes two bits and takes on one of four values, two positive and two negative.

Slide 42

Analog Data, Digital Signal

- Digitization is the 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 is 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. The section closes with a discussion of comparative performance.

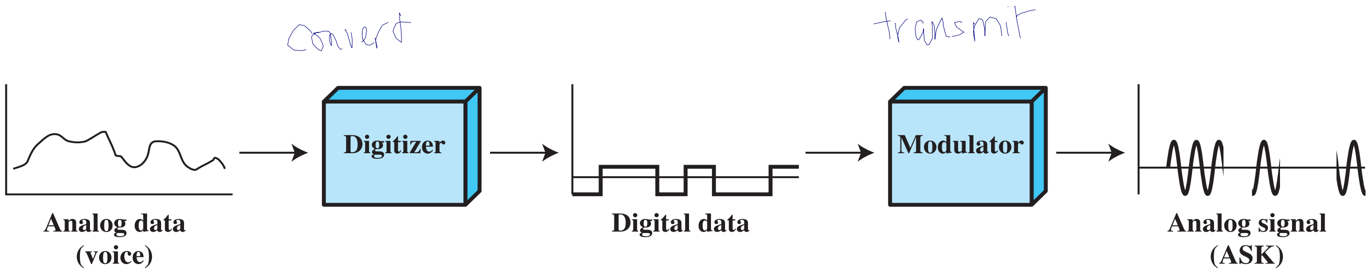
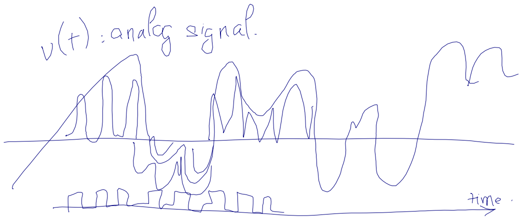
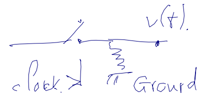


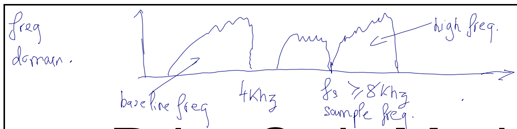
Figure 5.16 Digitizing Analog Data

Figure 5.16 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 (e.g., use of microwave) dictate that an analog signal be used.

Slide 44







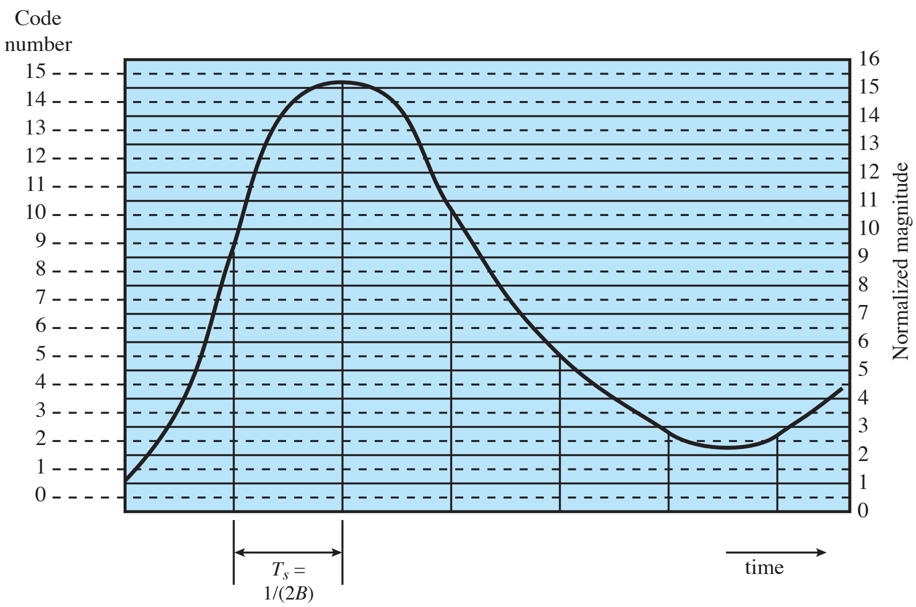
clock high: gate open => voltage follows input(v)
clock low: gate close => voltage follows ground (0).

Pulse Code Modulation (PCM)

- Based on the sampling theorem:
  - "If a signal *f(t)* is sampled at regular intervals of time and at a rate higher than twice the highest signal frequency, then the samples contain all the information of the original signal. The function *f(t)* may be reconstructed from these samples by the use of a lowpass filter."
- Pulse Amplitude Modulation (PAM)
  - Analog samples
  - To convert to digital, each of these analog samples must be assigned a binary code

For the interested reader, a proof is provided in Appendix G. If voice data are 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.



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

Figure 5.17 Pulse-Code Modulation Example

Figure 5.17 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 $2B$, 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.
The resulting digital code for this analog signal is
0001100111111010010100100010.

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 * 8 bits per sample = 64 kbps is needed for a single voice signal.

Slide 46

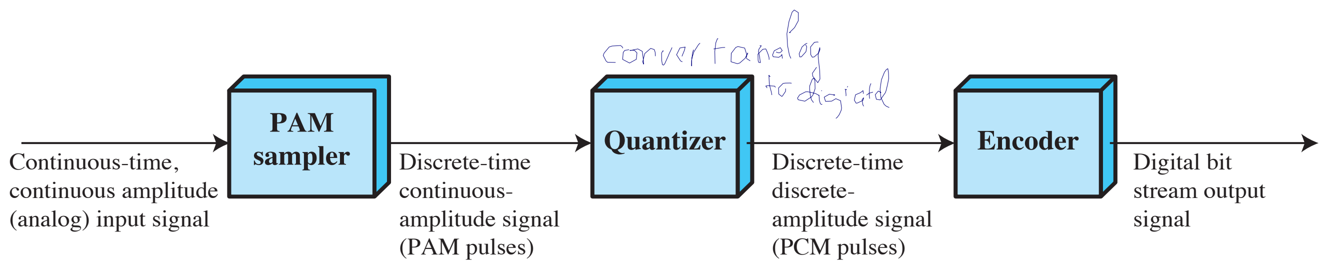
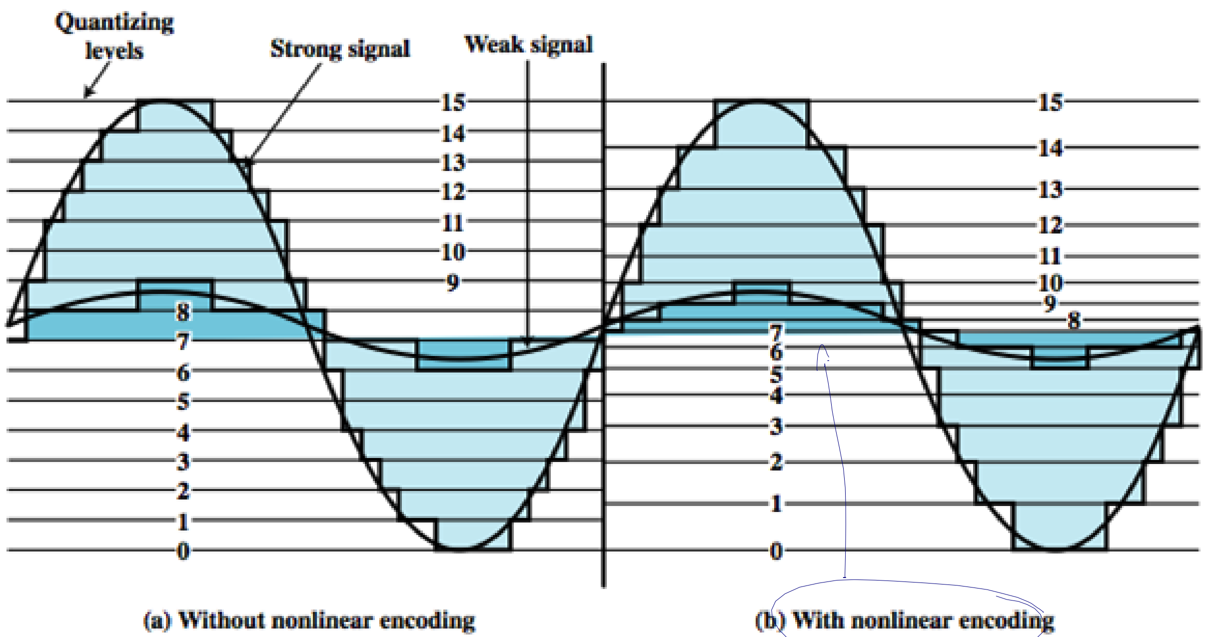


Figure 5.18 PCM Block Diagram

Thus, PCM starts with a continuous-time, continuous-amplitude (analog) signal, from which a digital signal is produced, as shown in Figure 5.18. 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



more levels of signal to reduce lost of signal.

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 Figure 5.19.

Nonlinear encoding can significantly improve the PCM SNR ratio. For voice signals, improvements of 24 to 30 dB have been achieved.

Slide 48

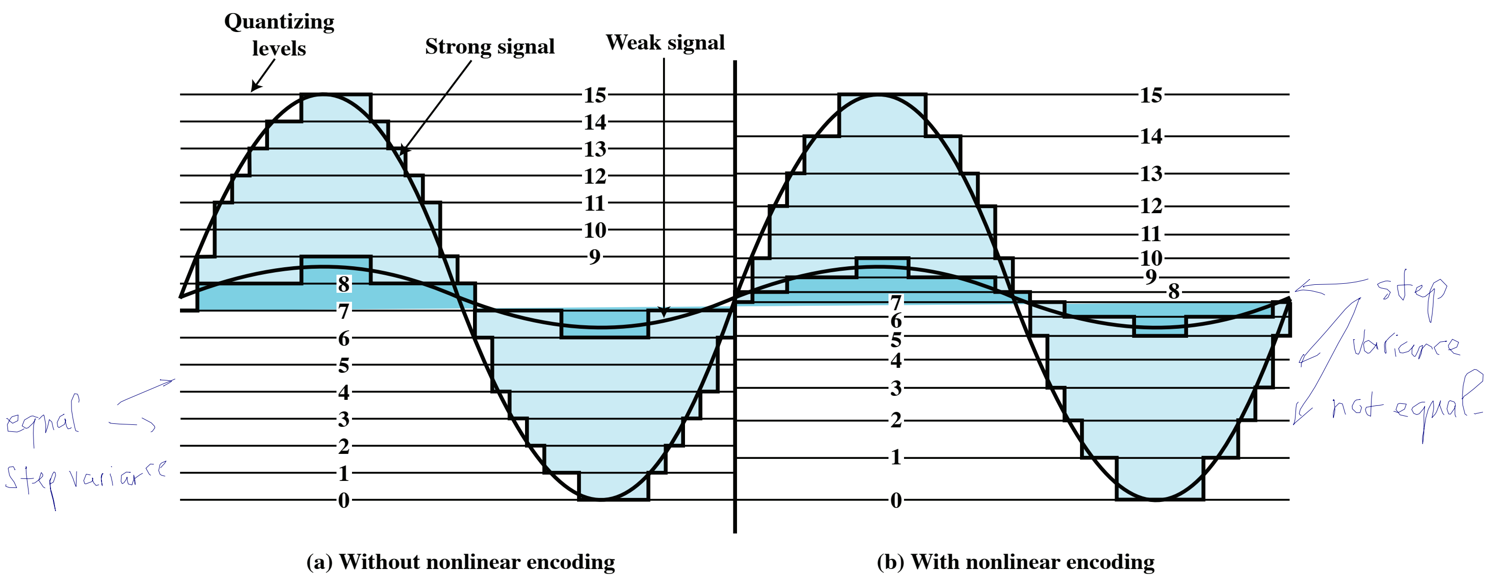


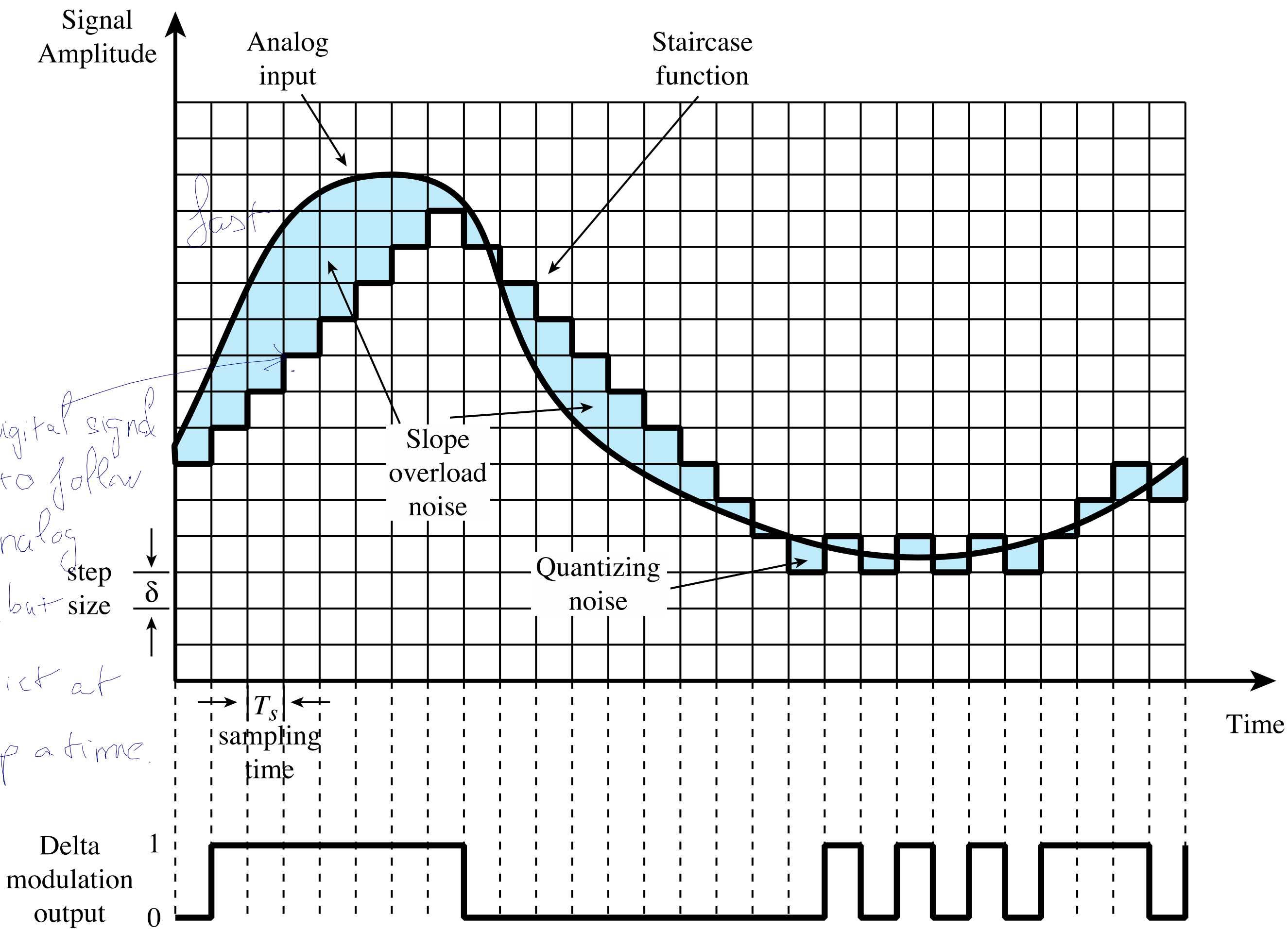
Figure 5.19 Effect of Nonlinear Coding

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. Figure 5.20 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 (DM)

- Analog input is approximated by a staircase function
  - Can move up or down one quantization level (δ) at each sampling interval
- Has binary behavior
  - Function only moves up or down at each sampling interval
  - Output of the delta modulation process can be represented as a single binary digit for each sample
  - 1 is generated if the staircase function is to go up during the next interval, otherwise a 0 is generated

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.



the digital signal tries to follow the analog signal, but restrict at 1 step a time.

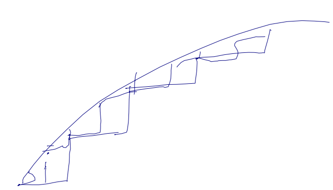
send 1 bit at a time (not 8 bit).

Figure 5.21 Example of Delta Modulation

Figure 5.21 shows an example, with the staircase function overlaid on the original analog waveform. The resulting digital code for this analog signal is 0111111100000000001010101110.

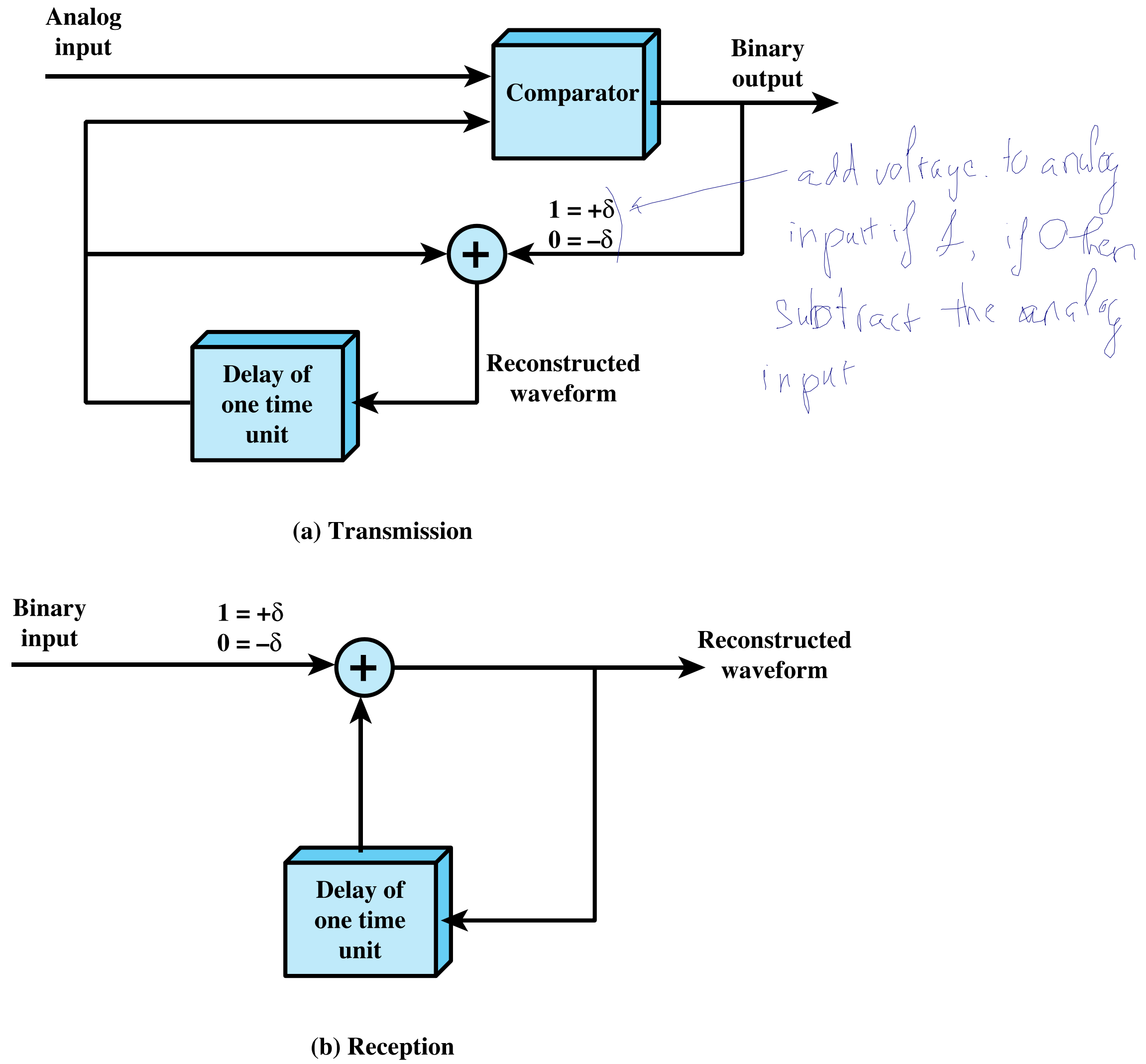
input signal fast → error high

input variation slow → error low.



pros: instead of sending 8 bit, only send 1 bit,

cons: noise margin is high. → solution: increase the rate (double, tripple, etc.)



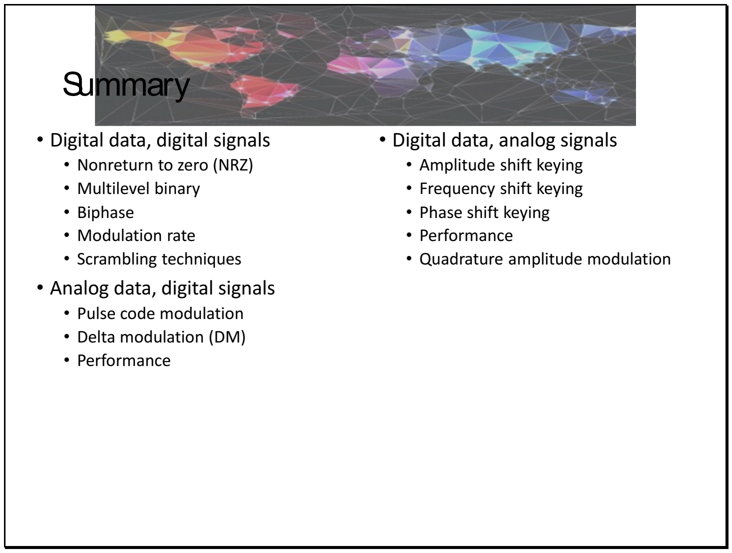
(a) Transmission

(b) Reception

Figure 5.22 Delta Modulation

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. Figure 5.22 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.

Slide 52



Chapter 5 summary.