Modern Academy for Engineering and Technology
Electronics Engineering and Communication Technology Dpt.

ELC 421

Communications (2)

Prof. Dr. Adel El-Sherif

Dr. Nelly Muhammad Hussein

Ch.[1] Analog Pulse Modulation

1.1 Introduction

In Continuous-Wave (CW) Modulation: (studied previously)

- Some parameter of a sinusoidal carrier wave is varied continuously in accordance with the message signal
- Types: Amplitude Modulation (AM), Frequency Modulation (FM), and Phase Modulation (PM)

In Pulse Modulation: (study in the present chapter)

- Some parameter of a pulse train is varied continuously in accordance with the message signal. We may distinguish two families of pulse modulation: analog pulse modulation and digital pulse modulation.

1.1.1Pulse Modulation Types

- Analog Pulse Modulation, and Digital Pulse Modulation
- In analog pulse modulation, a periodic pulse train is used as the carrier wave, and some characteristic feature of each pulse (e.g., amplitude, duration, or position) is varied in a continuous manner in accordance with the corresponding sample value of the message signal. Thus in analog pulse modulation, information is transmitted basically in analog form, but the transmission takes place at discrete times.
- In digital pulse modulation, on the other hand, the message signal is represented in a form that is discrete in both time and amplitude, thereby permitting its transmission in digital form as a sequence of coded pulses; this form of signal transmission has no CW counterpart

Two potential advantages of pulse modulating over CW modulation:

1. Transmitted Power, can be concentrated into short bursts instead of being generated continuously
2. Time Interval, between pulses can be filled with samples values from other signals, [Time-division Multiplexing (TDM)]

Disadvantage of Pulse Modulation, requiring very large transmission bandwidth compared to the message bandwidth

Pulse Modulation Types: Analog Pulse Modulation and Digital Pulse Modulation

→ **The three main types of analog pulse modulation:**

1. PAM: Pulse Amplitude Modulation.
2. PWM: Pulse Width Modulation.
3. PPM: Pulse Position Modulation.

→ **The four main types of digital pulse modulation:**

1. PCM: Pulse Code Modulation.
2. DPCM: Differential Pulse Code Modulation.
3. DM: Delta Modulation.
4. ΣDM: Sigma Delta Modulation.

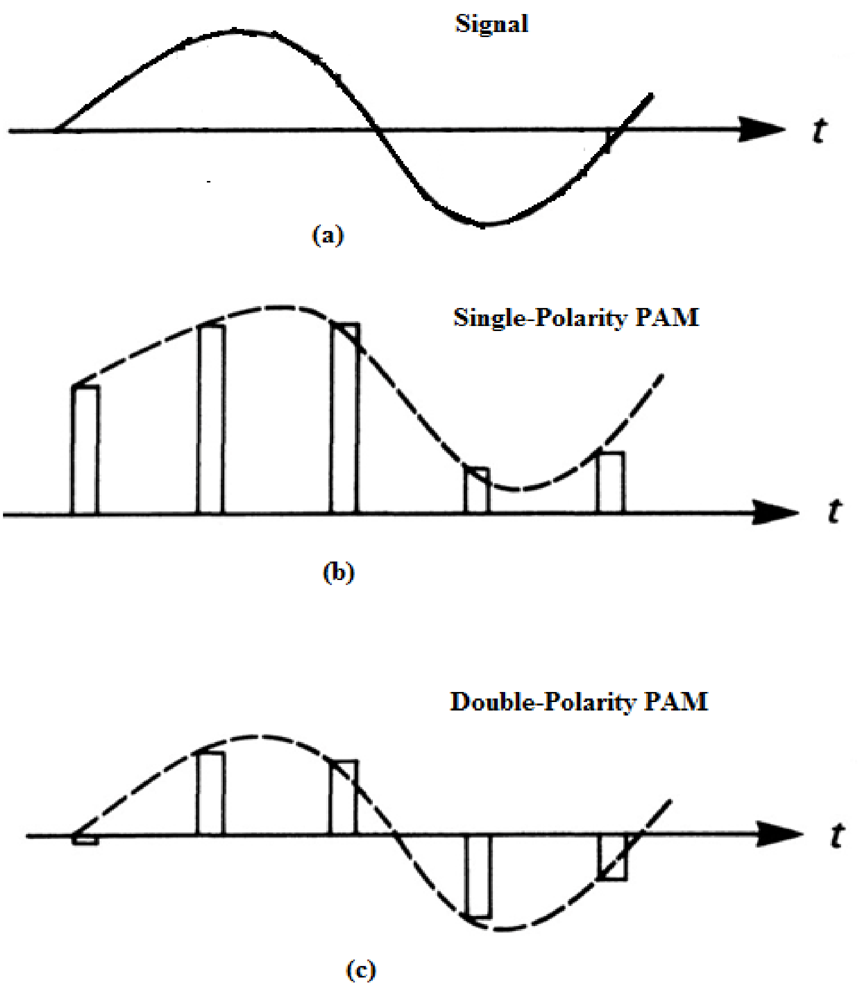
(Pulse-Modulation Types)

Analog — **(PAM, PWM, PPM)**

Digital — **(PCM, Diff. PCM, DM, ΣDM)**

1.1.2 Pulse Amplitude Modulation (PAM)

- The simplest and most basic form of analog pulse modulation
- In pulse-amplitude modulation (PAM), the amplitudes of regularly spaced pulses are varied in proportion to the corresponding sample values of a continuous message signal
- The waveform of a PAM signal is illustrated in Figure 1.1
- Two types of PAM: double-polarity PAM and single-polarity PAM
  - Double-polarity PAM, which is self-explanatory
  - Single-polarity PAM, in which a fixed dc level is added to the signal, to ensure that the pulses are always positive



bb

Fig. (1.1): Pulse-Amplitude Modulation. (a) Signal; (b) double-polarity PAM; (c) single-polarity PAM

[Refer to figure (1.1) in the text book. Page 3]

Generation of PAM signal

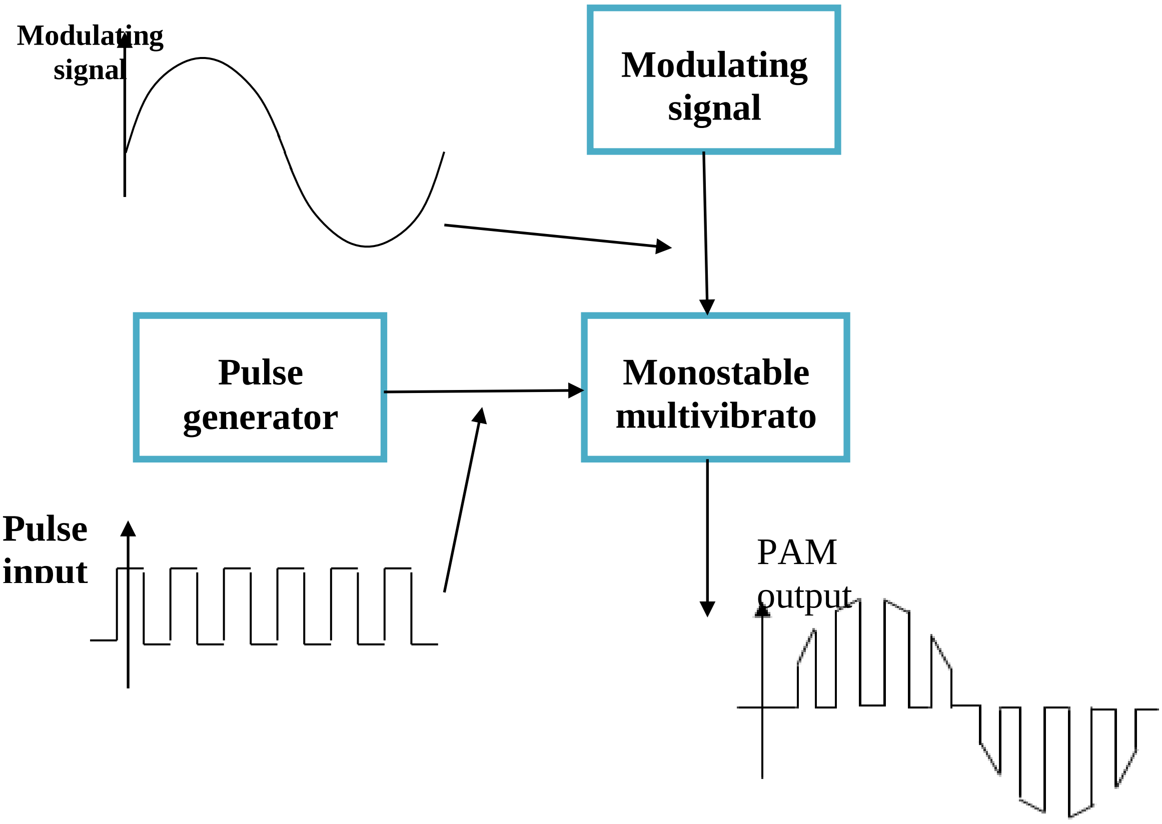


Fig. (1.2): PAM modulating system

[Refer to figure (1.3.a) in the text book. Page 6]

- A process for producing a PAM waveform is illustrated in the previous figure, in which a pulse generator triggers a monostable multivibrator at a sampling frequency.
- The output pulses from the multivibrator are made to increase and decrease in amplitude by the modulating signal.

→ The circuit of the monostable multivibrator

- From Figure 1.3, when Q3 is on, the output voltage is the saturation level of Q3 collector. When Q3 is switched off for the pulse time, the output voltage (pulse amplitude) is equal to the modulating signal level

- The actual voltage applied as a supply to R5 must have a dc component as well as the ac modulating signal. This is necessary to ensure correct operation of Q3

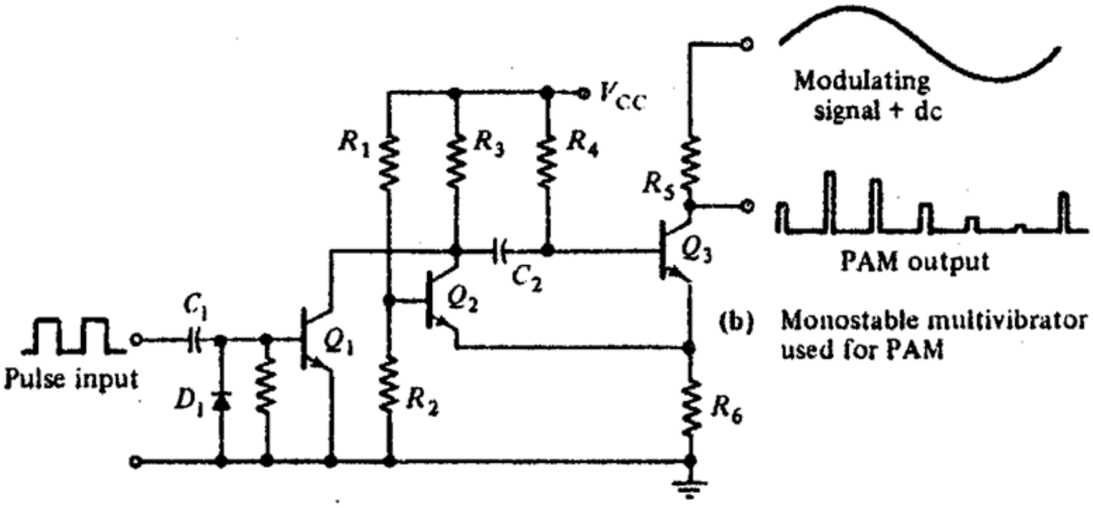


Fig. (1.3): PAM use of a monostable multivibrator

[Refer to figure (1.3.b) in the text book. Page 6]

Demodulation of PAM Signal

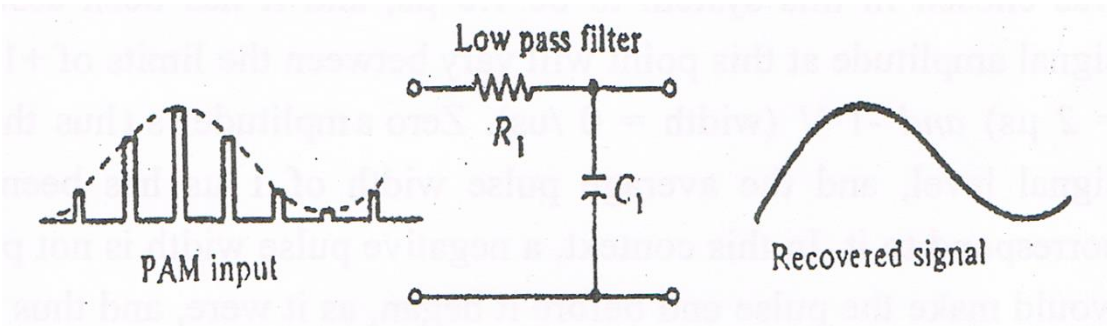


Fig. (1.4): PAM demodulation

[Refer to figure (1.4) in the text book. Page 7]

- This process is accomplished simply by passing the amplitude-modulated pulses through a LPF where the PAM waveform consists of the fundamental modulating frequency, and a number of high-frequency components which give the pulses their shape.

- The filter output is the low frequency component corresponding to the original baseband signal.

Advantages of PAM

- The simplest pulse modulation technique
- The possibility of sending more than one signal on the same channel ,each with certain time slot (using TDM)

Disadvantages of PAM

- Less noise immunity than the other types of analog modulation
- Not power saving

→ Pulse-Time Modulation (PTM)

- In PTM the signals have a constant amplitude with one of their timing characteristics is varied, being proportional to the sampled signal amplitude at that instant
- The variable characteristic may be the width, position or frequency of the pulses
- Pulse-frequency modulation has no significant practical applications and will be omitted
- In PWM: Sample values of the analog waveform are used to determine the width of the pulse signal, however in PPM: the analog sample values determine the position of a narrow pulse relative to the clocking time as shown in Figure 1.5
- The PTM have advantage over PAM: in all of them the pulse amplitude remains constant, so that amplitude limiters can be used to provide a good degree of noise immunity

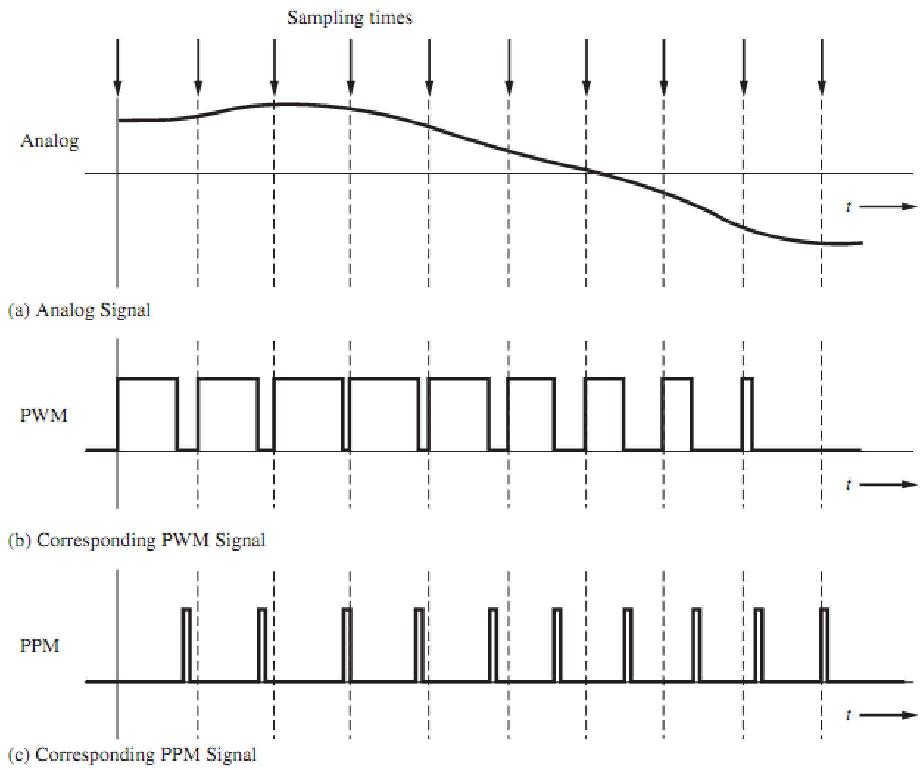


Fig. (1.5): Pulse time modulation signaling

1.1.3 Pulse Width Modulation (PWM)

- PWM is sometimes called pulse duration modulation (PDM) or pulse length modulation (PLM), as the width (active portion of the duty cycle) of a constant amplitude pulse is varied proportional to the amplitude of the analog signal at the time the signal is sampled.
- The maximum analog signal amplitude produces the widest pulse, and the minimum analog signal amplitude produces the narrowest pulse. Note, however, that all pulses have the same amplitude.

Generation of PWM signal

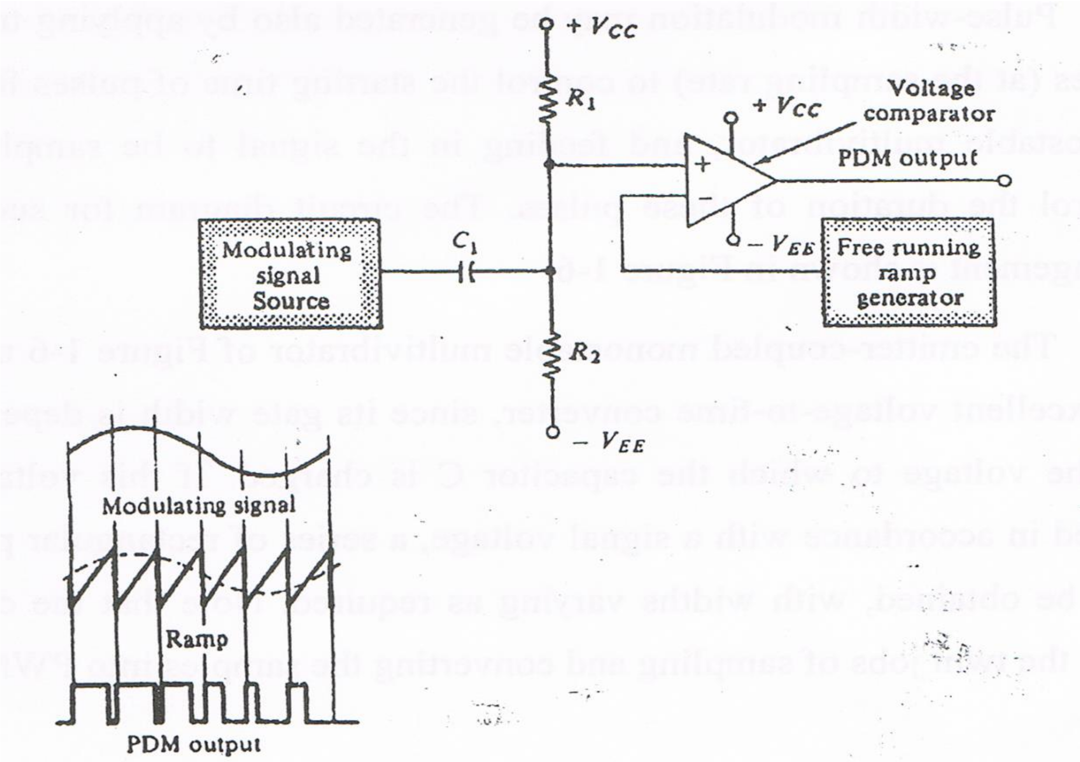


Fig. (1.6): PDM modulating system

[Refer to figure (1.5) in the text book. Page 9]

- We use free running ramp generator and voltage comparator:
  - When V^+(modulating signal) > V^-(ramp signal),, the comparator output = $+V_{cc}$

– When V^+(modulating signal) < V^-(ramp signal),, the comparator output = $-V_{EE}$

- So the output pulse width is made proportional to the amplitude of the modulating signal

→ **Another method to generate PWM**

- As shown in Figure 1.7, Appling trigger pulses (at the sampling rate) to control the starting time of pulses from a monostable multivibrator, and feeding in the signal to be sampled to control the duration of these pulses
- The emitter-coupled monostable mutivibrator in Figure makes excellent voltage-to-time converter, since its gate width is dependent on the voltage to which the capacitor C is charged

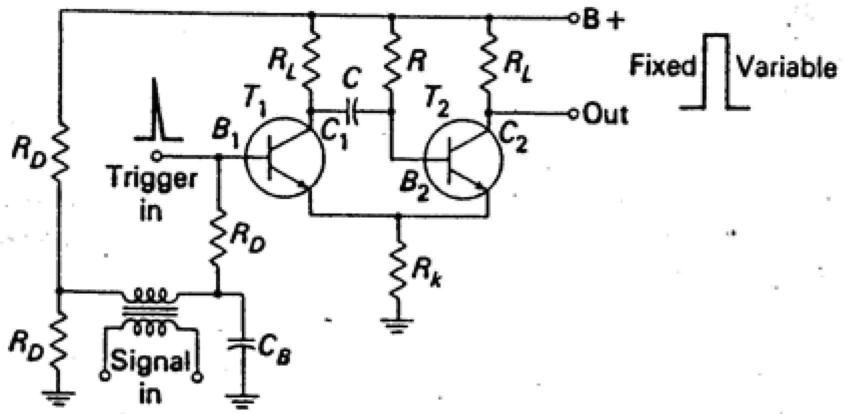


Fig. (1.7): Monostable multivibrator generating pulse-width modulation

[Refer to figure (1.6) in the text book. Page 11]

- If this voltage is varied in accordance with a signal voltage, a series of rectangular pulses will be obtained, with widths varying as required
  – The applied trigger pulses switches T_1 ON, C begins to charge up to the collector supply potential through R
  – After a time determined by the supply voltage and the RC time constant of the charging network, B_2 becomes sufficiently positive to switch T_2 ON

- T_1 is simultaneously switched OFF by regenerative action and stays OFF until the arrival of the next trigger pulse
- The applied modulation voltage controls the voltage to which B_2 must rise to switch T_2 ON. Since this voltage rise is linear, the modulation voltage is seen to control the period of time during which T_2 is OFF, that is, the pulse duration

Demodulation of PWM Signal

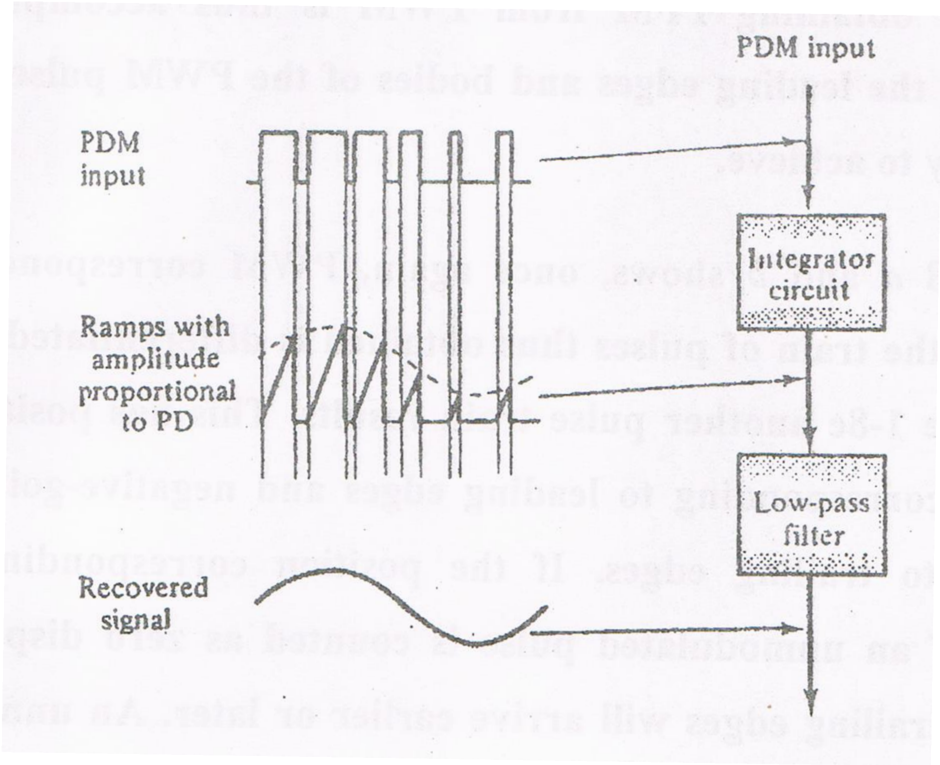


Fig. (1.8): PDM demodulating system

[Refer to figure (1.7) in the text book. Page 13]

- A miller integrator circuit is suitable for use with PWM demodulation
- The integrator circuit converts the PWM to PAM and using low-pass filter to recover the original signal

Advantages of PWM

– Has high noise immunity where modulated pulses are constant in amplitude.

Disadvantages of PWM

– Unstability in the transmitter power system due to variable time intervals

1.1.4 Pulse Position Modulation (PPM)

- With PPM, the position of a constant-width pulse within a prescribed time slot is varied according to the amplitude of the sample of the analog signal.
- The higher the amplitude of the sample, the farther to the right the pulse is positioned within the prescribed time slot. The highest amplitude sample produces a pulse to the far right, and the lowest amplitude sample produces a pulse to the far left

Generation of PPM signal

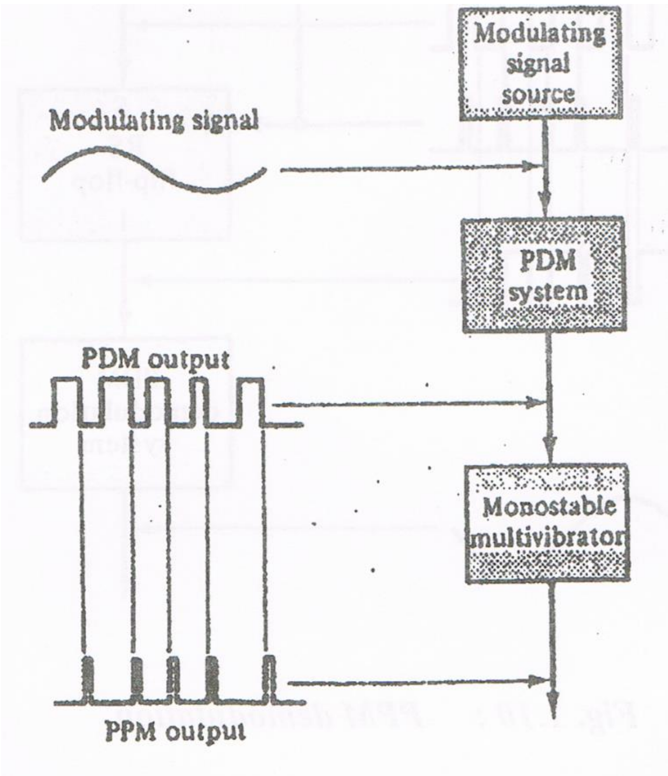


Fig. (1.9): PPM modulating system

[Refer to figure (1.9) in the text book. Page 15]

- The monostable multivibrator is arranged so that it is triggered by the trailing edges of the PWM pulses
- The monostable output is a series of constant-width, constant amplitude pulses which vary in position according to the modulating signal amplitude

Demodulation of PPM Signal

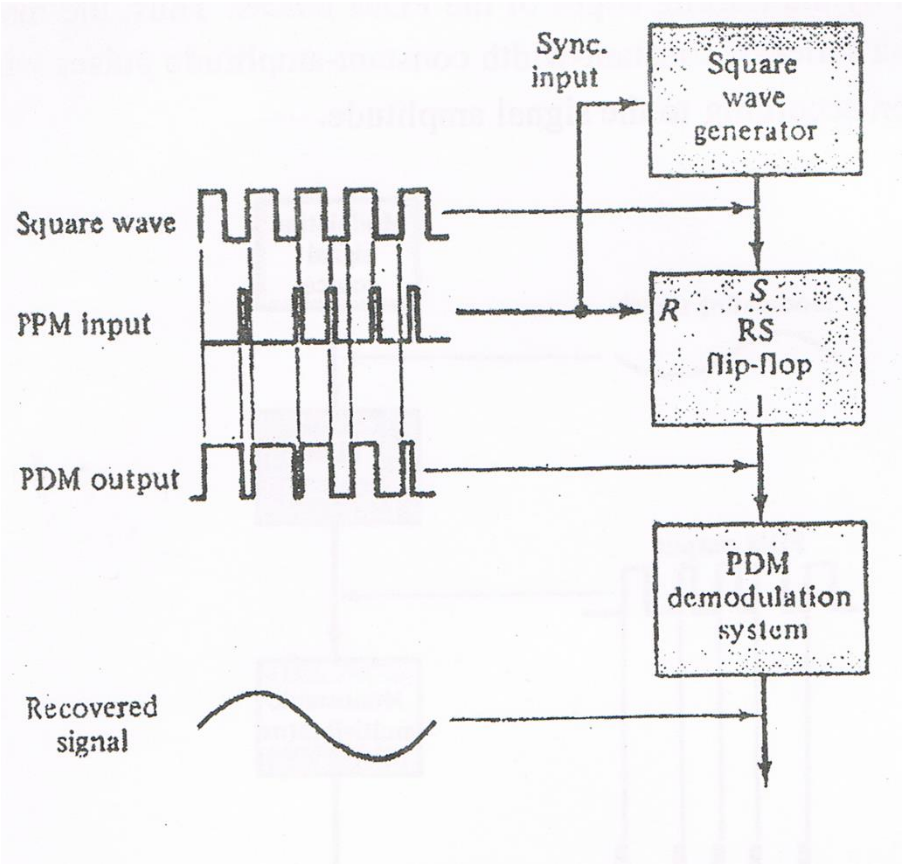


Fig. (1.10): PPM demodulation

[Refer to figure (1.10) in the text book. Page 16]

- First, PPM is converted to PWM by using RS flip-flop. Then using PWM demodulator to recover the modulating signal

Advantages of PPM

- High noise immunity where modulated pulses are constant in amplitude
- Stable in the transmitter power system
- Optimum power saving, since modulated pulses have fixed amplitude and duration

Disadvantages of PPM

- Required synchronization between modulator and demodulator

1.2 The Sampling Process

– Sampling is the process of converting continuous time, continuous amplitude analog signal into discrete time continuous amplitude signal

– Sampling can be achieved by taking samples values from the analog signal at an equally spaced time intervals T_s as shown graphically in Figure 1.11

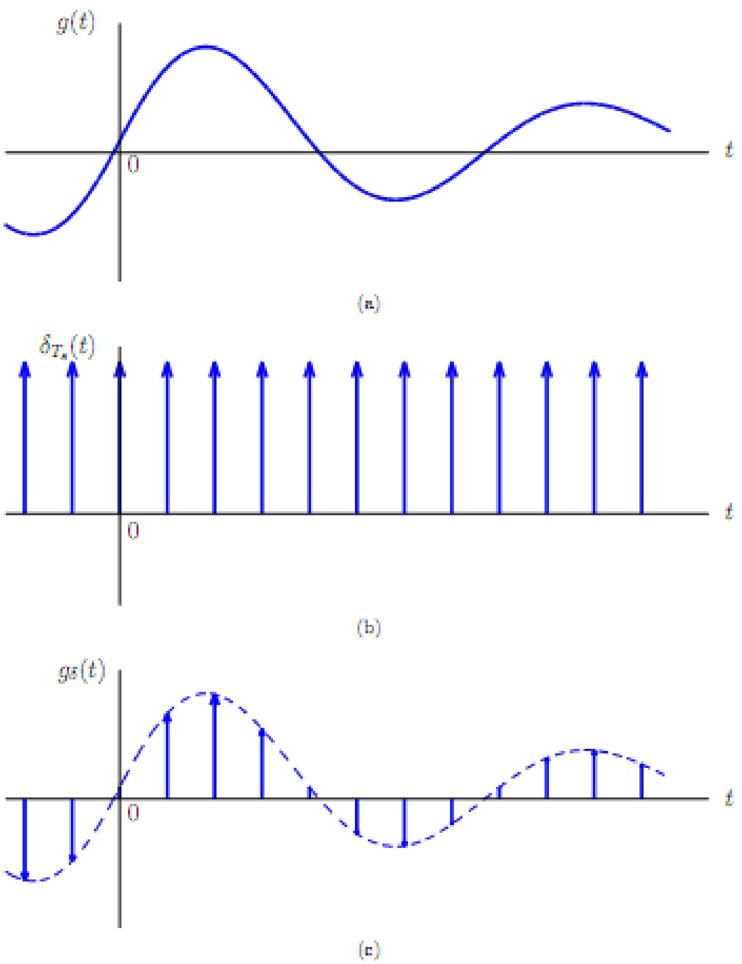


Fig. (1.11): A waveform illustration of the sampling process

→ Sampling Theorem

a. A band-limited signal of finite energy with bandwidth B.W = B hertz, can be represented by its values at instants of time separated by 1/2B sec

b. A band-limited signal of finite energy with bandwidth B.W = B hertz, can be completely recovered from its samples taken at a rate of 2B samples/sec

→ Advantages of Sampling

- It is the first step of digital communication and digital signal processing
- The transmitted power concentrated only at existence of samples (saved power)
- The time interval between samples can be filled with sample values from another signal [Time-division Multiplexing (TDM)]
- Less effect to noise and distortion
- Digital hardware can be used

→ Sampling Condition

- Sampling frequency ≥ 2B ,, where B is the maximum frequency of the message signal
- Nyquist rate [minimum sampling frequency] = 2B sample/sec
- Nyquist interval = $\frac{1}{2B}$ sec

→ Types of Sampling

- There are three sampling types available, these are
  a) Ideal sampling
  b) Natural sampling
  c) Flat top sampling

1.2.1 Ideal Sampling (Instantaneous Sampling)

- In ideal sampling the analog signal is multiplied by a delta impulse functions as shown in Figure 1.12
- Ideal sampling is used to explain the main concept of sampling theoretically

– In practical life Ideal sampling cannot be achieved, because there is no practical circuit which generates exact delta comb function

→ Figures in time and frequency domain

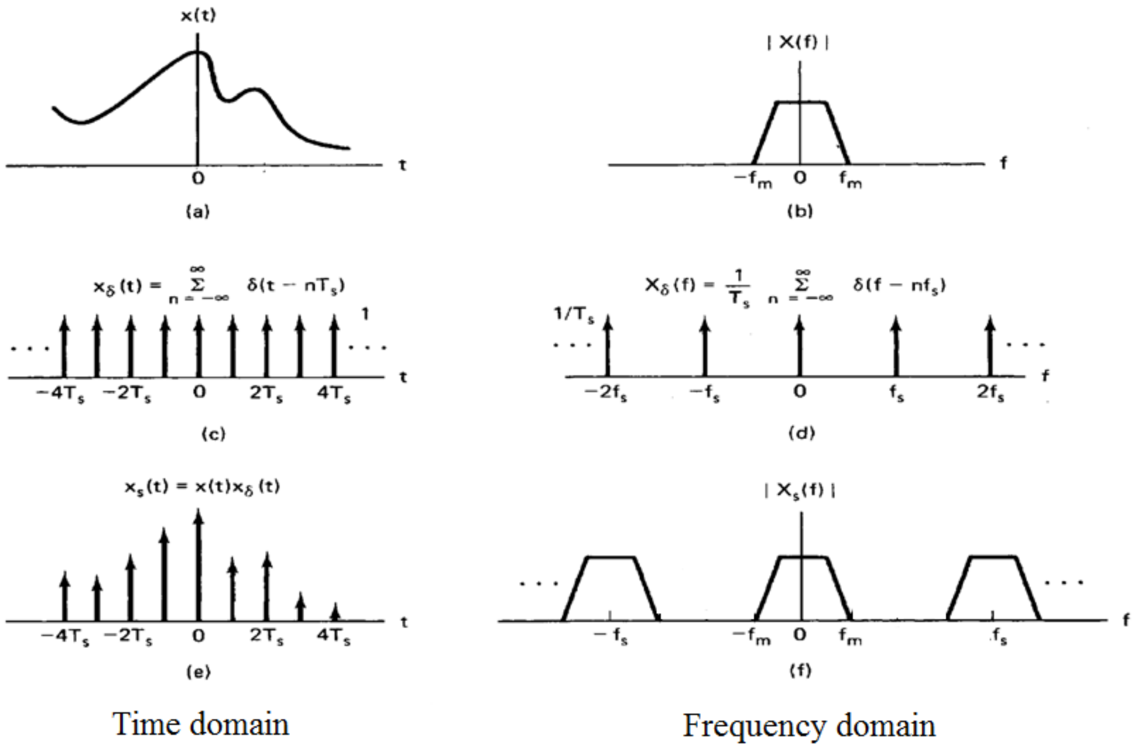


Fig. (1.12): Time domain: (a) Analog signal. (b) Sampling signal (c) Instantaneously sampled version of signal
Frequency domain: (a) Spectrum of a strictly band-limited signal (b) Spectrum of sampling signal (c) Spectrum of sampled version of the signal

[Refer to figure (1.11),fig(1.12) in the text book. Pages 19,20]

→ Let $g_\delta(t)$ denote the ideal sampled signal

$$-g_\delta(t) = \sum_{n=-\infty}^{\infty} g(nT_s)\,\delta(t - nT_s) \qquad \textbf{(1.1)}$$

– where T_s: sampling period

$-f_s = 1/T_s$:sampling rate

→ From previous, we have

$$g(t)\sum_{n=-\infty}^{\infty}\delta(t-nT_s)\Leftrightarrow G(f)*\frac{1}{T_s}\sum_{m=-\infty}^{\infty}\delta(f-\frac{m}{T_s})=\sum_{m=-\infty}^{\infty}f_sG(f-mf_s)$$

$$g_\delta(t)\Leftrightarrow f_s\sum_{m=-\infty}^{\infty}G(f-mf_s) \qquad (1.2)$$

→ or we may apply Fourier Transform on (1.1) to obtain

$$G_\delta(f)=\sum_{n=-\infty}^{\infty}g(nT_s)\exp(-j2\pi nfT_s) \qquad (1.3)$$

- The relations, as derived here, apply to any continuous-time signal g(t) of finite energy and infinite duration
- Suppose that the signal g(t) is strictly band-limited, with no frequency components higher than W hertz. That is, the Fourier transformation G(f) of the signal g(t) has the property that G(f) is zero for $|f|\geq W$ as illustrated in Figure

→ If $G(f)=0$ for $|f|\geq W$ and $T_s=1/2W$

$$G_\delta(f)=\sum_{n=-\infty}^{\infty}g(\frac{n}{2W})\exp(-\frac{j\pi nf}{W}) \qquad (1.4)$$

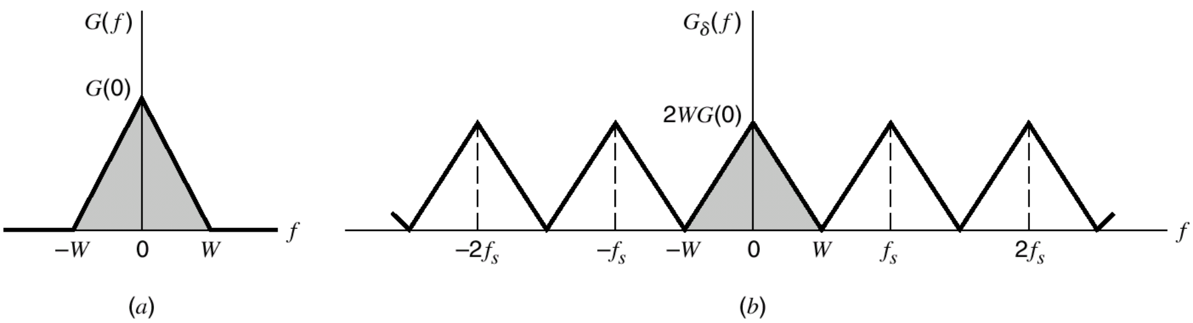


Fig. (1.13): (a) Spectrum of a strictly band-limited signal (b) Spectrum of sampled version of g(t) for a sampling period T_s = 1/2W

[Refer to figure (1.12) in the text book. Page 20]

$\rightarrow$ from Eq. (1.2) we readily see that the Fourier Transform of $g_{\delta}(t)$ may be expressed as

$$G_{\delta}(f) = f_s G(f) + f_s \sum_{\substack{m=-\infty \\ m \neq 0}}^{\infty} G(f - mf_s) \qquad (1.5)$$

$\rightarrow$ With

1. $G(f) = 0$ for $|f| \geq W$

2. $f_s = 2W$

$\rightarrow$ we find from Equation (1.5) that

$$G(f) = \frac{1}{2W} G_{\delta}(f) \ , \ -W < f < W \qquad (1.6)$$

$\rightarrow$ Substituting (3.4) into (3.6) we may rewrite $G(f)$ as

$$G(f) = \frac{1}{2W} \sum_{n=-\infty}^{\infty} g(\frac{n}{2W}) \exp(-\frac{j\pi nf}{W}) \ , \ -W < f < W \qquad (1.7)$$

$-g(t)$ is uniquely determined by $g(\frac{n}{2W})$ for $-\infty < n < \infty$

-or $\left\{ g(\frac{n}{2W}) \right\}$ contains all information of $g(t)$

$\rightarrow$ To reconstruct $g(t)$ from $\left\{ g(\frac{n}{2W}) \right\}$, we may have

$$g(t) = \int_{-\infty}^{\infty} G(f) \exp(j2\pi ft) df$$

$$= \int_{-W}^{W} \frac{1}{2W} \sum_{n=-\infty}^{\infty} g(\frac{n}{2W}) \exp(-\frac{j\pi nf}{W}) \exp(j2\pi f t) df \qquad (1.8)$$

$$= \sum_{n=-\infty}^{\infty} g(\frac{n}{2W}) \frac{1}{2W} \int_{-W}^{W} \exp\left[j2\pi f (t - \frac{n}{2W}) \right] df \qquad (1.9)$$

$$= \sum_{n=-\infty}^{\infty} g(\frac{n}{2W}) \frac{\sin(2\pi Wt - n\pi)}{2\pi Wt - n\pi}$$

$$= \sum_{n=-\infty}^{\infty} g(\frac{n}{2W}) \text{sinc}(2Wt - n) \ , \ -\infty < t < \infty \qquad (1.10)$$

-Eq. (1.10) is an interpolation formula of $g(t)$

○ **Conclusion:**

→ The sampled signal $x_\delta(t)$ can be expressed mathematically in time domain as:

$$\mathbf{x_\delta(t) = \sum_{n=-\infty}^{\infty} x(nT_s).\delta(t - nT_s)}$$

→ The frequency domain representation of the sampled signal is given by:

$$\mathbf{X_\delta(f) = f_s \sum_{n=-\infty}^{\infty} X(f - nf_s)}$$

→ Advantages:

- It is more power saving
- No distortion in the sampled signal spectrum

→ Disadvantages:

- The sampled signal has infinite bandwidth

→ **Aliasing effect**

- If the sampling frequency is selected below the Nyquist frequency $f_s < 2f_m$, then $f_s(t)$ is said to be under sampled and aliasing occurs as shown in Figure 1.14
- Aliasing refers to the phenomenon of a high-frequency component in the spectrum of the signal seemingly taking on the identity of a lower frequency in the spectrum of its sampled version
- The aliased shown by the solid curve in Figure 1.14 pertains to an "under-sampled" version of the message signal
- To combat the effects of aliasing in practice,

  1. Prior to sampling, a low-pass pre-alias filter is used to attenuate those high-frequency components of the signal that are not essential to the information being conveyed by the signal

2. The filtered signal is sampled at a rate slightly higher than the Nyquist rate

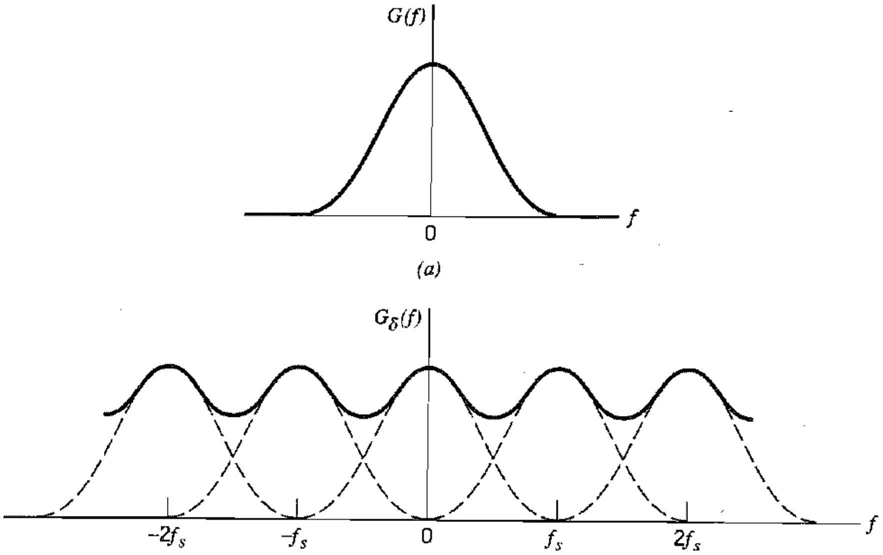


Figure 1.14: (a) Spectrum of a signal. (b) Spectrum of an undersampled version of the signal exhibiting the aliasing phenomenon

[Refer to figure (1.13) in the text book. Page 23]

– The use of a sampling rate higher than the Nyquist rate also has the beneficial effect of easing the design of the reconstruction filter used to recover the original signal from its sampled version

– Consider the example of a message signal that has been pre-alias (low-pass) filtered, resulting in the spectrum shown in Figure

– The corresponding spectrum of the instantaneously sampled version of the signal is shown in Figure, assuming a sampling rate higher than the Nyquist rate

– According to Figure 1.15, we readily see that the design of the reconstruction filter may be specified as follows:

→ The reconstruction filter is low-pass with a passband extending from –W to W, which is itself determined by the pre-alias filter

→ The filter has a transition band extending (for positive frequencies) from W to f_s - W , where f_s is the sampling rate

– The fact that the reconstruction filter has a well-defined transition band means that it is physically realizable

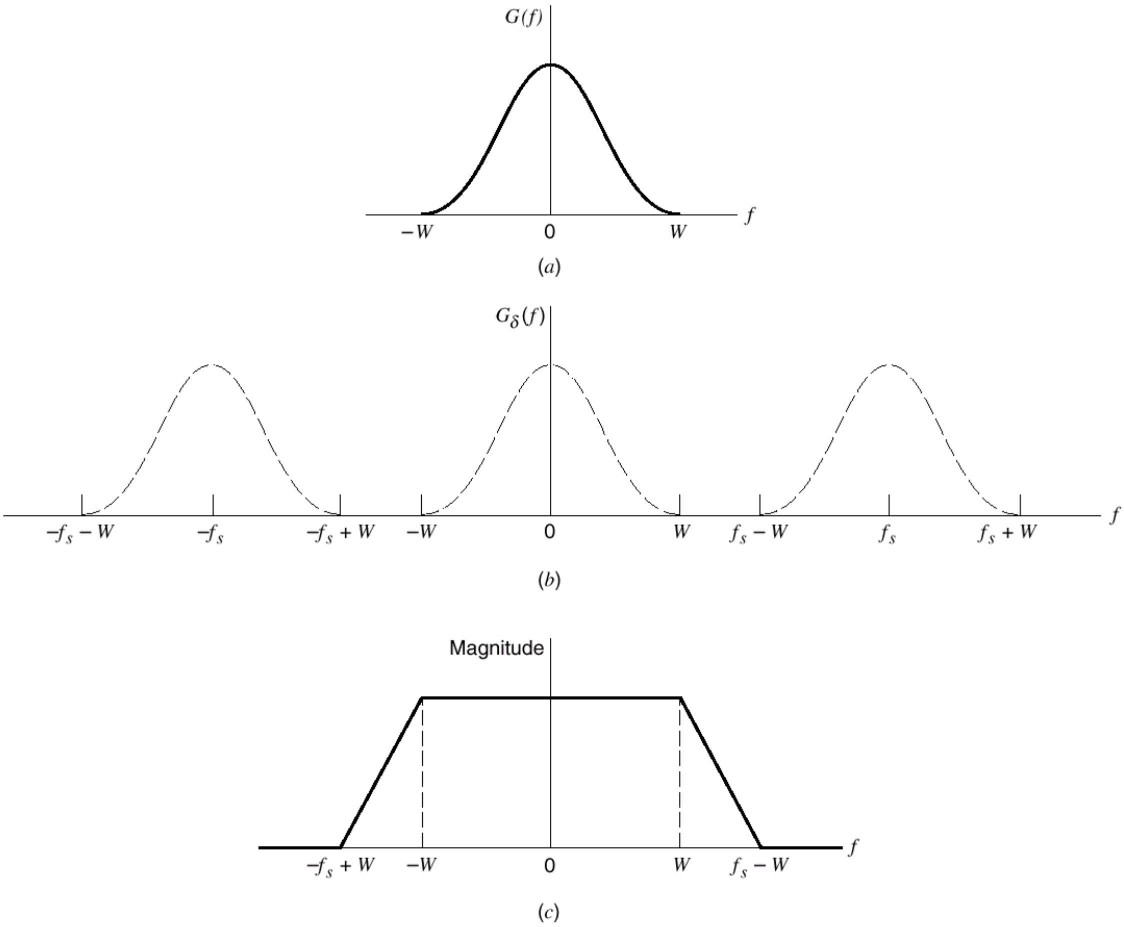


Figure 1.15: (a) Anti-alias filtered spectrum of an information-bearing signal. (b) Spectrum of instantaneously sampled version of the signal, assuming the use of a sampling rate greater than the Nyquist rate. (c) Magnitude response of reconstruction filter

[Refer to figure (1.14) in the text book. Page 24]

1.3 Pulse-Amplitude Modulation Analysis

1.3.1 Natural Sampling (Practical Sampling)

- In natural sampling the information signal $\boldsymbol{f(t)}$ is multiplied by a periodic pulse train with a finite pulse width τ as shown below
- As it can be seen from the figure shown, the natural sampling process produces a rectangular pulses whose amplitude and top curve depends on the amplitude and shape of the message signal $f(t)$

→ Figures in time and frequency domain

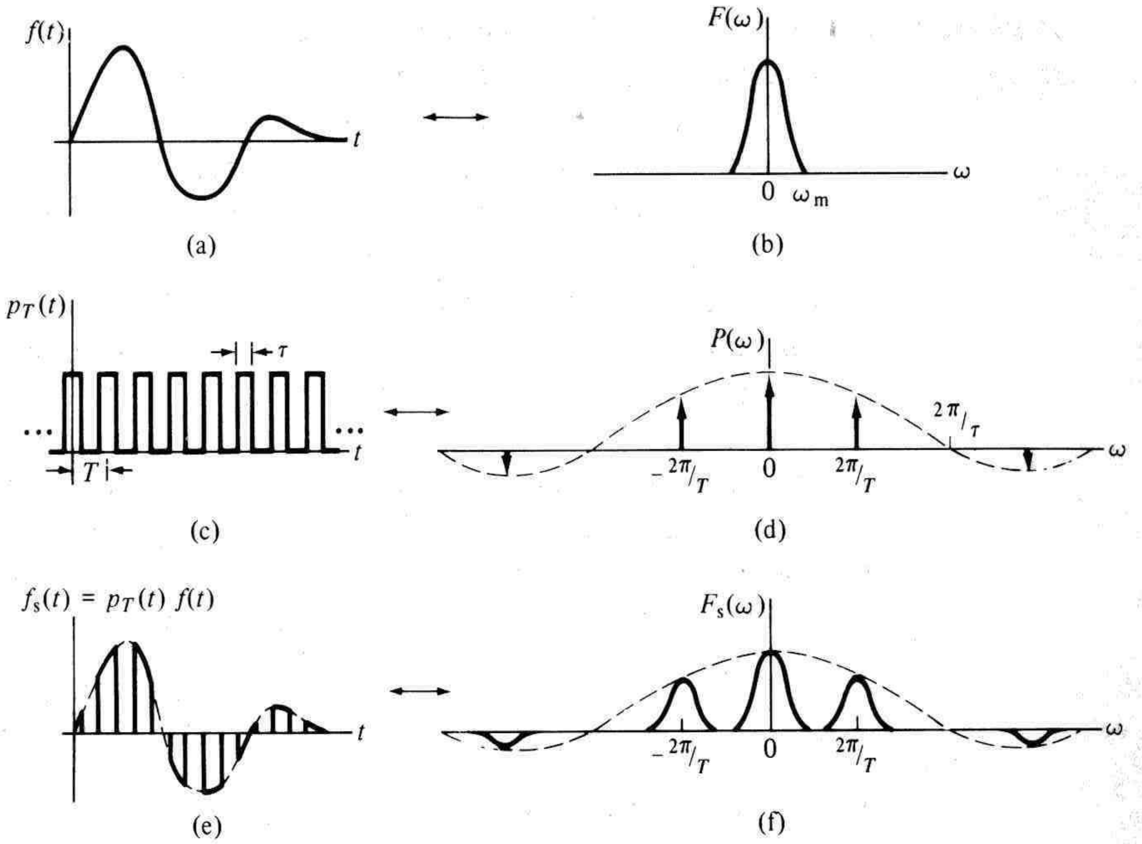


Fig. (1.16): Time domain: (a) Analog signal. (b) Sampling signal (c) Instantaneously sampled version of signal
Frequency domain: (a) Spectrum of a strictly band-limited signal (b) Spectrum of sampling signal (c) Spectrum of sampled version of the signal

→ Equation in time domain:

$$\mathbf{f_s(t)} = \boldsymbol{p_T}\mathbf{(t).f(t)} = \frac{\boldsymbol{T}}{\boldsymbol{T_s}}\sum_{\mathbf{n=-\infty}}^{\infty}\mathbf{Sinc(nf_s}\boldsymbol{T}\mathbf{).f(t)}\boldsymbol{e^{+j2\pi nf_s}}$$

→ Equation in frequency domain:

$$\boldsymbol{F_s}\mathbf{(f)} = \frac{\boldsymbol{T}}{\boldsymbol{T_s}}\sum_{\mathbf{n=-\infty}}^{\infty}\mathbf{Sinc(n}\boldsymbol{f_s}\mathbf{T).F(f-nf_s)}$$

→ Advantages:

- The bandwidth of the sampled signal could be limited due to "Sinc" function without distortion

→ Disadvantages:

- The top of the sample pulses are not flat
- It is not compatible with a digital system since the amplitude of each sample has infinite number of values

1.3.2 Flat-Top Sampling (Practical Sampling – Sample & Hold)

- Flat-top sampling is the most popular sampling method and involves two simple operations: sample and hold (Figure 1.17)

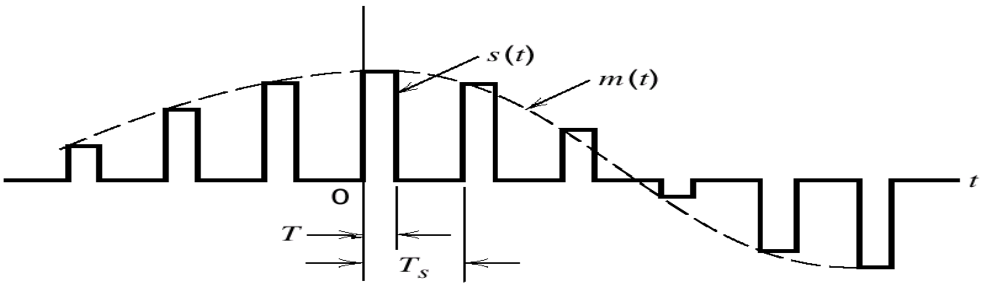


Fig. (1.17): Flat-top samples

[Refer to figure (1.15) in the text book. Page 26]

- The mathematical equations that describes the PAM in both time and frequency domain are described below

Let $s(t)$ denote the sequence of flat-top pulses as

$$s(t)=\sum_{n=-\infty}^{\infty} m(nT_s)\,h(t-nT_s) \qquad (1.11)$$

$$h(t)=\begin{cases} 1, & 0 < t < T \\ \frac{1}{2}, & t=0, t=T \\ 0, & \text{otherwise} \end{cases} \qquad (1.12)$$

The instantaneously sampled version of $m(t)$ is

$$m_\delta(t)=\sum_{n=-\infty}^{\infty} m(nT_s)\delta(t-nT_s) \qquad (1.13)$$

$$m_\delta(t)*h(t)=\int_{-\infty}^{\infty} m_\delta(\tau)h(t-\tau)d\tau$$

$$=\int_{-\infty}^{\infty}\sum_{n=-\infty}^{\infty} m(nT_s)\delta(\tau-nT_s)h(t-\tau)d\tau \qquad (1.14)$$

$$=\sum_{n=-\infty}^{\infty} m(nT_s)\int_{-\infty}^{\infty}\delta(\tau-nT_s)h(t-\tau)d\tau \qquad (1.15)$$

Using the sifting property , we have

$$m_\delta(t)*h(t)=\sum_{n=-\infty}^{\infty} m(nT_s)h(t-nT_s)=s(t) \qquad (1.16)$$

The PAM signal $s(t)$ is

$$s(t)=m_\delta(t)*h(t) \qquad (1.17)$$

$$\Leftrightarrow S(f)=M_\delta(f)H(f) \qquad (1.18)$$

Where $g_\delta(t)\Leftrightarrow f_s\sum_{m=-\infty}^{\infty} G(f-mf_s)$

$$M_\delta(f)=f_s\sum_{k=-\infty}^{\infty} M(f-k\,f_s) \qquad (1.19)$$

$$S(f)=f_s\sum_{k=-\infty}^{\infty} M(f-k\,f_s)H(f) \qquad (1.20)$$

→ **Recovering the original message signal m(t) from PAM signal**

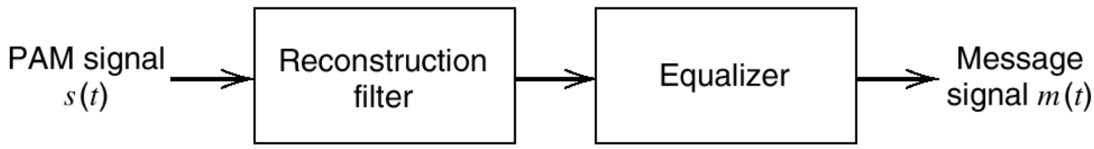


Fig. (1.18): Recovering m(t) from PAM signal s(t)

[Refer to figure (1.17) in the text book. Page 29]

Where the filter bandwidth is W

The filter output is $f_s M(f)H(f)$. Note that the

Fourier transform of $h(t)$ is given by

$$H(f) = T\ \mathrm{sinc}(fT)\exp(-j\pi fT) \tag{1.21}$$

amplitude distortion delay $= T/2$

$$\therefore S(f) = f_s T \sum_{k=-\infty}^{\infty} M(f - kf_s).\mathrm{sinc}(fT)\exp(-j\pi fT) \tag{1.22}$$

$\Rightarrow$ aparture effect

Let the equalizer response is

$$\left|H_{eq}(f)\right| = \frac{1}{\left|H(f)\right|} = \frac{1}{\left|T\ \mathrm{sinc}(fT)\right|} = \left|\frac{\pi f}{\sin(\pi fT)}\right| \tag{1.23}$$

Ideally the original signal $m(t)$ can be recovered completely.

→ Advantages:

- The bandwidth of the sampled signal could be limited due to "Sinc" function spectrum

→ Disadvantages:

- The distortion in the sampled signal due to "Sinc" function, solve it by connecting an equalizer after low-pass filter

1.4 Time Division Multiplexing (TDM)

- A sampled waveform is "off" most of the time, leaving the time between samples available for other purpose. In particular, sample values from several different signals can be interlaced into a single waveform. This is the principle of time division multiplexing (TDM) discussed here
- The sampling theorem provides the basis for transmitting the information contained in a band-limited message signal m(t) as a sequence of samples of m(t) taken uniformly at a rate that is usually slightly higher than the Nyquist rate
- An important feature of the sampling process is a conversion of time. That is, the transmission of the message samples engages the communication channel for use only a fraction of the sampling interval on a periodic basis, and in this way some of the time interval between adjacent samples is cleared for use by other independent message sources on a time-shared basis
- We thereby obtain a time-division multiplex (TDM) system, which enables the joint utilization of a common communication channel by a plurality of independent message sources without mutual interference among them

TDM Systems

- The simplified system in Figure 1.19 demonstrate the essential features of time-division multiplexing. Several input are pre-filtered by the bank of input LPFs and sampled sequentially
- The rotating sampling switch or commutator at the transmitter extracts one sample from each input per revolution. Hence, its output is a PAM waveform that contains the individual samples periodically interlaced in time

- A similar rotary switch at the receiver, called a de-commutator or distributor, separates the samples and distributes them to another bank of LPFs for reconstruction of the individual messages
- If all inputs have the same message bandwidth W, the commutator should rotate at rates $f_s \geq 2W$ so that successive samples from any one input are spaced by $T_s = 1/f_s \leq 1/2W$. The time interval T_s containing one sample from each input is called a frame
- If there are M input channels, the pulse-to-pulse spacing within a frame is $T_s/M = 1/Mf_s$. Thus the total number of pulses per second will be

$$r = Mfs \geq 2MW \qquad (1.24)$$

- Which represents the pulse rate or signaling rate of TDM signal

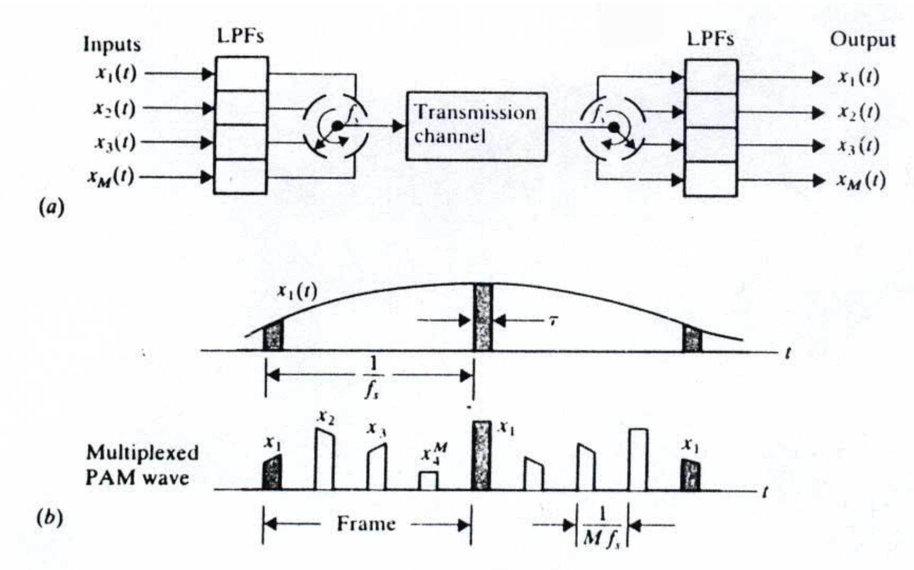


Fig. (1.19): TDM system. (a) Block diagram; (b) waveforms

[Refer to figure (1.19) in the text book. Page 31]

– TDM is a technique used for transmitting several message signals over a single communication channel by dividing the time frame into slots, one slot for each message signal.

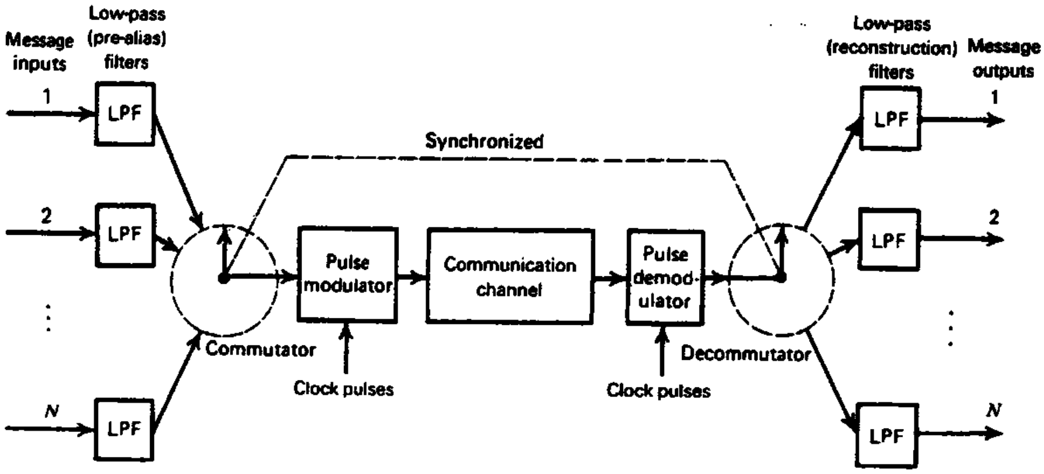


Fig. (1.20): Block diagram of TDM system

[Refer to figure (1.19) in the text book. Page 32]

→ In the time division multiplexing block diagram shown in Figure 1.20:

– Each message signal is first restricted in bandwidth be a low-pass pre-aliasing filter to remove the frequencies that are not essential which helps in reducing the aliasing problem. The outputs of these filters are then applied to a commutator. The functions of the commutator are:

I. allows narrow samples of each of the N input messages at a rate of fs and

II. Sequentially interleaves these N samples inside a sampling interval T_s.

– The multiplexed signal is then applied to a pulse amplitude modulator, which transforms the multiplexed signal into a form suitable for transmission over the communication channel.

– The time division scheme squeezes N samples derived from different N independent message signals into a time slot equal to one sampling interval. Thus the use of TDM introduces a bandwidth expansion factor N

– TDM-PAM Receiver: At the receiver end of the system, the received signal is applied to a pulse amplitude demodulator, which performs the reverse operation of the pulse amplitude modulator. The de-commutator distributes the appropriate pulses to the respective reconstruction filters. The de-commutator operates in synchronism with the commutator in the transmitter

→ The concept of TDM is indicated in the Figures 1.20 and Figure 1.21.

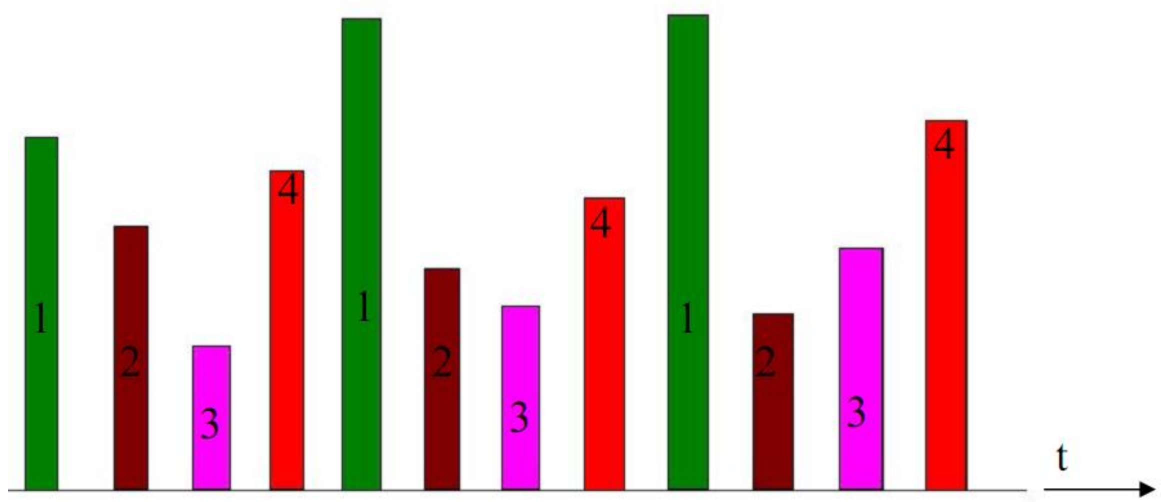


Fig. (1.21): Multiplexing of FOUR signals

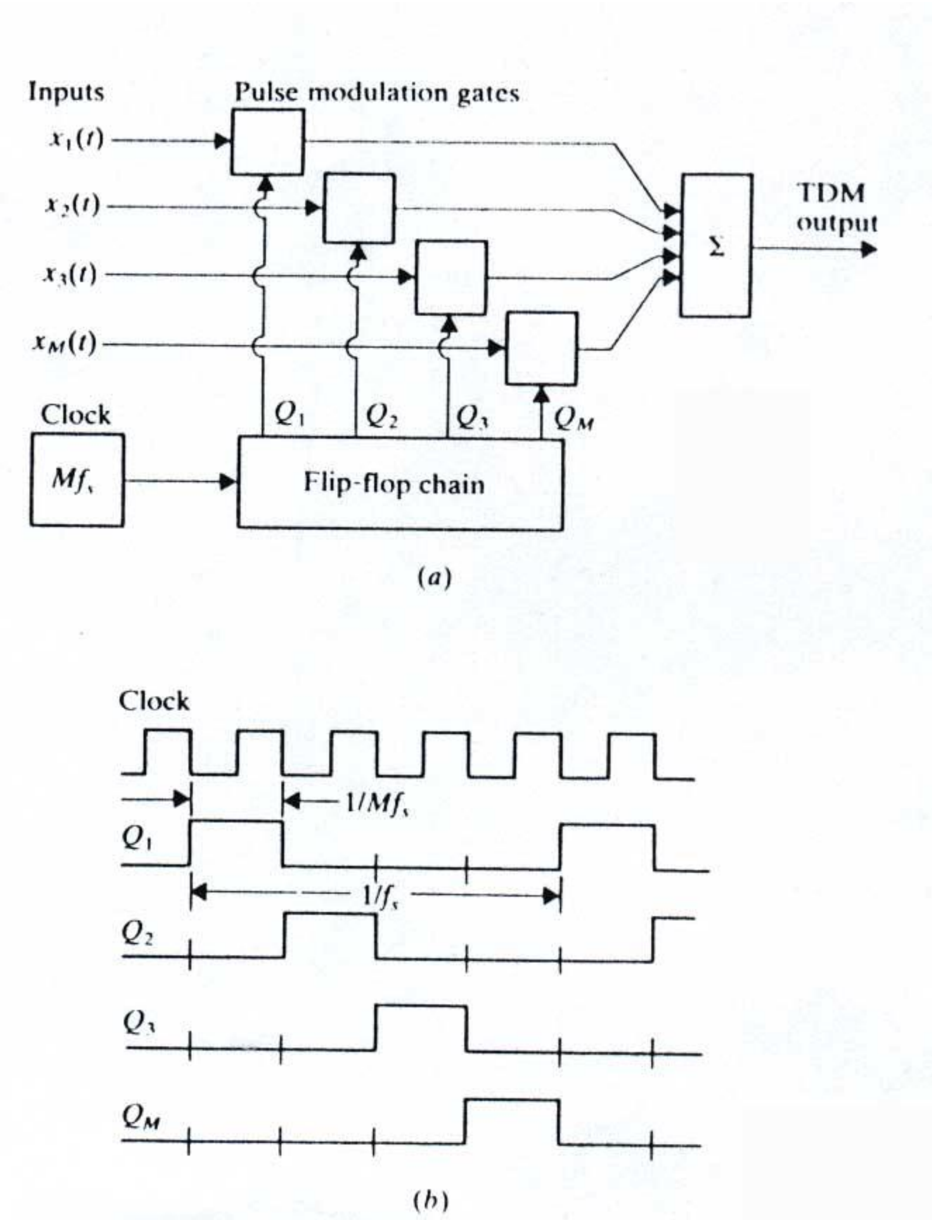


Fig. (1.22): (a) Electonic commutator for TDM; (b) Timing diagram

[Refer to figure (1.20) in the text book. Page 33]

- In FDM technique all multiplexing information sources emit there signals at the same time instant but each modulates different carrier frequency. C.W. modulation techniques are suitable for FDM system such as AM, FM, and PM for analog information signals. ASK, FSK, and PSK for digital information signals.

○ Figure 1.23 shows a block diagram of FDM

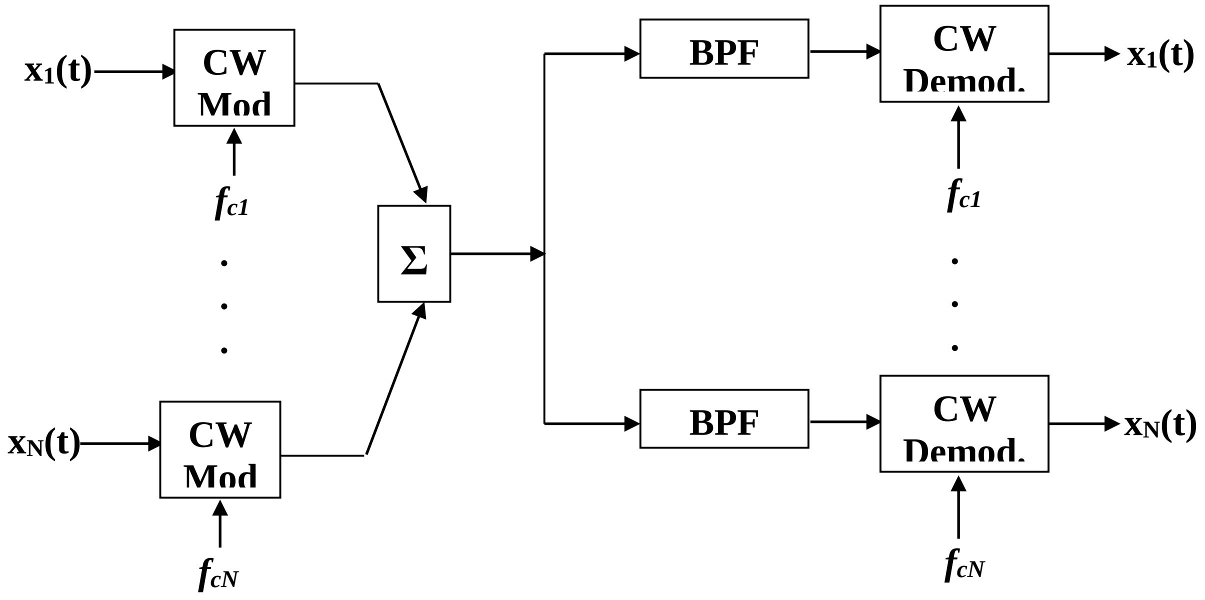


Fig. (1.23): FDM system

→ Comparison of TDM and FDM

Point of view	TDM	FDM
Main concept	The multiplexed signals have the same frequency but with different time slots	The multiplexed signals have the same time but with different frequency modulated carrier
Hardware implementation	Simpler implementation	Requires analog sub-carrier modulators, band-pass filter and demodulator for every channel
Techniques used	Pulse modulation techniques as PAM and PCM	Continuous wave techniques as AM-SSB and AM-DSB
Synchronization	Need synchronization between transmitter and receiver	Doesn’t need synchronization
Fading	Affected by slow narrow-band fading	Affected by rapid wide-band fading

1.5 Pulse-Width and Pulse-Position modulation analysis

- In a pulse modulation system, we may use the increased bandwidth consumed by pulses to obtain an improvement in noise performance by representing the sample values of the message signal by some property of the pulse other than amplitude
- In pulse-duration modulation (PDM), the samples of the message signal are used to vary the duration of the individual pulses. This form of modulation is also referred to as pulse-width modulation or pulse-length modulation. The modulating signal may vary the time of occurrence of the leading edge, the trailing edge, or both edges of the pulse. In Figure 1.24c the trailing edge of each pulse is varied in accordance with the message signal, assumed to be sinusoidal as shown in Figure 1.24a. The periodic pulse carrier is shown in Figure 1.24b

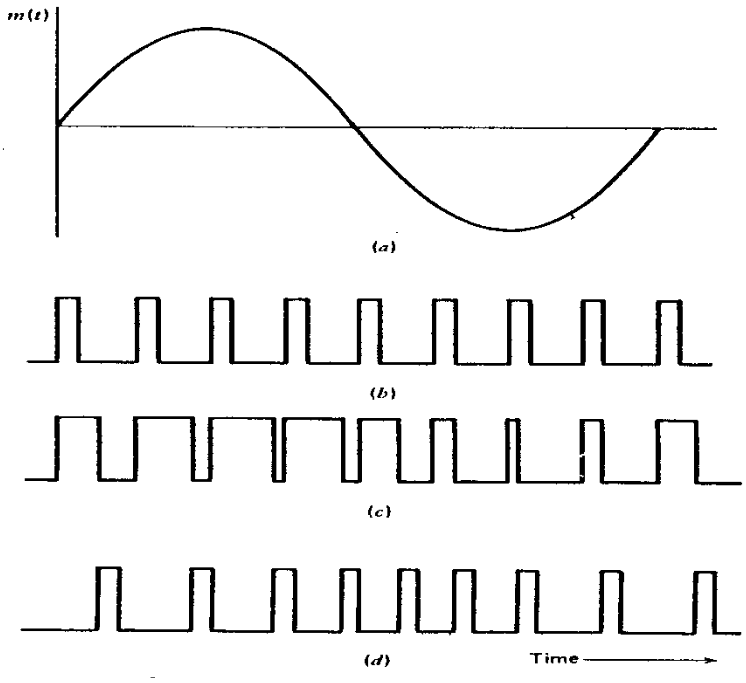


Fig. (1.24): Illustrating two different forms of pulse-time modulation for the case of a sinusoidal modulating wave. (a) Modulating wave. (b) Pulse carrier. (c) PDM wave. (d) PPM wave

[Refer to figure (1.22) in the text book. Page 36]

- In PDM, long pulses expend considerable power during the pulse while bearing no additional information. If this unused power is subtracted from PDM, so that only time transitions are preserved, we obtain a more efficient type of pulse modulation known as pulse-position modulation (PPM)
- In PPM, the position of a pulse relative to its unmodulated time of occurrence is varied in accordance with the message signal, as illustrated in Figure 1.24d for the case of sinusoidal modulation
- In other words we can lump PDM and PPM together under one heading for two reasons. First, in both cases a time parameter of the pulse is being modulated, and the pulses have constant amplitude. Second, a close relationship exists between the modulation methods for PDM and PPM
- To demonstrate these points, Figure 1.25 shows the block diagram and waveforms of a system that combines the sampling and modulation operations of either PDM or PPM. The system employs a comparator and a sawtooth-wave generator with period T_s. The output of the comparator is zero except when the message waveform x(t) exceeds the sawtooth wave, in which case the output is a positive constant A
- Hence, as seen in the figure, the comparator produces a PDM signal with trailing edge modulation of the pulse duration. (Reversing the sawtooth results in leading edge modulation on both edges)
- Position modulation is obtained by applying the PDM signal to a monostabel pulse generator that triggers on trailing edges at its input and produces short output pulses of fixed duration

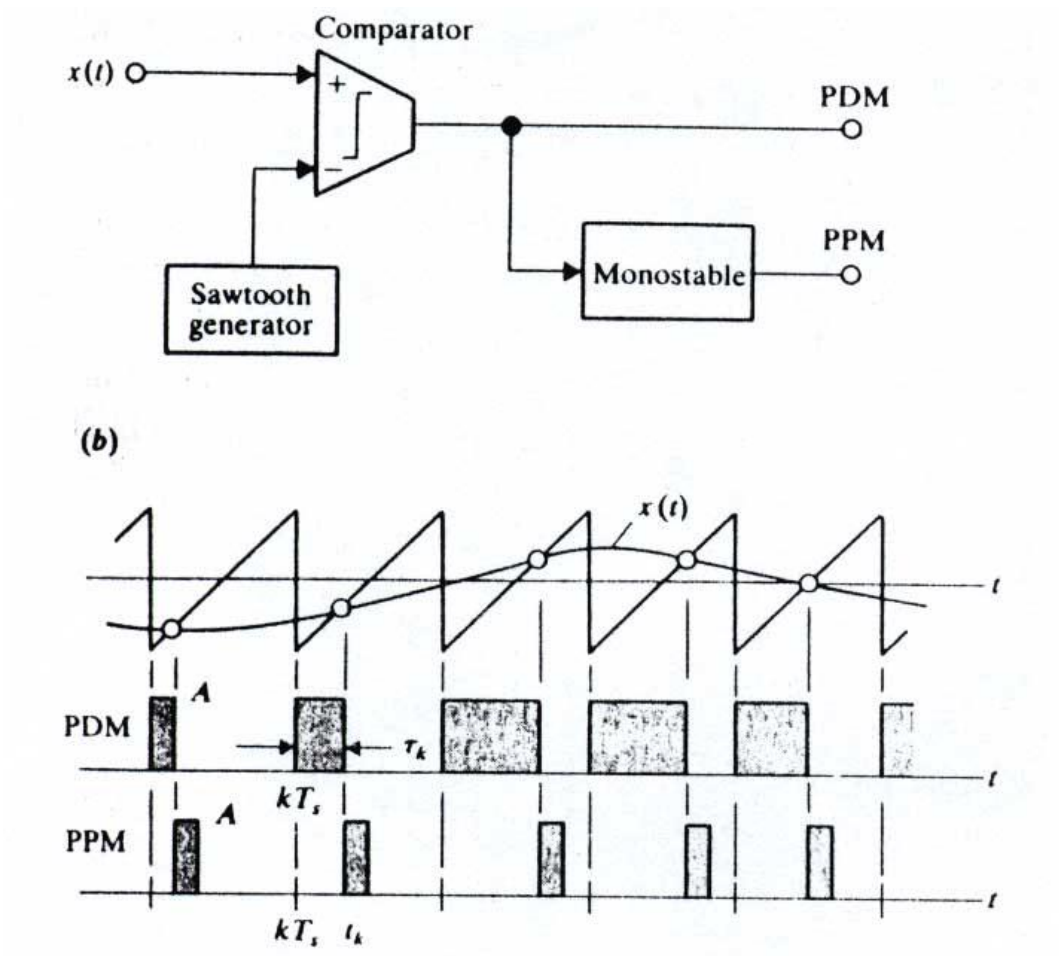


Fig. (1.25): Generation of PDM and PPM. (a) Block diagram. (b) waveforms

[Refer to figure (1.23) in the text book. Page 37]

- Let T_s denote the sample duration. Using the sample $m(nT_s)$ of a message signal $m(t)$ to modulate the position of the nth pulse, we obtain the PPM signal

$$\mathbf{s(t)} = \sum_{n=-\infty}^{\infty} \mathbf{g}\left(\mathbf{t} - \mathbf{nT_s} - \mathbf{K_p}\boldsymbol{m}(\mathbf{nT_s})\right) \qquad (1.25)$$

- Where K_p is the sensitivity of the pulse position modulator and $g(t)$ denotes a standard pulse of interest. Clearly, the different pulses constituting the PPM signal $s(t)$ must be strictly non-overlapping, a sufficient condition for this requirement to be satisfied is to have

$$G(t) = 0, \ |t| > \frac{T_s}{2} - K_p|m(t)|_{max} \qquad (1.26)$$

- Which, in turn, requires that

$$\boldsymbol{K_p |m(t)|_{max} < \frac{T_s}{2}} \quad (1.27)$$

- The closer $\boldsymbol{K_p|m(t)|_{max}}$ is to one half the sampling duration T_s, the narrower must the standard pulse g(t) be in order to ensure that the individual pulses of the PPM signal s(t) do not interfere with each other, and the wider will the bandwidth occupied by the PPM signal be
- Assuming that Eq. (1.26) is satisfied, and that there is no interference between adjacent pulses of the PPM signal s(t), then the signal samples $m(nT_s)$ can be recovered perfectly. Furthermore, if the message signal m(t) is strictly band limited, it follows from the sampling theorem that the original message signal m(t) can be recovered from the PPM signal s(t) without distortion

Generation of PPM waves

- The PPM signal described by Eq. (1.25) may be also generated using the system described in Figure 1.27 the message signal m(t) is first converted into a PAM signal by means as a sample-and-hold circuit, generating a staircase waveform u(t); note that the pulse duration T of the sample-and-hold circuit is the same as the sampling duration T_s. This operation is illustrated in Figure 1.27b for the message signal m(t) shown in Figure 1.27a.
- Next, the signal u(t) is added to a sawtooth wave shown in Figure 1.27c, yielding the combined signal v(t) shown in Figure 1.27d. The combined signal v(t) is applied to a threshold detector that produces a very narrow pulse (approximating an impulse) each time v(t) passes through a zero-crossing in the negative-going direction. The resulting sequence of "impulses" i(t) is shown in Figure 1.27e. Finally, the PPM signal s(t) is generated by using this sequence of impulses to excite a filter whose impulse response is defined by the standard pulse g(t)

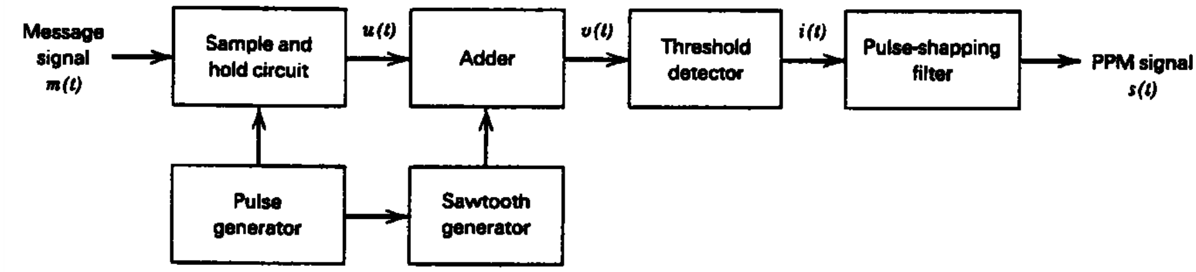


Fig. (1.26): Block diagram of PPM generator

[Refer to figure (1.24) in the text book. Page 39]

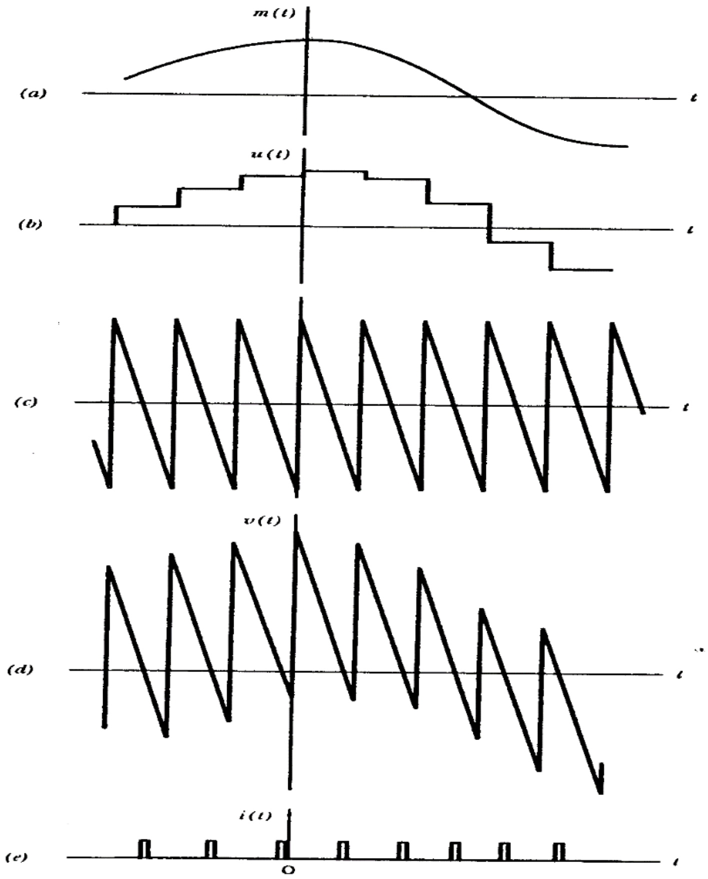


Fig. (1.27): Generation of PPM. (a) Message signal. (b) Staircase approximation of the message signal. (c) Sawtooth wave. (d) Composite wave obtained by adding (b) and (c). (e) Sequence of "impulses" used to generate the PPM signal

[Refer to figure (1.25) in the text book. Page 39]

Detection of PPM waves

– Another message reconstruction technique converts pulse-time modulation into pulse-amplitude modulation, and works for PDM and

PPM. To illustrate this technique the middle waveform in Figure 1.28 is produced by a ramp generator that starts at time KT_s, stops at t_K, restarts at $(K+1)T_s$, and so forth. Both the start and stop commands can be extracted from the edges of a PDM pulse, whereas PPM reconstruction must have an auxiliary synchronization signal for the start command

- Regardless of the particular details, demodulation of PDM or PPM requires received pulses with short rise time in order to preserve accurate message information

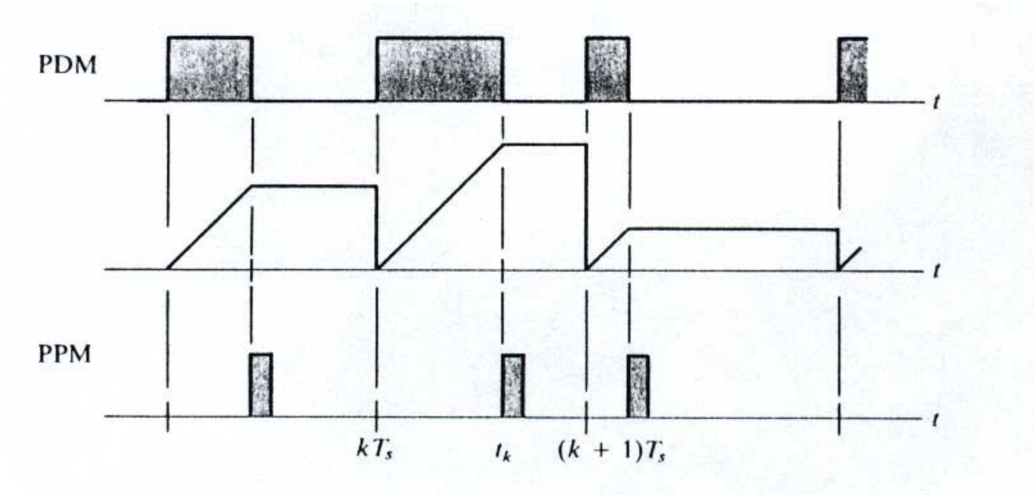


Fig. (1.28): Conversion of PDM or PPM into PAM

[Refer to figure (1.26) in the text book. Page 40]

- Additionally, like PM and FM CW modulation, PDM and PPM have the potential for wideband noise reduction – a potential more fully realized by PPM than by PDM. To appreciate why this is so, recall that the information resides in the time location of the pulse edges, not in the pulses themselves. Thus, somewhat like the carrier-frequency power of AM, the pulse power of pulse-time modulation is "wasted" power, and it would be more efficient to suppress the pulses and just transmit the edges. Of course we cannot transmit edges without transmitting pulses to define them. But we can send very short pulses indicating the position of the edges, a process equivalent to PPM. The reduced power required for PPM is a fundamental advantage over PDM, an advantage that becomes more apparent when we examine the signal-to-noise ratios

- As conclusion consider a PPM wave s(t) with uniform sampling, as defined by Eq. (1.25) and (1.27), and assume that the message (modulating) signal m(t) is strictly band-limited. The operation of one type of PPM receiver may proceed as follows
  - Convert the received PPM wave into a PDM wave with the same modulation
  - Integrate this PDM wave using a device with a finite integration time, thereby computing the area under each pulse of the PDM wave
  - Sample the output of the integrator at a uniform rate to produce a PAM wave, whose pulse amplitudes are proportional to the signal samples m(nTs) of the original PPM wave s(t)
  - Finally, demodulate the PAM wave to recover the message signal m(t)

Ch.[2] Pulse Code Modulation

2.1 Introduction

- Although PAM and PTM, studied in last chapter different from AM and FM because, unlike in those two continuous forms of modulation, the signal was sampled and sent in pulse form. But still like AM and FM, they were forms of analog communication in all these forms a signal is sent which has a characteristic that is infinitely variable and proportional to the modulating voltage concerning PCM it is in common with the other forms of pulse modulation, PCM also uses the sampling technique, but it differs from the others in that it is a digital process

- That is, instead of sending a pulse train capable of continuously varying on of the parameters, the PCM generator produces a series of numbers, or digits (hence the name digital process). Each one of these digits, almost always in binary code, represents the approximation amplitude of the signal sample at that instant. The approximation can be made as close as desired, but it is always just an approximation

2.2 Principle of PCM

Basic steps used to transform analog signal into PCM signal

1) Pre- Aliasing Filter: It is Low-pass filter to band limited the information signal to B Hz to prevent aliasing problem from appearing at receiving end

2) Sampling: Convert the continuous-time signal into a discrete-time signal by taking "samples" every fixed time interval called sampling time T_S where $T_s = 1/F_s$

3) Quantization: This is the conversion of a discrete-time continuous-valued signal into a discrete-time discrete-valued signal by making approximation to pre-determined quantization levels.

4) Coding: In the coding process, each discrete value is represented by an n-bit binary sequence.

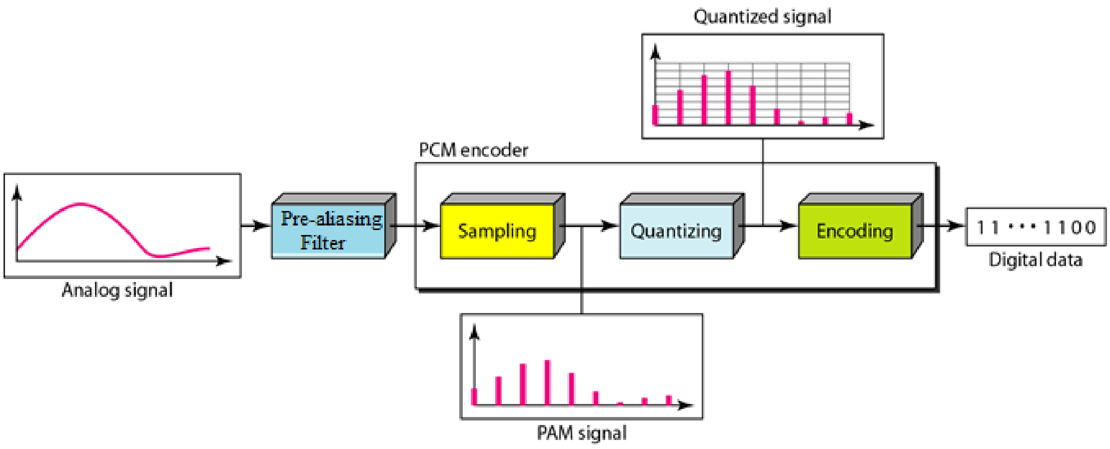


Figure (2.1): The basic elements of PCM

→ Major advantages of PCM modulation:

- Cheaper and high efficiency
- Less effect to noise and distortion
- Repeaters don't have to be replaced so close together
- Digital hardware can be used
- Using TDM multiplexing

→ Disadvantages of PCM modulation:

- High bandwidth
- Hardware is complex
- Quantization noise

2.2.1 Quantization noise (error)

- When a signal is quantized, we introduce an error - the coded signal is an approximation of the actual amplitude value.

- The difference between actual and coded value (midpoint) is referred to as the quantization noise and it is completely unpredictable, i.e. random

- The more levels, the smaller Δ which results in smaller errors. BUT, the more levels the more bits required to encode the samples (i.e. higher bit rate). In practical systems 128 levels for speech is considered quite adequate

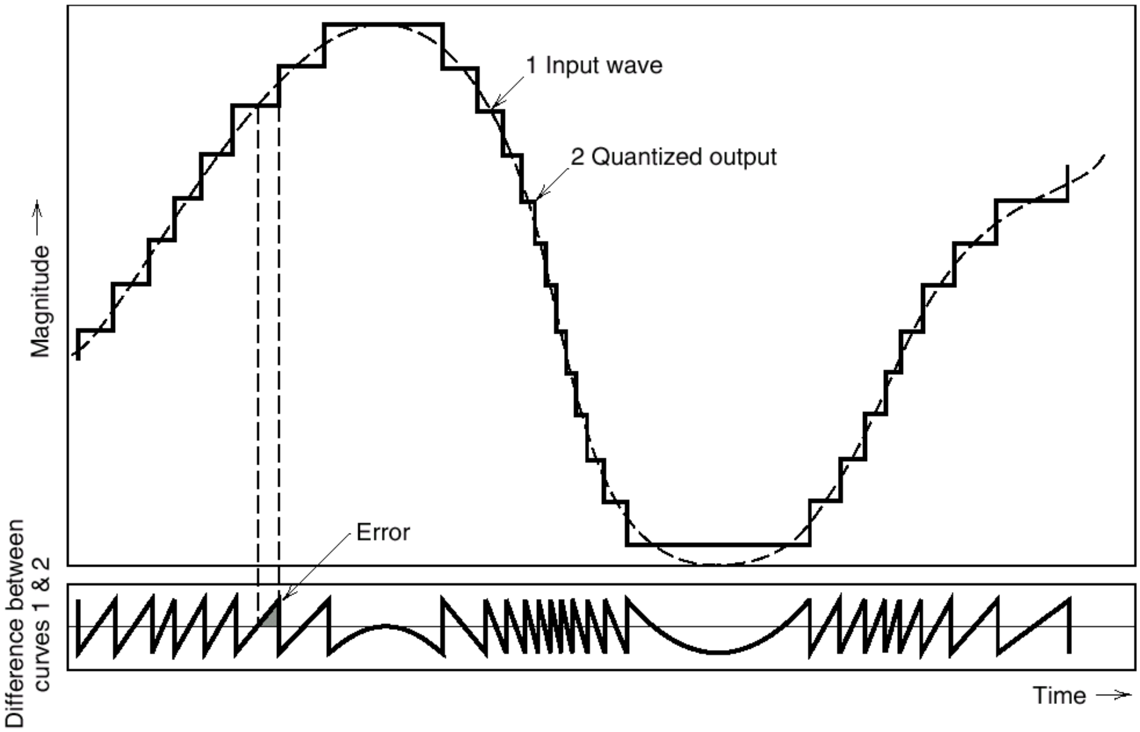


Figure (2.2): Illustration of the quantization process & quantization noise

- It will be noted that the biggest error that can be occur is equal to half the size of the sampling interval as shown in Figure 2.3

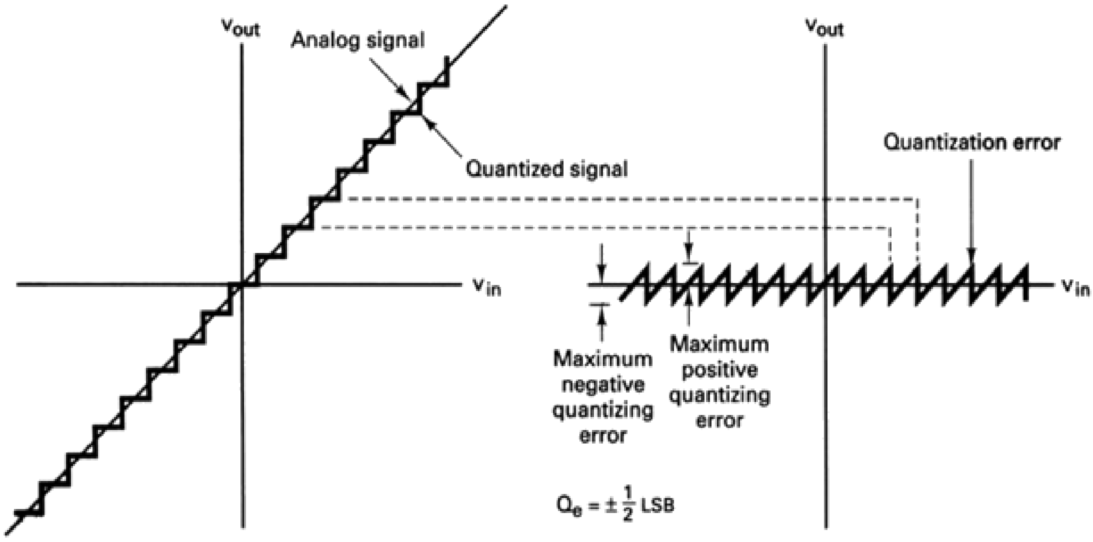


Figure (2.3): Maximum quantization noise

2.2.2 Generation and demodulation of PCM

- Essentially, the signal is sampled and converted to PAM, the PAM is quantized and encoded, and supervisory signals are added. The signal is then sent directly via cable, or modulated and transmitted. Because PCM is highly immune to noise, amplitude modulation may be used, so that PCM-AM is quite common

- At the receiver, the signaling information is extracted, and PCM is translated into corresponding PAM pulses which are then demodulated in the usual way. In fact, the quantized wave of Figure 2.2 would be the output for that signal from an ideal PCM receiver

2.2.3 Effects of noise

- Those forms of pulse modulation which, like FM, transmit constant-amplitude signals, are equally amenable to signal-to-noise-ratio improvement with amplitude limiters of one form or another

- Thus PWM, PPM and PCM have all the advantages of frequency modulation when it comes to noise performance; this is best illustrated by means of Figure 2.4

- Figure 2.4.a shows the effect of noise being superimposed on pulses with vertical sides. It is seen that noise will have no effect at all unless its peaks are so large that they can be mistaken for pulses, or so large negatively that they can mask legitimate pulses

- This is ensured by the slicer, or double clipper, which selects the amplitude range between x and y in Figure 2.4 for further transmission, thus removing all the effects of noise

- The transmitted pulses in a practical system cannot have sides with perfectly vertical slopes. These must "lean", as shown in an exaggerated fashion in Figure 2.4.b. Noise will now superimpose itself on the pulses sides, and the result may well be a change in width or position as shown. This will affect PWM and PPM, but not nearly as much as an amplitude change would have affected PAM

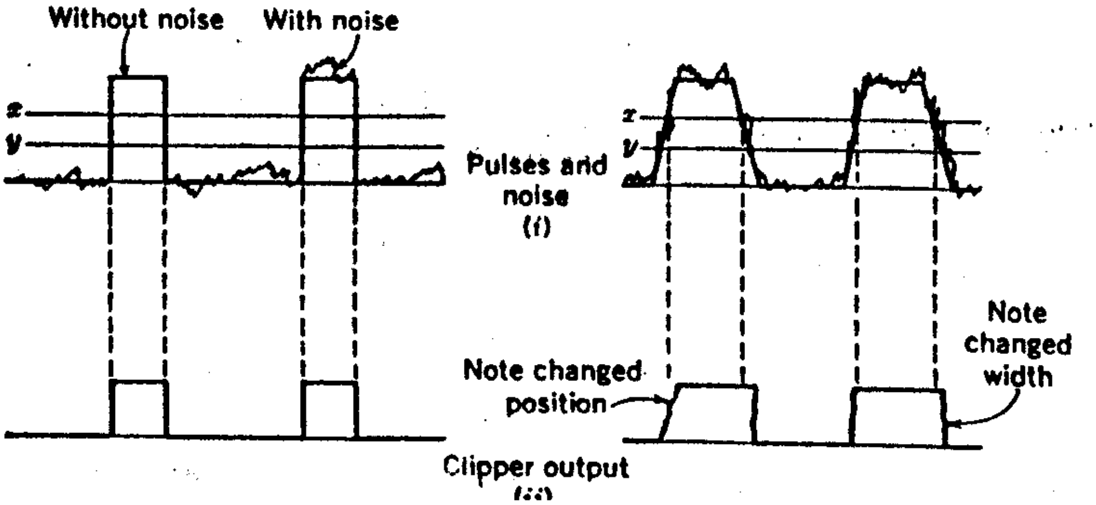


Figure (2.4): Effects of noise on pulses. (a) Vertical pulses; (b) Sloping pulses

[Refer to figure (2.2) in the text book. Page 49]

- Furthermore, the obvious method of reducing the effects of noise in PWM or PPM is to send pulses with steeper sides. This will increase the bandwidth required, so that these two forms of pulse modulation share with frequency modulation the ability to trade bandwidth for improved noise performance

- Good as these systems are, PCM is much better for noise immunity. As shown in Figure 2.4.b sloping pulses are affected by noise, but this does not matter in PCM at all. Provided that the signal-to-noise ratio is not so poor that noise pulses can be mistaken for normal pulses or can remove them, the effect of noise on PCM will be nil

- This is because PCM depends only on the presence or absence of pulses at any given time, not on any characteristic of the pulses which could be distorted. It is possible to predict statistically the error rate in PCM due to random noise

- Consider a channel signal-to-noise ratio of 17 db. For PCM in a 4-KHz channel, this would yield an error rate close to 1 error in 10,000 character sent. When S/N = 23 dB, the error rate falls to 1 in about 8 x 108. Such an error rate is negligible, corresponding to about one error every 30 hours in a system sampling 8000 times per second, using an 8-bit code, and operating 24 hours per day

- Since errors will occur in PCM only when noise pulses are large enough, it is seen that the digital modulation system does not suffer from a gradual, deterioration; Indeed, pulse-code modulation can be relayed without degradation when the signal-to-noise ratio exceeds about 21 dB. This gives PCM an enormous advantage over analog modulation in relay systems, since even in the best of analog systems some degradation will occur along every link and through every repeater

- Since the process is cumulative, any such system will have to start with a much higher S/N ratio and also use more low-noise equipment in route than would have been needed by PCM. The ability to be relayed without any distortion and to use poor-quality transmission paths is a very significant reason to use digital, rather than analog, modulation systems

2.2.4 Tapered Quantizer & Companding

→ A simple calculations showed that, with 16 standard levels, the maximum error is 1/32 of the total signal amplitude range. In a practical system with 128 levels, this maximum error is 1/256 of the total amplitude range. That is quite small and considered tolerable, provided the signal has an amplitude somewhere near to the maximum possible. If a small signal has a peak-to-peak amplitude which is 1/64 of the maximum possible, the quantizing error in the 128-level system could be as large as 1/256 – 1/4 ÷ 1/4 of the peak-to-peak value of this small signal. A value as high as that is not tolerable

→ Solving this problem by:

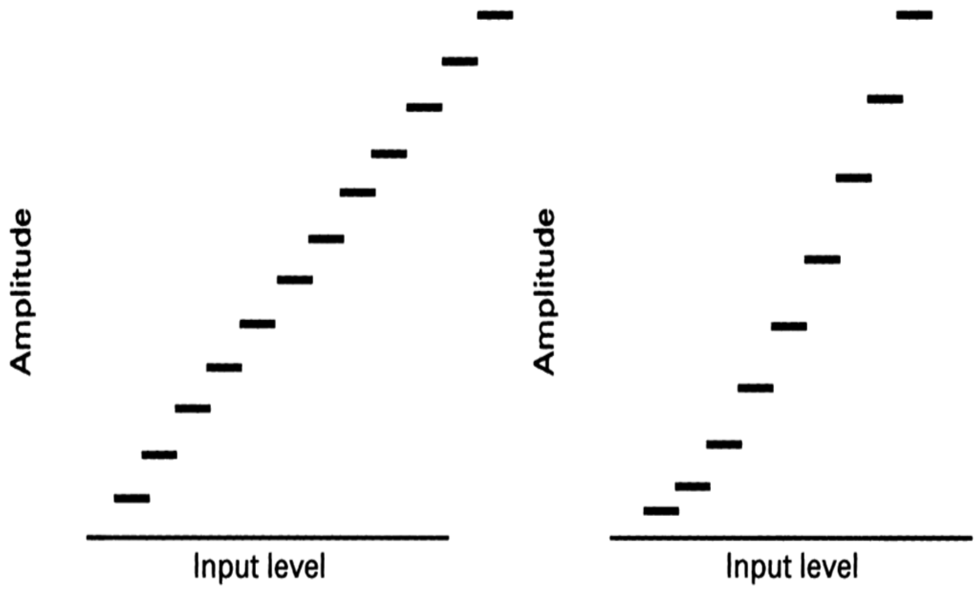
i. Tapered Quantizer
ii. Companding

i. Tapered Quantizer

- An obvious cure for the problem is to have tapered quantizing levels instead of constant-amplitude difference ones

- That is, the difference between adjoining levels can be made small signals, and gradually larger for larger signals as shown in Figure 2.5

- The quantizing noise could be "distributed" so as to affect small signals somewhat less and large signals somewhat more. In practice, this kind of tapered system is difficult to implement, because it would significantly complicate the (already complex) quantizer design. There is a suitable alternative



Linear **Nonlinear**

Figure (2.5): Linear versus Non-linear (tapered) quantization

ii. Companding

- It is possible to pre-distort the signal before it is modulated, and "un-distort" it after demodulation. This is always done in practice, and is shown in Figure 2.6

- The process is known as "companding", since it consists of compressing the signal at the transmitter and expanding it at the receiver

- With companding exactly the same results are obtained as with tapered quantizing, but much more easily. The signal to be transmitted is passes through an amplifier which has a correctly adjusted nonlinear transfer characteristic, favoring small-amplitude signals

- These are then artificially large when they are quantized, and so the effect of quantizing noise upon them is reduced. The correct amplitude relations are restored by the expander in the receiver

- It should be noted that disagreement about the companding law (between the United States and Europe) has been a real problem in establishing worldwide PCM standards for telephony. There has at least been agreement that companding should be used, and standard converters are available

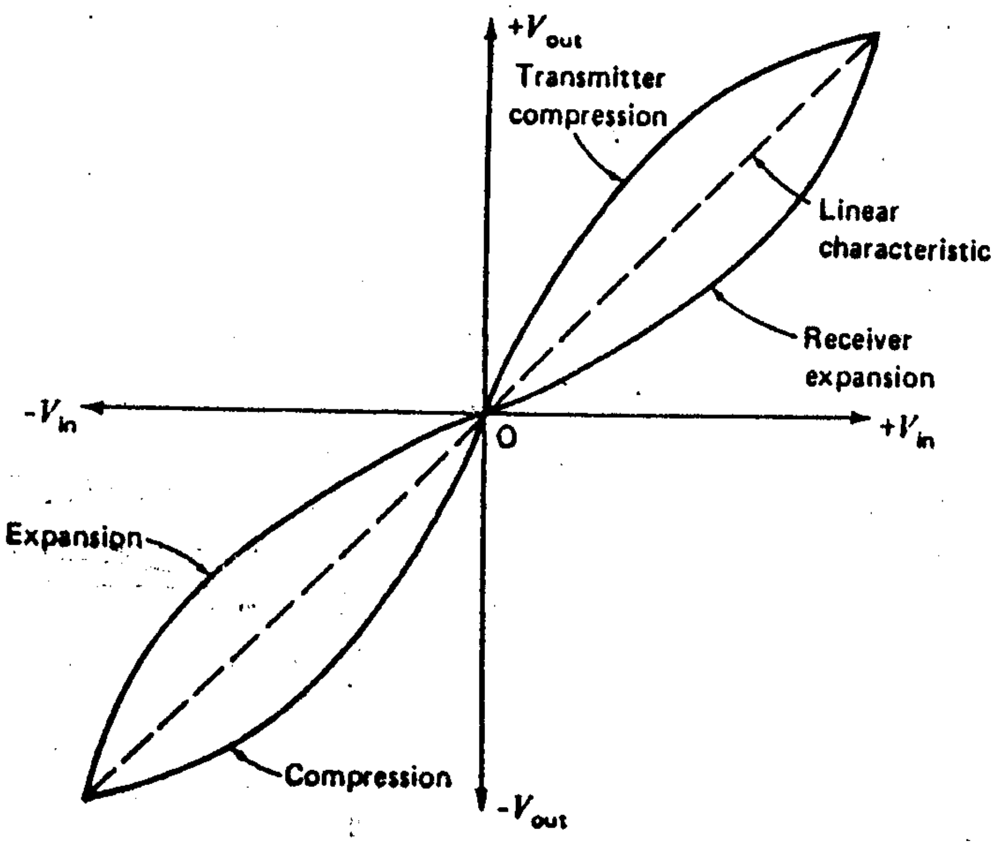


Figure (2.6): Companding curves for PCM

[Refer to figure (2.3) in the text book. Page 51]

2.2.5 Advantages and Applications of PCM

– A person may well ask, at this stage, "If PCM is so marvelous, why are any other modulation system used?", there three answers to this question, namely:

1. The other systems came first
2. PCM require very complex encoding and quantizing circuitry
3. PCM requires a large bandwidth compared to analog systems

– Concerning point (1): PCM was invented by Alex. H. Reeves in Great Britain, in 1937. Its first practical application in commercial telephony was in short-distance, medium-density work, in Great Britain and the United State in the early 1960s. Semiconductors and integration (it was not yet "large-scale" then made its use practicable. Quite a number of new communication facilities built around the world have used PCM, and its use has grown very markedly during the 1980s

– As regards the second point, it is perfectly true that PCM requires much more complex modulating procedures than analog systems. However, multiplexing equipment is very much cheaper, and repeaters do not have to be placed so close together because PCM tolerates much worse signal-to-noise ratios especially because of very large-scale integration, the complexity of PCM is no longer a significant cost penalty

– Concerning point (3): although the large bandwidth requirements still represent a problem, it is no longer as serious as it had earlier been, because of the advent of large-bandwidth fiber-optic systems. However, the large bandwidth requirements should be recognized. A typical first-level PCM system is the Bell T1 digital transmission system is use in North America

- It provides 24 PCM channels with time-division multiplexing. Each channel requires 8 bits per sample and thus 24 channels will need 24 x 8 + 1 = 193 bits, the extra bit is an additional sync signal. With a sampling rate of 8000 samples per second, a total of 8000 x 193 = 1,544,000 bps will be sent by using this system. Work showed that the bandwidth in hertz would have to be at least half that figure, but the practical system in fact uses a bandwidth of 1.5 MHz as an optimum figure.

- It can be shown that 24 channels correspond to two groups, requiring a bandwidth of 96 KHz if frequency-division multiplex is used. PCM is seen to require 16 times as much for the same number of channels. However, the situation in practice is not quite so bad, because economies of scale begin to appear when higher levels of digital multiplexing are used

- The following considerations ensured that, the main application of PCM for telephony was in 24-channel frames over wire pairs which previously had carried only one telephone conversation each. Their performance was not good enough to provide 24 FDM channels, but after a little modification 24 PCM channels could be carried over the one pair of wires
- Since the mid-1970s the picture has changed dramatically. First, very large-scale integration reduced costs significantly. Then came the advancement of digital systems, such as data transmission, which were clearly advantaged by not having to be converted into analog prior to transmission and reconverted to digital after reception

- Finally, fiber-optic systems became practical, with two effects. On the one hand, the current state of development of lasers and receiving diodes is such that digital operation is preferable to analog because of nonlinearities huge bandwidth, e.g., 565 Mbps per pair of fibers, have become available without attendant huge costs. The use of PCM in the broadband networks of advanced countries is increasing

– PCM also finds use in space communications. Indeed, the Mariner IV probe was an excellent example of the noise immunity of PCM, when, back in 1965, it transmitted the first pictures of Mars. Admittedly, each picture took 30 minutes to transmit, whereas it takes only 1/30 s in TV broadcasting. The Mariner IB transmitter was just over 200,000,000 Km away, and the transmitting power was only 10 W. PCM was used; no other system would have done the job

→ Applications of PCM modulation:

- Using in space communication due to it has high noise immunity
- Using in broad-band network with optical fiber
- Using in time-division multiplexing (Ex.: in telephone system)
- Using in data transmission

2.2.6 Other Digital Pulse Modulation Systems

PCM was the first digital system, but by now several others have been proposed. The major ones will now be mentioned, but it should be noted that none of them is in widespread use.

Differential PCM is quite similar to ordinary PCM. However, each word in this system indicates the difference in amplitude, positive or negative, between this sample and the previous sample. Thus the relative value of each sample is indicated, rather than the absolute value as in normal PCM. The rationale behind this system is that speech is redundant, to the extent that each amplitude is related to the previous amplitude, so that large variations from one sample to the next are unlikely. This being the case, it would take fewer bits to indicate the size of the amplitude change than the absolute amplitude, and so a smaller bandwidth would be required for the transmission. The differential PCM system has not found wide acceptance because complications in the encoding and decoding process appear to outweigh any advantages gained.

Delta modulation is a digital modulation system which has many forms, but at its simplest it may be equated with the basic form of differential PCM. In the simple form of delta modulation, there is just 1 bit sent per sample, to indicate whether the signal is larger or smaller than the previous sample. This system has the attraction of extremely simple coding and decoding procedures, and the quantizing process is also very simple. However, delta modulation cannot readily handle rapid amplitude variations, and so quantizing noise tends to be very high. Even with companding and more complex versions of delta modulation, it has been found that the transmission rate must be close to 100 kbits per second to give the same performance for a telephone channel as PCM gives with 64 kbits/s (8000 samples per second X 8 bits per sample). *Other digital systems* also exist.

2.3 Sampling theorem

- As was seen previously analog signals can be digitized through sampling and quantization. The sampling rate must be sufficiently large so that the analog signal can be reconstructed from the samples with sufficient accuracy

- The sampling theorem, which is the basis for determining the proper sampling rate for a given signal, has a deep significant in signal processing and communication theory

- We now show that a signal whose spectrum is band-limited to B Hz [G(ω) = 0 for │ω│ > 2πB can be reconstructed exactly (without an error) from its samples taken uniformly at a rate R > 2B Hz (samples per second). In other words, the minimum sampling frequency if f_s = 2B Hz

- To prove the ideal sampling theorem, consider a signal g(t) (Figure 2.7a) whose spectrum is band-limited to B Hz (Figure 2.7b). For convenience, spectra are shown as functions of ω as well as of f(Hz). Sampling g(t) at a rate of fs Hz (fs samples per second) can be accomplished by multiplying g(t) by an impulse train δ_{Ts} Figure 2.7c, consisting of unit impulses repeating periodically every T_s seconds, where $T_s = 1/f_s$

- This results in the sampled signal g(t) shown in Figure 2.7d. The sampled signal consists of impulses spaced every T_s seconds (the sampling interval). The nth impulse located at t = nT_s, has a strength $g(nT_s)$, the value of g(t) at t = nT_s. Thus,

$$\bar{g}(t) = g(t)\delta_{T_s} = \sum_n g(nT_s)\delta(t - nT_s) \qquad (2.1)$$

- Because the impulse train δ_{Ts} is a periodic signal of periodic Ts, it can be expressed as Fourier series. The trigonometric Fourier series, already found earlier is

$$\delta_{T_s}(t) = \frac{1}{T_s}[1 + 2cos\omega_s t + 2cos2\omega_s t + 2cos3\omega_s t + \cdots. \qquad (2.2)$$

$$\omega_s = \frac{2\pi}{T_s} = 2\pi f_s$$

- Therefore, for

$$\bar{g}(t) = g(t)\delta_{T_s}(t)$$

$$\therefore \bar{g}(t) = \frac{1}{T_s}[g(t) + 2g(t)cos\omega_s t + 2g(t)cos2\omega_s t + 2g(t)cos3\omega_s t + \cdots \quad (2.3)$$

- To find $\bar{G}(\omega)$, the Fourier transform og $\bar{g}(t)$, we take the Fourier transform of the right-hand side of Eq. (2.3), term by term. The transform of the first term in the brackets is G(ω). The transform of the second term 2g(t) cosω_st is G(ω-ω_s) + G(ω+ω_s). This represent spectrum G(ω) shifted to ωs and –ωs. Similarly, the transform of the third term 2g(t) 2cosωst is G(ω-2ω_s) + G(ω+ω_s), which represent the spectrum G(ω) shifted to 2ωs and –2ωs, and so on to infinity. This means that the spectrum $\bar{G}(\omega)$ consists of G(ω) repeating periodically with period ω_s= 2/T_s rad/s, of f_s = 1/ T_s Hz, as shown in Figure 2.7. There is also a constant multiplier 1/T_s in Eq. (2.3), Therefore

$$\bar{G}(\omega) = \frac{1}{T_s}\sum_{n=-\infty}^{\infty} G(\omega - n\omega_s) \quad (2.4)$$

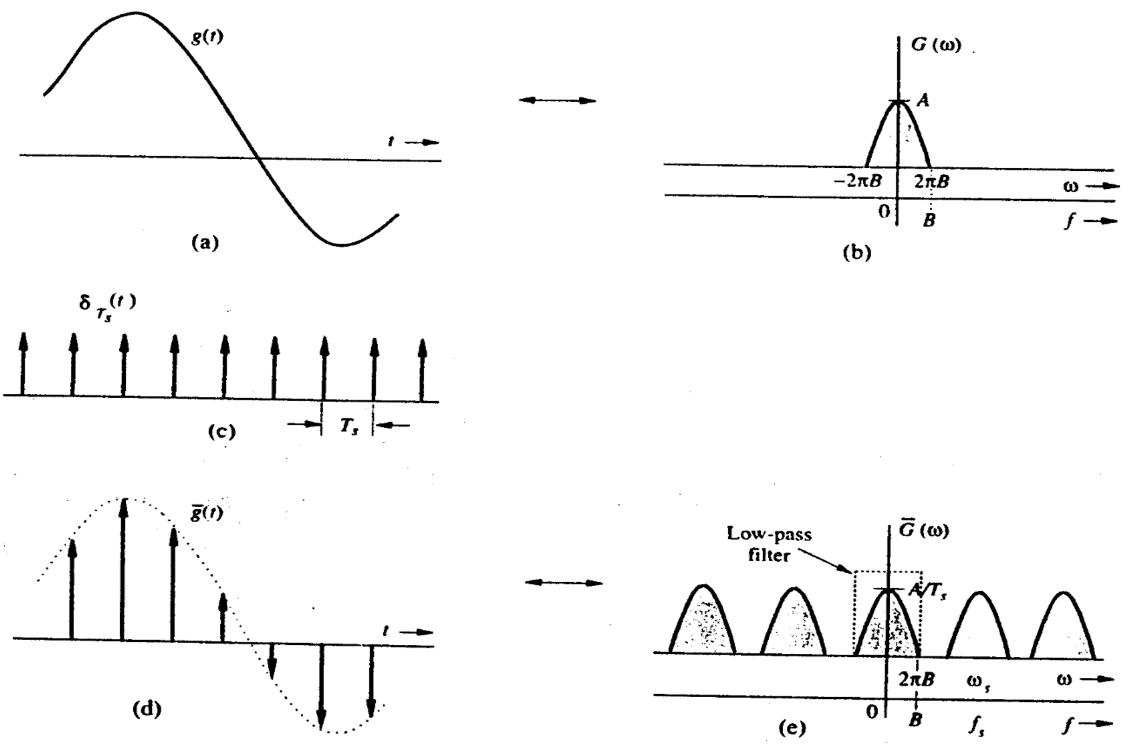


Figure (2.7): Sampled signal and its Fourier spectrum

[Refer to figure (2.4) in the text book. Page 56]

- If we are to reconstruct g(t) from $\bar{g}(t)$, we should be able to recover G(ω) from $\bar{G}(\omega)$. This is possible if there is no overlap between successive cycles of $\bar{G}(\omega)$. Figure 2.7e shows that this requires

$$F_s > 2B \tag{2.5}$$

- Also, the sampling interval Ts = 1/fs. Therefore,

$$T_s < 1/2B \tag{2.6}$$

- Thus as long as the sampling frequency fs is greater than twice the signal bandwidth B (in hertz), $\bar{G}(\omega)$ will consist of non-overlapping repetitions of G(ω). When this is true, Figure 2.7e shows that g(t) can be recovered from its samples g(t) by passing the sampled signal $\bar{g}(t)$,through an ideal low-pass filter of bandwidth B Hz. The minimum sampling rate fs = 2B required to recover g(t) from its samples $\bar{g}(t)$,is called the Nyquist rate for g(t), and the corresponding sampling interval T_s = 1/ 2B is called the Nyquist interval for g(t)

2.3.1 Signal Reconstruction: The interpolation Formula

- The process of reconstructing a continuous-time signal g(t) from its samples is also knows as interpolation. In sec 2.3, we saw that a signal g(t) band-limited to B Hz can be reconstructed (interpolated) exactly from its samples. This is done by passing the sampled signal through an ideal low-pass filter of bandwidth B Hz. As seen from Eq. (2.3), the sampled signal contain a component $(1/T_s)$g(t), and to recover g(t) [or G(ω)], the sampled signal must be passed through an ideal low-pass filter of bandwidth B Hz and gain T_s. Thus, the reconstruction (or interpolating) filter transfer function is

$$H(\omega) = T_s rect(\frac{\omega}{4\pi B}) \tag{2.7}$$

- The interpolation process here is expressed in the frequency domain as a filtering operation. Now, we shall examine this process from a different view point that of the time domain

- Let the signal interpolating (reconstruction) filter impulse response be h(t). Thus, if we were to pass the sampled signal $\bar{g}(t)$, through this filter, its response would be g(t). let us now consider a very simple interpolating filter whose impulse response is rect(t/Ts), as shown in Figure 2.8a this is a gate pulse of unit height, centered at the origin, and of width Ts (the sampling interval)

- Each sample in $\bar{g}(t)$, being an impulse, generates a gate pulse of the height equal to the strength of the sample. For instance, the kth sample is an impulse of strength g(kTs) located at t = kTs, and can be expressed as g(kTs)δ(t-kTs). When this impulse passes through the filter, it generates an output g(kTs) rect[(t-KTs)/Ts]. Thus is a gate pulse of height g(kTs), centered at t = kTs, (shown shaded in Figure 2.8b). Each sample in $\bar{g}(t)$,will generate a corresponding gate pulse resulting in an output

$$y(t) = \sum_k g(kT_s) rect(\frac{t-kT_s}{T_s}) \qquad (2.7')$$

- The filter output is a staircase approximation of g(t), shown dotted in Figure 2.8b. This filter thus gives a crude form of interpolation

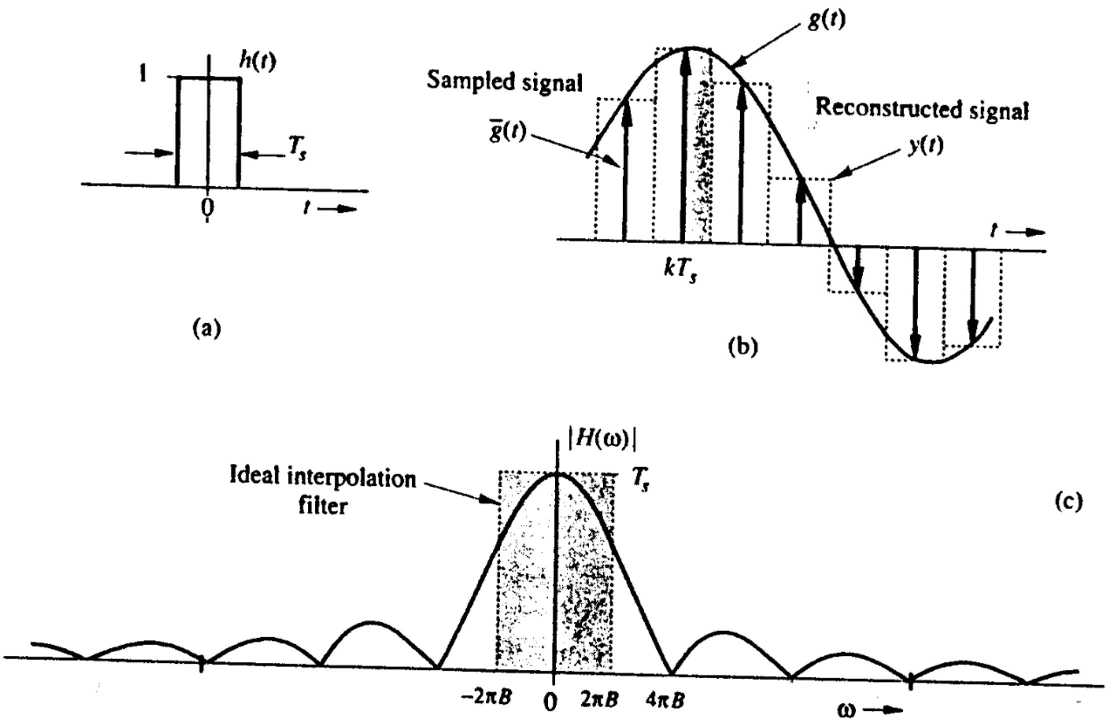


Figure (2.8): Simple interpolation using zero-order hold circuit.

[Refer to figure (2.5) in the text book. Page 58]

- The transfer function of this filter H(ω) is the Fourier transform of the impulse response rect(t/Ts). Assuming the Nyquist sampling rate, that is, Ts = 1/2B

$$h(t) = rect\left(\frac{t}{T_s}\right) = rect(2Bt)$$

- And

$$H(\omega) = T_s sinc\left(\frac{\omega T_s}{2\pi}\right) = \frac{1}{2B} sinc\left(\frac{\omega}{4\pi B}\right);\ T_s = \frac{1}{2B} \qquad (2.8)$$

- The amplitude response for this filter, shown in Figure 2.8c, explains the response for the crudeness of this interpolation. This filter, also known as the zero-order hold filter, is a poor approximation of the ideal low-pass filter (shown shaded in Figure 2.8c) required for exact interpolation

- We can improve on the zero-order hold filter by using the first-order hold filter, which results in a linear interpolation instead of the staircase interpolation. the linear interpolator, whose impulse response is a triangle pulse Δ(t/2Ts), results in an interpolation in which successive sample tops are connected by straight-line segments

- The ideal interpolation filter transfer function found in Eq. (2.7) is shown in Figure 2.9a the impulse response of this filter, the inverse Fourier transform of H(ω), is

$$h(t) = 2BT_s sinc(2Bt) \qquad (2.9a)$$

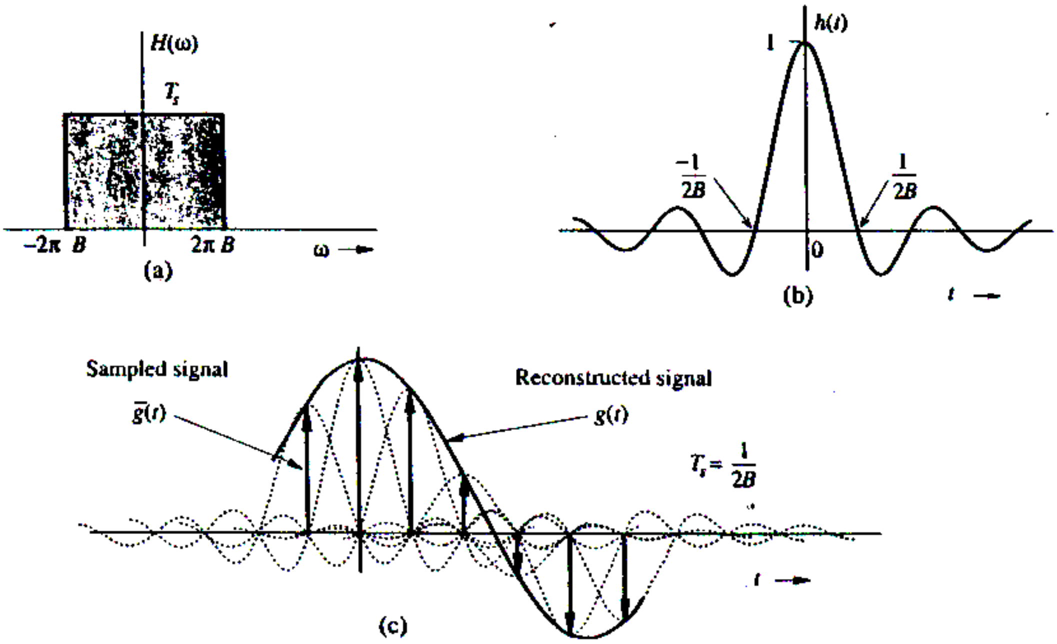


Figure (2.9): Ideal Interpolation

[Refer to figure (2.6) in the text book. Page 59]

- Assuming the Nyquist sampling rate, that is, $2BT_s = 1$, then

$$h(t) = sinc(2Bt) \qquad (2.9b)$$

- This h(t) is shown in Figure 2.9b. Observe the very interesting fact that h(t) = 0 at all Nyquist sampling instants (t = ± n/2B) except at t = 0. When the sampled signal $\bar{g}(t)$, is applied at the input of this filter, the output is g(t). Each sample in $\bar{g}(t)$, being an impulse, generates a sinc pulse of height equal to the strength of the sample, as shown in Figure 2.8b, except that h(t) is a sinc pulse instead of a gate pulse

- Addition of the sinc pulses generated by all the samples results in g(t). the kth sample of the input $\bar{g}(t)$, is the impulse $g(kTs)\delta(t-kTs)$; the filter output of this impulse is $g(kTs)h(t-kTs)$. Hence, the filter output to $\bar{g}(t)$,which is g(t), can now be expressed as a sum,

$$g(t) = \sum_k g(kT_s)h(t - kT_s) \qquad (2.10a)$$

$$= \sum_k g(kT_s)sinc[2B(t - kT_s)] \qquad (2.10b)$$

$$= \sum_k g(kT_s)sinc(2Bt - k) \qquad (2.10c)$$

- Eq. (2.10) is the interpolation formula, which yields values of g(t) between samples as a weighted sum of all the samples values

Example 2.1

- Find a signal g(t) that is band-limited to B Hz and whose samples are

$$G(0) = 1 \text{ and } g(\pm Ts) = g(\pm 2Ts) = g(\pm 3Ts) = \ldots = 0$$

- Where the sampling interval Ts is the Nyquist interval for g(t), that is Ts = 1/2B. We use the interpolation formula (2.10c) to reconstruct g(t) from its samples. Since all but one of the Nyquist samples are zero, only one term (corresponding to k = 0) in the summation on the right-hand side of Eq. (2.10c) survives. Thus,

$$g(t) = sinc(2Bt) \qquad (2.11)$$

- This signal is shown in Figure 2.10 observe that this is the only signal that has a bandwidth B Hz and with the sample values g(0) = 1 and g(nTs) = 0 (n≠0). No other signal satisfies these conditions

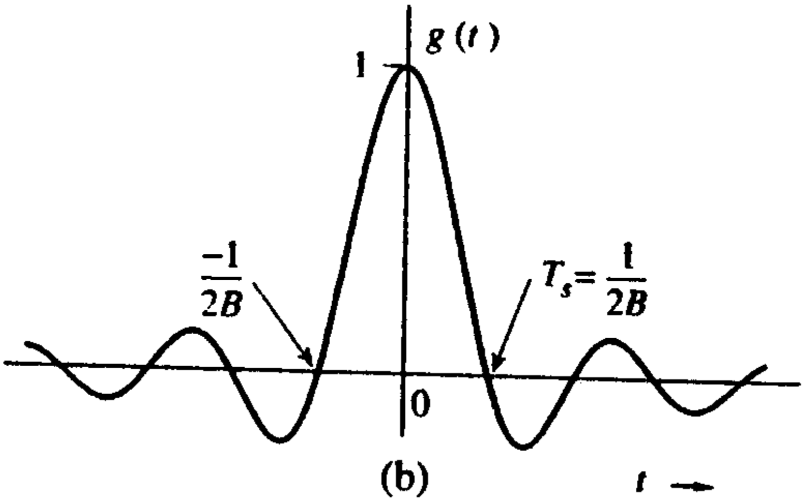


Figure (2.10): Signal reconstructed from the Nyquist samples in Example 2.1

[Refer to figure (2.7) in the text book. Page 61]

2.3.2 Practical Difficulties in Signal Reconstruction

- If a signal is sampled at the Nyquist rate fs = 2B Hz, the spectrum $\bar{G}(\omega)$, consists of repetitions of G(ω) without any gap between successive cycles, as shown in Figure 2.11 to recover g(t) from $\bar{g}(t)$,we need to pass the sampled. Signal $\bar{g}(t)$,through an ideal low-pass filter, shown dotted in Figure 2.11a

- As was seen previously this filter is unrealizable; it can be closely approximated only with infinite time delay in the response. This means that we can recover the signal g(t) from its samples with infinite time delay. A practical solution to this problem is to sample the signal at a rate higher than the Nyquist rate (fs > 2B or ωs > 4πB)

- This yields $\bar{G}(\omega)$, consisting of repetitions of G(ω) with a finite band gap between successive cycles, as shown in Figure 2.11b. we can now recover G(ω) from $\bar{G}(\omega)$, using a low-pass filter with a gradual cutoff characteristic, shown dotted in Figure 2.11b. But even in this case, the filter gain is required to be zero beyond the first cycle G(ω) (see Figure 2.11b)
- By the Paley-Wiener criterion, it is impossible to realize even this filter. The only advantage in this case is that the required filter can be closely approximated with a smaller time delay. Thus shows that it is impossible in practice to recover a band-limited signal g(t) exactly from its samples, even if the sampling rate is higher than the Nyquist rate. However, as the sampling rate increases, the recovered signal approaches the desired signal more closely

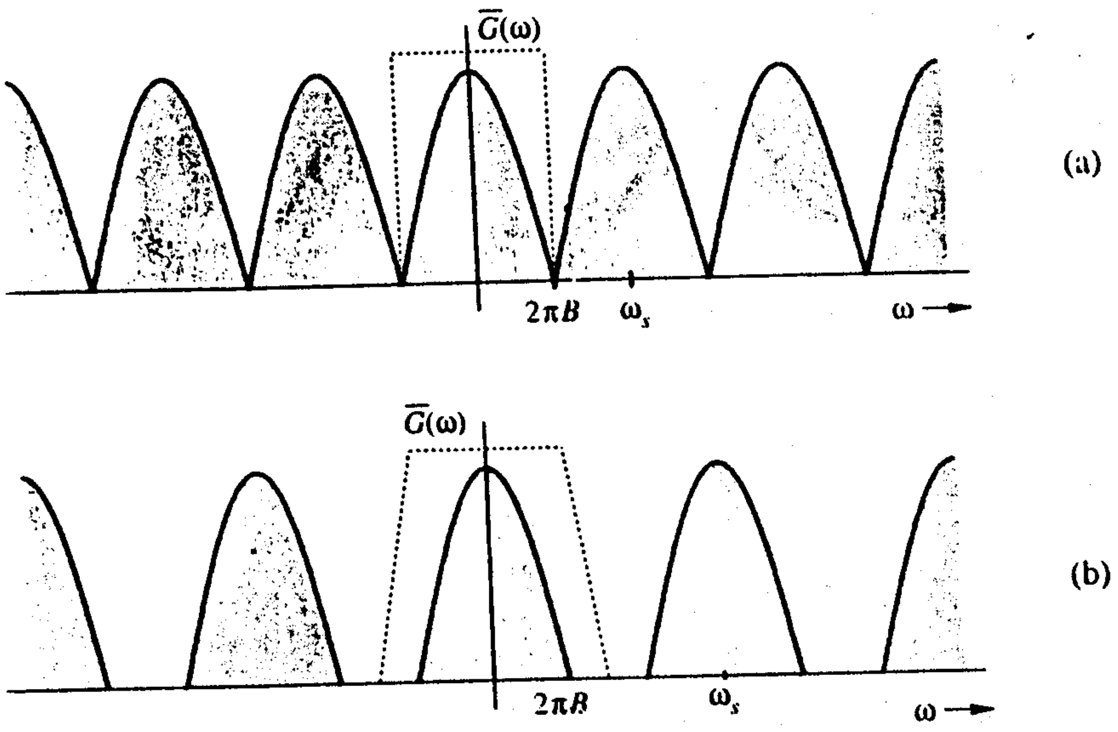


Figure (2.11): Spectra of a sampled signal. (a) At the Nyquist rate. (b) Above the Nyquist rate.

[Refer to figure (2.8) in the text book. Page 62]

2.3.2.1 Problems with Aliasing

- There is another fundamental practical difficulty in reconstructing a signal from its samples. The sampling theorem was proved on the assumption that the signal g(t) is band-limited. All practical signals are time-limited, that is, they are of finite duration or width

- It can be shown that a signal cannot be time-limited and band-limited simultaneously. If a signal is time-limited, it cannot be band-limited, and vice versa (but it can be simultaneously non-time-limited and non-band-limited). This means that all practical signals, which are time-limited, are non-band-limited; they have infinite bandwidth, and the spectrum $\bar{G}(\omega)$, consists of overlapping cycles of G(ω) repeating every fs Hz (the sampling frequency), as shown in Figure 2.12

- Because of infinite bandwidth in this case the spectral overlap is a constant feature, regardless of the sampling rate. Because of the overlapping tails, $\bar{G}(\omega)$, no longer has complete information about G(ω), and it is no longer possible, even theoretically, to recover g(t) from the sampled signal $\bar{g}(t)$,. If the sampled signal is passed through an low-pass filter, the output is not G(ω) but a version of G(ω) distorted as a result of two separate causes:

  1. The loss of the tail of G(ω) beyond $|f| > \frac{f_s}{2}$ Hz
  2. The reappearance of this tail inverted or folded onto the spectrum

- Note that the spectra cross at frequency fs/2 = 1/2Ts Hz. This frequency is called the folding frequency. The spectrum, therefore, folds onto itself at the folding frequency. For instant a component of frequency (fs/2)+fx shows up as a component of lower frequency (fs/2)-fx in the reconstructed signal

- Thus the components of frequencies above fs/2 reappear as components of frequencies below fs/2. This tail inversion known as spectral folding or aliasing, is shown shaded in Figure 2.12. In this process of aliasing, we are not only losing all the components of frequencies above fs/2 Hz, but these very components reappear (aliases) as lower frequency components. This destroys the integrity of the lower frequency components also, as shown in Figure 2.12

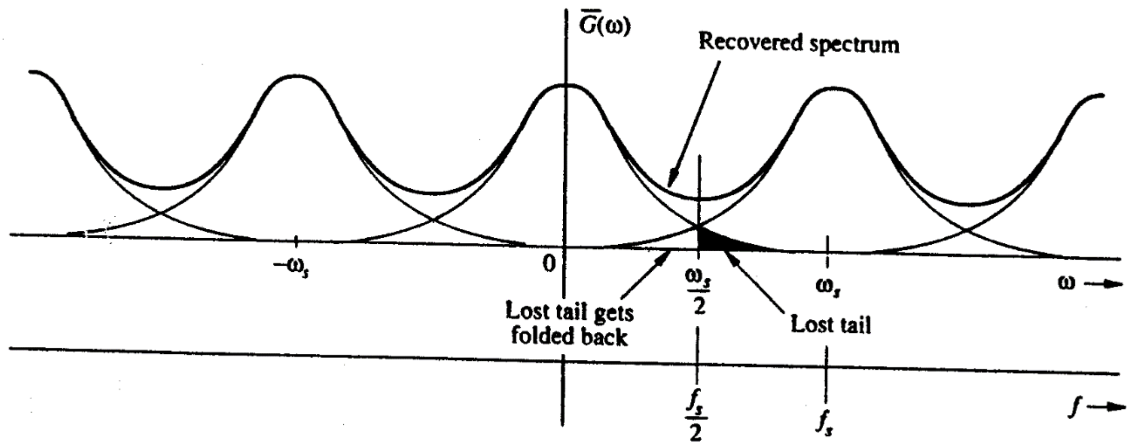


Figure (2.12): Aliasing effect

[Refer to figure (2.9) in the text book. Page 63]

2.3.2.2 A Solution: The Antialiasing Filter

- We present following procedures. The potential defectors are all the frequency components beyond fs/2 = 1/2Ts Hz. We should eliminate (suppress) these components from g(t) before sampling g(t). this way we lose only the components beyond the folding frequency fs/2 Hz

- These components now cannot reappear to corrupt the components with frequencies below the folding frequency. This suppression of higher frequencies cab be accomplished by an ideal low-pass filter of bandwidth fs/2 Hz. This filter is called the antialiasing filter. Note that the antialiasing operation must be performed before the signal is sampled

- The antialiasing filter, being an ideal filter, is unrealizable. In practice we use a steep cutoff filter, which leaves a sharply attenuated residual spectrum beyond the folding frequency fs/2

2.3.2.3 Practical Sampling

- In proving the sampling theorem, we assumed ideal samples obtained by multiplying a signal g(t) by an impulse train which is physically nonexistent. In practice, we multiply a signal g(t) by a train of pulses of finite width, shown in Figure 2.13b

- The sampled signal is shown in Figure 2.13c. It is possible to recover or reconstruct g(t) from the sampled signal g(t) in Figure 2.13c provided that the sampling rate is not below the Nyquist rate. The signal g(t) can be recovered by low-pass filtering $\bar{g}(t)$, as if it were sampled by impulse train

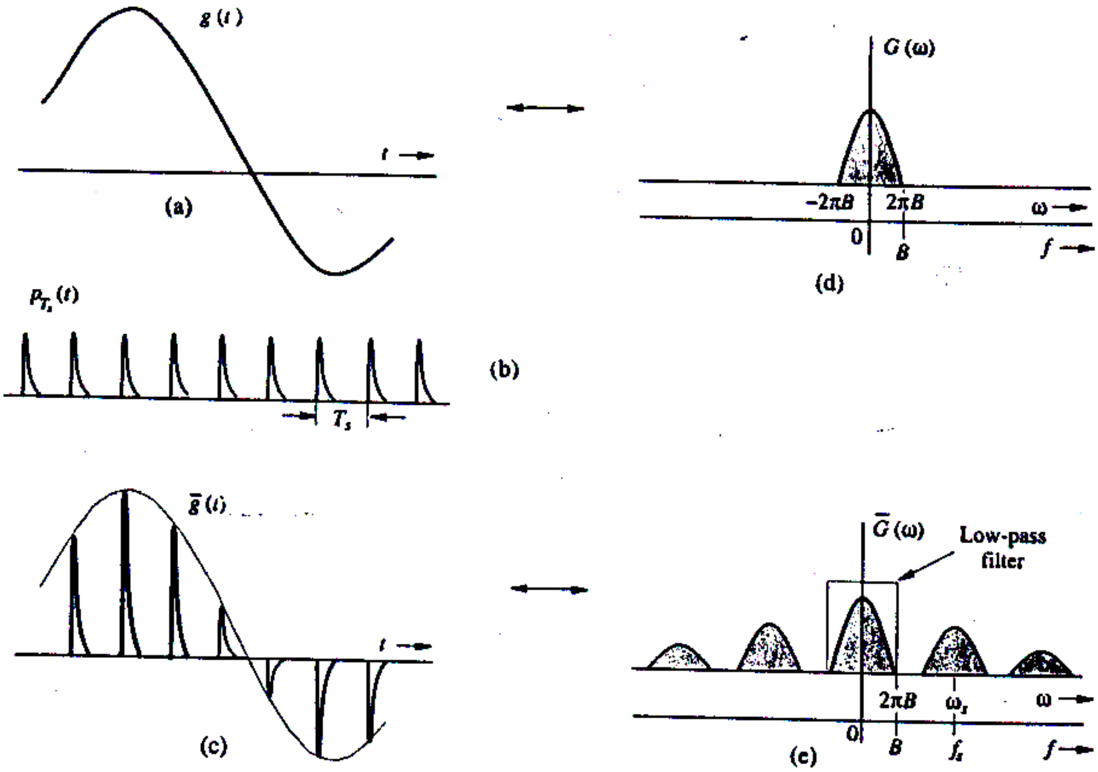


Figure (2.13): Sampled Signal and its Fourier Spectrum

[Refer to figure (2.10) in the text book. Page 64]

- The truth of this result can be seen from the fact that to reconstruct g(t), we need the knowledge of the Nyquist sample values. This information is available or built in the sampled signal $\bar{g}(t)$, in Figure 2.13c because the kth sampled pulse strength is g(kTs)

- To prove the result analytically, we observe that the sampling pulse train PTs(t) shown in Figure 2.13b being a periodic signal, can be expressed as a trigonometric Fourier series

$$P_{Ts}(t) = C_o + \sum_{n=1}^{\infty} C_n \cos(n\omega_s t + \theta_n)\,,\,,\omega_s = \frac{2\pi}{T_s}$$

- And

$$\bar{g}(t) = g(t)P_{Ts}(t) = g(t)[C_o + \sum_{n=1}^{\infty} C_n \cos(n\omega_s t + \theta_n)]$$

$$= g(t)C_o + \sum_{n=1}^{\infty} C_n \, g(t)\cos(n\omega_s t + \theta_n)]$$

- The sampled signal $\bar{g}(t)$ consists of C_o g(t), C_1 g(t) $\cos(\omega_s t+\theta_1)$, C_2 g(t) $\cos(2\omega_s t+\theta_2)$ ….

→ **Applying tha Fourier Transform (F.T)**

– And knowing the F.T properties especially the modulation properly. Thus we note that the first term C_og(t) is the desired signal and all the other terms are modulated signals with spectra centered at ±ωs, ±2ωs, ±3ωs, … as shown in Figure 2.13e clearly the signal g(t) can be recovered by low-pass filtering of $\bar{g}(t)$, provided that $\omega_s > 4\pi B$ (or $f_s > 2B$)

2.4 Some applications of the sampling theorem

– The sampling theorem is very important in signal analysis, processing, and transmission because it allows us to replace a continues-time signal by a discrete sequence of numbers. Processing a continues-time signal is therefore equivalent to processing a discrete sequence of numbers. This leads us directly into the area of digital filtering

– In the field of communication, transmission of a continuous-time message reduces to the transmission of a sequence of numbers. This opens doors to many new techniques of communicating continuous-time signals by pulse trains. The continuous-time signal g(t) is sampled, and sample values are used. To modify certain parameters of a periodic pulse train

- We may vary the amplitudes (Figure 2.14e) widths (Figure 2.14c), or positions (Figure 2.14d) of the pulses in proportion to the sample values of the signal g(t). Accordingly, we have pulse-amplitude modulation (PAM), pulse-width modulation (PWM), or pulse-position modulation (PPM). The most important form of pulse modulation today is pulse-code modulation (PCM) Figure 2.14f, introduced in this chapter

- In all these cases, instead of transmitting g(t), we transmit the corresponding pulse-modulated signal. At the receiver, we read the information of the pulse-modulated signal and reconstruct the analog signal g(t)

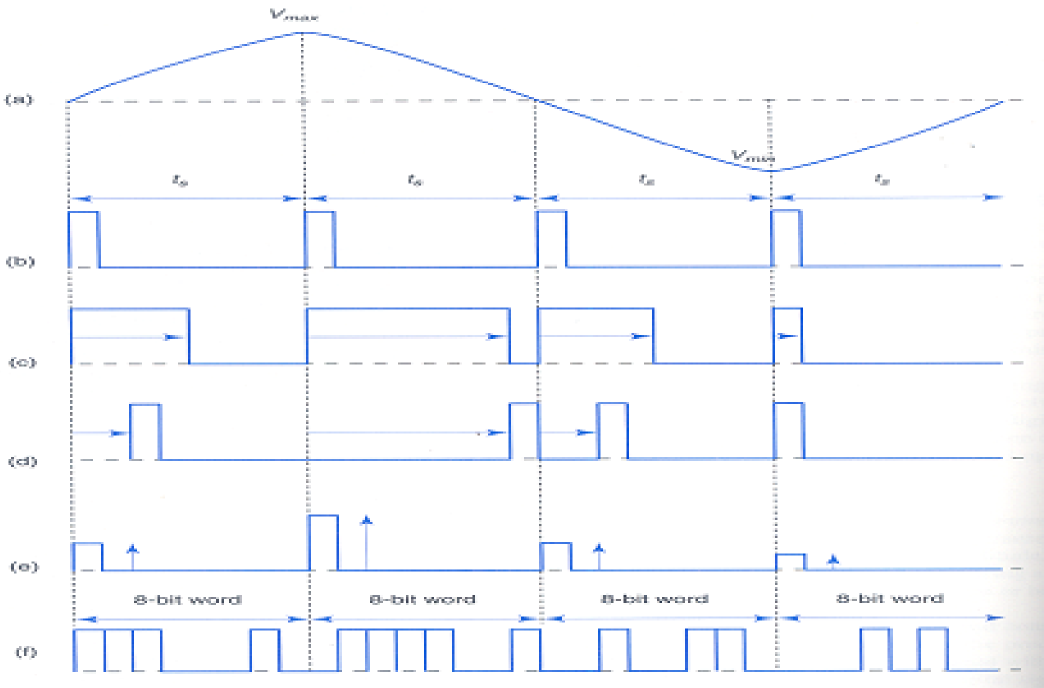


Figure (2.14): Pulse-modulated signal

[Refer to figure (2.11) in the text book. Page 66]

- One advantage of using pulse modulation is that it permits the simultaneous transmission of several signals on a time-sharing basis time-division multiplexing (TDM). Because a pulse-modulated signal occupies only a part of the channel time, we can transmit several pulse-modulated signals on the same channel by interleaving them. Figure 2.15 Shows the TDM of two PAM signals. In this manner we can multiplex several signals on the same channel by reducing pulse widths

- Another method of transmitting several baseband signals simultaneously is frequency division multiplexing (FDM), briefly discussed earlier. In FDM, various signals are multiplexed by sharing the channel bandwidth. The spectrum of each message is shifted to a specific band not occupied by any other signal. The information of various signals is located in non-overlapping frequency bands of the channel. in a way, TDM and FDM are the duals of each other

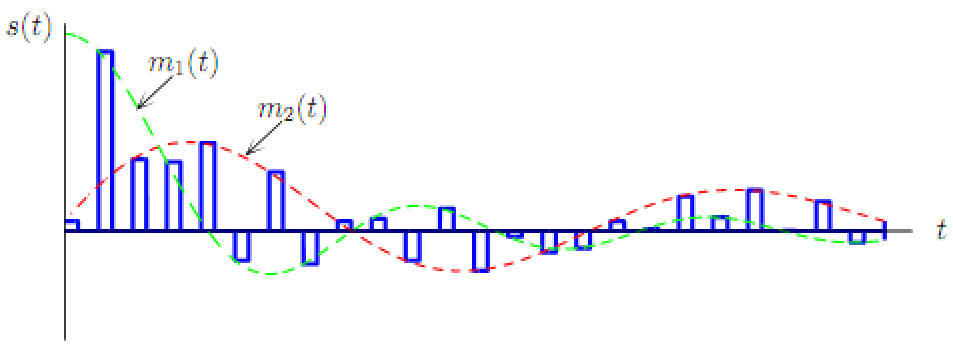


Figure (2.15): The basic elements of a PCM system.

[Refer to figure (2.12) in the text book. Page 66]

2.5 Pulse-Code Modulation (PCM)

- Pulse-code modulation is the most basic form of digital pulse modulation. In PCM a message signal is represented by a sequence of coded pulses, which is accomplished by representing the signal in discrete form in both time and amplitude

- The basic operations performed in the transmitter of a PCM system are sampling, quantization, and encoding, as shown in Figure 2.16a

→ **Advantages of digital communication**

– Some of the advantages of digital communication over analog communication are listed below:

1. Digital signals are very easy to receive. The receiver has to just detect whether the pulse is low or high.
2. AM & FM signals become corrupted over much short distances as compared to digital signals. In digital signals, the original signal can be reproduced accurately by using regenerative repeaters.
3. The signals lose power as they travel, which is called attenuation. When AM and FM signals are amplified, the noise also get amplified. But the digital signals can be cleaned up to restore the quality and amplified by the regenerators.
4. The noise may change the shape of the pulses but not the pattern of the pulses.

5. Digital hardware implementation is flexible and permits the use of microprocessors, mini-processors, digital switching, and large-scale integrated circuits

6. AM and FM signals can be received by any one by suitable receiver. But digital signals can be coded so that only the person, who is intended for, can receive them.

7. AM and FM transmitters are 'real time systems'. i.e. they can be received only at the time of transmission. But digital signals can be stored at the receiving end.

8. The digital signals can be stored relatively easy and inexpensive. It also has the ability to search and select information from distant electronic storehouses

9. It is easier and more efficient to multiplex several digital signals

10. Digital communication is inherently more efficient than analog in realizing the exchange of SNR for bandwidth

2.6 Pulse-Code Modulation Systems

- In the transmitter,
  - The band pass filter limits the frequency of the analog input signal to prevent aliasing of the message signal
  - The sample- and- hold circuit periodically samples the analog input signal and converts those samples to a multilevel PAM signal.
  - The analog-to-digital converter (ADC) converts the PAM samples to parallel PCM codes, which are converted to serial binary data in the parallel-to-serial converter and then outputted onto the transmission linear serial digital pulses, as shown in Figure 2.16a

- Transmission path,
  - The transmission line repeaters are placed at prescribed distances to regenerate the digital pulses, as shown in Figure 2.16b

- In the receiver,
  - The serial-to-parallel converter converts serial pulses received from the transmission line to parallel PCM codes.
  - The decoder converts the parallel PCM codes to multilevel PAM signals.
  - The low pass filter (reconstruction-filter) converts the PAM signals back to its original analog form, as shown in Figure 2.16c.

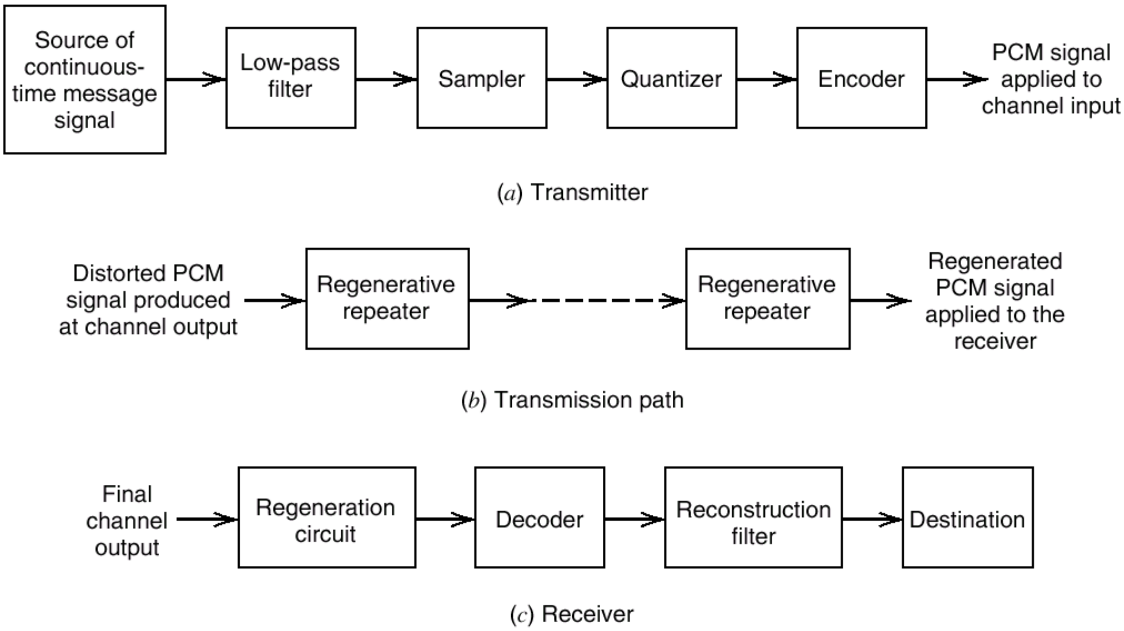


Figure (2.16): The basic elements of a PCM system.

[Refer to figure (2.14) in the text book. Page 71]

2.6.1 Sampling

- The incoming message signal is sampled with a train of narrow rectangular pulses so as to closely approximate the instantaneous sampling process
- In order to ensure perfect reconstruction of the message signal at the receiver, the sampling rate must be greater than twice the highest frequency component W of the message signal in accordance with the sampling theorem
- In practice, a pre-alias (low-pass) filter is used at the front end of the sampler in order to exclude frequencies greater than W before sampling
- Thus the application of sampling permits the reduction of the continuously varying message signal (of some finite duration) to a limited number of discrete values per second

2.6.2 Quantization

- The sampled version of the message signal is then quantized, thereby providing a new representation of the signal that is discrete in both time and amplitude
- The quantization process may follow a uniform law as described earlier in certain applications, however, it is preferable to use a variable separation between the representation levels
- For example, the range of voltages covered by voice signals, from the peaks of loud talk to the weak passages of weak talk, is on the order of 1000 to 1
- By using a non-uniform quantizer called tapered quantization levels instead of constant
- The use of a non-uniform quantizer is equivalent to passing the baseband signal through a compressor and then applying the compressed signal to a uniform quantizer
- In order to restore the signal samples to their correct relative level, we must, of course, use a device in the receiver with a characteristic complementary to the compressor. Such a device is called an expander
- Ideally, the compression and expansion laws are exactly inverse so that, except for the effect of quantization, the expander output is equal to the compressor input. The combination of a compressor and an expander is called a compander

2.6.3 Encoding

- In combining the processes of sampling and quantizating, the specification of a continuous message (baseband) signal becomes limited to a discrete set of values, but not in the form best suited to transmission over a line or radio path

- To exploit the advantages of sampling and quantizing for the purpose of making the transmitted signal more robust to noise, interference and other channel degradations, we require the use of an encoding process to translate the discrete set of sample values to a more appropriate form of signal

- Any plan for representing each of this discrete set of values as a particular arrangement of discrete events is called a code. One of the discrete events in a code is called a code element or symbol

- For example, the presence or absence of a pulse is a symbol. A particular arrangement of symbols used in a code to represent a single value of the discrete set is called a code word or character

- In a binary code, each symbol may be either of two distinct values or kinds, such as the presence or absence of a pulse. The two symbols of a binary code are customarily denoted as 0 and 1

- In a ternary code, each symbol may be one of three distinct values or kinds, and so on for other codes. However, the maximum advantage over the effects of noise in a transmission medium is obtained by using a binary code, because a binary symbol withstands a relatively high level of noise and is easy to regenerate
- Suppose that, in a binary code, each code word consists of R bits: the bit is an alternative to binary digit; thus R denotes the number of bits per sample. Then, using such a code, we may represent a total of 2^R distinct numbers. For example, a sample quantized into one of 256 levels may be represented by an 8-bit code word

- There are several ways of establishing a one-to-one correspondence between representation levels and code words. A convenient method is to express the ordinal number of the representation level as a binary number

- There are several line codes that can be used for the electrical representation of binary symbols 1 and 0, as described here:
    1. On-off signaling (Unipolar NRZ), in which symbol 1 is represented by transmitting a pulse of constant amplitude for the duration of the symbol, and symbol 0 is represented by switching off the pulse, as in Figure 2.17a
    2. Non-return-to-zero signaling (Polar NRZ), in which symbol 1 and 0 are represented by pulses of equal positive and negative amplitudes, as illustrated in Figure 2.17b
    3. Return-to-zero signaling (Unipolar RZ), in which symbol 1 is represented by a positive rectangular pulse of half-symbol width, and symbol 0 is represented by transmitting no pulse, as illustrated in Figure 2.17c
    4. Return-to-zero signaling (Bipolar RZ), in which symbol 1 is represented by a positive rectangular pulse of half-symbol width, and symbol 0 is represented by a negative rectangular pulse of half-symbol width, as illustrated in Figure 2.17d
    5. Split-phase (Manchester code), which is illustrated in Figure 2.17e. In this method of signaling, symbol 1 is represented by a positive pulse followed by a negative pulse, with both pulses being of equal amplitude and half-symbol width. For symbol 0, the polarities of these two pulses are reversed. The Manchester code suppresses the dc component and has relatively insignificant low-frequency components, regardless of the signal statistics. This property is essential in some applications

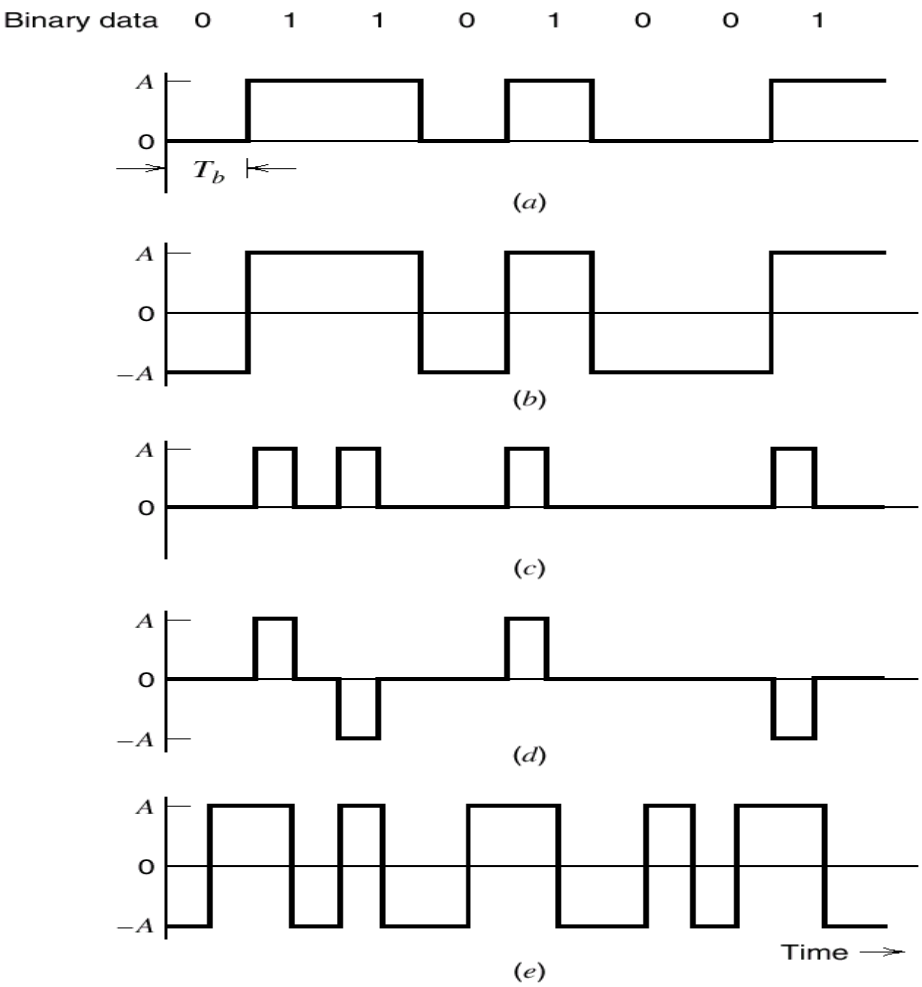


Figure (2.17): Line codes for the electrical representations of binary data

(a) Unipolar NRZ signaling. (b) Polar NRZ signaling.

(c) Unipolar RZ signaling. (d) Bipolar RZ signaling.

(e) Split-phase or Manchester code

[Refer to figure (2.15) in the text book. Page 75]

2.6.4 Regeneration

Regeneration (re-amplification, re-timing, re-shaping)

- The most important feature of PCM system lies in the ability to control the effects of distortion and noise produced by transmitting a PCM signal through a channel. This capability is accomplished by reconstructing the

PCM signal by means of a chain of regenerative repeaters located at sufficiently close spacing along the transmission route

- As illustrated in Figure 2.18 three basic functions are performed by a regenerative repeater: equalization, timing, and decision making

- The equalizer shapes the received pulses so as to compensate for the effects of amplitude and phase distortions produced by the transmission characteristics of the channel

- The timing circuitry provides a periodic pulse train, derived from the received pulses, for sampling the equalized pulses at the instant of time where the signal-to-noise ratio is a maximum

- The sample so extracted is compared to a predetermined threshold in the decision-making device. In each bit interval a decision is then made whether the received symbol is a 1 or a 0 on the basis of whether the threshold is exceeded or not

- If the threshold is exceeded, a clean new pulse representing symbol 1 is transmitted to the next repeater. Otherwise, another clean new pulse representing symbol 0 is transmitted
- In this way, the accumulation of distortion and noise in a repeater span is completely removed, provided that the disturbance is not too large to cause an error in the decision-making process

- Ideally, except for delay, the regenerated signal is exactly the same as the signal originally transmitted. In practice, however, the regenerated signal departs from the original signal for two main reasons:

  1. The unavoidable presence of channel noise and interference causes the repeaters to make wrong decisions occasionally, thereby introducing bit errors into the regenerated signal

  2. If the spacing between received pulses deviates from its assigned value, a jitter is introduced into the regenerated pulses position, thereby causing distortion

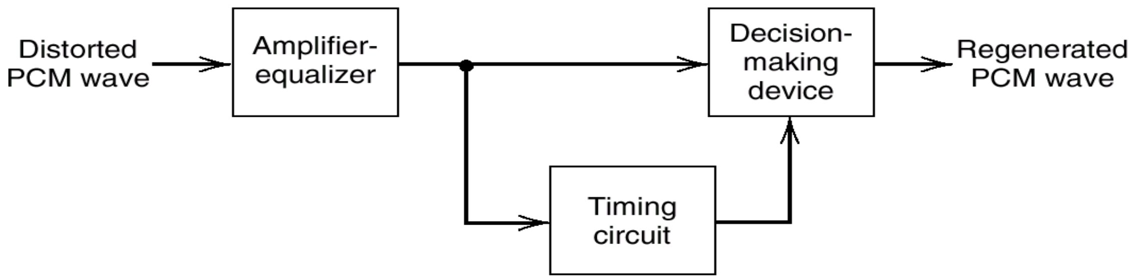


Figure (2.18): Block diagram of a regenerative repeater

[Refer to figure (2.16) in the text book. Page 76]

2.6.5 Decoding

- The first operation in the receiver is to generate (i.e., reshape and clean up) the received pulses one last time. These clean pulses are then regrouped into code words and decoded (i.e., mapped back) into a quantized PAM signal
- The decoding process involves generating a pulse the amplitude of which is the linear sum of all the pulses in the code word, with each pulse being weighted by its place value (2^0, 2^1, 2^2, 2^3, ..., 2^{R-1}) in the code, where R is the number of bits per sample

2.6.6 Filtering

- The final operation in the receiver is to recover the message signal wave by passing the decoder output through a low-pass reconstruction filter whose cutoff frequency is equal to the message bandwidth W.

- Assuming that the transmission path is error free, the recovered signal includes no noise with the exception of the initial distortion introduced by the quantization process

2.6.7 Multiplexing

- In applications using PCM, it is natural to multiplex different message sources by time division, whereby each source keeps its individuality throughout the journey from the transmitter to the receiver

- This individuality accounts for the comparative case with which message sources may be dropped or reinserted in a time-division multiplex system. As the number of independent message sources is increased, the time interval that may be allotted to each source has to be reduced, since all of them must be accommodated into a time interval equal to the reciprocal of the sampling rate
- This in turn means that the allowable duration of a code word representing a single sample is reduced. However, pulses tend to became more difficult to generate and to transmit as their duration is reduced. Furthermore, if the pulses become too short, impairments in the transmission medium begin to interfere with the proper operation of the system. Accordingly, in practice, it is necessary to restrict the number of independent message sources that can be included within a time-division group

2.6.8 Synchronization

- For a PCM system with time-division multiplexing to operate satisfactorily, it is necessary that the timing operations at the receiver, except for the time lost in transmission and regenerative repeating, follow closely the corresponding operations at the transmitter

- In a general way, this amounts to requiring a local clock at the receiver to keep the same time as a distant standard clock at the transmitter, except that the local clock is somewhat slower by an amount corresponding to the time required to transport the message signals from the transmitter to the receiver

- One possible procedure to synchronize the transmitter and receiver clocks is to set aside a code element or pulse at the end of a frame (consisting of a code word derived from each of the independent message sources in succession) and to transmit this pulse every other frame only

- In such a case, the receiver includes a circuit that would search for the pattern of Is and Os alternating at half the frame rate, and thereby establish synchronization between the transmitter and receiver

- When the transmission path is interrupted, it is highly unlikely that transmitter and receiver clocks will continue to indicate the same time for long. Accordingly, in carrying out a synchronization process, we must set up an orderly procedure for detecting the synchronizing pulse

- The procedure consists of observing the code elements one by one until the synchronizing pulse is detected. That is, after observing a particular code element long enough to establish the absence of the synchronizing pulse, the receiver clock is set back by one code element and the text code element is observed

- This searching process is repeated until the synchronizing pulse is detected. Clearly, the time required for synchronization depends on the epoch at which proper transmission is reestablished

2.7 Delta Modulation (DM)

- In delta modulation (DM), an incoming message signal is oversampled (i.e., at a rate much higher than the Nyquist rate) to purposely increase the correlation between adjacent samples of the signal
- In its basic form, DM provides a staircase approximation to the oversampled version of the message signal, as illustrated in Figure 2.19

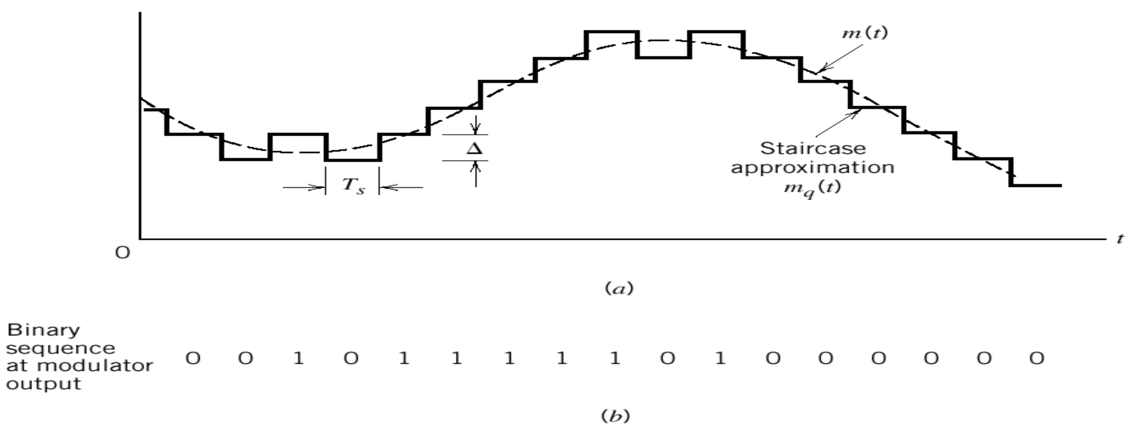


Figure (2.19): Illustration of delta system

[Refer to figure (2.17) in the text book. Page 79]

- The difference between the input and the approximation is quantized into only two levels, namely, $\pm\Delta$, corresponding to positive and negative differences, respectively

- Thus, if the approximation falls below the signal at any sampling epoch, it is increased by Δ. If, on the other hand, the approximation lies above the signal, it is diminished by Δ
- Provided that the signal doesn't change too rapidly from sample to sample, we find that the staircase approximation remains within ±Δ of the input signal

Modulator & Demodulator of DM signal

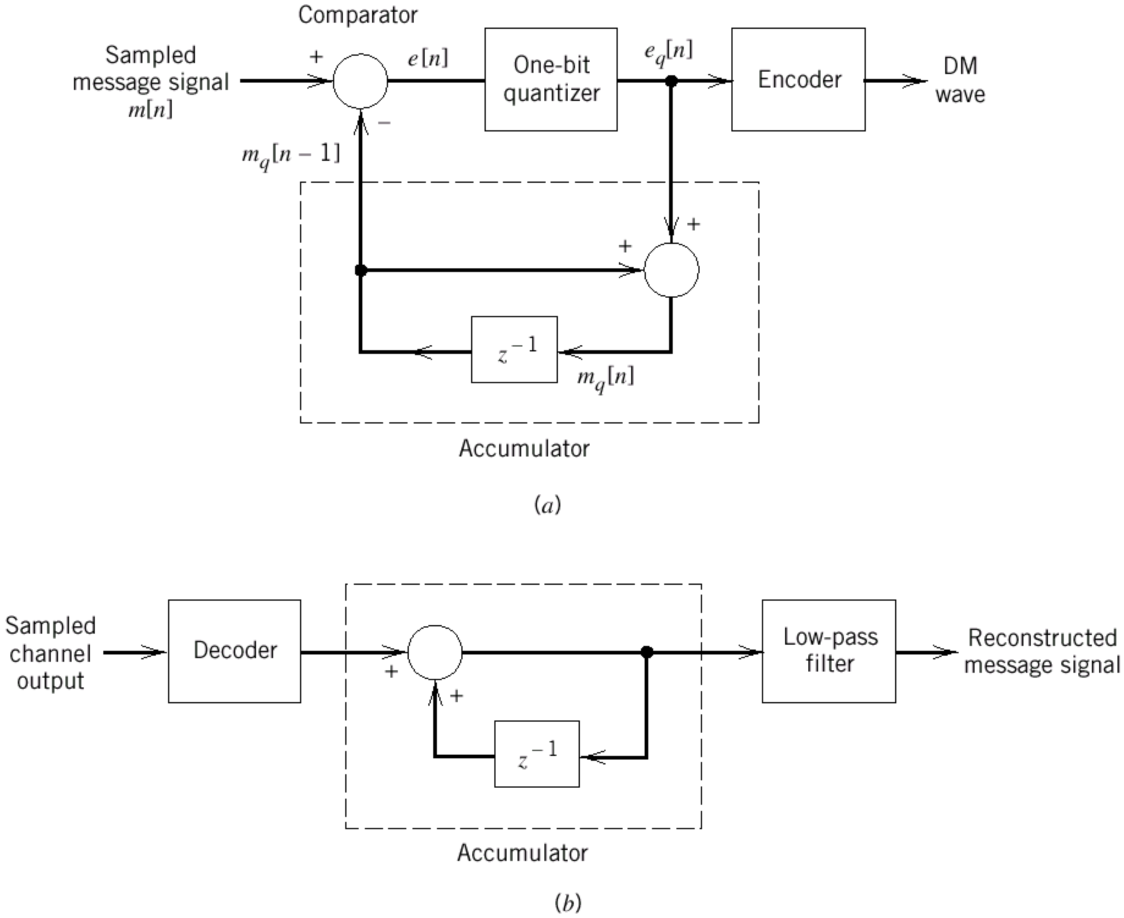


Figure (2.20): DM system. (a) Transmitter. (b) Receiver

[Refer to figure (2.18) in the text book. Page 81]

$-T_s$ is the sampling period and $m(nT_s)$ is a sample of $m(t)$.

⇒ The error signal is

$$-e[\mathrm{nT_s}] = m[\mathrm{nT_s}] - m_q[\mathrm{nT_s}\text{-}\mathrm{T_s}] \quad (2.12)$$

$$-e_q[\mathrm{nT_s}] = \Delta\,\mathrm{sgn}(e[\mathrm{nT_s}]) \quad (2.13)$$

$$-m_q[\mathrm{nT_s}] = m_q[\mathrm{nT_s} - \mathrm{T_s}] + e_q[\mathrm{nT_s}] \quad (2.14)$$

–Where $m_q[\mathrm{nT_s}]$ is the quantizer output ,

$e_q[\mathrm{nT_s}]$ is the quantized version of $e[\mathrm{nT_s}]$,

and Δ is the step size

→ In the transmitter:

- The modulator consists of a comparator, a quantizer, and an accumulator as shown in Figure 2.20(a). The output of the accumulator is

$$m_q[n] = \Delta\sum_{i=1}^{n}\mathrm{sgn}(e[i])$$
$$= \sum_{i=1}^{n} e_q[i] \quad (2.15)$$

- At the sampling instant $\mathrm{nT_s}$, the accumulator increments the approximation by a step Δ in a positive or negative direction, depending on the algebraic sign of the error signal $e(\mathrm{nT_s})$.
- If the input signal $m(\mathrm{nT_s})$ is greater than the most recent approximation $m_q(\mathrm{nT_s})$, a positive increment $+\Delta$ is applied to the approximation
- If, on the other hand, the input signal is smaller, a negative increment -Δ is applied to the approximation
- In this way, the accumulator does the best it can to track the input samples by one step (of amplitude $+\Delta$ or $-\Delta$) at a time

→ In the receiver shown in Figure 2.20b,

- the staircase approximation $m_q(t)$ is reconstructed by passing the sequence of positive and negative pulses, produced at the decoder output, through an accumulator in a manner similar to that used in the transmitter
- The out-of-band quantization noise in the high-frequency staircase waveform $m_q(t)$ is rejected by passing it through a low-pass filter, as in Figure 2.20b, with a bandwidth equal to the original message bandwidth

○ **Drawback of Delta-Modulation**

- Two types of quantization errors: Slope overload distortion and granular noise as shown in Figure
- This drawback can be overcome by integrating the message signal prior to delta modulation to give delta-sigma modulation

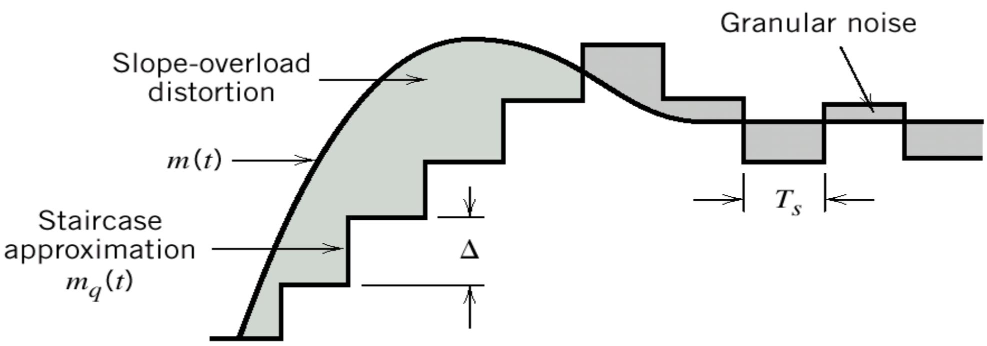


Figure (2.21): Drawback of Delta-Modulation

2.8 Delta-Sigma Modulation (D-ΣM)

- The Δ-Σ modulation which has an integrator can relieve the drawback of delta modulation
- Beneficial effects of using integrator:
  1. Pre-emphasize the low-frequency content
  2. Increase correlation between adjacent samples (reduce the variance of the error signal at the quantizer input)
  3. Simplify receiver design

- Figure 2.22 Shows the block diagram of a delta-sigma modulation system

- The message signal m(t) is defined in its continuous-time form, which means that the pulse modulator now consists of a hard-limiter followed by a multiplier; the latter comonenr is also fed from an external pulse generator (clock) to produce a 1-bit encoded signal

- Because the transmitter has an integrator, the receiver consists simply of a low-pass filter as shown in Figure

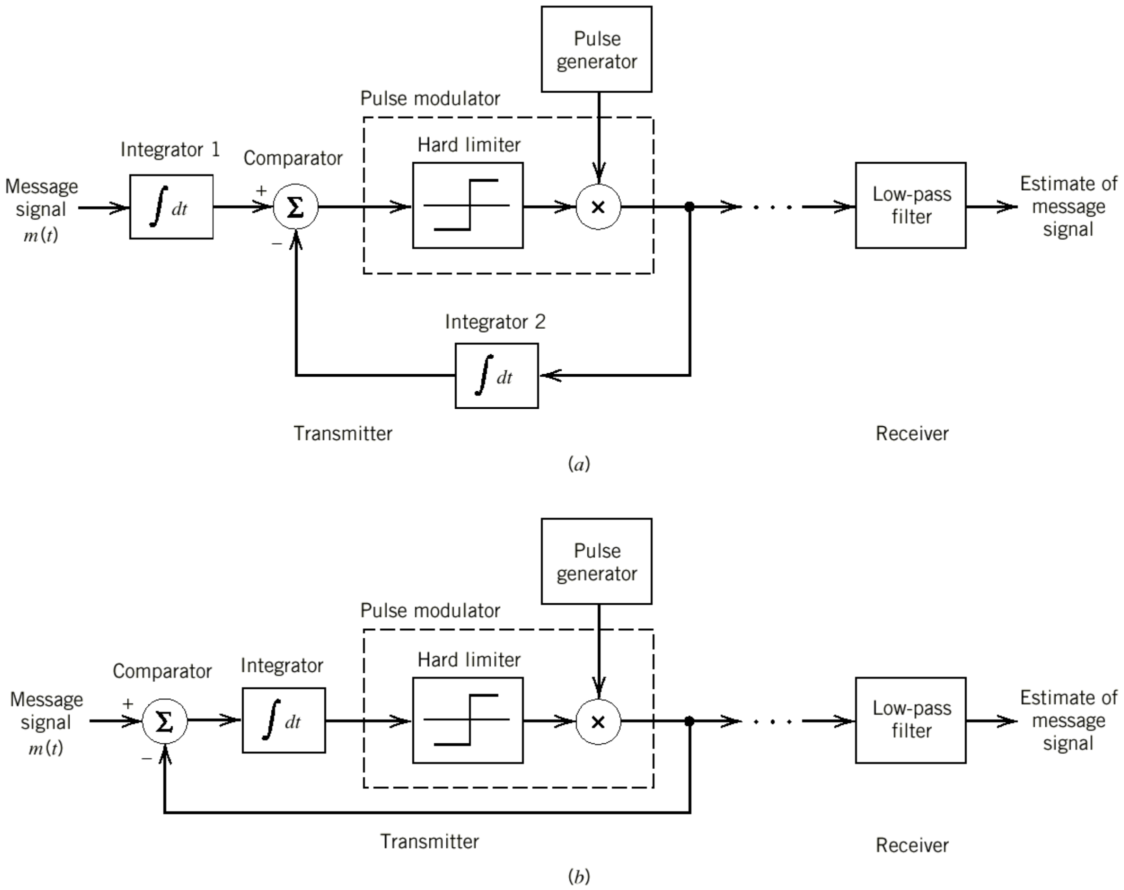


Figure (2.22): Two equivalent versions of delta-sigma modulation system

[Refer to figure (2.19) in the text book. Page 83]

- We may simplify the design of the transmitter by combining the two integrators 1 and 2 of Figure 2.22a Into a single integrator placed after the comparator, as shown in Figure 2.22b
- This later form of delta-sigma modulation system is not only simpler than that of Figure 2.22a, but it also provides an interesting interpretation of delta-sigma modulation as a "smoothed" version of 1-bit pulse-code modulation
- The term smoothness refers to the fact that the comparator output is integrated prior to quantization, and the term 1-bit merely restates that the quantizer consists of a hard-limiter with only two representation levels

2.9 Differential Pulse-Code Modulation (DPCM)

– When a voice or video signal is sampled at a rate slightly higher than the Nyquist rate, the resulting sampled signal is found to exhibit a high correlation between adjacent samples. The meaning of this high correlation is that, in an average sense, the signal does not change rapidly from one sample to the next, with the result that the difference between adjacent samples has a variance that is smaller than the variance of the signal itself

– Usually PCM has the sampling rate higher than the Nyquist rate .The encode signal contains redundant information. DPCM can efficiently remove this redundancy

– With Differential Pulse Code Modulation (DPCM), the difference in the amplitude of two successive samples is transmitted rather than the actual sample. Because the range of sample differences is typically less than the range of individual samples, fewer bits are required for DPCM than conventional PCM.

Modulator & Demodulator of DPCM signal

– The analog signal is sampled and then the difference between the sample value and its predict value (previous sample value) is quantized and then encoded forming a digital value

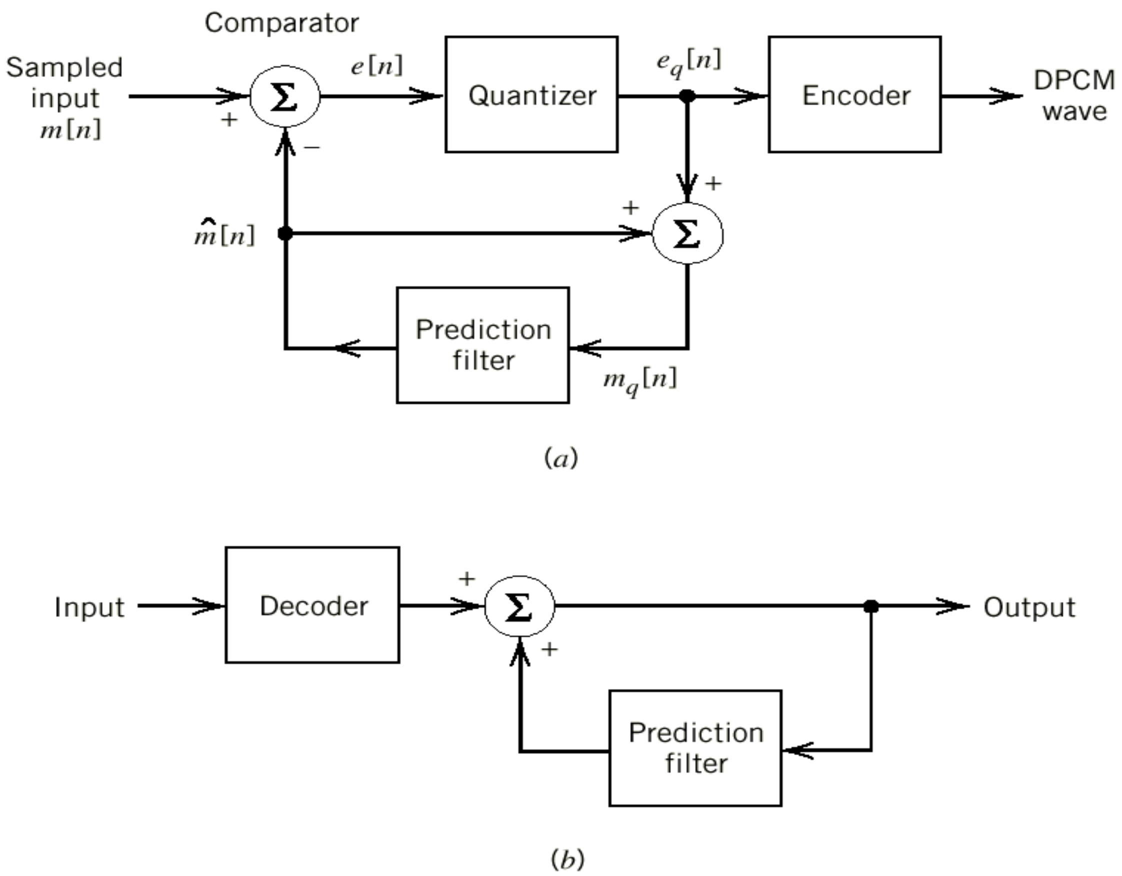


Figure (2.23): DPCM system. (a) Transmitter. (b) Receiver

[Refer to figure (2.20) in the text book. Page 85]

→ Input signal to the quantizer is defined by:

$$\Rightarrow e[nT_S] = m[nT_S] - \hat{m}[nT_S] \quad (2.16)$$

–*Where* : $\hat{m}[nT_S]$ is a prediction value.

⇒ The quantizer output is

$$e_q[nT_S] = e[nT_S] + q[nT_S] \quad (2.17)$$

– Where $q[nT_S]$ is quantization error.

$$m[nT_S] = \hat{m}[nT_S] + e[nT_S] \quad (2.18)$$

⇒ The prediction filter input is

$$m_q[nT_S] = \underline{\hat{m}[nT_S] + e[nT_S]} + q[nT_S] \quad (2.19)$$

$$\Rightarrow m_q[nT_S] = m[nT_S] + q[nT_S] \quad (2.20)$$

– In the demodulator: the decoder is used to reconstruct the quantized error signal. Using the same prediction filter in the transmitter, the receiver output is equal to quantized value of input signal $m_q(nT_s)$

→ **The prediction filter**

- A simple and yet effective approach to implement the prediction filter is to use a tapped-delay-line filter, with the basic delay set equal to the sampling period
- The block diagram of this filter is shown in Figure 2.24, according to which the prediction is modeled as a linear combination of p past sample values of the quantized input, where p is the prediction order

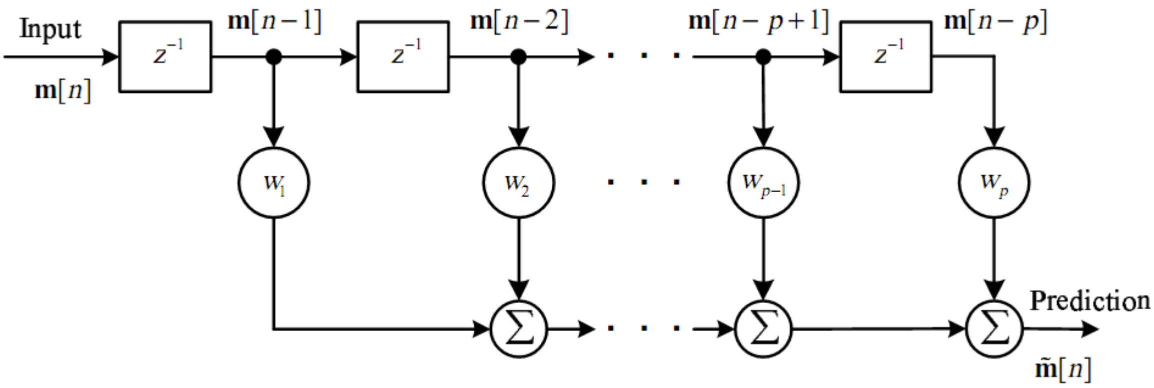


Figure (2.24): Tapped-delay-line filter used as a prediction filter

[Refer to figure (2.21) in the text book. Page 86]

Advantages of DPCM

- Less bandwidth than PCM

Disadvantages of DPCM

- Less quality than PCM
- Complex implementation
- Quantization noise

Ch.[3] Digital Radio communications

3.1 Introduction

- Traditional electronic communications systems that use conventional analog modulation techniques, such as amplitude modulation (AM), frequency modulation (FM), and phase modulation (PM), are gradually being replaced with more modern digital communications systems. Digital communications systems offer several advantages over traditional analog systems: ease of processing, ease of multiplexing, and noise immunity

- Electronic communications is the transmission, reception, and processing of information with the use of electronic circuits. Information is defined as knowledge or intelligence communicated or received

- Information is propagated through a communications system in the form of symbols which can be analog (proportional), such as the human voice, video picture information, or music, or digital (discrete), such as binary-coded numbers, alpha/numeric codes, graphic symbols, microprocessor op-codes, or database information

- However, very often the source information is unsuitable for transmission in its original form and must be converted to a more suitable form prior to transmission. For example, with digital communications systems, analog information is converted to digital form prior to transmission, and with analog communications systems, digital data are converted to analog signals prior to transmission

- The term digital communications covers a board area of communications techniques, including digital transmission and digital radio

- Digital transmission is the transmittal of digital pulses between two or more points in a communications system

- Digital radio is the transmittal of digitally modulated analog carriers between two or more points in a communications system

- Digital transmission systems require a physical facility between the transmitter and receiver, such as a metallic wire pair, a coaxial cable, or an optical fiber cable

- In digital radio systems, the transmission medium is free space or Earth's atmosphere

- Figure 3.1 shows simplified block diagram of both a digital transmission system and a digital radio system. In a digital transmission system, the original source information may be in digital or analog form. If it is in analog form, it must be converted to digital pulses prior to transmission and converted back to analog form at the receiver end. In a digital radio system, the modulating input signal and the demodulated output signal are digital pulses. The digital pulses could originate from a digital transmission system, for a digital source such as a mainframe computer, or from the binary encoding of an analog signal

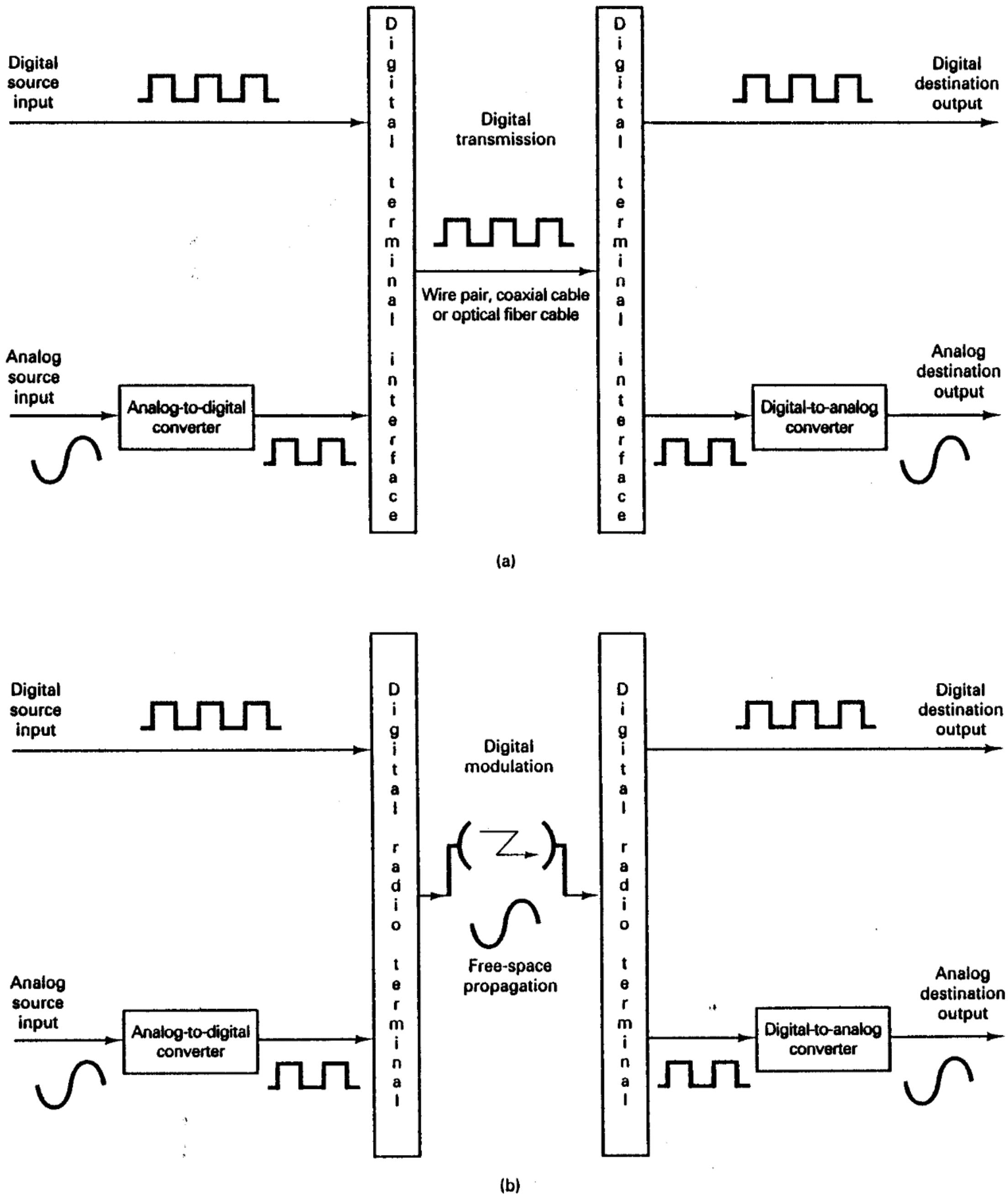


Figure (3.1): Digital communications systems: (a) Digital transmission; (b) Digital radio

[Refer to figure (3.1) in the text book. Page 92]

3.2 Shannon limit for Information Capacity

- The information capacity of a communications system represents the number of independent symbols that can be carried through the system in a given unit

- The most basic symbol is the binary digit (bit). Therefore, it is often convenient to express the information capacity of a system in bits per second (bps). A useful relationship among bandwidth, transmission time, and information capacity is simply stated, in Hartley's law as:

$$I \propto B \times t \qquad (3.1)$$

- Where I= information capacity (bits per second)
  B = bandwidth (HZ)
  t = transmission time (seconds)

- From Equation 3.1, it can be seen that information capacity is a linear function of bandwidth and transmission time and is directly proportional to both

- If either the bandwidth or the transmission time changes, a directly proportional change occurs in the information capacity.

- The higher the signal-to-noise ratio, the better the performance and the higher the information capacity.

- Shannon related the information capacity of a communications channel to bandwidth and signal-to-noise ratio. Mathematically stated, the Shannon limit for information capacity is

$$I = B \log_2\left(1 + \frac{S}{N}\right) \tag{3.2a}$$

or

$$I = 3.32B \log_{10}\left(1 + \frac{S}{N}\right) \tag{3.2b}$$

- Where: I = information capacity (bps)

  B = bandwidth (hertz)

  $\frac{S}{N}$ = signal-to-noise power ratio (unitless)

- For a standard telephone circuit with a signal-to-noise power ratio of 1000 (30 dB) and a bandwidth of 2.7 kHz, the Shannon limit for information capacity is

$$I = (2700) \log_2 (1 + 1000) = 26.9 \text{ kbps}$$

- Shannon's formula is often misunderstood. The results of the preceding example indicate that 26.9 kbps can be propagated through a 2.7-kHz communications channel. This may be true, but it cannot be done with a binary system. To achieve an information transmission rate of 26.9 kbps through a 2.7-kHz channel, each symbol transmitted must contain more than one bit.

3.3 Digital Radio

- The property that distinguishes a digital radio system from a conventional AM, FM, or PM radio system is that in a digital radio system the modulating and demodulated signals are digital pulses rather than analog waveform

- Digital radio uses analog carriers just as conventional systems do. Essentially, there are three digital modulation techniques that are commonly used in digital radio systems: frequency shift keying (FSK), phase shift keying (PSK), and quadrature amplitude modulation (QAM)

→ it rate & Baud rate

- In digital modulation, the rate of change at the input to the modulator is called the bit rate and has the units of bits per second (bps)
- Baud rate refers to the rate of change at the output of the modulator
- Hence, baud is a unit of transmission rate, modulation rate, or symbol rate and, therefore, the terms symbols per second and baud are often used interchangeably.
- Mathematically, baud is the reciprocal of the time of one output signaling element, and a signaling element may represent several information bits. Baud is expressed as

$$\text{Baud} = \frac{1}{t_s} \tag{3.3}$$

- Where baud = symbol rate (baud per second)

  t_s = time of one signaling element (seconds)

- In addition, since baud is the encoded rate of change, it also equals the bit rate divided by the number of bits encoded into one signaling element. Thus,

$$\text{Baud} = \left(\frac{f_b}{N}\right) \tag{3.4}$$

3.4 Frequency Shift Keying

- FSK is a form of constant-amplitude angle modulation similar to standard frequency modulation (FM) except the modulating signal is a binary signal that varies between two discrete voltage levels rather than a continuously changing analog waveform.

- Consequently, FSK is sometimes called binary FSK (BFSK). The general expression for FSK is

$$v(t) = V_c \cos[\left(\omega_c + \frac{v_m(t)\Delta\omega}{2}\right) t] \quad (3.5)$$

- Where:
  - v (t) = binary FSK waveform
  - V_c = peak unmodulated carrier amplitude (volts)
  - ω_c = analog carrier center frequency (hertz)
  - $v_m(t)$ = binary digital (modulating) signal (volts)
  - $\Delta\omega$ = peak change (shift) in the analog carrier frequency (hertz)

- From Equation 3.5, it can be seen that the peak shift in the carrier radian frequency ($\Delta\omega$) is proportional to the amplitude of the binary input signal ($v_m[t]$), and the direction of the shift is determined by the polarity.

- The modulating signal is a normalized binary waveform where a logic 1 = + 1 V and a logic 0 = -1 V producing frequency shifts of $+\Delta\omega/2$ and $-\Delta\omega/2$

3.4.1 FSK Transmitter

- With binary FSK, the carrier center frequency (f_c) is shifted (deviated) up and down in the frequency domain by the binary input signal as shown in Figure 3.2.

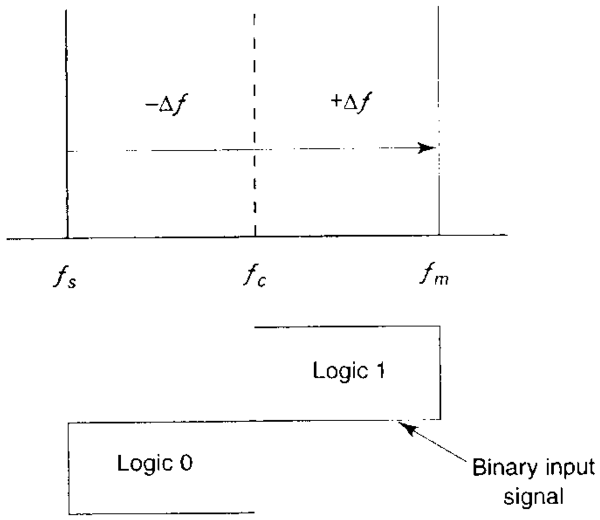


Figure (3.2): FSK in the frequency domain

- As the binary input signal changes from a logic 0 to a logic 1 and vice versa, the output frequency shifts between two frequencies: a mark, or logic 1 frequency (f_m), and a space, or logic 0 frequency (f_s). The mark and space frequencies are separated from the carrier frequency by the peak frequency deviation (Δf) and from each other by $2\ \Delta f$.

- Figure 3.3a shows in the time domain the binary input to an FSK modulator and the corresponding FSK output.

- When the binary input (f_b) changes from a logic 1 to a logic 0 and vice versa, the FSK output frequency shifts from a mark (f_m) to a space (f_s) frequency and vice versa.

- In Figure 3.3a, the mark frequency is the higher frequency ($f_c + \Delta f$) and the space frequency is the lower frequency ($f_c - \Delta f$), although this relationship could be just the opposite.

- Figure 3.3b shows the truth table for a binary FSK modulator. The truth table shows the input and output possibilities for a given digital modulation scheme.

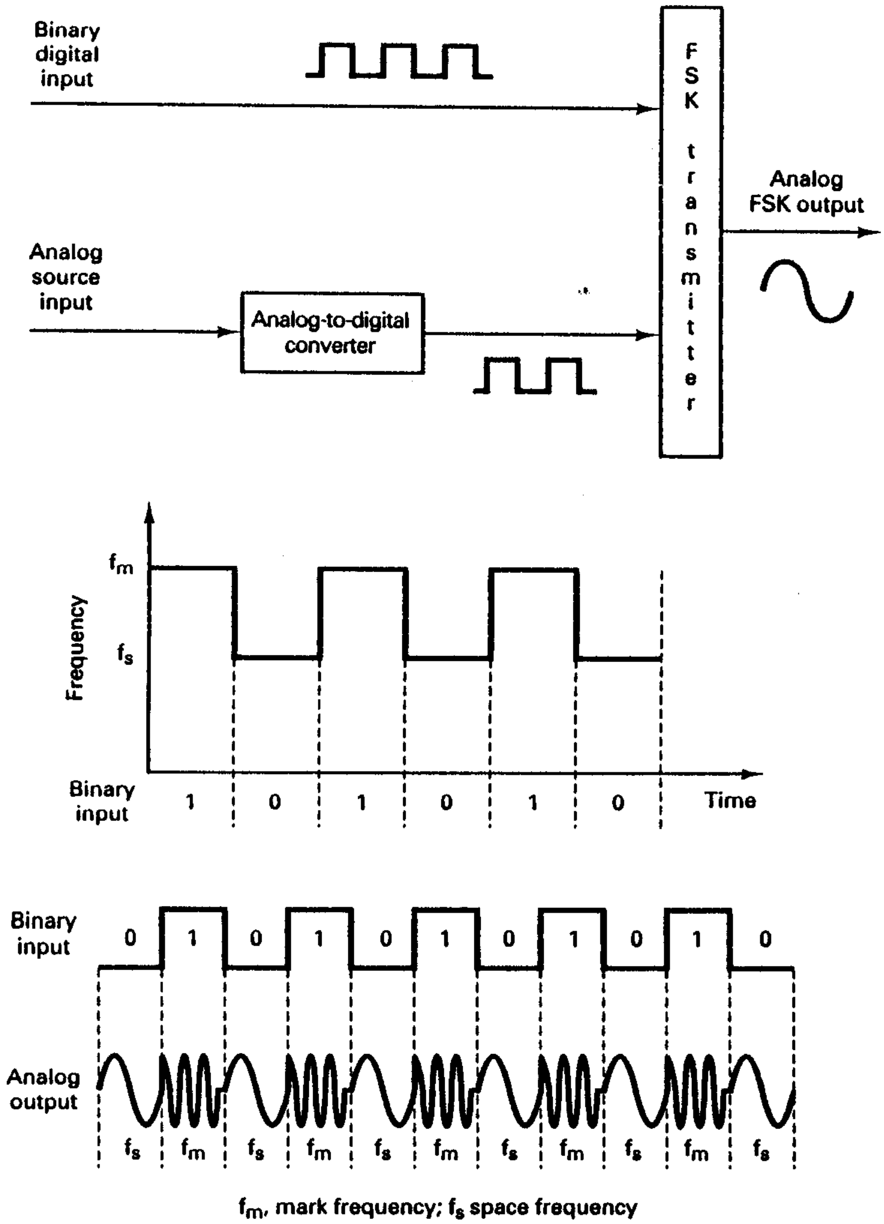


Figure (3.3): FSK in the time domain: (a) waveform: (b) truth table

[Refer to figure (3.2) in the text book. Page 95]

→ FSK Modulator Bandwidth

- Figure 3.4 shows a simplified binary FSK modulator, which is very similar to a conventional FM modulator and is very often a voltage-controlled oscillator (VCO).

- The center frequency (f_c) is chosen such that it falls halfway between the mark and space frequencies.

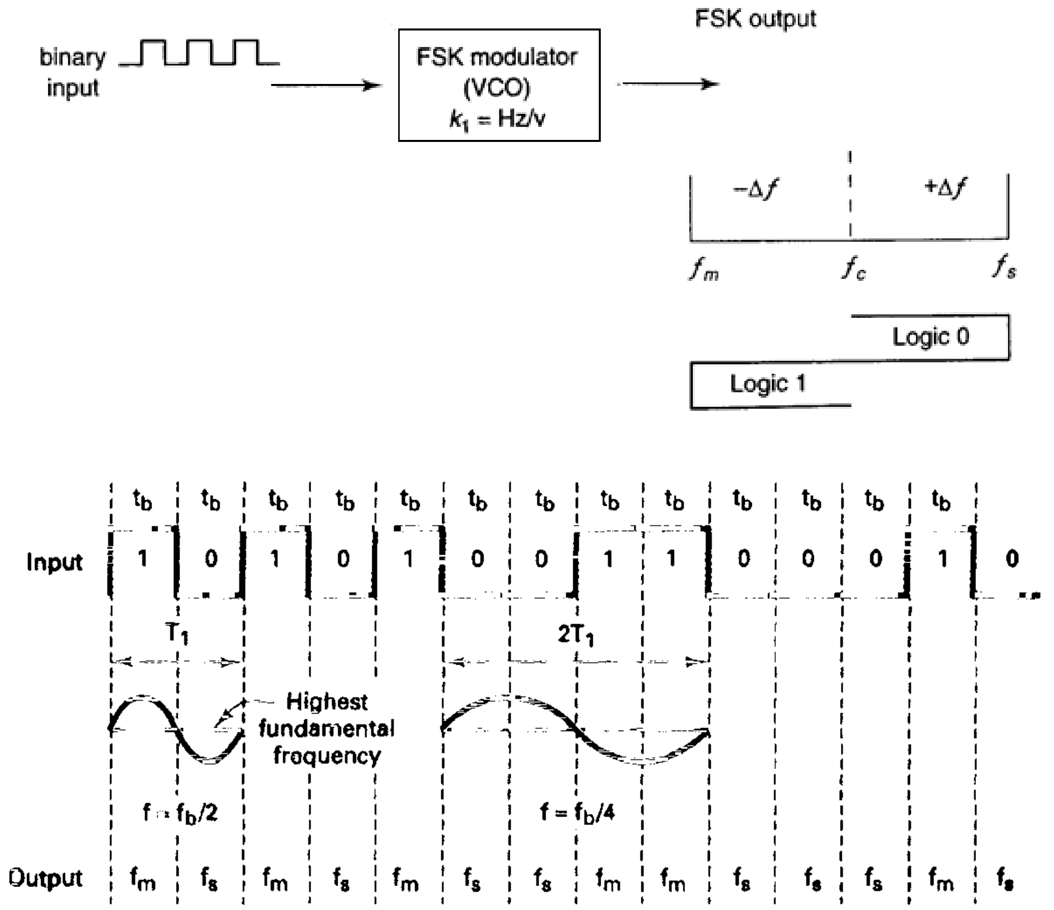


Figure (3.4): FSK modulator, t_b, time of one bit = $1/f_b$; f_m mark frequency; f_s, space frequency; T_1, period of shortest cycle; $1/T_1$, fundamental frequency of binary square wave; f_b, input bit rate (bps)

[Refer to figure (3.3) in the text book. Page 97]

- A logic 1 input shifts the VCO output to the mark frequency, and a logic 0 input shifts the VCO output to the space frequency.

- Consequently, as the binary input signal changes back and forth between logic 1 and logic 0 conditions, the VCO output shifts or deviates back and forth between the mark and space frequencies.

- A VCO-FSK modulator can be operated in the sweep mode where the peak frequency deviation is simply the product of the binary input voltage and the deviation sensitivity of the VCO.
- With the sweep mode of modulation, the frequency deviation is expressed mathematically as

$$\Delta f = v_m(t)k_l \qquad (3.6)$$

– $v_m(t)$ = peak binary modulating-signal voltage (volts)

– k_l = deviation sensitivity (hertz per volt).

3.4.2 FSK Bit Rate, Baud, and Bandwidth

- In Figure 3.3a, it can be seen that the time of one bit *(t_b)* is the same as the time the FSK output is a mark of space frequency *(t_s)*. Thus, the bit time equals the time of an FSK signaling element, and the bit rate equals the baud.
- Since it takes a high and a low to produce a cycle, the highest fundamental frequency present in a square wave equals the repetition rate of the square wave, which with a binary signal is equal to half the bit rate. Therefore,

$$f_a = f_b / 2 \qquad (3.7)$$

- Where:
  - f_a = highest fundamental frequency of the binary input signal (hertz)
  - f_b = input bit rate (bps)
- The formula used for modulation index in FM is also valid for FSK; thus,

$$MI = \Delta f / f_a \quad (\text{unitless}) \qquad (3.8)$$

- Where
  - *h* = FM modulation index called the h-factor in FSK
  - f_o = fundamental frequency of the binary modulating signal (hertz)
  - Δf = peak frequency deviation (hertz)
- Frequency deviation is illustrated in Figure 3.4 and expressed mathematically as

$$\Delta f = |f_m - f_s| / 2 \qquad (3.9)$$

- Where
  - Δf = frequency deviation (hertz)
  - $|f_m - f_s|$ = absolute difference between the mark and space frequencies (hertz)

- The peak frequency deviation in FSK is constant and always at its maximum value, and the highest fundamental frequency is equal to half the incoming bit rate. Thus,

$$MI = \frac{\frac{|f_m - f_s|}{2}}{\frac{f_b}{2}}$$

or

$$MI = \frac{|f_m - f_s|}{f_b} \qquad (3.10)$$

- Where
  - h = h-factor (unitless)
  - f_m = mark frequency (hertz)
  - f_s = space frequency (hertz)

- f_b = bit rate (bits per second)

- With conventional narrowband FM, the bandwidth is a function of the modulation index. Consequently, in binary FSK the modulation index is generally kept below 1, thus producing a relatively narrow band FM output spectrum.
- The minimum bandwidth required to propagate a signal is called the minimum bandwidth (f_N). When modulation is used and a double-sided output spectrum is generated, the minimum bandwidth is called the minimum double-sided Nyquist bandwidth or the minimum IF bandwidth

Example 3.1

- Determine (a) the output baud, and (b) minimum required bandwidth, for a binary FSK signal with a mark frequency of 80 MHz, a space frequency of 60 MHz, a rest frequency 70 MHZ and an input bit rate of 20 Mbps.

Solution

a. For FSK, N = 1, and the baud is determined from Equation 3.4 as

Baud = Bit rate / 1 = Bit rate = 20 Mbaud

b. The modulation index is found by substituting into Equation 3.10:

$$MI = \frac{|f_m - f_s|}{f_b} = |80\ \text{MHz} - 60\ \text{kHz}| / 20\ \text{Mbps} = 1$$

$$= 20\ \text{MHz} / 20\ \text{Mbps}$$

→ The bandwidth determined using the Bessel table

Table 3.1 Bessel Function Chart

MI	J_0	J_1	J_2	J_3	J_4
0.0	1.00				
0.25	0.98	0.12			
0.5	0.94	0.24	0.03		
1.0	0.77	0.44	0.11	0.02	
1.5	0.51	0.56	0.23	0.06	0.01
2.0	0.22	0.58	0.35	0.13	0.03

Modulation Index	Carrier	Side Frequency Pairs													
m	J_0	J_1	J_2	J_3	J_4	J_5	J_6	J_7	J_8	J_9	J_{10}	J_{11}	J_{12}	J_{13}	J_{14}
0.00	1.00	—	—	—	—	—	—	—							
0.25	0.98	0.12	—	—	—	—	—	—	—	—	—	—	—	—	—
0.5	0.94	0.24	0.03	—	—	—	—	—	—	—	—	—	—	—	—
1.0	0.77	0.44	0.11	0.02	—	—	—								
1.5	0.51	0.56	0.23	0.06	0.01	—	—	—	—	—	—	—	—	—	—
2.0	0.22	0.58	0.35	0.13	0.03	—	—	—	—	—	—	—	—	—	—
2.4	0	0.52	0.43	0.20	0.06	0.02	—	—	—	—	—	—			
2.5	−0.05	0.50	0.45	0.22	0.07	0.02	0.01	—	—	—	—	—	—	—	—
3.0	−0.26	0.34	0.49	0.31	0.13	0.04	0.01	—	—	—	—	—	—	—	—
4.0	0.40	0.07	0.36	0.43	0.28	0.13	0.05	0.02							
5.0	−0.18	−0.33	0.05	0.36	0.39	0.26	0.13	0.05	0.02	—	—	—	—	—	—
5.45	0	−0.34	−0.12	0.26	0.40	0.32	0.19	0.09	0.03	0.01	—	—	—	—	—
6.0	0.15	−0.28	0.24	0.11	0.36	0.36	0.25	0.13	0.06	0.02	—	—	—	—	
7.0	0.30	0.00	−0.30	−0.17	0.16	0.35	0.34	0.23	0.13	0.06	0.02	—	—	—	—
8.0	0.17	0.23	−0.11	−0.29	−0.10	0.19	0.34	0.32	0.22	0.13	0.06	0.03	—	—	—
8.65	0	0.27	0.06	−0.24	−0.23	0.03	0.26	0.34	0.28	0.18	0.10	0.05	0.02		
9.0	−0.09	0.25	0.14	−0.18	−0.27	−0.06	0.20	0.33	0.31	0.21	0.12	0.06	0.03	0.01	—
10.0	−0.25	0.05	0.25	0.06	−0.22	−0.23	−0.01	0.22	0.32	0.29	0.21	0.12	0.06	0.03	0.01

- The output spectrum for this modulator is shown in Figure 3.5 which shows that the minimum double-sided Nyquist bandwidth is 60 MHz

- Because binary FSK is a form of narrowband frequency modulation, the minimum bandwidth is dependent on the modulation index. For a modulation index between 10.5 and 1, either two or three sets of significant side frequencies are generated. Thus the minimum bandwidth is two or three times the input bit rate

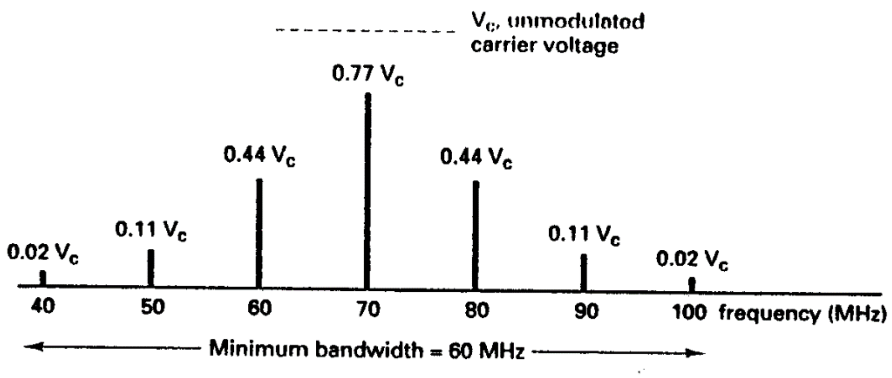


Figure (3.5): **FSK output spectrum for Example 3.1**

[Refer to figure (3.4) in the text book. Page 99]

3.4.3 FSK Receiver

- The most common circuit used for demodulating binary FSK signals is the phase-locked loop (PLL), which is shown in block diagram form in Figure 3.6.

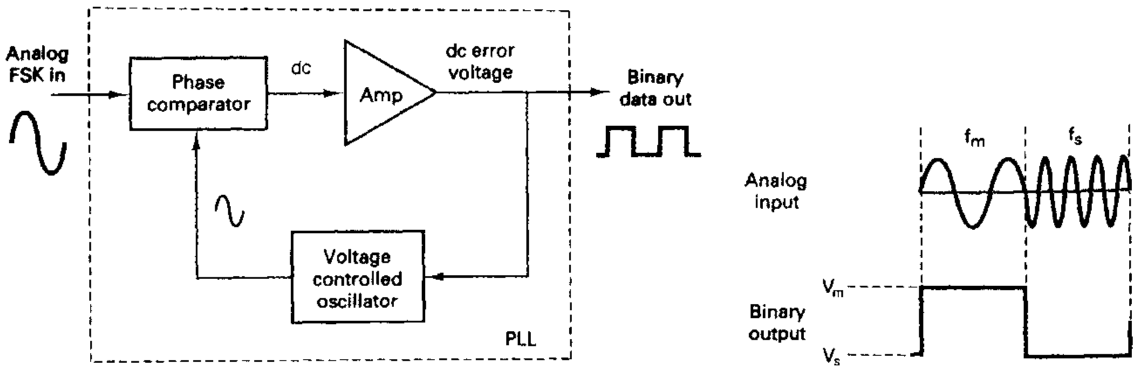


Figure (3.6): **PLL-FSK demodulator**

[Refer to figure (3.5) in the text book. Page 100]

- As the input to the PLL shifts between the mark and space frequencies, the dc error voltage at the output of the phase comparator follows the frequency shift.

- Because there are only two input frequencies (mark and space), there are also only two output error voltages. One represents a logic 1 and the other a logic 0.

- Binary FSK has a poorer error performance than PSK or QAM and, consequently, is seldom used for high-performance digital radio systems.

- Its use is restricted to low-performance, low-cost, asynchronous data modems that are used for data communications over analog, voice-band telephone lines.

3.4.4 Minimum Shift-Keying FSK

- Minimum shift-keying FSK (MSK) is a form of Continuous-phase frequency-shift keying (CP-FSK). Essentially, MSK is binary FSK except the mark and space frequencies are synchronized with the input binary bit rate.

- With CP-FSK, the mark and space frequencies are selected such that they are separated from the center frequency by an exact multiple of one-half the bit rate (f_m and f_s = $n[f_b / 2]$), where n = any odd integer).

- This ensures a smooth phase transition in the analog output signal when it changes from a mark to a space frequency or vice versa.

- Figure 3.7 shows a non-continuous FSK waveform. It can be seen that when the input changes from a logic 1 to a logic 0 and vice versa, there is an abrupt phase discontinuity in the analog signal. When this occurs, the demodulator has trouble following the frequency shift; consequently, an error may occur.

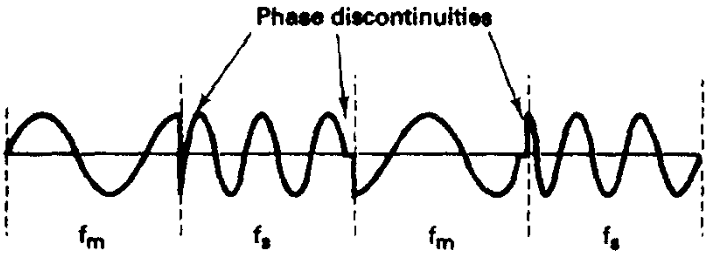


Figure (3.7): Non-continuous FSK waveform

[Refer to figure (3.6) in the text book. Page 101]

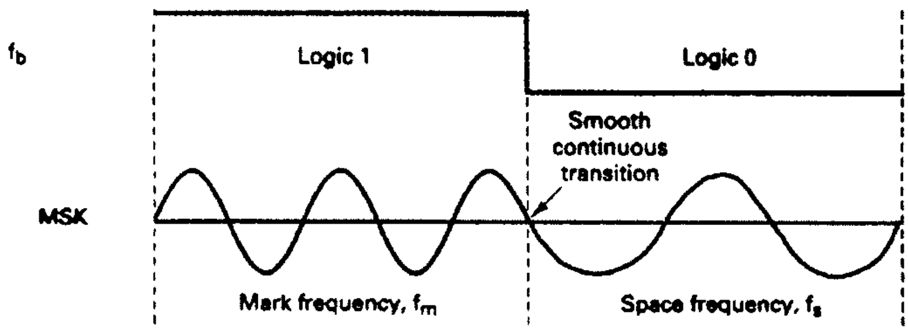


Figure (3.8): Continuous-phase MSK waveform

[Refer to figure (3.7) in the text book. Page 101]

- Figure 3.8 shows a continuous phase MSK waveform. Notice that when the output frequency changes, it is a smooth, continuous transition. Consequently, there are no phase discontinuities.

- CP-FSK has a better bit-error performance than conventional binary FSK for a given signal-to-noise ratio.

- The disadvantage of CP-FSK is that it requires synchronization circuits and is, therefore, more expensive to implement.

3.5 Phase-Shift Keying

- Phase-shift keying (PSK) is another form of angle-modulated, constant-amplitude digital modulation. PSK is similar to conventional phase modulation except that with PSK the input signal is a binary digital signal and a limited number of output phases are possible

3.6 Binary Phase-Shift Keying

- The simplest form of PSK is binary phase-shift keying (BPSK), where N = 1 and M = 2. Therefore, with BPSK, two phases ($2^1 = 2$) are possible for the carrier.

- One phase represents a logic 1, and the other phase represents a logic 0. As the input digital signal changes state (i.e., from a 1 to a 0 or from a 0 to a 1), the phase of the output carrier shifts between two angles that are separated by 180°.

- Hence, other names for BPSK are phase reversal keying (PRK) and bi-phase modulation. BPSK is a form of square-wave modulation of a continuous wave (CW) signal.

3.6.1 BPSK transmitter

- Figure 3.9 shows a simplified block diagram of a BPSK transmitter.

- The balanced modulator acts as a phase reversing switch. Depending on the logic condition of the digital input, the carrier is transferred to the output either in phase or 180° out of phase with the reference carrier oscillator.

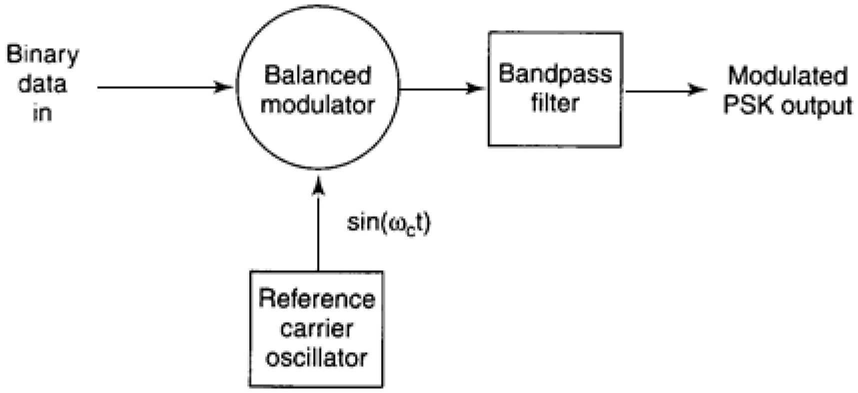


Figure (3.9): BPSK transmitter

[Refer to figure (3.8) in the text book. Page 102]

- Figure 3.10 shows the schematic diagram of a balanced ring modulator. The balanced modulator has two inputs: a carrier that is in phase with the reference oscillator and the binary digital data.

- For the balanced modulator to operate properly, the digital input voltage must be much greater than the peak carrier voltage.

- This ensures that the digital input controls the on/off state of diodes D1 to D4. If the binary input is a logic 1(positive voltage), diodes D1 and D2 are forward biased and on, while diodes D3 and D4 are reverse biased and off (Figure 3.10b). With the polarities shown, the carrier voltage is developed across transformer T2 in phase with the carrier voltage across T1. Consequently, the output signal is in phase with the reference oscillator.

- If the binary input is a logic 0 (negative voltage), diodes D_l and D_2 are reverse biased and off, while diodes D_3 and D_4 are forward biased and on (Figure 3.10c). As a result, the carrier voltage is developed across transformer T_2 180° out of phase with the carrier voltage across T_1.

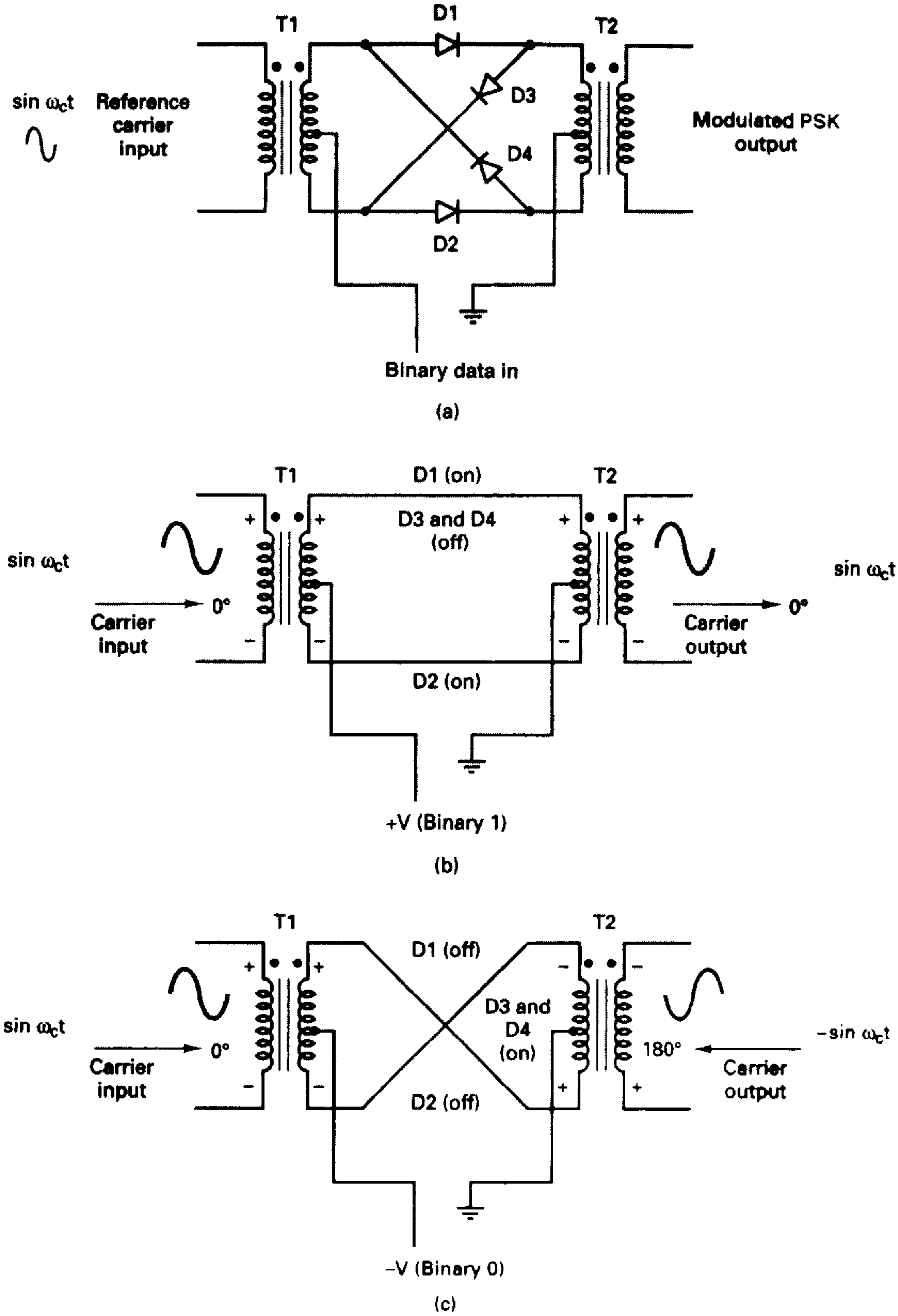


Figure (3.10): **(a) Balanced ring modulator; (b) logic 1 input; (c) logic 0 input**

[Refer to figure (3.9) in the text book. Page 103]

- Consequently, the output signal is 180° out of phase with the reference oscillator. Figure 3.11 shows the truth table, phasor diagram, and constellation diagram for a BPSK modulator

- A constellation diagram which is sometimes called a signal state-space diagram, is similar to a phasor diagram except that the entire phasor is not drawn. In a constellation diagram, only the relative positions of the peaks of the phasor are shown

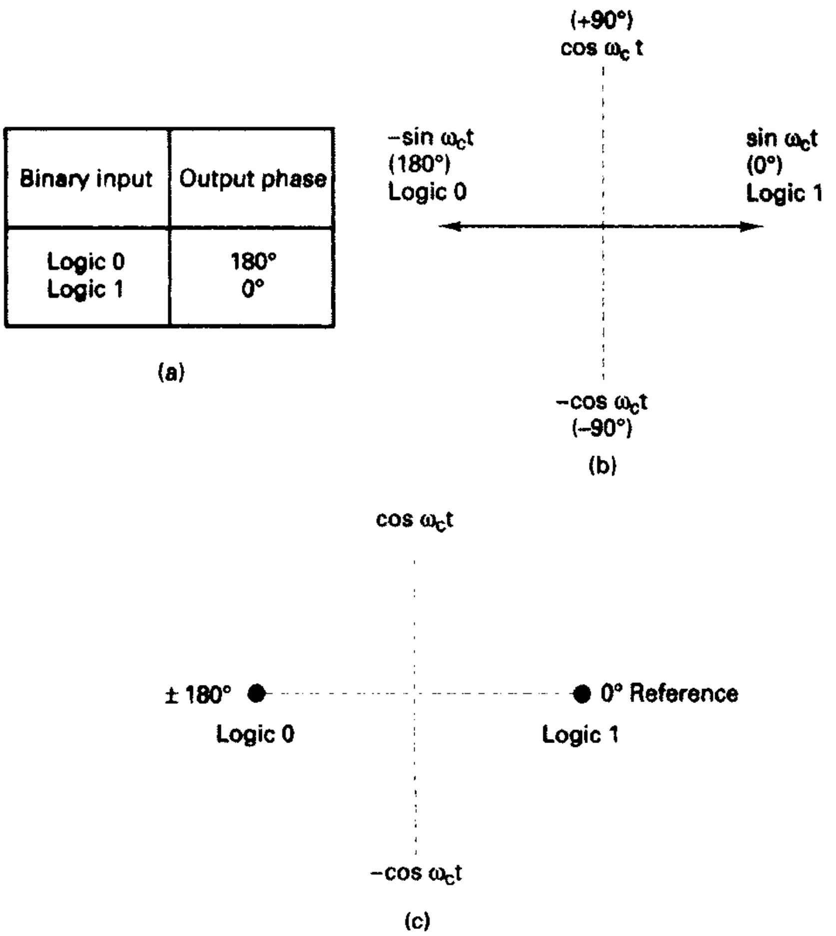


Figure (3.11): **BPSK modulator: (a) truth table; (b) phasor diagram; (c) constellation diagram**

[Refer to figure (3.10) in the text book. Page 104]

3.6.2 Bandwidth considerations of BPSK

- In a BPSK modulator. The carrier input signal is multiplied by the binary data. If + 1 V is assigned to a logic 1 and -1 V is assigned to a logic 0, the input carrier ($\sin \omega_c t$) is multiplied by either a + or - 1.

- The output signal is either $+1 \sin \omega_c t$ or $-1 \sin \omega_c t$ the first represents a signal that is *in phase* with the reference oscillator, the latter a signal that is 180° out of phase with the reference oscillator.

- Each time the input logic condition changes, the output phase changes. Mathematically, the output of a BPSK modulator is proportional to

$$\text{BPSK output} = [\sin(\omega_a t)] \times [\sin(\omega_c t)] \qquad (3.11)$$

- Where
  - ω_a = maximum fundamental radian frequency of binary input (rad/sec)
  - ω_c = reference carrier radian frequency (rad/sec)

- Solving for the trig identity for the product of two sine functions,

$$0.5\cos[(\omega_c - \omega_a)t] - 0.5\cos[(\omega_c + \omega_a)t]$$

- Thus, the minimum double-sided Nyquist bandwidth (f_N) is

$$\begin{array}{ccc} \omega_c + \omega_a & & \omega_c + \omega_a \\ -(\omega_c + \omega_a) & \text{or} & \underline{-\omega_c + \omega_a} \\ & & 2\omega_a \end{array}$$

  - and because $f_a = f_b / 2$, where f_b = input bit rate,

- Where f_N is the minimum double-sided Nyquist bandwidth.

$$f_n = 2\left(\frac{f_b}{2}\right) = f_b$$

- Figure 3.12 shows the output phase-versus-time relationship for a BPSK waveform. Logic 1 input produces an analog output signal with a 0° phase angle, and a logic 0 input produces an analog output signal with a 180° phase angle.

- As the binary input shifts between a logic 1 and a logic 0 condition and vice versa, the phase of the BPSK waveform shifts between 0° and 180°, respectively.

- BPSK signaling element *(t_s)* is equal to the time of one information bit *(t_b)*, which indicates that the bit rate equals the baud.

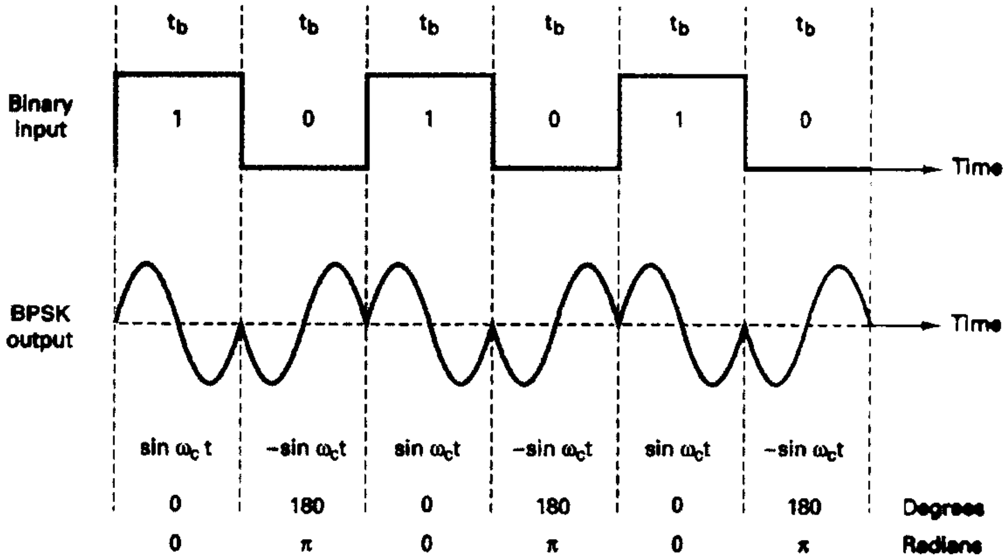


Figure (3.12): Output phase-versus-time relationship for a BPSK modulator

[Refer to figure (3.11) in the text book. Page 106]

Example 3.2

– For a BPSK modulator with a carrier frequency of 70 MHz and an input bit rate of 10 Mbps, determine the maximum and minimum upper and lower side frequencies, draw the output spectrum, determine the minimum Nyquist bandwidth, and calculate the baud.

Solution

– Substituting into Equation 3.11 yields

– output = [sin ($2\pi f_a t$)] x [sin ($2\pi f_c t$)] ; $f_a = f_b / 2 = 5$ MHz

= [sin 2π(5MHz)t)] x [sin 2π(70MHz)t)]

= 0.5cos[2π(70MHz – 5MHz)t] – 0.5cos[2π(70MHz + 5MHz)t]

lower side frequency upper side frequency

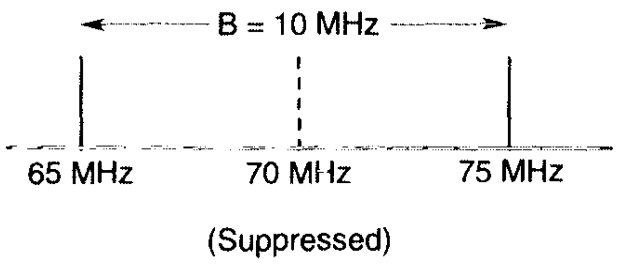
→ Minimum lower side frequency (LSF):

LSF=70MHz - 5MHz = 65MHz

→ Maximum upper side frequency (USF):

USF = 70 MHz + 5 MHz = 75 MHz

- Therefore, the output spectrum for the worst-case binary input conditions is as follows: The minimum Nyquist bandwidth (B) is



→ The minimum Nyquist bandwidth (f_N)= 75 MHz - 65 MHz = 10 MHz

→ And the baud = f_b or 10 megabaud.

3.6.3 BPSK receiver

- Figure 3.13 shows the block diagram of a BPSK receiver. The input signal may be + sin $\omega_c t$ or - sin $\omega_c t$. The coherent carrier recovery circuit detects and regenerates a carrier signal that is both frequency and phase coherent with the original transmit carrier.

- The balanced modulator is a product detector; the output is the product of the two inputs (the BPSK signal and the recovered carrier).

- The low-pass filter (LPF) operates the recovered binary data from the complex demodulated signal.

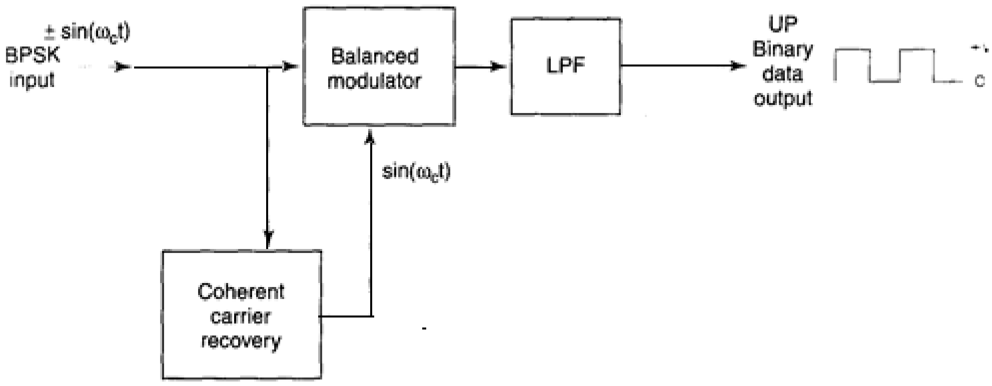


Figure (3.13): Block diagram of a BPSK receiver

[Refer to figure (3.12) in the text book. Page 108]

- Mathematically, the demodulation process is as follows.
  - For a BPSK input signal of + sin $\omega_c t$ (logic 1), the output of the balanced modulator is

$$\text{output} = (\sin \omega_c t)(\sin \omega_c t) = \sin^2\omega_c t \qquad (3.12)$$

or

$$\sin^2\omega_c t = 0.5(1 - \cos 2\omega_c t) = 0.5 - 0.5\cos 2\omega_c t$$

Filtered out

leaving

$$\text{output} = +\,0.5\text{ V} = \text{logic } 1$$

- It can be seen that the output of the balanced modulator contains a positive voltage (+[1/2]V) and a cosine wave at twice the carrier frequency ($2\ \omega_c t$).

- The LPF has a cutoff frequency much lower than $2\ \omega_c t$, and, thus, blocks the second harmonic of the carrier and passes only the positive constant component. A positive voltage represents a demodulated logic 1.

  – For a BPSK input signal of $-\sin \omega_c t$ (logic 0), the output of the balanced modulator is

$$\text{output} = (-\sin \omega_c t)(\sin \omega_c t) = -\sin^2 \omega_c t$$

or

$$-\sin^2 \omega_c t = -0.5(1 - \cos 2\omega_c t) = -0.5 + 0.5\cos 2\omega_c t$$

Filtered out

leaving

$$\text{output} = -\,0.5\text{ V} = \text{logic } 0$$

- The output of the balanced modulator contains a negative voltage (-[l/2]V) and a cosine wave at twice the carrier frequency ($2\omega_c t$).

- Again, the LPF blocks the second harmonic of the carrier and passes only the negative constant component. A negative voltage represents a demodulated logic 0.

3.6.4 M-ary Encoding

- M-ary is a term derived from the word binary.

- M simply represents a digit that corresponds to the number of conditions, levels, or combinations possible for a given number of binary variables.

- For example, a digital signal with four possible conditions (voltage levels, frequencies, phases, and so on) is an M-ary system where M = 4. If there are eight possible conditions, M = 8 and so forth.

- The number of bits necessary to produce a given number of conditions is expressed mathematically as

$$N=\log_2 M \tag{3.13}$$

- Where N = number of bits necessary

  M = number of output conditions possible with N bits

- Equation 3.13 can be simplified and rearranged to express the number of conditions possible with N bits as

$$2^N=M \tag{3.14}$$

- For example, with one bit, only $2^1 = 2$ conditions are possible. With two bits, $2^2 = 4$ conditions are possible, with three bits, $2^3 = 8$ conditions are possible, and so on.

3.7 Quaternary Phase Shift Keying

- Quaternary Phase Shift Keying (QPSK), or quadrature PSK as it is sometimes called, is another form of angle-modulated, constant-amplitude digital modulation

- QPSK is an M-ary encoding technique where M = 4 (hence the name "quaternary," meaning "4"). With QPSK four output phases are possible for a single carrier frequency. Because there are four different output phases, there must be four different input conditions

- Because the digital input to a QPSK modulator is a binary (base 2) signal, to produce four different input conditions it takes more than a single input bit. With 2 bits, there are four possible conditions: 00, 01, 10, 11. Therefore, with QPSK, the binary input data are combined into groups of 2 bits called dibits

- Each dibit code generates one of the four possible output phases. Therefore, for each 2-bit dibit clocked into the modulator, a single output change occurs. Therefore, the rate of change at the output (baud rate) is one-half of the input bit rate

3.7.1 QPSK transmitter

- A block diagram of a QPSK modulator is shown in Figure 3.14. Two bits (a dibit) are clocked into the bit splitter. After both bits have been serially inputted, they are simultaneously parallel outputted.
- The I bit modulates a carrier that is in phase with the reference oscillator (hence the name "I" for "in phase" channel), and the Q bit modulate, a carrier that is 90° out of phase.
- For a logic 1 = + 1 V and a logic 0= - 1 V, two phases are possible at the output of the I balanced modulator (+sin $\omega_c t$ and - sin $\omega_c t$), and two phases are possible at the output of the Q balanced modulator (+cos $\omega_c t$), and (-cos $\omega_c t$).
- When the linear summer combines the two quadrature (90° out of phase) signals, there are four possible resultant phasors given by these expressions: + sin $\omega_c t$ + cos $\omega_c t$, + sin $\omega_c t$ - cos $\omega_c t$, -sin $\omega_c t$ + cos $\omega_c t$, and -sin $\omega_c t$ - cos $\omega_c t$.

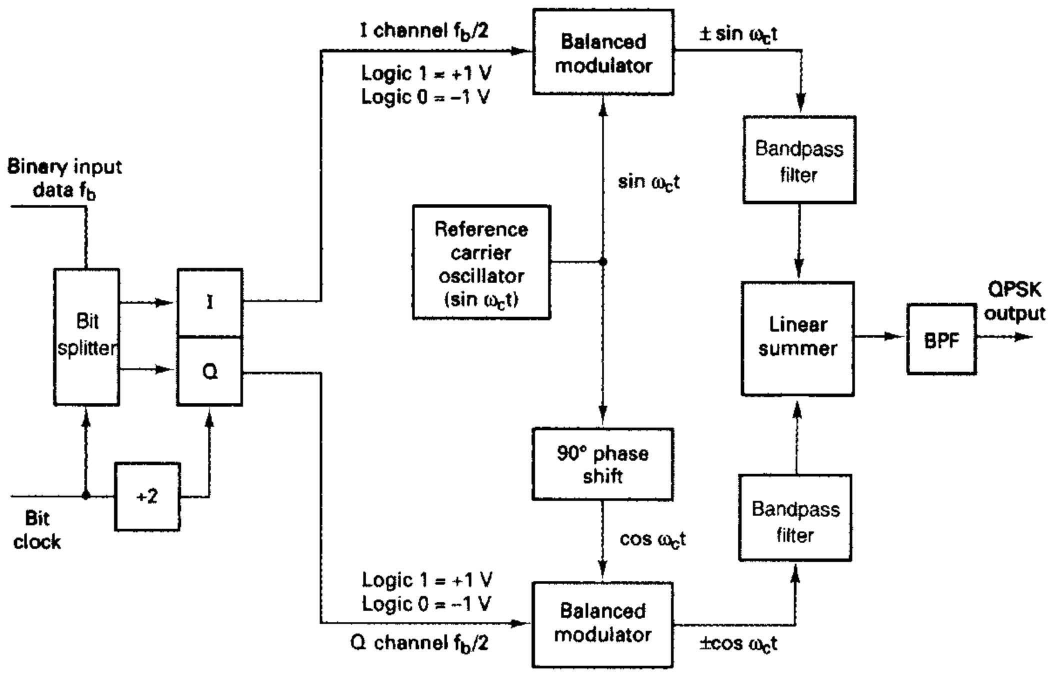


Figure (3.14): QPSK modulator

[Refer to figure (3.13) in the text book. Page 110]

Example 3.3

- For the QPSK modulator shown in Figure 3.14, construct the truth table, phasor diagram, and constellation diagram.

Solution

- For a binary data input of Q = O and I= 0, the two inputs to the I balanced modulator are -1 and sin $\omega_c t$, and the two inputs to the Q balanced modulator are -1 and cos $\omega_c t$.

→ Consequently, the outputs are

$$\text{I balanced modulator} = (-1)(\sin \omega_c t) = -1 \sin \omega_c t$$

$$\text{Q balanced modulator} = (-1)(\cos \omega_c t) = -1 \cos \omega_c t$$

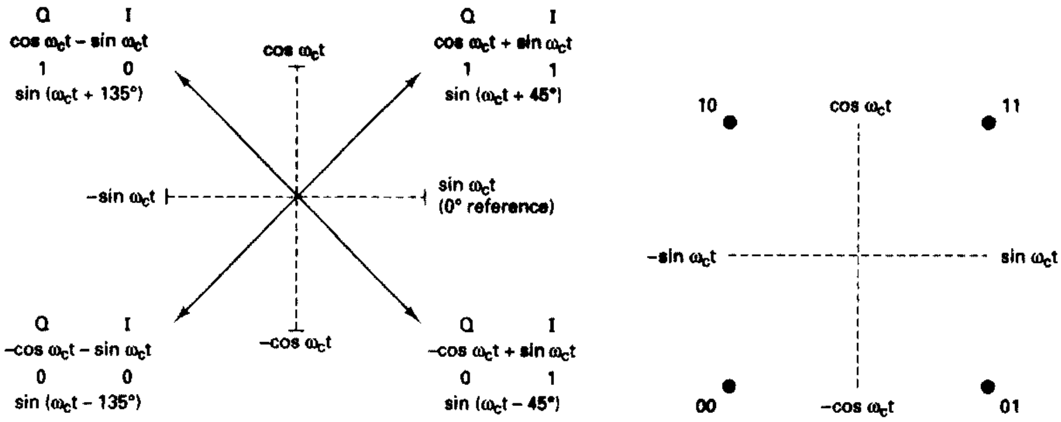
- and the output of the linear summer is

$$-1 \cos \omega_c t - 1 \sin \omega_c t = 1.414 \sin(\omega_c t - 135°)$$

- For the remaining dibit codes (01, 10, and 11), the procedure is the same. The results are shown in Figure 3.15a.

Binary input		QPSK output phase
Q	I	
0	0	−135°
0	1	−45°
1	0	+135°
1	1	+45°

(a)



(b) (c)

Figure (3.15): QPSK modulator: (a) truth table; (b) phasor diagram; (c) constellation diagram

[Refer to figure (3.14) in the text book. Page 111]

- In Figures 3.15b and c, it can be seen that with QPSK each of the four possible output phasors has exactly the same amplitude. Therefore, the binary information must be encoded entirely in the phase of the output signal.

- Figure 3.15b, it can be seen that the angular separation between any two adjacent phasors in QPSK is 90°.

- Therefore, a QPSK signal can undergo almost a+45° or -45° shift in phase during transmission and still retain the correct encoded information when demodulated at the receiver.

- Figure 3.16 shows the output phase-versus-time relationship for a QPSK modulator.

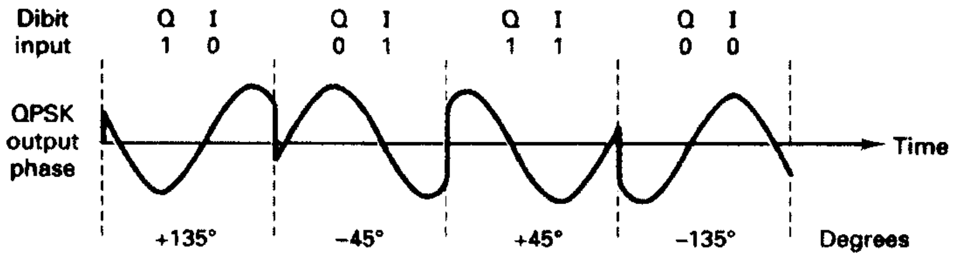


Figure (3.16): Output phase-versus-time relationship for a PSK modulator

[Refer to figure (3.15) in the text book. Page 112]

3.7.2 Bandwidth considerations of QPSK

- With QPSK, because the input data are divided into two channels, the bit rate in either the I or the Q channel is equal to one-half of the input data rate ($f_b/2$) (one-half of $f_b/2 = f_b/4$).
- This relationship is shown in Figure 3.17.

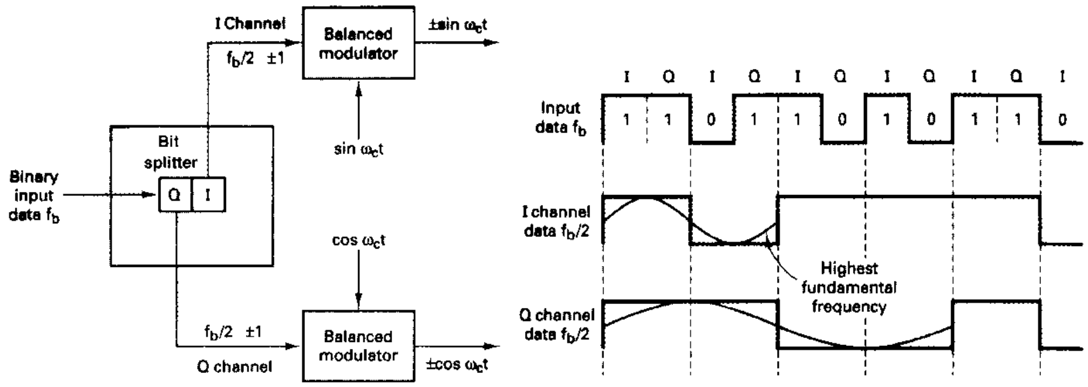


Figure (3.17): Bandwidth considerations of a QPSK modulator

[Refer to figure (3.16) in the text book. Page 113]

- In Figure 3.17, it can be seen that the worse-case input condition to the I or Q balanced modulator is an alternative 1/0 pattern, which occurs when the binary input data have a 1100 repetitive pattern. One cycle of the fastest binary transition (a 1/0 sequence in the I or Q channel takes the same time as four input data bits.

- Consequently, the highest fundamental frequency at the input and fastest rate of change at the output of the balance.: modulators is equal to one-fourth of the binary input bit rate. The output of the balanced modulators can be expressed mathematically as

$$\text{output} = (\sin \omega_a t)(\sin \omega_c t) \qquad (3.15)$$

- where

$$\underbrace{\omega_a t = 2\pi \frac{f_b}{4} t}_{\text{modulating signal}} \quad \text{and} \quad \underbrace{\omega_c t = 2\pi f_c}_{\text{carrier}}$$

$$\text{output} = \left(\sin 2\pi \frac{f_b}{4} t\right)(\sin 2\pi f_c t)$$

$$\frac{1}{2} \cos 2\pi\left(f_c - \frac{f_b}{4}\right)t - \frac{1}{2} \cos 2\pi\left(f_c + \frac{f_b}{4}\right)t$$

- The output frequency spectrum extends from $f'_c + f_b / 4$ to $f'_c - f_b / 4$ and the minimum bandwidth (f_N) is

$$\left(f_c + \frac{f_b}{4}\right) - \left(f_c - \frac{f_b}{4}\right) = \frac{2f_b}{4} = \frac{f_b}{2}$$

Example 3.4

- For a QPSK modulator with an input data rate (f_b) equal to 10 Mbps and a carrier frequency 70 MHz, determine the minimum double-sided Nyquist bandwidth *(f_N)* and the baud. Also, compare the results with those achieved with the BPSK modulator in Example 3.2. Use the QPSK block diagram shown in Figure 3.17 as the modulator model.

Solution

- The bit rate in both the I and Q channels is equal to one-half of the transmission bit rate, or

$$f_{bQ} = f_{b1} = f_b / 2 = 10 \text{ Mbps} / 2 = 5 \text{ Mbps}$$

- The highest fundamental frequency presented to either balanced modulator is

$$f_a = f_{bQ} / 2 = 5 \text{ Mbps} / 2 = 2.5 \text{ MHz}$$

- The output wave from each balanced modulator is

$$(\sin 2\pi f_a t)(\sin 2\pi f_c t)$$

$$0.5 \cos 2\pi(f_c - f_a)t - 0.5 \cos 2\pi(f_c + f_a)t$$

$$0.5 \cos 2\pi[(70 - 2.5)\text{MHz}]t - 0.5 \cos 2\pi[(70 - 2.5)\text{MHz}]t$$

$$0.5 \cos 2\pi(67.5\text{MHz})t - 0.5 \cos 2\pi(72.5\text{MHz})t$$

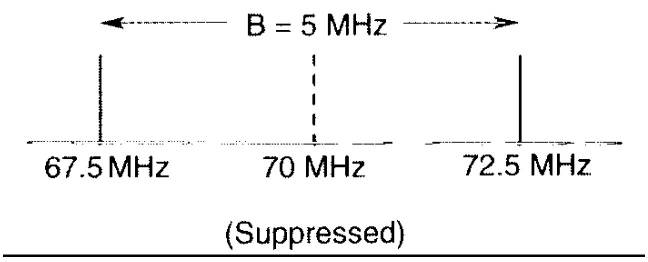
- The minimum Nyquist bandwidth is

$$f_N = B = (72.5\text{-}67.5)\text{MHz} = 5\text{MHz}$$

- The symbol rate equals the bandwidth: thus,

Symbol rate = 5 megabaud

- The output spectrum is as follows:



f_N = 5 MHZ

- It can be seen that for the same input bit rate the minimum bandwidth required to pass the output of the QPSK modulator is equal to one-half of that required for the BPSK modulator

3.7.3 QPSK receiver

- The block diagram of a QPSK receiver is shown in Figure 3.18. The power splitter directs the input QPSK signal to the I and Q product detectors and the carrier recovery circuit. The carrier recovery circuit reproduces the original transmit carrier oscillator signal. The recovered carrier must be frequency and phase coherent with the transmit reference carrier. The QPSK signal is demodulated in the I and Q product detectors and LPF, which generate the original I and Q data bits. The outputs of the product detectors and LPF are fed to the bit combining circuit, where they are converted from parallel I and Q data channels to a single binary output data stream.

- The incoming QPSK signal may be any one of the four possible output phases shown in Figure 3.18. To illustrate the demodulation process, let the incoming QPSK signal be $-\sin \omega_c t + \cos \omega_c t$. Mathematically, the demodulation process is as follows.

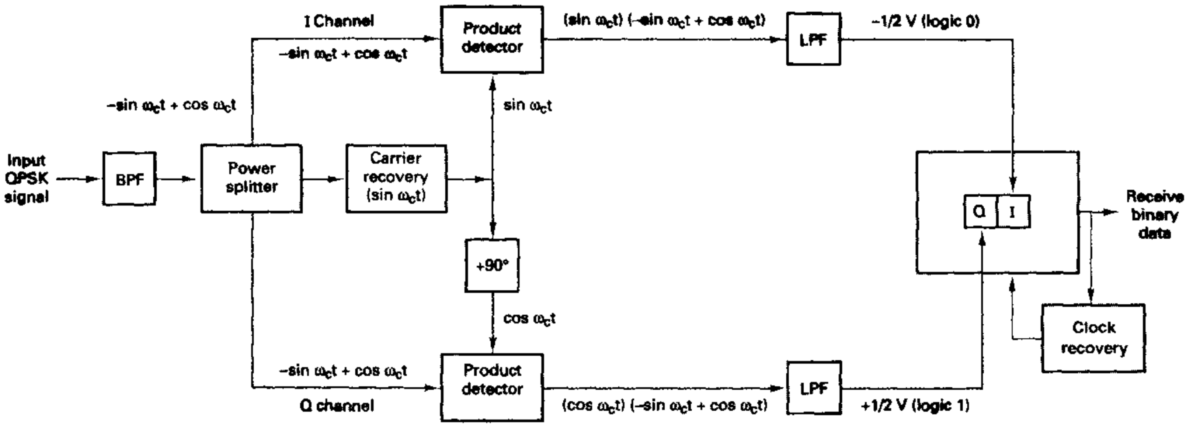


Figure (3.18): QPSK receiver

[Refer to figure (3.17) in the text book. Page 117]

- The receive QPSK signal ($-\sin \omega_c t + \cos \omega_c t$) is one of the inputs to the I product detector. The other input is the recovered carrier ($\sin \omega_c t$). The output of the I product detector is

$$I = \underbrace{(-\sin \omega_c t + \cos \omega_c t)}_{\text{QPSK input signal}} \underbrace{(\sin \omega_c t)}_{\text{carrier}}$$

$$= (-\sin \omega_c t)(\sin \omega_c t) + (\cos \omega_c t)(\sin \omega_c t)$$

$$= -\sin^2 \omega_c t + (\cos \omega_c t)(\sin \omega_c t)$$

$$= -\frac{1}{2}(1 - \cos 2\omega_c t) + \frac{1}{2}\sin(\omega_c + \omega_c)t + \frac{1}{2}\sin(\omega_c - \omega_c)t$$

$$I = -\frac{1}{2} + \underbrace{\frac{1}{2}\cos 2\omega_c t + \frac{1}{2}\sin 2\omega_c t}_{\text{(filtered out)}} + \underbrace{\frac{1}{2}\sin 0}_{\text{(equals 0)}}$$

$$= -\frac{1}{2}\text{V (logic 0)} \qquad (3.16)$$

- Again, the receive QPSK signal ($-\sin \omega_c t + \cos \omega_c t$) is one of the inputs to the Q product detector. The other input is the recovered carrier shifted 90° in phase ($\cos \omega_c t$). The output of the Q product detector is

$$Q = \underbrace{(-\sin \omega_c t + \cos \omega_c t)}_{\text{QPSK input signal}}\underbrace{(\cos \omega_c t)}_{\text{carrier}}$$

$$= \cos^2 \omega_c t - (\sin \omega_c t)(\cos \omega_c t)$$

$$= \frac{1}{2}(1 + \cos 2\omega_c t) - \frac{1}{2}\sin(\omega_c + \omega_c)t - \frac{1}{2}\sin(\omega_c - \omega_c)t$$

$$Q = \frac{1}{2} + \underbrace{\frac{1}{2}\cos 2\omega_c t - \frac{1}{2}\sin 2\omega_c t}_{\text{(filtered out)}} - \underbrace{\frac{1}{2}\sin 0}_{\text{(equals 0)}}$$

$$= \frac{1}{2}V(\text{logic } 1) \tag{3.17}$$

- The demodulated I and Q bits (0 and 1, respectively) correspond to the constellation diagram and truth table for the QPSK modulator shown in Figure 3.18.

3.7.4 Offset QPSK

- Offset QPSK (OQPSK) is a modified form of QPSK where the bit waveforms on the I and Q channels are offset or shifted in phase from each other by one-half of a bit time.

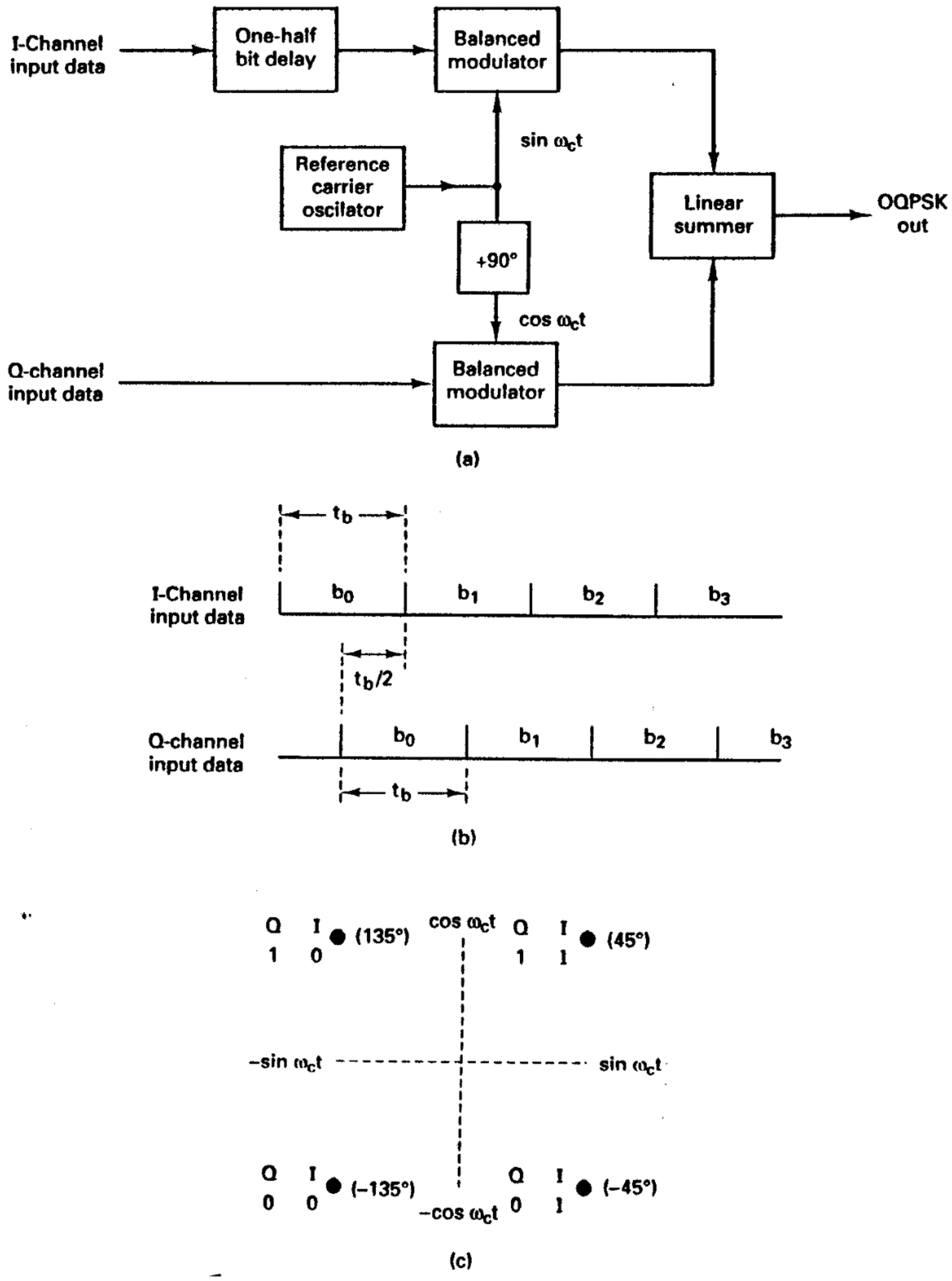


Figure (3.19): Offset keyed (OQPSK): (a) block diagram; (b) bit alignment; (c) Constellation diagram

[Refer to figure (3.18) in the text book. Page 118]

- Because changes in the I channel occur at the midpoints of the Q channel bits and vice versa, there is never more than a single bit change in the dibit code and, therefore, there is never more than a 90° shift in the output phase. In conventional QPSK, a change in the input dibit from 00 to 11 or 01 to 10 causes a corresponding 180° shift in the output phase.

- Therefore, an advantage of OQPSK is the limited phase shift that must be imparted during modulation.

- A disadvantage of OQPSK is that changes in the output phase occur at twice the data rate in either the I or Q channel".

- Consequently, with OQPSK the baud and minimum bandwidth are twice that of conventional QPSK for a given transmission bit rate. OQPSK is sometimes called OKQPSK (offset-keyed QPSK).

3.8 Eight-Phase PSK

- With 8-PSK, three bits are encoded, forming tribits and producing eight different output phases. To encode eight different phases, the incoming bits are encoded in groups of three, called tribits ($2^3 = 8$).

3.8.1 8-PSK transmitter

- A block diagram of an 8-PSK modulator is shown in Figure 3.20.

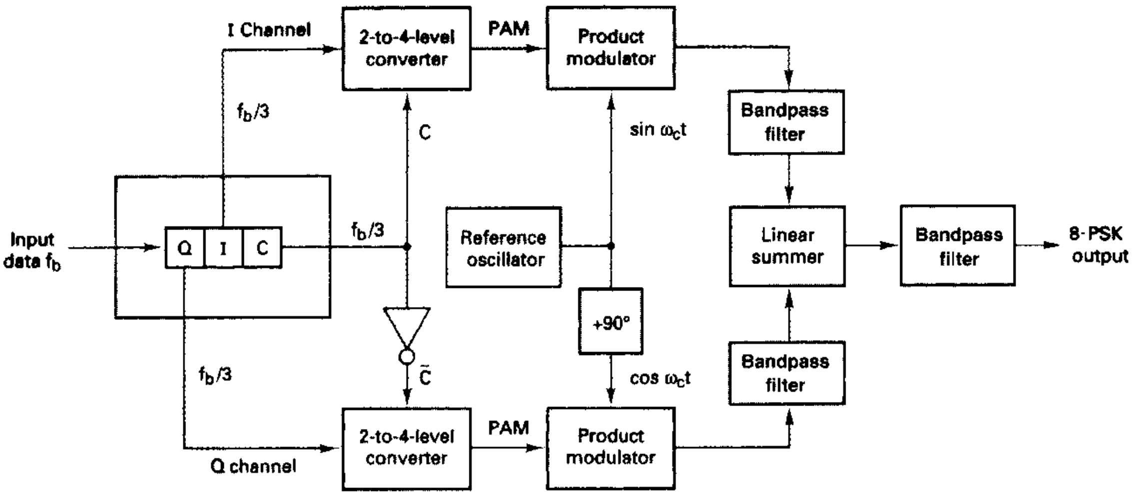


Figure (3.20): **8-PSK modulator**

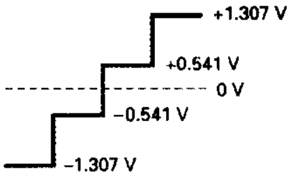
[Refer to figure (3.19) in the text book. Page 119]

I	C	Output
0	0	−0.541 V
0	1	−1.307 V
1	0	+0.541 V
1	1	+1.307 V

(a)

Q	$\bar{C}$	Output
0	1	−1.307 V
0	0	−0.541 V
1	1	+1.307 V
1	0	+0.541 V

(b)



(c)

Figure (3.21): **I- and Q-channel 2-to-4-level converters: (a) 1-channel truth table; (b) D-channel truth table; (c) PAM levels**

[Refer to figure (3.20) in the text book. Page 120]

- The bit rate in each of the three channels is f_b,/3. The bits in the I and C channels enter the I channel 2-to-4-level converter and the bits in the Q and C channels enter the Q channel 2-to-4-level converter.

- Essentially, the 2-to-4-level converters are parallel-input digital-to-analog converter, (DACs). With two input bits, four output voltages are possible.

- The I or Q bit determines the polarity of the output analog signal (logic 1= +V and logic 0 = -V), whereas the C or C bit determines the magnitude (logic 1= 1.307 V and logic 0 = 0.541 V).

- Figure 3.21 shows the truth table and corresponding output conditions for the 2-to4-level converters. Because the C and $\bar{C}$ bits can never be the same logic state, the outputs from the I and Q 2-to-4-level converters can never have the same magnitude, although they can have the same polarity. The output of a 2-to-4-level converter is an M-ary, pulse amplitude-modulated (PAM) signal where M = 4.

Example 3.5

- For a tribit input of Q = 0, 1 = 0, and C = 0 (000), determine the output phase for the 8-PSK modulator shown in Figure 3.20.

Solution

- The inputs to the I channel 2-to-4-level converter are I = 0 and C = 0. From Figure 2-24 the output is -0.541 V. The inputs to the Q channel 2-to-4-level converter are Q = 0 and $\bar{C}$ = 1.

- Again from Figure 2-24, the output is - 1.307 V.

- Thus, the two inputs to the I channel product modulators are -0.541 and sin $\omega_c t$. The output is

$$I = (-0.541)(\sin \omega_c t) = -0.541 \sin \omega_c t$$

- The two inputs to the Q channel product modulator are - 1.307 V and cos $\omega_c t$. The output is

$$Q = (-1.307)(\cos \omega_c t) = -1.307 \cos \omega_c t$$

- The outputs of the I and Q channel product modulators are combined in the linear summer and produce a modulated output of

- summer output = $-0.541 \sin \omega_c t - 1.307 \cos \omega_c t$

$$= 1.41 \sin(\omega_c t - 112.5°)$$

- For the remaining tribit codes (001, 010, 011, 100, 101, 110, and 111), the procedure is the same. The results are shown in Figure 3-25.

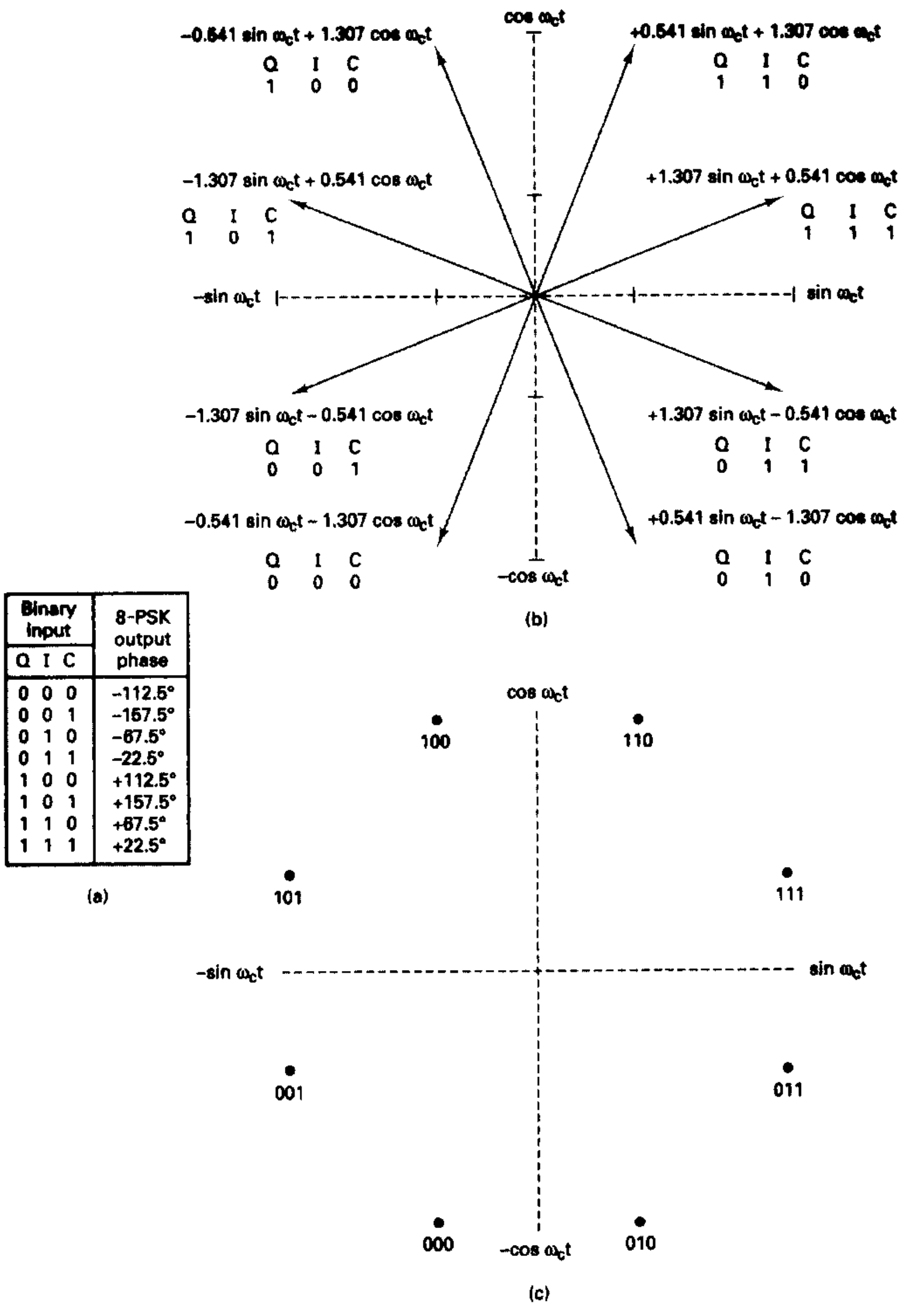


Figure (3.22): **8-PSK modulator: (a) truth table; (b) phasor diagram; (c) constellation diagram**

[Refer to figure (3.21) in the text book. Page 121]

- From Figure 3.22, it can be seen that the angular separation between any two adjacent phasors is 45°, half what it is with QPSK.

- Therefore, an 8-PSK signal can undergo almost a ± 22.5° phase shift during transmission and still retain its integrity. Also, each phasor is of equal magnitude; the tribit condition (actual information) is again contained only in the phase of the signal.

- The PAM levels of 1.307 and 0.541 are relative values. Any levels may be used as long as their ratio is 0.541/1.307 and their arc tangent is equal to 22.5°. For example, if their values were doubled to 2.614 and 1.082, the resulting phase angles would not change, although the magnitude of the phasor would increase proportionally.

- Figure 3.23 shows the output phase-versus-time relationship of an 8-PSK modulator.

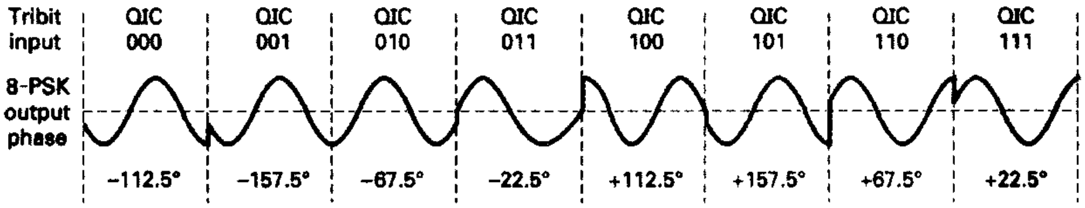


Figure (3.23): Output phase-versus-time relationship for an 8-PSK modulator

[Refer to figure (3.22) in the text book. Page 122]

3.8.2 Bandwidth considerations of 8-PSK

- With 8-PSK, because the data are divided into three channels, the bit rate in the I, Q, or C channel is equal to one-third of the binary input data rate (fb /3).
- Figure 3.24 shows that the highest fundamental frequency in the I, Q, or C channel is equal to one-sixth the bit rate of the binary input (one cycle in the I, Q, or C channel takes the same amount of time as six input bits). Also, the highest fundamental frequency in either PAM signal is equal to one-sixth of the binary input bit rate

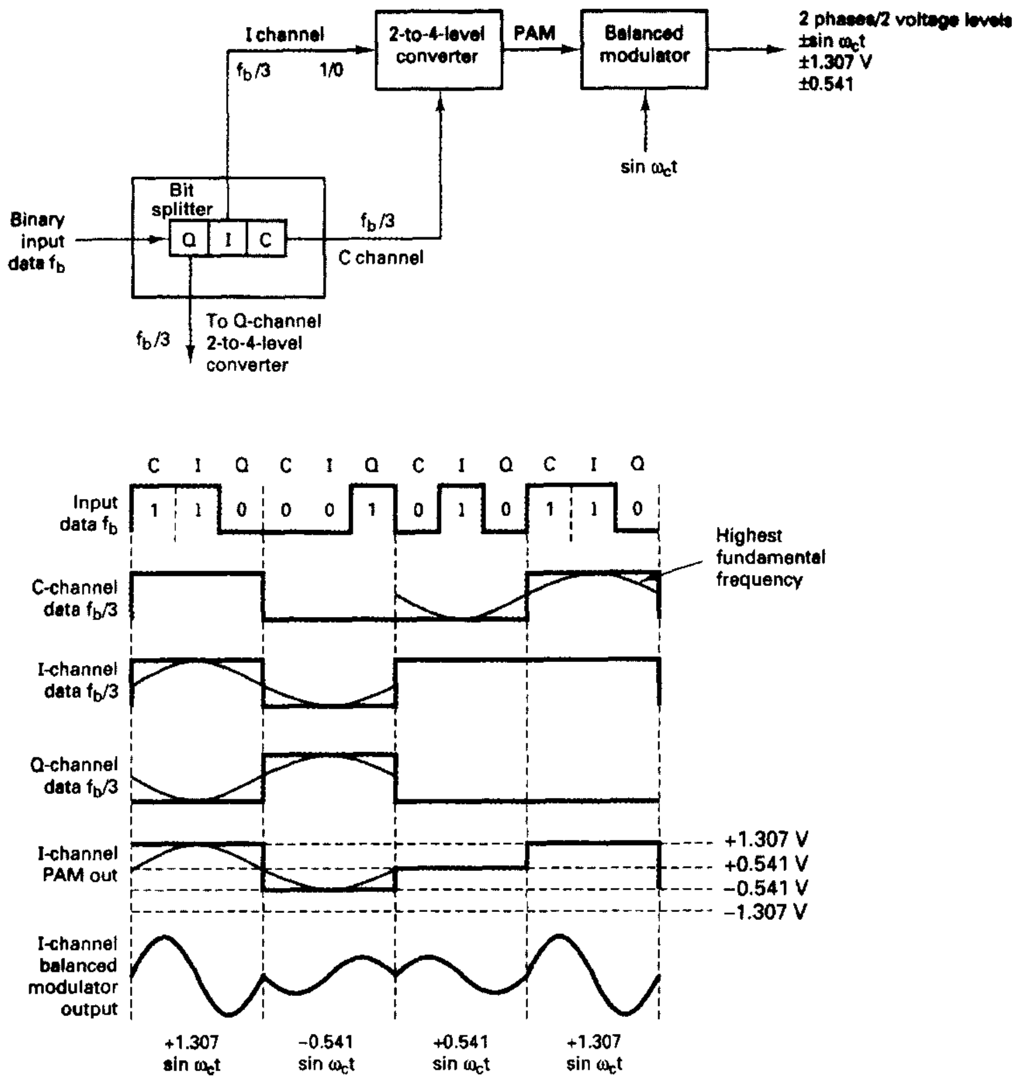


Figure (3.24): Bandwidth considerations of an 8-PSK modulator

[Refer to figure (3.23) in the text book. Page 123]

- With an 8-PSK modulator, there is one change in phase at the output for every three data input bits. Consequently, the baud for 8-PSK equals f_b / 3, the same as the minimum bandwidth. Again, the balanced modulators are product modulators; their outputs are the product of the carrier and the PAM signal. Mathematically, the output of the balanced modulators is

$$\theta = (X \sin \omega_a t)(\sin \omega_c t) \tag{3.17}$$

  - Where

$$\underbrace{\omega_a t = 2\pi \frac{f_b}{6} t}_{\text{modulating signal}} \quad \text{and} \quad \underbrace{\omega_c t = 2\pi f_c t}_{\text{carrier}}$$

  - And $X = \pm 1.307$ or ± 0.541
  - Thus

$$\theta = \left(X \sin 2\pi \frac{f_b}{6} t\right)(\sin 2\pi f_c t)$$

$$= \frac{X}{2} \cos 2\pi\left(f_c - \frac{f_b}{6}\right)t - \frac{X}{2} \cos 2\pi\left(f_c + \frac{f_b}{6}\right)t$$

- The output frequency spectrum extends from $f_c + f_b / 6$ to $f_c - f_b / 6$, and the minimum bandwidth (f_N) is

$$\left(f_c + \frac{f_b}{6}\right) - \left(f_c - \frac{f_b}{6}\right) = \frac{2f_b}{6} = \frac{f_b}{3}$$

Example 3.6

- For an 8-PSK modulator with an input data rate *(fb)* equal to 10 Mbps and a carrier frequency of 70 MHz, determine the minimum double-sided Nyquist bandwidth *(f_N)* and the baud. Also, compare the results with those achieved with the BPSK and QPSK modulators in Examples 3.2 and 3.4. If the 8-PSK block diagram shown in Figure 3.19 as the modulator model.

Solution

- The bit rate in the I, Q, and C channels is equal to one-third of the input bit rate, or 10 Mbps

$$f_{bc} = f_{bQ} = f_{b1} = 10 \text{ Mbps} / 3 = 3.33 \text{ Mbps}$$

- Therefore, the fastest rate of change and highest fundamental frequency presented to either balanced modulator is

$$f_a = f_{bc} / 2 = 3.33 \text{ Mbps} / 2 = 1.667 \text{ Mbps}$$

- The output wave from the balance modulators is

$$(\sin 2\pi f_a t)(\sin 2\pi f_c t)$$

$$0.5 \cos 2\pi(f_c - f_a)t - 0.5 \cos 2\pi(f_c + f_a)t$$

$$0.5 \cos 2\pi[(70 - 1.667)\text{MHz}]t - 0.5 \cos 2\pi[(70 + 1.667)\text{MHz}]t$$

$$0.5 \cos 2\pi(68.333\text{MHz})t - 0.5 \cos 2\pi(71.667\text{MHz})t$$

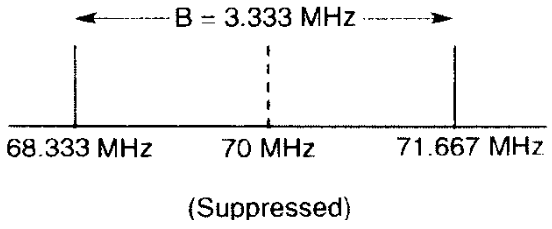
- The minimum Nyquist bandwidth is

$$f_N = (71.667 - 68.333) \text{ MHz} = 3.333 \text{ MHz}$$

- Again, the baud equals the bandwidth; thus,

$$\text{baud} = 3.333 \text{ megabaud}$$

– The output spectrum is as follows:



f_N = 3.333 MHz

- It can be seen that for the same input bit rate the minimum bandwidth required to pass the output of an 8-PSK modulator is equal to one-third that of the BPSK modulator in Example 2-4 and 50% less than that required for the QPSK modulator in Example 2-6. Also, in each case the baud has been reduced by the same proportions.

3.8.3 8-PSK receiver

- Figure 3.25 shows a block diagram of an 8-PSK receiver. The power splitter directs the input 8-PSK signal to the I and Q product detectors and the carrier recovery circuit.

- The carrier recovery circuit reproduces the original reference oscillator signal. The incoming 8-PSK signal is mixed with the recovered carrier in the I product detector and with a quadrature carrier in the Q product detector.

- The outputs of the product detectors are 4-level PAM signals that are fed to the 4-to-2-level analog-to-digital converters (ADCs). The outputs from the I channel 4-to-2-level converter are the I and C_bits, whereas the outputs from the Q channel 4-to-2-level converter are the Q and $\bar{C}$ bits. The parallel-to-serial logic circuit converts the I/C and Q/$\bar{C}$ bit pairs to serial I, Q, and C output data streams.

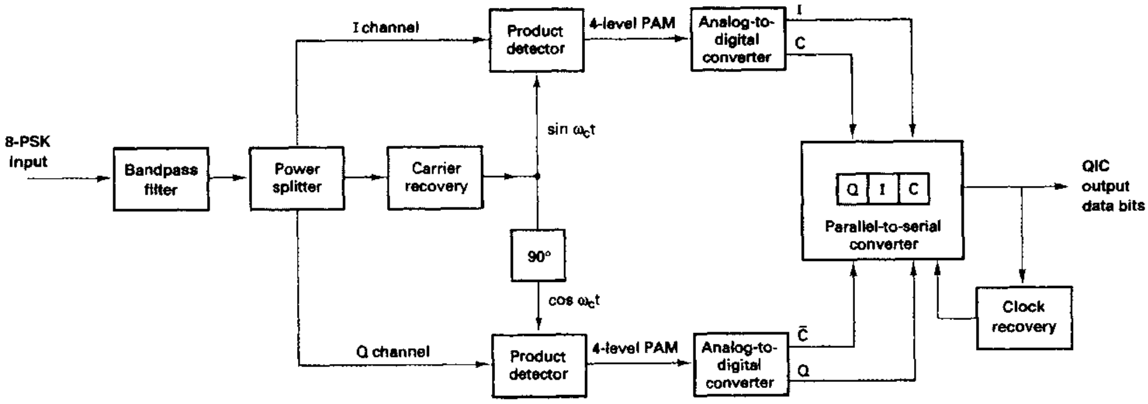


Figure (3.25): 8-PSK receiver

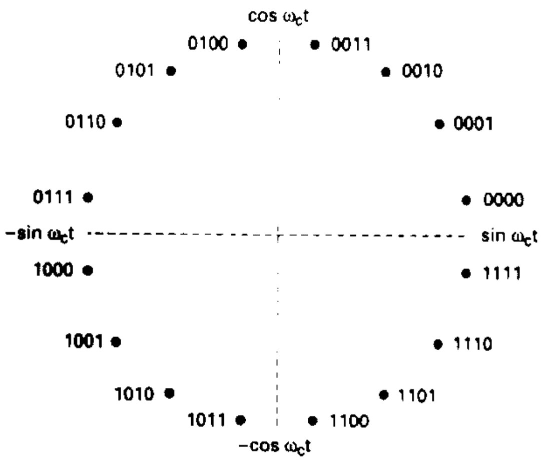
[Refer to figure (3.24) in the text book. Page 126]

3.9 Sixteen-Phase PSK

- *16-PSK* is an M-ary encoding technique where M = 16; there are 16 different output phases possible. With 16-PSK, four bits (called *quadbits)* are combined, producing 16 different output phases. With 16-PSK, $n = 4$ and $M = 16$; therefore, the minimum bandwidth and baud equal one-fourth the bit rate ($f_b/4$).

Bit code	Phase	Bit code	Phase
0000	11.25°	1000	191.25°
0001	33.75°	1001	213.75°
0010	56.25°	1010	236.25°
0011	78.75°	1011	258.75°
0100	101.25°	1100	281.25°
0101	123.75°	1101	303.75°
0110	146.25°	1110	326.25°
0111	168.75°	1111	348.75°

(a)



(b)

Figure (3.26): 16-PSK: (a) truth table; (b) constellation diagram

[Refer to figure (3.25) in the text book. Page 127]

- Figure 3.26 shows the truth table and constellation diagram for 16-

PSK, respectively. Comparing Figures 3.18, 3.23, and 3.26 shows that as the level of encoding increases (i.e., the values of n and M increase), more output phases are possible and the closer each point on the constellation diagram is to an adjacent point. With 16-PSK, the angular separation between adjacent output phases is only 22.5° (1800 / 8). Therefore, 16-PSK can undergo only a 11.25° (1800 / 16) phase shift during transmission and still retain its integrity.

- Because of this, 16-PSK is highly susceptible to phase impairments introduced in the transmission medium and is therefore seldom used

3.10 QUADRATURE – AMPLITUDE MODULATION

- Quadrature amplitude modulation (QAM) is a form of digital modulation where the digital information is contained in both the amplitude and phase of the transmitted carrier

3.11 Eight QAM

- 8-QAM is an M-ary encoding technique where M = 8. Unlike 8-PSK, the output signal from an 8-QAM modulator is not a constant-amplitude signal.

3.11.1 8-QAM transmitter

- Figure 3.27a shows the block diagram of an 8-QAM transmitter. As you can see, the only difference between the 8-QAM transmitter and the 8-PSK transmitter shown in Figure 3.23 is the omission of the inverter between the C channel and the Q product modulator.

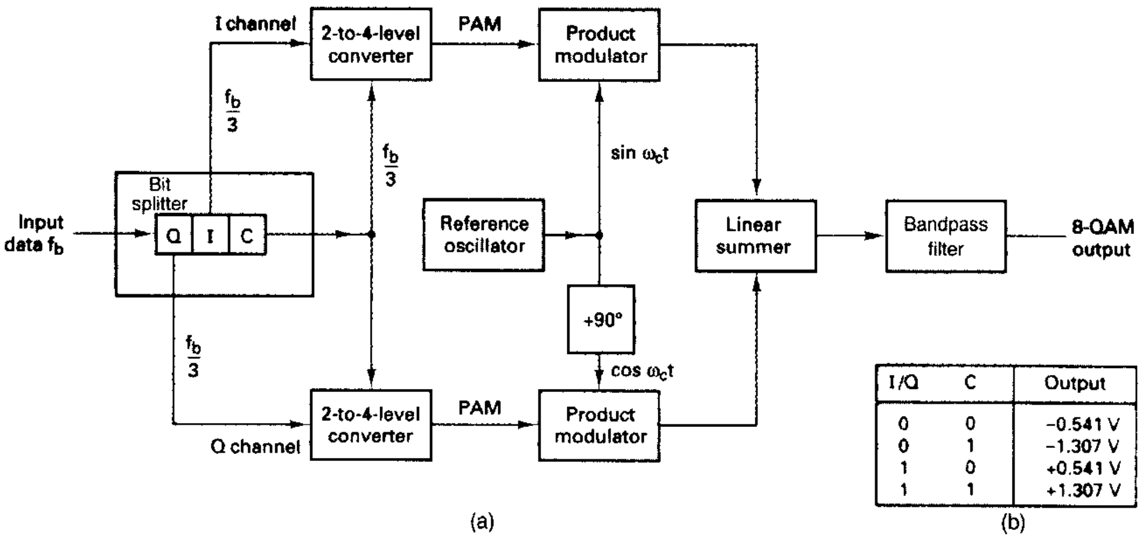


Figure (3.27): 8-OAM transmitter: (a) block diagram; (b) truth table 2-4 for the I and Q channel 2-to-4 level converters

[Refer to figure (3.26) in the text book. Page 128 & figure (3.27) in Page 129]

- As with 8-PSK, the incoming data are divided into groups of three bits (tribits): the I, Q, and C bit streams, each with a bit rate equal to one-third of the incoming data rate. Again, the I and Q bits determine the polarity of the PAM signal at the output of the 2-to-4-level converters, and the C channel determines the magnitude. Because the C bit is fed uninverted to both the I and the Q channel 2-to-4-level converters. the magnitudes of the I and Q PAM signals are always equal. Their polarities depend on the logic condition of the I and Q bits and, therefore, may be different. Figure 2-30b shows the truth table for the I and Q channel 2-to-4-level converters; they are identical.

Example 3-7

– For a tribit input of Q = 0, I= 0, and C = 0 (000), determine the output amplitude and phase for the 8-QAM transmitter shown in Figure 3.27a.

Solution

– The inputs to the I channel 2-to-4-level converter are I= 0 and C = 0.

From Figure 2-30b, the output is -0.541 V. The inputs to the Q channel 2-to-4-level converter are Q = 0 and C = 0. Again from Figure 9-30b, the output is -0.541 V.

- Thus, the two inputs to the I channel product modulator are -0.541 and sin $\omega_c t$. The output is

$$I = (-0.541)(\sin \omega_c t) = -0.541 \sin \omega_c t.$$

- The two inputs to the Q channel product modulator are -0.541 and cos $\omega_c t$.. The output is

$$Q = (-0.541)(\cos \omega_c t.) = -0.541 \cos \omega_c t.$$

I/Q	C	Output
0	0	−0.541
0	1	−1.307 V
1	0	+0.541
1	1	+1.307 V

Fig.3.27 Truth table for the I- and Q- channel 2-to-4-level converters.

- The outputs from the I and Q channel product modulators are combined in the linear summer and produce a modulated output of

$$\text{summer output} = -0.541 \sin \omega_c t. -0.541 \cos \omega_c t.$$

$$= 0.765 \sin(\cos - 135°)$$

- For the remaining tribit codes (001, 010, 0ll, 100, 101, 110, and 111), the procedure is the same. The results are shown in Figure 3.28.

Binary input			8-QAM output	
Q	I	C	Amplitude	Phase
0	0	0	0.765 V	−135°
0	0	1	1.848 V	−135°
0	1	0	0.765 V	−45°
0	1	1	1.848 V	−45°
1	0	0	0.765 V	+135°
1	0	1	1.848 V	+135°
1	1	0	0.765 V	+45°
1	1	1	1.848 V	+45°

(a)

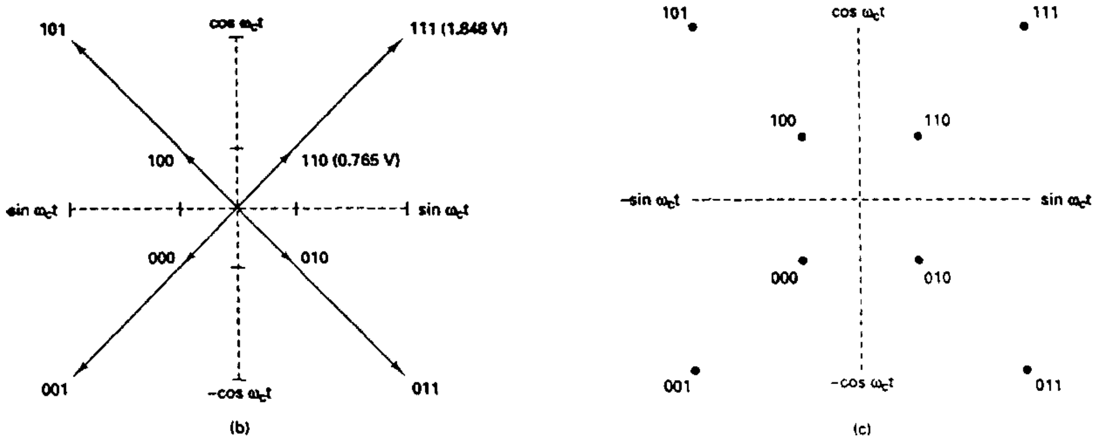


Figure (3.28): 8-QAM modulator: (a) truth table; (b) phasor diagram; (c) constellation diagram

[Refer to figure (3.28) in the text book. Page 130]

- Figure 3.29 shows the output phase-versus-time relationship for an 8-QAM modulator. Note that there are two output amplitudes, and only four phases are possible.

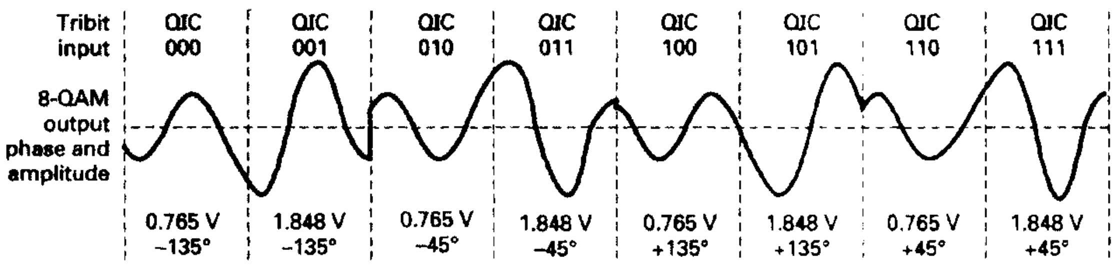


Figure (3.29): Output phase and amplitude-versus-time relationship for 8-QAM

[Refer to figure (3.29) in the text book. Page 130]

3.11.2 Bandwidth considerations of 8-QAM

- In 8-QAM, the bit rate in the I and Q channels is one-third of the input binary rate, the same as in 8-PSK. As a result, the highest fundamental modulating frequency and fastest output rate of change in 8-QAM are the same as with 8-PSK

- The minimum bandwidth required for 8-QAM is $f_b / 3$, the same as in 8-PSK.

3.11.3 8-QAM receiver

- An 8-QAM receiver is almost identical to the 8-PSK receiver shown in Figure 3.25. The difference are the PAM levels at the output of the product detectors and the binary signals at the output of the analog-to-digital converters

- Because there are two transmit amplitudes possible with 8-QAM that are different from those achievable with 8-PSK, the four demodulated PAM levels in 8-QAM are different from those in 8-PSK. Therefore, the conversion factor for the analog-to-digital converters must also be different

- Also, with 8-QAM the binary output signals from the I-channel analog-to-digital converter are the I and C bits, and the binary output signals from the Q-channel analog-to-digital converter are the Q and C bits

3.12 Sixteen-QAM

- As with the 16-PSK, 16-QAM is an M-ary system where $M = 16$. The input data are acted on in groups of four ($2^4 = 16$). As with 8-QAM, both the phase and the amplitude of the transmit carrier are varied.

3.12.1 16-QAM transmitter

- The block diagram for a 16-QAM transmitter is shown in Figure 3.30.

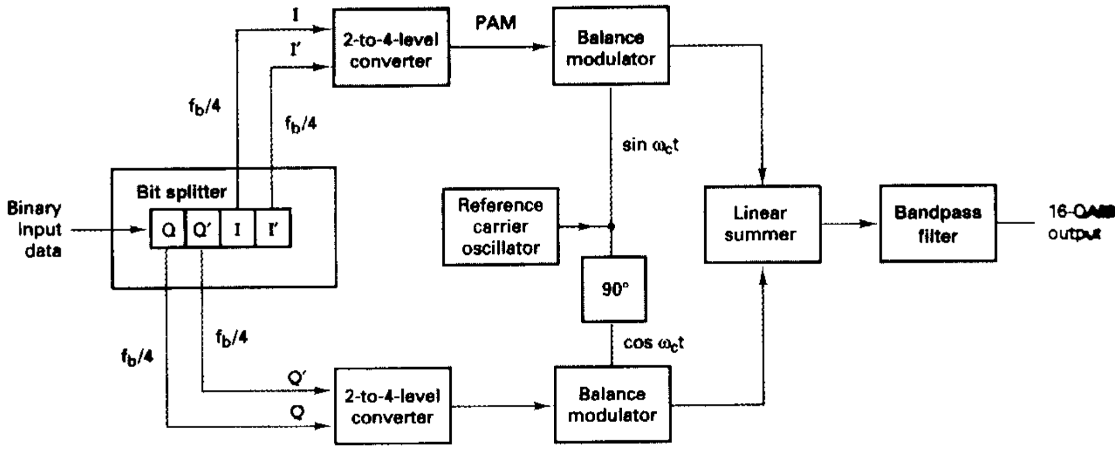


Figure (3.30): 16-QAM transmitter block diagram

[Refer to figure (3.30) in the text book. Page 132]

- The input binary data are divided into four channels: I, I', Q, and Q'. The bit rate in each channel is equal to one-fourth of the input bit rate (fb/4). The I and Q bits determine the polarity at the output of the 2-to-4-level converters (a logic 1 = positive and a logic 0 = negative). The I' and Q' buy determine the magnitude (a logic 1 = 0.821 V and a logic 0 = 0.22 V).

- For the I product modulator they are +0.821 sin ωct, -0.821 sin ωct, +0.22 sin ω_ct, and -0.22 sin ωct.

- For the Q product modulator, they are +0.821 cos ωct, +0.22 cos ωct, -0.821 cos ω_ct, and -0.22 cos ωct.
- The linear summer combines the outputs from the I and Q channel product modulators and produces the 16 output conditions necessary for 16-QAM. Figure 3.31 shows the truth table for the I and Q channel 2-to-4-level converters.

I	I'	Output
0	0	−0.22 V
0	1	−0.821 V
1	0	+0.22 V
1	1	+0.821 V

(a)

Q	Q'	Output
0	0	−0.22 V
0	1	−0.821 V
1	0	+0.22 V
1	1	+0.821 V

(b)

Figure (3.31): Truth tables for the I- and Q-channel 2-to-4evel converters: (a) I channel; (b) Q channel

[Refer to figure (3.31) in the text book. Page 132]

Example 3.8

- For a quadbit input of I= 0, I' = 0, Q = 0, and Q' = 0 (0000), determine the output amplitude and phase for the 16-QAM modulator shown in Figure 3.30.

Solution

- The inputs to the I channel 2-to-4-level converter are I = 0 and I' = 0. From Figure 2-34, the output is -0.22 V. The inputs to the Q channel 2-to-4-level converter are Q= 0 and Q' = 0. Again from Figure 3.30, the output is -0.22 V.

- Thus, the two inputs to the I channel product modulator are -0,22 V and sin $\omega_c t$. The output is

$$I = (-0.22)(\sin \omega_c t) = -0.22 \sin \omega_c t$$

- The two inputs to the Q channel product modulator are -0.22 V and cos $\omega_c t$. The output is

$$Q = (-0.22)(\cos \omega_c t) = -0.22 \cos \omega_c t$$

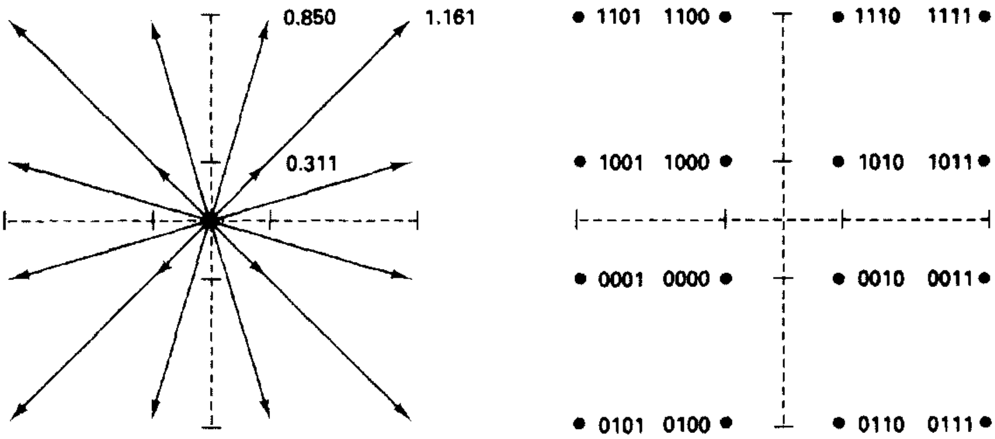
- The outputs from the I and Q channel product modulators are combined in the linear summer and produce a modulated output of

$$\text{summer output} = -0.22 \sin \omega_c t - 0.22 \cos \omega_c t$$
$$= 0.311 \sin(\omega_c t - 135°)$$

– For the remaining quadbit codes, the procedure is the same. The results are shown in Figure 3.32.

Binary input				16-QAM output	
Q	Q′	I	I′		
0	0	0	0	0.311 V	−135°
0	0	0	1	0.850 V	−165°
0	0	1	0	0.311 V	−45°
0	0	1	1	0.850 V	−15°
0	1	0	0	0.850 V	−105°
0	1	0	1	1.161 V	−135°
0	1	1	0	0.850 V	−75°
0	1	1	1	1.161 V	−45°
1	0	0	0	0.311 V	135°
1	0	0	1	0.850 V	165°
1	0	1	0	0.311 V	45°
1	0	1	1	0.850 V	15°
1	1	0	0	0.850 V	105°
1	1	0	1	1.161 V	135°
1	1	1	0	0.850 V	75°
1	1	1	1	1.161 V	45°

(a)



(b) (c)

Figure (3.32): 16-QAM modulator: (a) truth table; (b) phasor diagram; (c) constellation diagram.

[Refer to figure (3.32) in the text book. Page 134]

3.12.2 Bandwidth considerations of 16-QAM.

- With a 16-QAM, the bit rate in the I, I', Q, or Q' channel is equal to one-fourth of the binary input data rate ($f_b/4$).

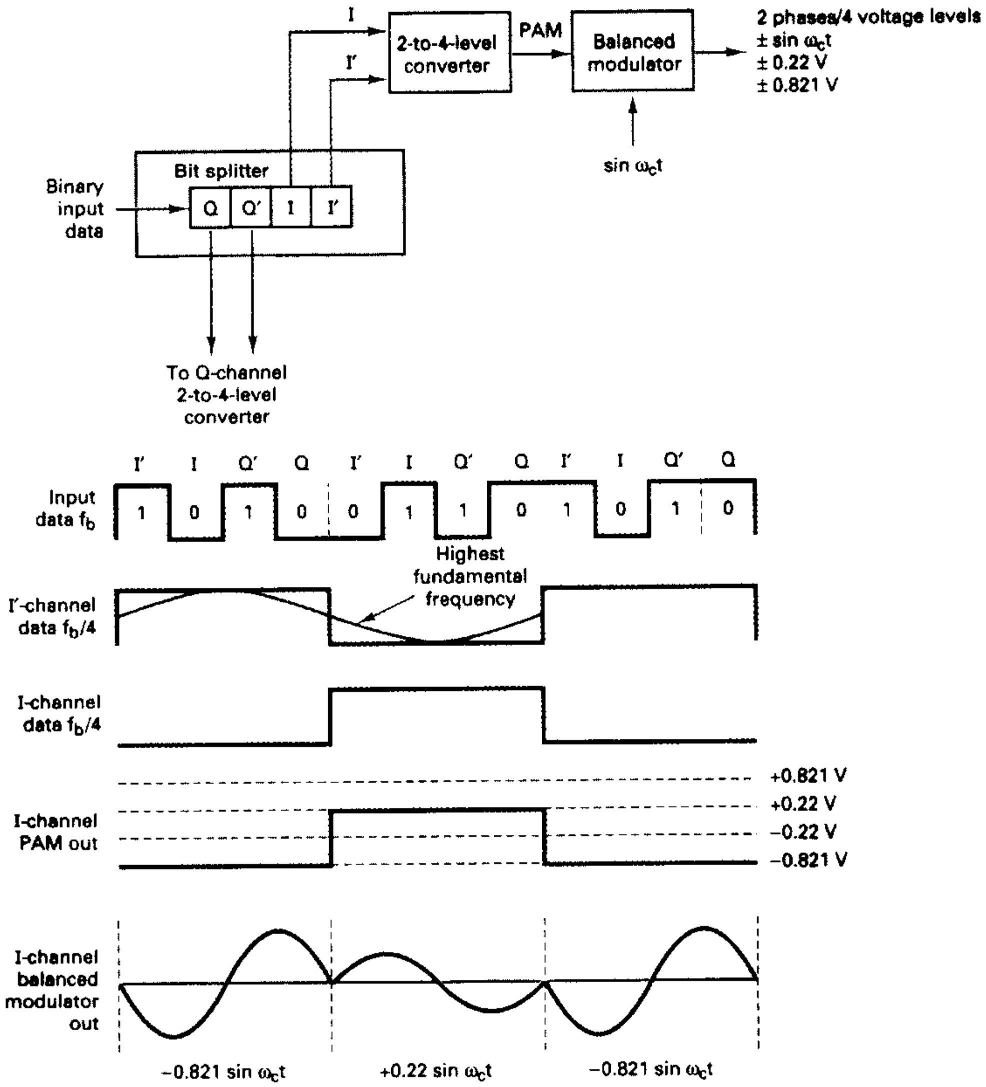


Figure (3.33): Bandwidth considerations of a 16-QAM modulator

[Refer to figure (3.33) in the text book. Page 136]

- Figure 3.33 shows the bit timing relationship between the binary input data; the I, I'. Q, and Q' channel data; and the I PAM signal. It can be seen that the highest fundamental frequency in the I, I', Q, or Q' channel is equal to one-eighth of the bit rate of the binary input data (one cycle in the I, I', Q, or Q' channel takes the same amount of time as eight input bits).
- Also, the highest fundamental frequency of either PAM signal is equal to one-eighth of the binary input bit rate. With a 16-QAM modulator, there is one change in the output signal (either its phase, amplitude, or both) for every four

input data bits. Consequently, the baud equals fb/4, the same as the minimum bandwidth.

- Again, the balanced modulators are product modulators and their outputs can be represented mathematically as

$$\text{output} = (X \sin \omega_a t)(\sin \omega_c t) \qquad (3.18)$$

  - where

$$\underbrace{\omega_a t = 2\pi \frac{f_b}{8} t}_{\text{modulating signal}} \quad \text{and} \quad \underbrace{\omega_c t = 2\pi f_c t}_{\text{carrier}}$$

  - and

$$X = \pm 0.22 \text{ or } \pm 0.821$$

  - Thus,

$$\text{output} = \left(X \sin 2\pi \frac{f_b}{8} t\right)(\sin 2\pi f_c t)$$

$$= \frac{X}{2} \cos 2\pi \left(f_c - \frac{f_b}{8}\right)t - \frac{X}{2} \cos 2\pi \left(f_c + \frac{f_b}{8}\right)t$$

- The output frequency spectrum extends from $f_c + f_b / 8$ and $f_c - f_b / 8$ the minimum bandwidth (f_N) is

$$\left(f_c + \frac{f_b}{8}\right) - \left(f_c - \frac{f_b}{8}\right) = \frac{2f_b}{8} = \frac{f_b}{4}$$

Example 3.9

- For a 16-QAM modulator with an input data rate (f_b) equal to 10 Mbps and a carrier frequency of 70 MHz, determine the minimum double-sided Nyquist frequency (f_N) and the baud. Also, compare the results with those achieved with the BPSK, QPSK, and 8-PSK modulators in Examples 3.4, 3.6, and 3.8. Use the 16-QAM block diagram shown in Figure 3.30 as the modulator model.

Solution

- The bit rate in the I, I', Q, and Q' channels is equal to one-fourth of the input bit rate,

$$f_{bI} = f_{bI'} = f_{bQ} = f_{bQ'} = f_b / 4 = 10 \text{ Mbps} / 4 = 2.5 \text{ Mbps}$$

- Therefore, the fastest rate of change and highest fundamental frequency presented to either balanced modulator is

$$f_a = f_{bI} / 2 = 2.5 \text{ Mbps} / 2 = 1.25 \text{ MHz}$$

- The output wave from the balanced modulator is

$$(\sin 2\pi f_a t)(\sin 2\pi f_c t)$$

$$0.5 \cos 2\pi(f_c - f_a)t - 0.5 \cos 2\pi(f_c + f_a)t$$

$$0.5 \cos 2\pi[(70 - 1.25)\text{MHz}]t - 0.5 \cos 2\pi[(70 + 1.25)\text{MHz}]t$$

$$0.5 \cos 2\pi(68.75\text{MHz})t - 0.5 \cos 2\pi(71.25\text{MHz})t$$

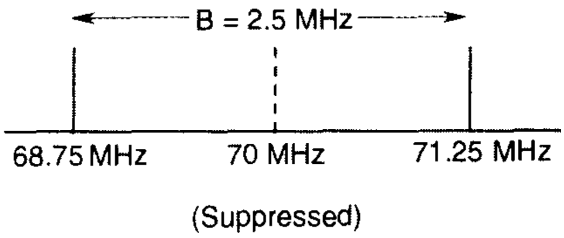
- The minimum Nyquist bandwidth is

$$f_N = (71.25 - 68.75) \text{ MHz} = 2.5 \text{ MHz}$$

- The symbol rate equals the bandwidth; thus,

symbol rate = 2.5 megabaud

- The output spectrum is as follows:



FN = 2.25 MHz

- For the same input bit rate, the minimum bandwidth required to pass the output of a 16-QAM modulator is equal to one-fourth that of the BPSK modulator, one-half that of QPSK, and 25% less than with 8-PSK. For each modulation technique, the baud is also reduced by the same proportions.

3.13 BANDWIDTH EFFICIENCY

- Bandwidth efficiency (sometimes called information density or spectral efficiency, often used to compare the performance of one digital modulation technique to another.

- Mathematical bandwidth efficiency is

$$\text{Bandwidth efficiency} = B\eta = \frac{transmission\ bit\ rate\ (bps)}{\min imum\ bandwidth\ (Hz)} = \frac{bits/s}{Hertz} = \frac{bits/s}{cycle/s} = \frac{bits}{cycle} \tag{3.19}$$

Example 3.10

– Determine the bandwidth efficiencies for the following modulation schemes: BPSK, QPSK, 8-PSK, and 16-QAM.

Solution:

– Recall from Example 3.2, 3.4, 3.6, and 3.9, the minimum bandwidths required to propagate a 10-Mbps transmission rate with the following modulation schemes:

Modulation Scheme	Minimum Bandwidth (MHz)
BPSK	10
QPSK	5
8-PSK	3.33
16-QAM	2.56

– Substituting into Equation 3.19, the bandwidth efficiencies are determined as follows:

$$\text{BPSK : BW efficiency} = \frac{10Mbps}{10MHz} = \frac{1\ bps}{Hz} = \frac{1\ bit}{cycle}$$

$$\text{QPSK : BW efficiency} = \frac{10Mbps}{5MHz} = \frac{2\ bps}{Hz} = \frac{2\ bit}{cycle}$$

$$\text{8-PSK : BW efficiency} = \frac{10Mbps}{3.33MHz} = \frac{3\ bps}{Hz} = \frac{3\ bit}{cycle}$$

$$\text{16-QAM : BW efficiency} = \frac{10Mbps}{2.5MHz} = \frac{4\ bps}{Hz} = \frac{4\ bit}{cycle}$$

– The result indicate the BPSK is the least efficient and 16-QAM is the most efficient , 16-QAM requires one-fourth as much bandwidth as BPSK for the same input rate

3.14 PSK and QAM Summary

- Table 3.2 summarizes the relationship between the number of bits encoded, the number of output conditions possible, the minimum bandwidth, and the baud for ASK, FSK. PSK, and QAM.

- When data compression is performed, higher data transmission rates are possible for a given bandwidth.

Table 3.2 ASK, FSK, PSK AND QAM summary

Modulation	Encoding	Bandwidth (Hz)	Baud	Bandwidth Efficiency (bps/Hz)
FSK	*Single bit*	$\geq f_b$	f_b	≤ 1
BPSK	*Single bit*	f_b	f_b	*1*
QPSK	*Dibit*	$f_b/2$	$f_b/2$	*2*
8-PSK	*Tribit*	$f_b/3$	$f_b/3$	*3*
8-QAM	*Tribit*	$f_b/3$	$f_b/3$	*3*
16-PSK	*Quadbit*	$f_b/4$	$f_b/4$	*4*
16-QAM	*Quadbit*	$f_b/4$	$f_b/4$	*4*

Modulation	Encoding Scheme	Outputs Possible	Minimum Bandwidth	Baud
ASK	Single bit	2	f_b	f_b
FSK	Single bit	2	f_b	f_b
BPSK	Single bit	2	f_b	f_b
QPSK	Dibits	4	$f_b/2$	$f_b/2$
8-PSK	Tribits	8	$f_b/3$	$f_b/3$
8-QAM	Tribits	8	$f_b/3$	$f_b/3$
16-QAM	Quadbits	16	$f_b/4$	$f_b/4$
16-PSK	Quadbits	16	$f_b/4$	$f_b/4$
32-PSK	Five bits	32	$f_b/5$	$f_b/5$
32-QAM	Five bits	32	$f_b/5$	$f_b/5$
64-PSK	Six bits	64	$f_b/6$	$f_b/6$
64-QAM	Six bits	64	$f_b/6$	$f_b/6$
128-PSK	Seven bits	128	$f_b/7$	$f_b/7$
128-QAM	Seven bits	128	$f_b/7$	$f_b/7$

Note: f_b indicates a magnitude equal to the input bit rate.

3.15 Carrier Recovery

- Carrier recovery is the process of extracting a phase-coherent reference carrier from a received signal. This is sometimes called phase referencing. In the phase modulation techniques described thus far, the binary data were encoded as a precise phase of the transmitted carrier. (This is referred to as absolute phase encoding).

- Depending on the encoding method, the angular separation between adjacent phasor varied between 30 and 180. To correctly demodulate the data, a phase-coherent carrier was recovered and compared with the received carrier in a product detector

- To determine the absolute phase of the received carrier, it is necessary to produce a carrier at the receiver that is phase coherent with the transmit reference oscillator. This is the function of the carrier recovery circuit

- With PSK and QAM, the carrier is suppressed in the balanced modulators and is therefore not transmitted. Consequently, at the receiver the carrier cannot simply be tracked with a standard phase-locked loop. With suppressed carrier systems, such as PSK and QAM, sophisticated methods of carrier recovery are required such as a squaring loop, a Costas loop, or a remodulator

3.15.1 Squaring Loop

- As a common method of achieving carrier recovery for BPSK is the squaring loop. Figure 3.34 shows the block diagram of a squaring loop. The received BPSK waveform is filtered and then squared. The filtering reduces the spectral width of the received noise
- The squaring circuit removes the modulation and generates the second harmonic of the carrier frequency. This harmonic is phase tracked by the PLL. The VCO output frequency from the PLL is then divided by 2 and used as the phase reference for the product detectors

- With BPSK, only two output phases are possible: $+\sin \omega_c t$ and $-\sin \omega_c t$. Mathematically, the operation of the squaring circuit can be described as follows. For a receive signal of $+\sin\omega_c t$ the output of the squaring circuit is

$$output = (+\sin\omega_c t)(+\sin\omega_c t) = [+\sin\omega_c t]^2$$
$$= \frac{1}{2}(1-\cos 2\omega_c t) = \frac{1}{2} - \frac{1}{2}\cos 2\omega_c t$$

Filtered out

- For a received signal of $-\sin\omega_c t$ the output of the squaring circuit is

$$output = (-\sin\omega_c t)(-\sin\omega_c t) = [\sin\omega_c t]^2$$
$$= \frac{1}{2}(1-\cos 2\omega_c t) = \frac{1}{2} - \frac{1}{2}\cos 2\omega_c t$$

Filtered out

Figure (3.34): Squaring loop carrier recovery circuit for a BPSK receiver

[Refer to figure (3.34) in the text book. Page 141]

- It can be seen that in both cases the output from the squaring circuit contained a constant voltage $(+\frac{1}{2}V)$ and a signal at twice the carrier frequency ($\cos 2\omega_c t$). The constant voltage is removed by filtering, leaving only $\cos 2\omega_c t$, which is fed to a frequency divider producing at its output a signal with carrier frequency

3.15.2 Costas Loop

- A second method of carrier recovery is the Costas, or quadrature loop shown in figure 3.35. The Costas loop produces the same results as a squaring circuit followed by an ordinary PLL in place of the BPF. This recovery scheme uses two parallel tracking loops (I and Q) simultaneously to drive the product of the I and Q components of the signal that drives the VCO

- The in-phase (I) loop uses the VCO as in a PLL, and the quadrature (Q) loop uses a 90^o shifted VCO signal. Once the frequency of the VCO is equal to the suppressed carrier frequency, the product of the I and Q signals will produce an error voltage proportional to any phase error in the VCO. The error voltage controls the phase and thus the frequency of the VCO

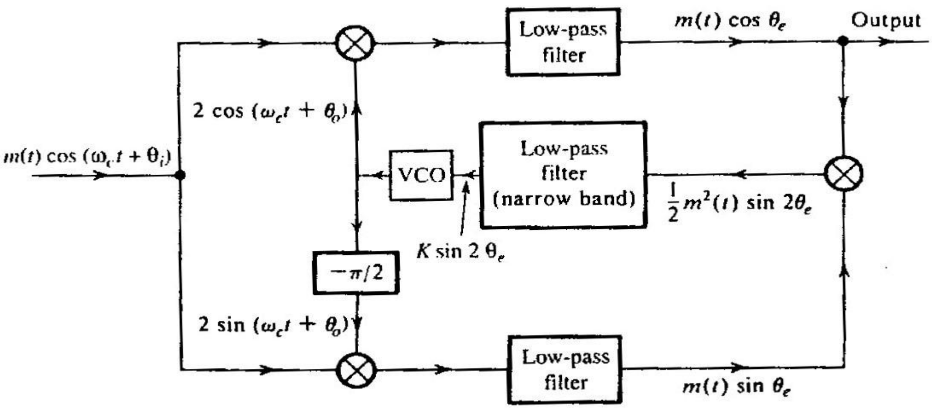


Figure (3.35): Costas loop carrier recovery circuit

[Refer to figure (3.35) in the text book. Page 141]

3.15.3 Remodulator

- A third method of achieving recovery of a phase and frequency coherent carrier is the remodulator shown in figure 3.36. The remodulator produces a loop error voltage that is proportional to twice the phase error between the incoming signal and the VCO signal. The remodulator has a faster acquisition time than either the squaring or the coasts loops

- Carrier recovery circuits for higher than binary encoding techniques are similar to BPSK except that circuits which raise the receive signal to the fourth, eighth, and higher powers are used

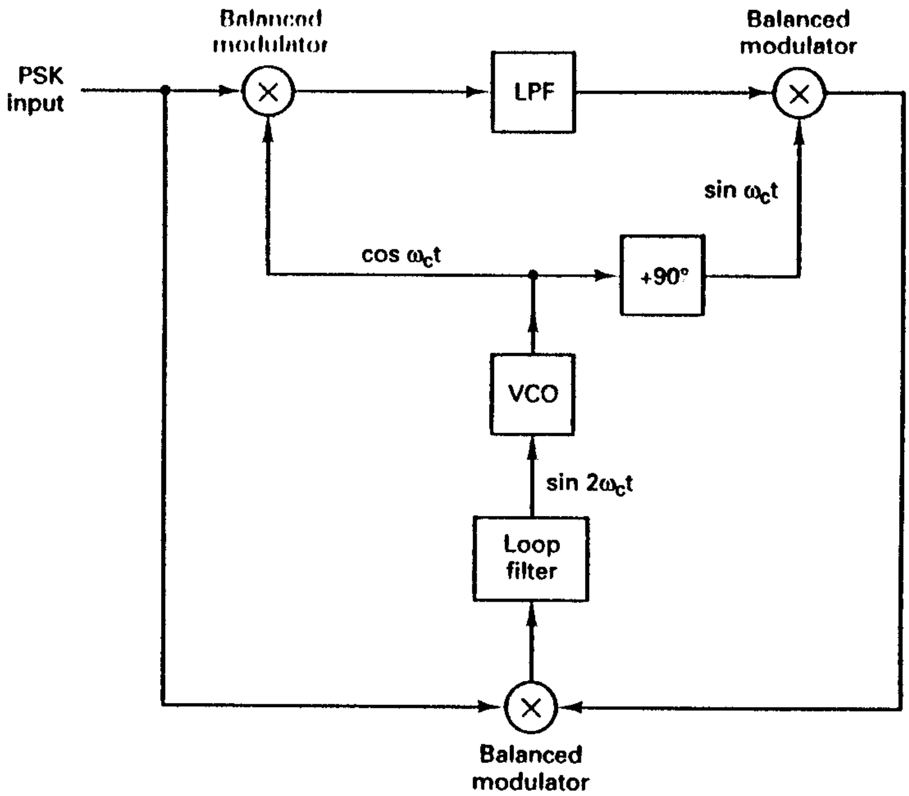


Figure (3.36): Remodulator loop carrier recovery circuit

[Refer to figure (3.36) in the text book. Page 143]

3.16 Differential Phase-Shift Keying

- Differential phase-shift keying (DPSK) is an alternative form of digital modulation where the binary input information is contained in the difference between two successive signaling elements rather than the absolute phase.

3.17 Differential BPSK

- Differential phase shift keying (DPSK) is an alternative form of digital modulation where the binary input information is contained in the difference between two successive signaling elements rather than the absolute phase

- With DPSK it is not necessary to recover a phase-coherent carrier. Instead, a received signaling element is delayed by one signaling element time slot and then compared to the next received signaling element. The difference in the phase of the two signaling elements determines the logic condition of the data

3.17.1 DBPSK transmitter

- Figure 3.37a shows a simplified block diagram of a differential binary phase-shift keying (DBPSK) transmitter. An incoming information bit is XNORed with the preceding bit prior to entering the BPSK modulator (balanced modulator).
- For the first data bit, there is no preceding bit with which to compare it. Therefore, an initial reference bit is assumed. Figure 3.37b shows the relationship between the input data, the XNOR output data, and the phase at the output of the balanced modulator. If the initial reference bit is assumed a logic 1, the output from the XNOR circuit is simply the complement of that shown.
- In Figure 3.37b, the first data bit is XNORed with the reference bit. If they are the same, the XNOR output is a logic 1; if they are different, the XNOR output is a logic 0. The balanced modulator operates the same as a conventional BPSK modulator; a logic 1 produces $+\sin \omega_c t$ at the output, and a logic 0 produces $-\sin \omega_c t$ at the output.

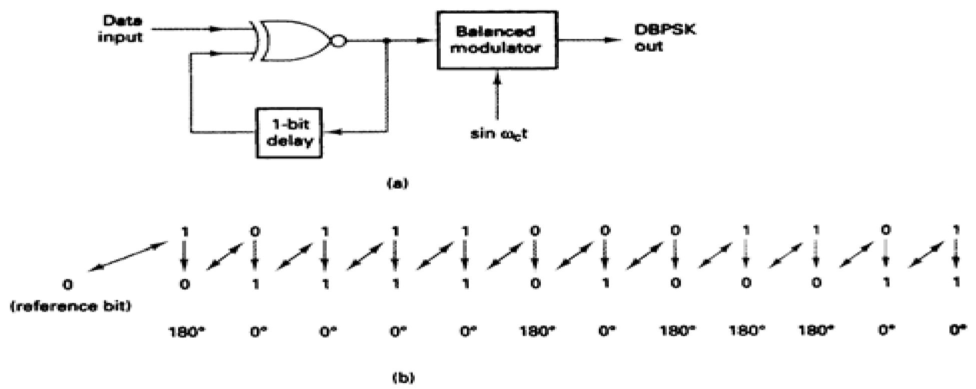


Figure (3.37): DBPSK modulator (a) block diagram (b) timing diagram

[Refer to figure (3.37) in the text book. Page 144]

3.17.2 DBPSK receiver

- Figure 3.38 shows the block diagram and timing sequence for a DBPSK receiver. The received signal is delayed by one bit time, then compared with the next signaling element in the balanced modulator. If they are the same. A logic 1(+ voltage) is generated. If they are different, a logic 0 (- voltage) is generated. [f the reference phase is incorrectly assumed, only the first demodulated bit is in error. Differential encoding can be implemented with higher-than-binary digital modulation schemes, although the differential algorithms are much more complicated than for DBPSK.

- The primary advantage of DBPSK is the simplicity with which it can be implemented. With DBPSK, no carrier recovery circuit is needed. A disadvantage of DBPSK is, that it requires between 1 dB and 3 dB more signal-to-noise ratio to achieve the same bit error rate as that of absolute PSK.

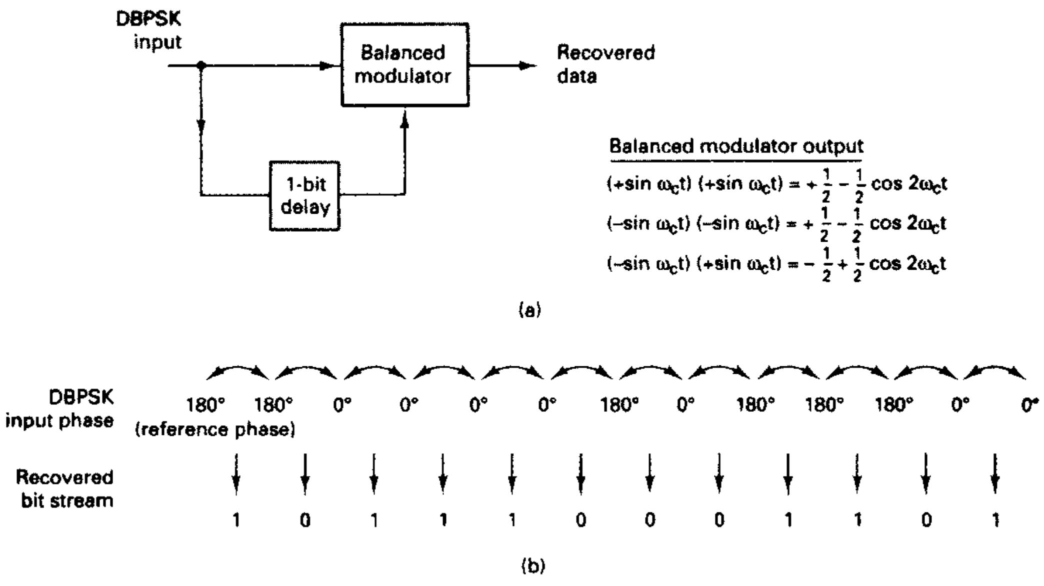


Figure (3.38): DBPSK demodulator: (a) block diagram; (b) timing sequence

[Refer to figure (3.38) in the text book. Page 145]

3.18 Clock Recovery

- As with any digital system, digital radio requires precise timing or clock synchronization between the transmit and the receive circuitry. Because of this, it is necessary to regenerate clocks at the receiver that are synchronous with those at the transmitter

- Figure 3.39a shows a simple circuit that is commonly used to recover clocking information from the received data. The recovered data are delayed by one-half a bit time and then compared with the original data in an XOR circuit. The frequency of the clock that is recovered with this method is equal to the received data rate (fb)

- Figure 3.39b shows the relationship between the data and the recovered clock timing. From figure 3.39b it can be seen that as long as the receive data contains a substantial number of transitions (1/0 sequence), the recovered clock is maintained

- If the receive data were to undergo an extended period of successive 1's or 0's, the recovered clock would be lost. To prevent this from occurring, the data are scrambled at the transmit end and descrambled at the receive end. Scrambling introduces transitions (pulses) into the binary signal using a prescribed algorithm, and the descrambler uses the same algorithm to remove the transitions

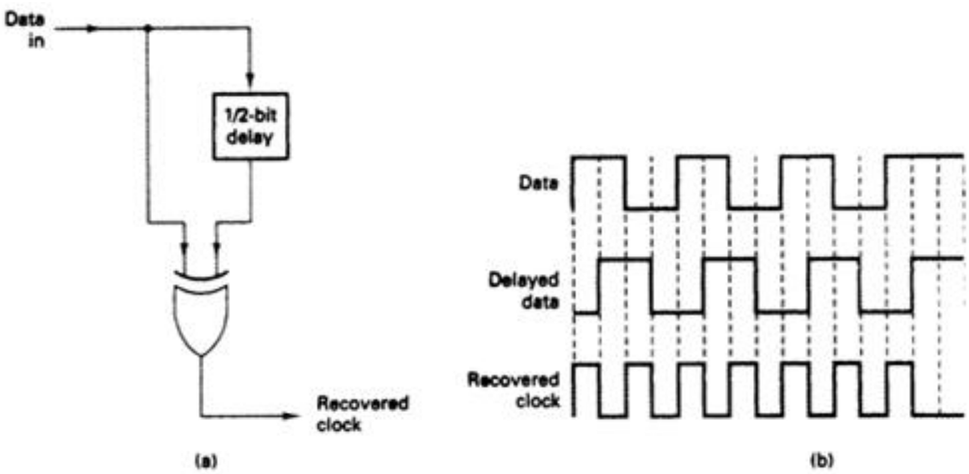


Figure (3.39): (a) Clock recovery circuit, (b) Timing diagram

[Refer to figure (3.35) in the text book. Page 141]

3.19 Applications for Digital Modulation

- A digitally modulated transceiver (transmitter-receiver) that uses FSK, PSK, or QAM has many applications. They are used in digitally modulation microwave radio and satellite systems with carrier frequencies from tens of megahertz to several gigahertz, and they are also used for voice band data modems with carrier frequencies between 300 and 3000 Hz

Ch.[4] Random Noise

- Noise is probably the only topic in electronics and telecommunications with which every-one must be familiar, no matter what his or her specialization. Electrical disturbances interfere with signals, producing ‘noise. It is ever present and limits the performance of most systems. Measuring it is very contentious almost everybody has a different method of quantifying noise and its effects

- After studying this chapter, you should be familiar with the types and sources of noise. The methods of calculating the noise produced by various sources will be learned, and so will be the ways of adding such noise. The very important noise quantities, signal-to-noise ratio, noise figure, and noise temperature, will have been covered in detail, as methods of measuring noise

4.1 Introduction

– Noise may be defined, in electrical terms, as any unwanted introduction of energy tending to interfere with the proper reception and reproduction of transmitted signals. Many disturbances of an electrical nature produce noise in receivers, modifying the signal in an unwanted manner.

– In radio receivers, noise may produce hiss in the loudspeaker output. In television receivers “snow”, or “confetti” (colored snow) be-comes superimposed on the picture. In pulse communications systems, noise may produce unwanted pulses or perhaps cancel out the wanted ones. It may cause serious mathematical errors.

- Noise can limit the range of systems, for a given transmitted power. It affects the sensitivity of receivers, by placing a limit on the weakest signals that can be amplified. It may sometimes even force a reduction in the bandwidth of a system.

- There are numerous ways of classifying noise. It may be subdivided according to type, source, effect, or relation to the receiver, depending on circumstances. It is most convenient hereto divide noise into two broad groups: noise whose sources are external to the receiver and noise created within the receiver itself. External noise is difficult to treat quantitatively, and there is often little that can be done-about it, short of moving the system to another location.
- Note how radio telescopes are always located away from industry, whose processes create so much electrical noise. International satellite earth stations are also located in noise-free valleys, where possible. Internal noise is both more quantifiable and capable of being reduced by appropriate receiver design
- Because noise has such a limiting effect, and also because it is often possible to reduce its effects through intelligent circuit use and design, it is most important for all those connected with communications to be well informed about noise and its effects

4.2 External Noise

- The various forms of noise created outside the receiver come under the heading of external noise and include atmospheric and extraterrestrial noise and industrial noise

4.2.1 Atmospheric Noise

- Perhaps the best way to become acquainted with atmospheric noise is to listen to shortwaves on a receiver, which is not well equipped to receive them. An astonishing variety of strange sounds will be heard, all tending to interfere with the program. Most of these sounds are the result of spurious radio waves, which induce voltages in the antenna. The majority of these radio waves come from natural sources of disturbance. They represent atmospheric noise, generally called static.

- Static is caused by lightning discharges in thunderstorms and other natural electric disturbances occurring in the atmosphere. It originates in the form of amplitude-modulated impulses, and because such processes are random in nature, it is spread over most of the RF spectrum normally used for broadcasting .Atmospheric noise consists of spurious radio signals with components distributed over a wide range of frequencies.

- It is propagated over the earth in the same way as ordinary radio waves of the same frequencies, so that at any point on the ground, static will be received from all thunderstorms, local and distant. Field strength is inversely proportional to frequency, so that this noise will interfere more with the reception of radio than that of television.

- Such noise consists of impulses, and these non-sinusoidal waves have harmonics whose amplitude falls off with increase in the harmonic. Static from distant Sources will vary in intensity according to the variations in propagating conditions. The usual increase in its level takes place at night, at both broadcast and shortwave frequencies

– Atmospheric noise becomes less severe at frequencies above about 30 MHz because of two separate factors. First, the higher frequencies are limited to line-of-sight propagation, i.e., less than 80 kilometers or so. Second, the nature of the mechanism generating this noise is such that very little of it is created in the VHF range and above

4.2.2 Extraterrestrial Noise

– It is safe to say that there are almost as many types of space noise as there are sources. For convenience, a division into two subgroups will suffice.

→ **Solar noise:** The sun radiates so many things noise is noticeable among them, again there are two types. Under normal "quiet" conditions, there is a constant noise radiation from the sun, simply because it is a large body at a very high temperature (over 6000°C on the surface). It therefore radiates over a very broad frequency spectrum, which includes the frequencies we use for communications.

- However, the sun is a constantly changing star, which undergoes cycles of peak activity from which electrical disturbances erupt, such as corona flares and sunspots. Even though the additional noise produced comes from a limited portion of the sun's surface, it may still be orders of magnitude greater than that received during periods of quiet sun.
- The solar cycle disturbances repeat themselves approximately every 11 years. In addition, if a line is drawn to join these 11 years peaks, it is seen that a super-cycle is in operation with the peaks reaching even higher maximum every 100 years or so. Finally, these 100 year peaks appear to be increasing in, intensity

→ **Cosmic noise**: Since distant stars are also suns and have high temperatures, they radiate RF noise, in the same manner as our sun, and what they lack in nearness they nearly make up in numbers which in combination can become significant. The noise received is called thermal (or blackbody) noise and is distributed fairly uniformly over the entire sky. We also receive noise from the center of our own galaxy (the Milky Way), from other galaxies, and from other virtual point sources such as "quasars" and "pulsars." This galactic noise is very intense, but it comes from sources, which are only points in the sky.

– Summary Space noise is observable at frequencies in the range from about 8 MHz to somewhat above 1.43 gigahertz (1.43GHz), the latter frequency corresponding to the 21-cm hydro-gen "line." Apart from man-made noise it is the strongest component over the range of about 20 to 120 MHz. Not very much of it below 20 MHz penetrates down through the ionosphere, while its eventual disappearance at frequencies in excess of 1.5 GHz is probably governed by the mechanisms generating it, and its absorption by hydrogen in interstellar space

4.2.3 Industrial Noise

- Between the frequencies of 1 to 600 MHz (in urban, sub-urban and other industrial areas) the intensity of noise made by humans easily outstrips that created by any other source, internal or external to the receiver. Under this heading, sources such as auto- mobile and aircraft ignition, electric motors and switching equipment, leakage from high-voltage lines and a multitude of other heavy electric machines are all included.

- Fluorescent lights are another powerful source of such noise and therefore should not be used where sensitive receiver reception or testing is being conducted. The noise is produced by the arc discharge present in all these operations, and under these circumstances it is not surprising that this noise should be most intense in industrial and densely populated areas

- The nature of industrial noise is so variable that it is difficult to analyze it on any basis other than the statistical. It does, however, obey the general principle that received noise increases as the receiver bandwidth is increased

4.3 Internal Noise

- Under the heading of internal noise, we discuss noise created by any of the active or passive devices found in receivers. Such noise is generally random, impossible to treat on an individual voltage basis, but easy to observe and describe statistically. Because the noise is randomly distributed over the entire radio spectrum there is on the average, as much of it at one frequency as at any other, Random noise power is proportional to the bandwidth over which it is measured

4.3.1 Thermal Agitation Noise

- The noise generated in a resistance or the resistive component is random and is referred to as thermal, agitation, white or Johnson noise. It is due to the rapid and random motion of the molecules (atoms and electrons) inside the component itself

- In thermodynamics, kinetic theory shows that the temperature of a particle is a way of expressing its kinetic energy. Thus the "temperature" of a body is the statistical root mean square (rms) value of the velocity of motion of the particles in the body
- As the theory states, the kinetic energy of these particles becomes approximately zero(i.e., their motion ceases) at the temperature of absolute zero, which is 0 K (Kelvins, formerly called degrees Kelvin) and very nearly equals -273ºC
- It becomes apparent that the noise generated by a resistor is proportional to its absolute temperature, in addition to being proportional to the bandwidth over which the noise is to be measured. Therefore

$$P_n \propto T\delta f = KT\delta f \qquad (4.1)$$

– Where
  - K = Boltzmann's constant = 1.38 x 10^{-23} J (Joules)/K the appropriate proportionality constant in this case
  - T = absolute temperature, K = 273 + C°
  - δf = bandwidth of interest
  - Pn = maximum noise power output of a resistor
  - α = varies directly

– If an ordinary resistor at the standard temperature of 17 C° (290 K) is not connected to any voltage source, it might at first be thought that there is no voltage to be measured across it. That is correct if the measuring instrument is a direct current (dc) voltmeter, but it is incorrect if a very sensitive electronic voltmeter is used

– The resistor is a noise generator, and there may even be quite a large voltage across it. Since it is random and therefore has a finite rms value but no dc component, only the alternating current (ac) meter will register at a reading. This noise voltage is caused by the random movement of electrons within the resistor, which constitutes a current

– It is true that as many electrons arrive at one end of the resistor as at the other over any long period of time. At any instant of time, there are bound to be more electrons arriving at one particular end than at the other because their movement is random. The rat of the arrival of electrons at either end of the resistor therefore varies randomly, and so does the potential difference between the two ends. A random voltage across the resistor definitely exists and may be both measured and calculated

- It must be realized that all formulas referring to random noise are applicable only to the rms value of such noise, not to its instantaneous value, which is quit unpredictable. So far as peak noise voltages are concerned, all that may be stated is that they are unlikely to have values in excess of 10 times the rms value
- Using Eq. (4.1), the equivalent circuit of a resistor as a noise generator may be drawn as in Figure 4.1, and from this the resistor's equivalent noise voltage V_n may be calculated. Assume that R_L is noiseless and is receiving the maximum noise power generated by R; under these conditions of maximum power transfer, R_L must equal to R. Then

$$P_n = \frac{V^2}{R_L} = \frac{V^2}{R} = \frac{(V_n/2)^2}{R} = \frac{V_n^2}{4R}$$

$$V_n^2 = 4RP_n = 4RKT\delta f$$

- And

$$V_n = \sqrt{4KT\delta fR} \tag{4.2}$$

- It is seen from Eq. (4.2) that the square of the rms noise voltage associated with a resistor is proportional to the absolute temperature, the value or its resistance, and the bandwidth over which the noise is measured. Note especially that the generated noise voltage is quite independent of the frequency at which it is measured. This stems from the fact that it is random and therefore evenly distributed over the frequency spectrum

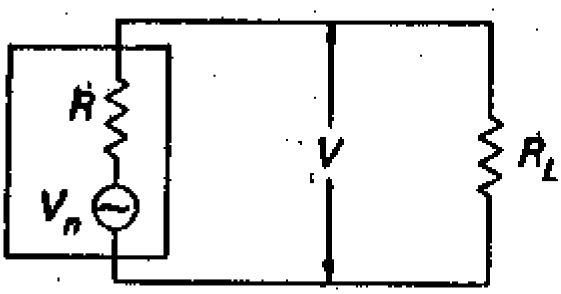


Figure (4.1): Resistance noise generator

[Refer to figure (4.1) in the text book. Page 156]

Example 4.1

- An amplifier operating over the frequency range from 18 to 20 MHz has a 10-Kilohm (10=KΩ) input resistor. What is the rms noise voltage at the input to this amplifier if the ambient temperature is 27 C° ?

Solution

$$V_n = \sqrt{4KT\delta fR}$$

$$= \sqrt{4x1.38x10^{-23}x(27+273)x(20-18)x10^6x10^4}$$

$$= \sqrt{4x1.38x3x2x10^{-11}} = 1.82x10^{-5} = 18.2 \text{ microvolt}(18.2\mu V)$$

– As we can see from this example, it would not be possible to expect this amplifier to handle signals unless they were considerably larger than 18.2 µV. A low voltage fed to this amplifier would be masked by the noise and lost

4.3.2 Shot Noise

– Thermal agitation is by no means the only source of noise in receivers. The most important of all the other sources is the shot effect, which leads to shot noise in all amplifying devices and virtually all active devices. Shot noise is caused by random variations in the arrival of electrons (or holes) at the output electrode of an amplifying device and appears as a randomly varying noise current superimposed on the output

- Although the average output current of a device is governed by the various bias voltages, at any instant of time there may be more or fewer electrons arriving at the output electrode. In bipolar transistors, this is mainly a result of the random drift of the discrete current carriers across the junctions.

- The path taken are random and therefore unequal, so that although the average collector current is constant, minute variations nevertheless occur. Shot noise behaves in a similar manner to thermal agitation noise, apart from the fact that it has a different sources

- Many variables are involved in the generation of this noise in the various amplifying devices, and so it is customary to use approximate equations for it. In addition, shot noise current is a little difficult to add to thermal noise voltage in calculations, so that for all devices with the exception of the diode, shot noise formulas used are generally simplified

- For a diode, the formula is exactly

$$i_n = \sqrt{2ei_p\delta f} \tag{4.2}$$

- Where
    - i_n = rms shot noise current
    - e = charge of an electron = 1.6 x 10^{-19} C
    - i_p = direct diode current
    - δf = bandwidth of system

- Note: it may be shown that, for a vacuum tube diode, Eq. (4.3) applies only under so-called temperature limited conditions, under which the "virtual cathode" has not been formed

- In all other instant not only is the formula simplified but it is not even a formula shot noise current. The most convenient method of dealing with shot noise is to find the value or formula for an equivalent input-noise resistor. This precedes the device, which is now assumed to be noiseless, and has a value such that the same amount of noise is present at the output of the equivalent system as in the practical amplifier.

- The noise current has been replaced by a resistance so that it is now easier to add shot noise to thermal noise. It has also been referred to the input of the amplifier, which is a much more convenient place, as will be seen

- The value of the equivalent shot noise resistance R_{eq} of a device is generally quoted in the manufacturer's specifications. Approximate formulas for equivalent shot noise resistances are also available. They all show that such noise is inversely proportional to transconductance and also directly proportional to output current
- So far as the use of R_{eq} is concerned, the important thing to realize is that it is a completely fictitious resistance, whose sole function is to simplify calculations involving shot noise. For noise only, this resistance is treated as though it were an ordinary noise creating resistor, at the same temperature as all the other resistors, and located in series with the input electrode of the device

4.3.3 Transit Time Noise

- If the time taken by an electron to travel from the emitter to the collector of a transistor becomes significant to the period of the signal being amplified, i.e., at frequencies in the upper VHF range and beyond, the so-called transit-time effect take place, and the noise input admittance of the transistor increases. The minute currents induced in the input of the device by random fluctuations in the output current become of a great importance at such frequencies and create random noise (frequency distortion)

- Once this high frequency noise makes its presence felt, it goes on increasing with frequency at a rate that soon approaches 6 decibels (6 dB) per octave, and this random noise then quickly predominates over the other forms. The result of all this is that it is preferable to measure noise at such high frequencies, instead of trying to calculate an input equivalent noise resistance for it

- Radio frequency (RF) transistors are remarkably low noise. A noise figure as low as 1 dB is possible with transistor amplifiers well into the UHF range

4.3.4 Miscellaneous Noise

- **Flicker** At low audio frequencies, a poorly understood form of noise called flicker or modulation noise is found in transistors. It is proportional to emitter current and junction temperature, but since it is inversely proportional to frequency, it may be completely ignored above about 500 Hz. It is no longer very serious

- **Resistance** Thermal noise, sometimes called resistance noise, is also present in transistors. It is due to the base, emitter, and collector internal resistances, and in most circumstances the base resistances makes the largest contribution
- From above 500 Hz up to about $f_{ab}/5$, transistor noise remains relatively constant, so that an equivalent input resistance for shot and thermal noise may be freely used

- **Noise in mixers** Mixers (nonlinear amplifying circuits) are much noisier than amplifiers using identical devices, except at microwave frequencies, where the situation is rather complex. This high value of noise in mixers is caused by two separate effects. First, conversion transconductance of mixers is much lower than the transconductance of amplifiers. Second, if image frequency rejection is inadequate, as often happens at shortwave frequencies, noise associated with the image frequency will also be accepted

4.4 Noise Calculations

4.4.1 Addition of Noise due to Several Sources

- Let's assume there are two sources of thermal agitation noise generators in series:

$V_{n1} = \sqrt{4KT\delta f R_1}$ and $V_{n2} = \sqrt{4KT\delta f R_2}$ The sum of two such rms voltages in series is given by the square root of their squares, so that we have

$$V_{n,tot} = \sqrt{V_{n1}^2 + V_{n2}^2} = \sqrt{4KT\delta f R_1 + 4KT\delta f R_2}$$

$$= \sqrt{4KT\delta f (R_1 + R_2)} = \sqrt{4KT\delta f R_{tot1}} \qquad (4.4)$$

– Where

$$R_{tot} = R_1 + R_2 + \cdots \tag{4.5}$$

– It is seen from the previous equations that in order to find the total noise voltage caused by several sources of thermal noise in series, the resistances are added and the noise voltage is calculated using this total resistance. The same procedure applies if one of those resistances is an equivalent input-noise resistance

Example 4.2

- Calculate the noise voltage at the input of a television RF amplifier, using a device that has a 200-ohm (200Ω) equivalent noise resistance and a 300Ω input resistor. The bandwidth of the amplifier is 6 MHz, and the temperature is 17 C^{o}

Solution

$$V_{n,tot} = \sqrt{4KT\delta f R_{tot}}$$

$$= \sqrt{4\text{x}1.38\text{x}10^{-23}\text{x}(17+273)\text{x}6\text{x}10^{6}\text{x}(300+200)}$$

$$= \sqrt{4\text{x}1.38\text{x}2.9\text{x}6\text{x}5\text{x}10^{-13}} = 6.93\text{x}10^{-6} = 6.93\ \mu\text{V}$$

– To calculate the noise voltage due to several resistors in parallel, find the total resistance by standard methods, and the substitute this resistance into Eq. (4.4) as before. This means that the total noise voltage is less than that due to any of the individual resistors, but, as shown in Eq. (4.1), the noise power remains constant

4.4.2 Addition of Noise due to Several Amplifiers in Cascade

- The situation that occurs in receivers is illustrated in Figure 4.2. It shows a number of amplifying stages in cascade, each having a resistance at its input and output. The first stage is very often an RF amplifier, while the second is a mixer. The problem is to find their combined effect on receiver noise

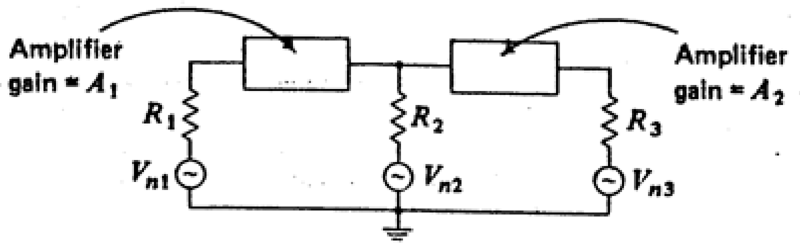


Figure (4.2): Noise of several amplifying stages in cascade

[Refer to figure (4.2) in the text book. Page 161]

- It may appear logical to combine all the noise resistances at the input, calculate their noise voltage, multiply it by the gain of the first stage and add this voltage to the one generated at the input of the second stage. The process might then be continued, and the noise voltage at the output, due to all the intervening noise sources, would be found

- Admittedly, there is nothing wrong with such a procedure. The result is useless because the argument assumed that it is important to find the total output noise voltage, whereas the important thing is to find the equivalent input noise voltage. It is even better to go one step further and find an equivalent resistance for such an input voltage. i.e., the equivalent noise resistance for the whole receiver

- This is the resistance that will produce the same random noise at the output of the receiver as does the actual receiver, so that we have succeeded in replacing an actual receiver amplifier by an ideal noiseless one with equivalent noise resistance R_{eq} located across its input. This greatly simplifies subsequent calculations, gives a good figure for comparison with other receivers, and permits a quick calculation of the lowest input signal which this receiver may amplify without drowning it with noise

- Consider the two stages amplifier of Figure 4.2. The gain of the first stage is A_1, and that of the second is A_2. The first stage has a total input noise resistance R_1, the second R_2 and the output resistance is R_3. The rms noise voltage at the output due to R_3 is

$$V_{n3} = \sqrt{4KT\delta f R_3}$$

- The same noise voltage would be present at the output if there were no R_3 there. Instead R_3' was present at the input of stage 2, such that

$$V_{n3}' = \frac{V_{n3}}{A_2} = \frac{\sqrt{4KT\delta f R_3}}{A_2} = \sqrt{4KT\delta f R_3'}$$

- Where R_3' is the resistance which if placed at the input of the second stage would produce the same noise voltage at the output as does R_3. Therefore

$$R_3' = \frac{R_3}{A_2^2} \tag{4.6}$$

- Eq. (4.6) shows that when a noise resistance is "transferred" from the output of a stage to its input, it must be divided by the square of the voltage gain of the stage. Now the noise resistance actually present at the input of the second stage R_2, so that the equivalent noise resistance at the input of the second stage, due to the second stage and the output resistance is,

$$R'_{eq} = R_2 + R'_3 = R_2 + \frac{R_3}{A_2^2}$$

– Similarly, a resistor R_2 may be placed at the input of the first stage to replace R'_{eq}, both naturally producing the same noise voltage at the output. Using Eq.(4.6) and its conclusion, we have

$$R'_2 = \frac{R'_{eq}}{A_1^2} = \frac{R_2 + R_3/A_2^2}{A_1^2} = \frac{R_2}{A_1^2} + \frac{R_3}{A_1^2 A_2^2}$$

– The noise resistance actually present at the input of the first stage is R_1, so that the equivalent noise resistance of the whole cascaded amplifier, at the input of the first stage, will be

$$\begin{aligned} R_{eq} &= R_1 + R'_2 \\ &= R_1 + \frac{R_2}{A_1^2} + \frac{R_3}{A_1^2 A_2^2} \end{aligned} \quad (4.7)$$

– It is possible to extend Eq.(4.7) by introduction to apply to an n-stage cascaded amplifier, but this is not normally necessary. As example 4.3 will show, the noise resistance located at the input of the first stage is by far the greatest contributor to the total noise, and only in broadband, i.e., low gain amplifiers is it necessary to consider a resistor past the output of the second stage

Example 4.3

- The first stage of a two-stage amplifier has a voltage gain of 10, a 600 input resistor, a 1600Ω equivalent noise resistance and a 27-KΩ output resistor. For the second stage, these values are 25, 81 KΩ, 10 KΩ and 1 megaohm (1 MΩ), respectively. Calculate the equivalent input-noise resistance of this two-stage amplifier

Solution

$$\mathrm{R}_1 = 600 + 1600 = 2200\ \Omega$$

$$R_2 = \frac{27x81}{27 + 81} + 10 = 20.2 + 10 = 30.2\ K\Omega$$

$$R_3 = 1M\Omega(\text{as given})$$

$$R_{eq} = 2200 + \frac{30{,}200}{10^2} + \frac{1{,}000{,}000}{10^2 x 25^2} = 2200 + 302 + 16 = 2518\Omega$$

- Note that the 1-MΩ output resistor has the same noise effect as a 16-Ω resistor at the input

4.4.3 Noise in Reactive Circuits

- If a resistance is followed by a tuned circuit which is theoretically noiseless, then the presence of the tuned circuit does not affect the noise generated by the resistance at the resonant frequency. To either side of resonance the presence of the tuned circuit affects noise in just the same way as any other voltage, so that the tuned circuit limits the bandwidth of the noise source by not passing noise outside its own bandpass. The more interesting case is a tuned circuit which is not ideal, i.e., one in which the inductance has a resistive component, which naturally generates noise

- In the preceding sections dealing with noise calculations, an input (noise) resistance has been used. It must be stressed here that this need not necessarily be an actual resistor. If all the resistors shown in Figure 4.2 had been tuned circuits with equivalent parallel resistances equal to R_1, R_2, and

R_3, respectively, the results obtained would have been identical. Consider Figure 4.3, which shows a parallel-tuned circuit

- The series resistance of the coil, which is the noise source here, is shown as a resistor in series with a noise generator and with the coil. It is required to determine the noise voltage across the capacitor, i.e., at the input to the amplifier. This will allow us to calculate the resistance which may be said to be generating the noise

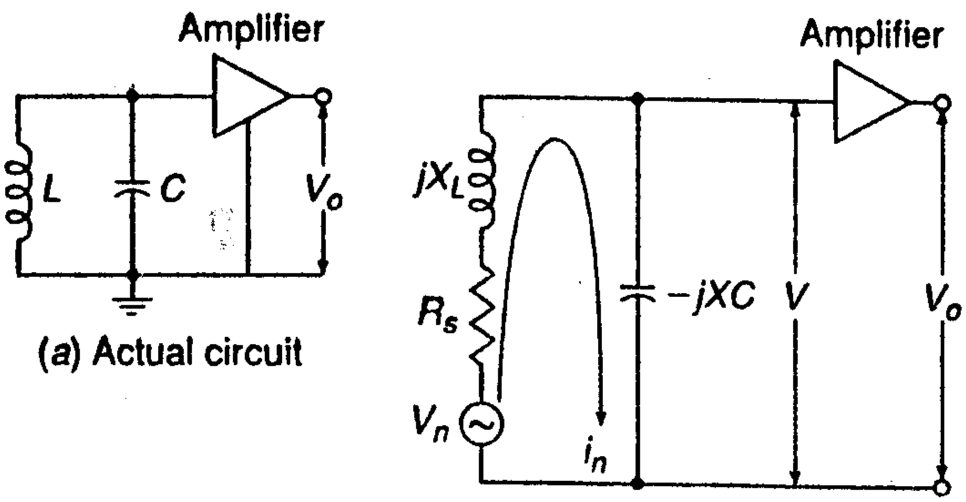


Figure (4.3): Noise in a tuned circuit

[Refer to figure (4.3) in the text book. Page 164]

- The noise current in the circuit will be

$$i_n = \frac{v_n}{Z}$$

- Where $Z = R_s + j(X_L - X_c)$. Thus $i_n = v_n/R_n$ at resonance

- The magnitude of the voltage appearing across the capacitor, due to v_n, will be

$$v = i_n X_c = \frac{v_n X_c}{R_s} = \frac{v_n Q R_s}{R_s} = Q v_n \tag{4.8}$$

- Since Xc = QRs at resonance, Eq. (4.8) should serve as a further reminder that Q is called the magnification factor. Continuing, we have

$$v^2 = Q^2 v_n^2 = Q^2 4KT\delta f R_s = 4KT\delta f(Q^2 R_s) = 4KT\delta f R_p$$

$$v = \sqrt{4KT\delta f R_p} \tag{4.9}$$

- Where v is the noise voltage across a tuned circuit due to its internal resistance, and R_p is the equivalent parallel impedance of the tuned circuit at resonance

- Eq. (4.9) shows that the equivalent parallel impedance of a tuned circuit is its equivalent resistance for noise (as well as for other purposes)

4.5 Noise Figure

4.5.1 Signal-to- Noise Ratio

- The calculation of the equivalent noise resistance of an amplifier, receiver or device may have one of two purpose or sometimes both. The first purpose is comparison of two kinds of equipment in evaluating their performance

- The second is comparison of noise and signal at the same point to ensure that the noise is not excessive. In the second instance, and also when equivalent noise resistance is difficult to obtain, the signal-to-noise ratio S/N is very often used. It is defined as the ratio of signal power to noise power at the same point. Therefore

$$\frac{S}{N} = \frac{X_s}{X_n} = \frac{V_s^2/R}{V_n^2/R} = (\frac{V_s}{V_n})^2 \quad (4.10)$$

- Where: S = Signal power ,, N = Noise power

- Eq. (4.10) is a simplification that applies whenever the resistance across which the noise is developed is the same as the resistance across which signal is developed, and this is almost invariable. An effort is naturally made to keep the signal-to-noise ratio as high as practicable under a given set of conditions

4.5.2 Definition of Noise Figure

- For comparison of receivers or amplifiers working at different impedance levels the use of the equivalent noise resistance is misleading. For example, it is hard to determine at a glance whether a receiver with an input impedance of 50Ω and R_{eq} = 90Ω is better, from the point of view of noise, than another receiver whose input impedance is 300Ω and R_{eq} = 400Ω

- As a matter of fact, the second receiver is the better one, as will be seen. Instead of equivalent noise resistance, a quantity known as noise figure, sometimes known as noise factor, is defined and used. The noise figure F is defined as the ratio of the signal-to noise power supplied to the input terminals of a receiver or amplifier to the signal-to-noise power supplied to the output or load resistor. Thus

$$F = \frac{input\ S/N}{output\ S/N} \quad (4.11)$$

- It can be seen immediately that a practical receiver will generate some noise, and the S/N will deteriorate as one moves toward the output. Consequently, in a practical receiver, the output S/N will be lower than the input value, and so the noise figure will exceed 1 , However, the noise figure will be 1 for an ideal receiver, which introduces no noise of its own

- Hence, we have the alternative definition of noise figure, which states that F is equal to the S/N of and ideal system divided by the S/N at the output of the receiver or amplifier under test, both working at the same temperature over the same bandwidth and fed from the same source
- In addition, both must be linear. The noise figure may be expressed as an actual ratio or in decibels. The noise figure of practical receivers can be kept to below a couple of decibels up to frequencies in the lower gigahertz range by a suitable choice of the first transistor, combined with proper circuit design and low-noise resistors

- At frequencies higher than that, equally low-noise figures may be achieved (lower, in fact) by devices which use the transit-tome or are relatively independent of it. This will be shown latter using Tunnel effect (Tunnel diode amplifiers), molecular amplifiers (Maser amplifiers), parametric devices, and MESFET transistor amplifiers

4.5.3 Calculations of Noise Figure

- Noise figure may be calculated for an amplifier or receiver in the same way by treating either as a whole. Each is treated as a four-terminal network having an input impedance R_b, an output impedance R_L, and an overall voltage gain A. It is fed from a source (antenna) of internal impedance Ra, which may or may not be equal to R_b. A block diagram of such a four-terminal network (with the source feeding it) is shown in figure 4.4

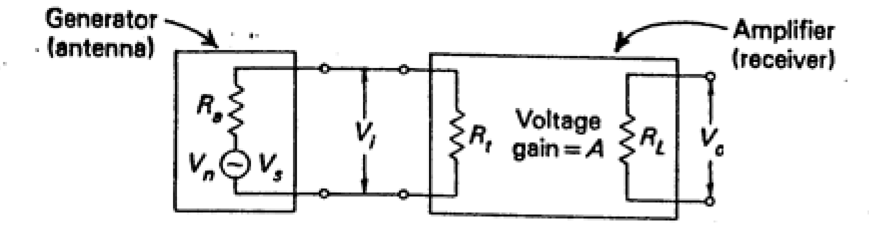


Figure (4.4): Block diagram for noise figure calculation

[Refer to figure (4.4) in the text book. Page 166]

- The calculation procedure may be broken down into a number of general steps. Each is now shown, followed by the number of the corresponding equation(s) to follow:

1. Determine the signal input power P_{si}(4.12,4.13)
2. Determine the noise input power P_{ni}(4.14,4.15)
3. Calculate the input signal-to-noise ratio S/N_i from the ratio of P_{si} and P_{ni}(4.16)
4. Determine the signal output power P_{so}(4.17)
5. Write P_{no} for the noise output power to be determine later(4.18)
6. Calculate the output signal-to-noise ratio S/No from the ratio of P_{so} and P_{no}(4.19)
7. Calculate the generalized form of noise figure from steps 3 and 6 (4.20)
8. Calculate P_{no} from R_{eq} if possible (4.21,4.22), and substitute into the general equation for F to obtain the actual formula (4.23,4.24), or determine P_{no} from measurement (4.3,4.25,4.26), and substitute to obtain the formula for F (4.27,4.28,4.29)

- It is seen from figure 4.4 that the signal input voltage and power will be

$$v_{si} = \frac{V_s R_t}{R_a+R_b} \tag{4.12}$$

$$P_{si} = \frac{V_{si}^2}{R_t} = (\frac{V_s R_t}{R_a+R_t})^2 \frac{1}{R_t} = \frac{V_s^2 R_t}{(R_a+R_t)^2} \tag{4.13}$$

– Similarly, the noise input voltage and power will be

$$V_{ni}^2 = 4KT\delta f \frac{R_a R_t}{R_a+R_t} \tag{4.14}$$

$$P_{ni} = \frac{V_{ni}^2}{R_t} = 4KT\delta f \frac{R_a R_t}{R_a+R_t} \frac{1}{R_t} = \frac{4KT\delta f R_a}{R_a+R_t} \tag{4.15}$$

– The input signal-to-noise ratio will be

$$\frac{S}{N_i} = \frac{P_{si}}{P_{ni}} = \frac{V_s^2 R_t}{(R_a+R_t)^2} \div \frac{4KT\delta f R_a}{R_a+R_t} = \frac{V_s^2 R_t}{4KT\delta f R_a(R_a+R_t)} \tag{4.16}$$

– The output power will be

$$P_{so} = \frac{V_{so}^2}{R_L} = \frac{(AV_{si})^2}{R_L}$$

$$= (\frac{AV_s R_t}{R_a+R_t})^2 \frac{1}{R_L} = \frac{A^2 V_s^2 R_t^2}{(R_a+R_t)^2 R_L} \tag{4.17}$$

– The noise output power may be difficult to calculate. For the time being, it may simply be written as

$$P_{no} = \text{noise output power} \tag{4.18}$$

– The output signal-to-noise ratio will be

$$\frac{S}{N_o} = \frac{P_{so}}{P_{no}} = \frac{A^2 V_s^2 R_t^2}{(R_a+R_t)^2 R_L P_{no}} \tag{4.19}$$

– Finally, the general expression for the noise figure is

$$F = \frac{S/N_i}{S/N_o} = \frac{V_s^2 R_t}{4KT\delta f R_a(R_a + R_t)} \div \frac{A^2 V_s^2 R_t^2}{(R_a + R_t)^2 R_L P_{no}}$$

$$= \frac{R_L P_{no}(R_a + R_t)}{4KT\delta f A^2 R_a R_t} \tag{4.20}$$

– Note that Eq. (4.20) is an intermediate result only. An actual formula for F may now be obtained by substitution for the output noise power, or from a knowledge of the equivalent noise resistance, or from measurement

4.5.4 Noise Figure from Equivalent Noise Resistance

– As derived in Eq. (4.7), the equivalent noise resistance of an amplifier or receiver is the sum of the input terminating resistance and the equivalent noise resistance of the first stage, together with the noise resistances of the following stages referred to the input. Putting it in another way, we see that all these resistances are added to R_t, giving a lumped resistance which is then said to concentrate all the "noise making" of the receiver

– The rest of it is now assumed to be noiseless. All this applies here, with the minor exception that these noise resistances must now be added to the parallel combination of R_a and R_t. In order to correlate noise figure and equivalent noise resistance, it is convenient to define R'_{eq}, which is a noise resistance that does not incorporate R_t and which is given by

$$R'_{eq} = R_{eq} - R_t$$

– The total equivalent noise resistance for this receiver will now be

$$R = R'_{eq} + \frac{R_a R_t}{R_a + R_t} \tag{4.21}$$

- The equivalent noise voltage generated at the input of the receiver will be

$$v_{ni} = \sqrt{4KT\delta fR}$$

- Since the amplifier has an overall voltage gain A and may now be treated as though it were noiseless, the noise output will be

$$P_{no} = \frac{V_{no}^2}{R_L} = \frac{(AV_{ni})^2}{R_L} = \frac{A^2 4KT\delta fR}{R_L} \tag{4.22}$$

- When Eq. (4.22) is substituted into the general Eq. (4.20), the result is an expression for the noise figure in terms of the equivalent noise resistance, namely,

$$F = \frac{R_L(R_a + R_t)}{4KT\delta fA^2R_aR_t}P_{no} = \frac{R_L(R_a + R_t)}{4KT\delta fA^2R_aR_t}\frac{A^2 4KT\delta fR}{R_L}$$

$$= R\frac{(R_a + R_t)}{R_aR_t} = (R'_{eq} + \frac{R_aR_t}{(R_a + R_t)})\frac{R_a + R_t}{R_aR_t}$$

(4.23) $= 1 + \frac{R'_{eq}(R_a+R_t)}{R_aR_t}$

- It can be seen from Eq. (4.23) that if the noise is to be a minimum for any given value of the antenna resistance R_a, the ratio $(R_a + R_t)/R_t$ must also be minimum, so that R_t must be much larger than Ra. This is a situation exploited very often in practice, and it may now be applied to eq. (4.23). under these mismatched conditions, $(R_a + R_t)/R_t$ approaches unity, and the formula for the noise figure reduces to

$$F = 1 + \frac{R'_{eq}}{R_a} \tag{4.24}$$

– This is a most important relationship, but it must be remembered that it applies under mismatched conditions only. Under matched conditions ($R_t = R_a$) or when the mismatch is not serve, Eq. (4.23) must be used instead

Example 4.4

- Calculate the noise figure of the amplifier of Example 4.3 if it is driven by a generator whose output impedance is 50Ω. Note that this constitutes a large enough mismatch temperature is 17 C°

Solution

$$R'_{eq} = R_{eq} - R_t = 2518 - 600 = 1918\Omega$$

$$F = 1 + \frac{R'_{eq}}{R_a} = 1 + 38.4 = 39.4\ (= 15.84\ dB)$$

– Note that if an "equivalent noise resistance" is given without any other comment in connection with noise figure calculations, it may be assumed to be R_{eq}

4.5.5 Noise Figure from Measurement

– The preceding section showed how the noise figure may be computed if the equivalent noise resistance is easy to calculate. When this is not practicable, as under transit-time conditions, it is possible to make measurements that lead to the determination of the noise figure. A simple method, using the diode noise generator is often employed. It is shown in figure 4.5 in circuit block form

– Eq. (4.3) gave the formula of the exact plate noise current of a vacuum-tube diode and this can be used. As shown, the anode current is controlled by means of the potentiometer which varies filament voltage, and that is how shot noise current is adjusted

– The output capacitance of the diode and its associated circuits is resonated at the operating frequency of the receiver by means of the variable inductance, so that it may be ignored. The output impedance of the noise generator will now simply be R_a. The noise voltage supplied to the input of the receiver by the diode will be given by

$$V_n = i_n Z_n = i_n \frac{R_a R_t}{R_a + R_t} = \frac{R_a R_t \sqrt{2 V i_p \delta f}}{R_a + R_t} \tag{4.23}$$

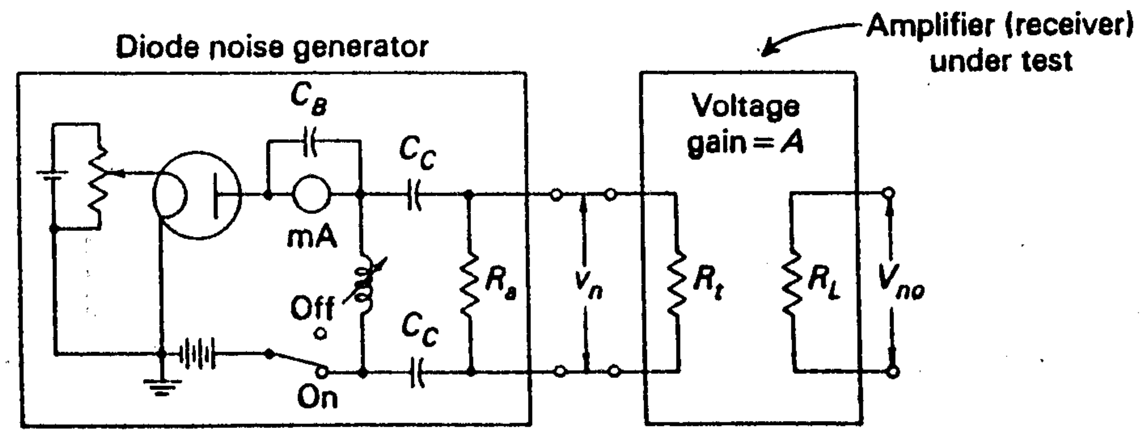


Figure (4.5): Noise Figure measurement

[Refer to figure (4.5) in the text book. Page 171]

– The noise generator is connected to the receiver (or amplifier) under test, and the noise output power of the receiver is measured with zero diode plate current, i.e., with the diode plate voltage supply switched off

– The diode plate voltage supply is now switched on, and the filament potentiometer is adjusted so that diode plate current begins to flow. It is further adjusted until the noise power devolved in R_L is twice as large as the noise power in the absence of diode plate current

- The plate current at which this happens, i_p is measured with the millimeter and noted. The additional noise power output is now equal to the normal noise power output, so that the latter can be expressed in terms of the diode plate current. We now have

$$P_{no} = \frac{V_{no}^2}{R_L} = \frac{(AV_n)^2}{R_L} = \frac{A^2 R_a^2 R_t^2 2V i_p \delta f}{R_L (R_a + R_t)^2} \quad (4.26)$$

- As already outlined, Eq. (4.26) may be substituted into Eq. (4.20). This yields

$$F = \frac{R_L(R_a + R_t)}{4KT\delta f A^2 R_a R_t} P_{no} = \frac{R_L(R_a + R_t)}{A^2 4KT\delta f R_a R_t} \frac{A^2 R_a^2 R_t^2 2V i_p \delta f}{R_L (R_a + R_t)^2}$$

$$= \frac{V i_p R_a R_t}{2KT(R_a + R_t)} \quad (4.27)$$

- If it is assumed once again that the system is mismatched and $R_t >> R_a$, Eq. (4.27) is simplified to

$$F = \frac{R_a V i_p}{2KT} \quad (4.28)$$

- If the above procedure is repeated right from the beginning for a system under matched conditions, it may then be proved that Eq. (4.28) applies exactly to such a system instead of being merely a good approximation as it is here. Such a result emphasizes the value of the noise diode measurement
- As a final simplification, we substitute into Eq. (4.28) the value of the various constants it contains. These include the standard temperature at

which such measurements are made, which is 17°C or 290 K. this gives a formula which is very often quoted:

$$F = \frac{R_a V i_p}{2KT} = \frac{(R_a i_p)(1.6x10^{-19})}{2x290x1.38x10^{-23}}$$
$$= (R_a i_p)(2x10)$$
$$=20 R_a V i_p \qquad (4.29)$$

– Where R_a is measured in ohms and i_p in amperes

4.6 Noise Temperature

– The concept of noise figure, although frequently used, is not always the most convenient measure of noise, practically in dealing with UHF and microwave low-noise antennas, receivers and devices. There exists another measure called effective noise temperature or equivalent noise temperature derived from early work in radio astronomy, is employed extensively for antennas and low-noise microwave amplifiers

– Not the least reason for its use is convenience, in that it is an additive like noise power. This may be seen from reexamining Eq. (4.1), as follows:

$$P_t = KT\delta f$$
$$= P_1 + P_2 = KT_1\delta f + KT_2\delta f$$
$$KT_t\delta f = KT_1\delta f + KT_2\delta f$$
$$T_t = T_1 + T_2 \qquad (4.30)$$

– Where

- P_1 and P_2 = two individual noise powers (e.g., received by the antenna and generated by the antenna, respectively)

- P_t is their sum
- T_1 and T_2 = the individual noise temperatures
- T_t = the "total" noise temperature

– Another advantage of the use of noise temperature for low noise levels is that it shows a greater variations for any given noise-level change than does the noise figure, so changes are easier to track for comparison or different systems

– It will be recalled that the equivalent noise resistance introduced in section 4.4 is quite fictitious, but it is often employed because of its convenience. Similarly, T_{eq}, the equivalent noise temperature, may also be utilized if it proves convenient

– In defining the equivalent noise temperature of a receiver or amplifier, it is assumed that $R_{eq} = R_a$. If this is to lead to the correct value of noise output power, then obviously R_{eq} must be at temperature other than the standard one at which all the components (including R_a) are assumed to be. It is then possible to use eq. (4.24) to equate noise figure and equivalent noise temperature, as follows

$$F = 1 + \frac{R'_{eq}}{R_a} = 1 + \frac{KT_{eq}\delta f R'_{eq}}{KT_o \delta f R_a}$$

$$= 1 + \frac{T_{eq}}{T_o} \qquad (4.31)$$

– Where $R'_{eq} = R_a$ as postulated in the definition of T_{eq}

– $T_o = 17^oC = 290$ K

– T_{eq} = equivalent noise temperature of the amplifier or receiver whose noise figure is F

– There is another approach to reach the same result as follows

$$F = \frac{S_{in}/N_{in}}{S_{out}/N_{out}} \qquad \text{as in Eq. (4.11)}$$

– Or equivalently

$$F = \frac{N_{out}}{KTB_nG} \tag{4.32}$$

– Where:

- S_{in} = available input signal power
- N_{in} = available input noise power (equal to KTB_n)
- S_{out} = available output signal power (equal to Sin.G)
- G = the power gain (A^2)
- N_{out} = available output noise power
- B_n = Noise bandwidth

– Available power refers to the power which would be delivered to matched load

– The noise figure may also be written as

$$F = \frac{KTB_nG + \Delta N}{KTB_nG} = 1 + \frac{\Delta N}{KTB_nG} \tag{4.33}$$

– ΔN is the additional noise introduced by the network itself the temperature T is taken standard as $T_o = 290^oK$

– Then using the concept of effective noise temperature defined as that temperature T_e at the input of the network which would account for the noise ΔN at the output therefore

$$\Delta N = KT_e B_n G$$

$$F = 1 + \frac{T_e}{T_o} = 1 + \frac{T_{eq}}{T_o} \tag{4.34}$$

– Note that F here is a ratio and is not expressed in decibels. Also T_{eq} may be influenced by (but is certainly not equal to) the actual ambient temperature of the receiver or amplifier. It must be repeated that the equivalent noise temperature is just a convenient fiction. If all the noise of the receiver were generated by Ra, its temperature would have to be T_{eq}. Finally we have, from Eq. (4.34)

$$T_o F = T_o + T_{eq}$$

$$T_{eq} = T_o(F - 1) = T_e \tag{4.35}$$

– Once noise figure is known, equivalent noise temperature may be calculated from Eq. (4.35), or a monograph may be constructed if use is frequent enough to justify it. Graphs of noise temperature verses frequency are also available. Also we can calculate the output noise power of amplifier or receiver, given us noise figure and gain from Eq. (4.32)

$$N_{out} = FKTBG$$

Example 4.5

- A receiver connected to an antenna whose resistance is 50Ω has an equivalent noise resistance of 30Ω. Calculate the receiver's noise figure in decibels and its equivalent noise temperature

Solution

$$F = 1 + \frac{R'_{eq}}{R_a} = 1 + \frac{30}{50} = 1 + 0.6 = 1.6$$

$$= 10 \log 1.6 = 10 \text{ x } 0.204 = 2.04 \text{ dB}$$

$$T_{eq} = T_o(F - 1) = 290(1.6 - 1) = 290x0.6 = 174K$$

→ Effective noise temperature and noise figure of cascade of networks

$$T_e = T_1 + \frac{T_2}{G_1} + \frac{T_3}{G_1 G_2} + \cdots \ldots \ldots \ldots + \frac{T_N}{G_1 G_2 \ldots\ldots G_{N-1}} \qquad (4.36)$$

– If we use Eq. (4.35) into Eq. (4.36), we get the formula for noise figure of cascade of networks

$$F_o = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \cdots \ldots \ldots \ldots + \frac{F_{N-1}}{G_1 G_2 \ldots\ldots G_{N-1}} \qquad (4.37)$$

– Both noise figure and effective noise temperature are used in practice. The effective noise temperature is preferred for describing low-noise devices, and the noise figure is preferred for conventional receivers

Example 4.6

- The signal input to an RF amplifier is 40 microwatts and the noise power at amplifier input is 2 microwatts. The signal-to-noise ratio at the output is measured to be 14. Assume T = 290 K. Compute the noise factor of this amplifier

Solution

$$\frac{\mathrm{S_i}}{\mathrm{N_i}} = \frac{40\mathrm{x}10^{-6}}{2\mathrm{x}10^{-6}} = 20$$

$$F = \frac{S_i/N_i}{S_o/N_o} = \frac{20}{14} = 1.43$$

Example 4.7

- A microwave amplifier has a gain of 20 dB, a bandwidth of 30 MHz, and a noise factor of 2.5. Calculate the noise power at the output of the amplifier if it is operated at 30°C

Solution

$$\mathrm{F} = \frac{\mathrm{N_o}}{\mathrm{KTBG}}$$

$$\mathrm{N_o} = \mathrm{FKTBG}$$

$$\mathrm{dB} = 10\log\frac{\mathrm{P_o}}{\mathrm{P_i}} = 10\log\mathrm{G}$$

$$\mathrm{G} = \log^{-1}\frac{\mathrm{dB}}{10}$$

$$\mathrm{G} = \log^{-1}2 = 100$$

$$\mathrm{N_o} = (2.5)(1.38\mathrm{x}10^{-23})(303)(3\mathrm{x}10^{7})(100)$$

$$\mathrm{N_o} = 3.14\mathrm{x}10^{-11}\mathrm{watts}$$

Example 4.8

- A 8-18 GHz microwave amplifier's specification sheet indicates that its effective noise temperature is 150 K when it is operated at 17°C. Calculate the amplifier's noise figure

Solution

$$T_e = (F-1)T_o \qquad ,, \quad (T_o = 290K)$$

$$F = \frac{T_e}{T_o} + 1$$

$$F = \frac{150}{290} + 1 = 1.52$$

NF = 10 log F

NF = 10 log (1.52)

NF = 1.82 dB

Example 4.9

- A 2-4 GHz amplifier with a noise figure of 2 dB and a gain of 20 dB is cascaded with another 2-4 GHz amplifier with a noise figure of 6 dB and a gain of 30 dB. Calculate the overall gain and noise figure for the cascaded system

Solution

NF = 10 log F

$$F = log^{-1}\frac{NF}{10}$$

$$F_1 = log^{-1}(0.2) = 1.58$$

$$F_2 = log^{-1}(0.6) = 3.98$$

$$G_{dB} = 10\log G$$

$$G = log^{-1}(\frac{G_{dB}}{10})$$

$$G_1 = log^{-1}(2) = 100$$

$$G_2 = log^{-1}(3) = 1000$$

$$F_{12} = F_1 + \frac{F_2 - 1}{G_1}$$

$$F_{12} = 1.58 + \frac{3.98 - 1}{100} = 1.58 + 0.03 = 1.61$$

$$NF_{12} = 10\log F_{12} = 10\log(1.61) = 2.07\ dB$$

$$G_{12} = G_1 G_2 = (100)(1000) = 10^5 = 50\ dB$$

Or

$$G_{12}(dB) = G_1(dB) + G_2(dB) = 20(dB) + 30(dB) = 50\ dB$$

– Note that the rather noisy second stage has little effect upon the overall figure because of the high-gain G_1 of the relatively low-noise preamplifier

Example 4.10

- Two identical 4-8 GHz amplifiers with a gain of 20 dB each are cascaded an overall effective noise temperature of the two-amplifier system is measured 150 K. Calculate the noise figure of an individual amplifier

Solution

$$T_{in} = T_1 + \frac{T_2}{G_1} \quad ,, \quad G_1 = 20dB = 100 \quad ,, T_1 = T_2 = T$$

$$150 = T + \frac{T}{100}$$

$$150 = 1.01\ T$$

$$T = 148.5\ K = T_1 = T_2$$

$$T_1 = (F_1 - 1)T_o$$

$$F_1 = \frac{T_1}{T_o} + 1$$

$$F_1 = \frac{148.5}{290} + 1 = 1.51$$

$$NF_1 = 10 \log F_1$$

$$NF_1 = 10 \log (1.51)$$

$$NF_1 = 1.79\ dB$$

Ch.[5] Analog & Digital Communication Systems Behavior In Presence of Noise

5.1 Introduction

- In this chapter, we analyze the behavior of analog and digital communication systems in the presence of noise to facilitate the comparison of various systems

- Figure 5.1 shows a schematic of a communication system. A certain signal power ST is transmitted over a channel. The transmitted signal is corrupted by channel noise during transmission. We shall assume channel noise to be additive. The channel will attenuate (and may also distort) the signal

- At the receiver input, we have a signal mixed with noise. The signal and noise powers at the receiver input Si and Ni, respectively. The receiver processes (filters, demodulates, etc.) the signal to yield the desired signal plus noise. The signal and noise powers at the receiver output are So and No, respectively

- In analog systems, the quality of the received signal is determined by So/No, the output SNR. Hence, we shall focus our attention on this parameters. But So/No can be increased as much as desired simply by increasing the transmitted power ST. In practice, however, the maximum value of ST is limited by other considerations, such as transmitter cost, channel capability, interference with other channels, and so on.
- Hence, the value of So/No for a given transmitted power is an appropriate figure of merit in an analog communication system. In

practice, it is more convenient to deal with the received power Si rather than the transmitted power ST. From Figure 5.1, it is apparent that Si is proportional to ST. hence, the value of So/No for a given Si will serve equally well as a figure of merit

- Hence, we can say that in analog systems, the chief objective is the fidelity of reproduction of waveforms and hence, the suitable performance criterion is the output signal-to-noise ratio. The choice of this criterion stems from the fact that the signal-to-noise ratio is related to the ability of the listener to interpret a message

- In digital communication systems, the transmitter input is chosen from a finite set of possible symbols, or messages. The objective at the receiver is not to reproduce the waveform with fidelity, because the possible waveforms are already known exactly

- Our goal is to decide from the noisy received signal, which of the waveforms has been transmitted. Logically, the appropriate figure of merit in a digital communication system is the probability of error in making such a decision at the receiver

5.2 Baseband Systems

- In baseband systems, the signal is transmitted directly without any modulation. This mode of communication is suitable over a pair of wires, optical fiber, or coaxial cables. It is mainly used in short-haul links
- The study of baseband system is important because many of the basic concepts and parameters encountered in baseband systems are carried over directly to modulated systems. Second, baseband systems serve as

a basis against which other systems may be compared

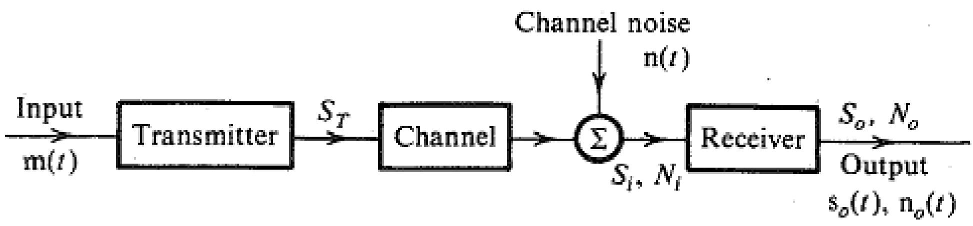


Figure (5.1): Communication System Model

[Refer to figure (5.1) in the text book. Page 184]

- For a baseband system, the transmitter and the receiver are ideal baseband filters (Figure 5.2). The low-pass filter $H_p(\omega)$ at the transmitter limits the input signal spectrum to a given bandwidth. The low-pass filter $H_d(\omega)$ at the receiver eliminates the out-of-band noise and other channel interference

- These filters can also serve an additional purpose, that of pre-emphasis and de-emphasis, which optimizes the signal-to-noise ratio (SNR) at the receiver (or minimize the channel noise interference)

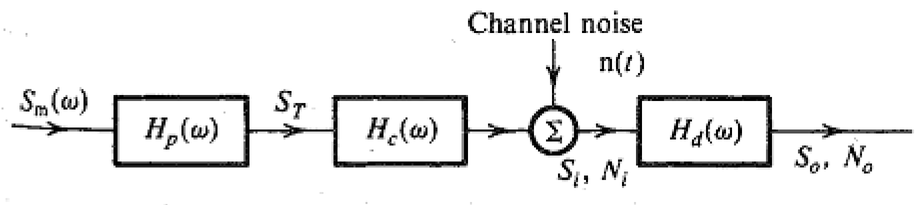


Figure (5.2): Baseband System

[Refer to figure (5.2) in the text book. Page 184]

- The baseband signal m(t) is assumed to be a zero mean, wide-sense stationary random process band-limited to B Hz. To begin with, we shall consider the case of ideal low-pass (or baseband) filters with bandwidth B at the transmitter and the receiver (Figure 5.2). The channel is assumed to be distortionless. For this case,

$$S_o = S_i \tag{5.1a}$$

- And

$$N_o = 2\int_0^B S_n(\omega)df \tag{5.1b}$$

- Where $S_n(\omega)$ is the PSD of the channel noise. For the case of a white noise, $S_n(\omega) = N/2$,

$$N_o = 2\int_0^B \frac{N}{2}df = NB \tag{5.1c}$$

- And

$$\frac{S_o}{N_o} = \frac{S_i}{NB} \tag{5.1d}$$

- We define a parameter γ as

$$\gamma = \frac{S_i}{NB} \tag{5.2}$$

- From Eqs. (5.1d) and (5.2), we have

$$\frac{S_o}{N_o} = \gamma \tag{5.3}$$

- The parameter γ is directly proportional to Si and therefore, directly proportional to ST. Hence, a given ST (or Si) implies a given γ. Equation

(5.3) is precisely the result we are looking for. It gives the receiver output SNR for a given ST (or Si). The value of the SNR in Eq. (5.3) will serve as a standard against which the output SNR of other systems will be measured

- The power, or the mean square value, of m(t) is,

$$\overline{m^2} = 2\int_0^B S_m(\omega)df \tag{5.4}$$

- In analog signals, the SNR is basic in specifying the signal quality. For voice signals, an SNR of 5 to 10 dB at the receiver implies a barely intelligible signal. Telephone quality signals have an SNR of 25 to 35 dB, whereas for television, an SNR of 45 to 55 is required

5.3 Amplitude-Modulated Systems

– We shall analyze DSB, SSB-SC, and AM systems separately

5.3.1 DSB-SC

- A basic DSB-SC system is shown in Figure 5.3. The modulated signal is a bandpass signal centered at ω_c with a bandwidth 2B. The channel noise is assumed to be additive. The channel and the filters in Figure 5.3 are assumed to be ideal

- Let S_i and S_o represent the useful signal powers at the input and the output of the demodulator, and let No represent the noise power at the demodulator output. The signal at the demodulator input is $\sqrt{2}m(t)cos\omega_c + n(t)$, where n(t) is the bandpass channel noise. Its spectrum is centered at ω_c and has a bandwidth 2B. The input signal power S_i is the power of the modulated signal $\sqrt{2}m(t)cos\omega_c$. If we take the average,

$$S_i = [\sqrt{2}m(t).cos\omega_c t]^2 = (\sqrt{2})^2[m(t).cos\omega_c t]^2 = m^2(t) = m^2 \quad (5.5)$$

- The reader may now appreciate our use of $\sqrt{2}cos\omega_c$ (rather than $cos\omega_c$) in the modulator Figure 5.3. This was done to make the received power equal to that in the baseband system in order to facilitate comparison. We shall use a similar approach in our analysis of the SSB system

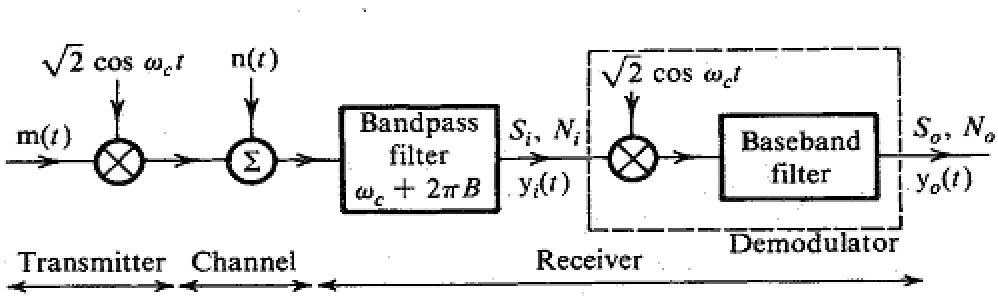


Figure (5.3): Baseband System

[Refer to figure (5.3) in the text book. Page 186]

- To determine the output powers S_o and N_o, we note that the signal at the demodulator input is

$$y_i(t) = \sqrt{2}m(t).cos\omega_c t + n(t)$$

- Because n(t) is a bandpass signal centered at ω_c we can express it in terms of quadrature components. This gives

$$y_i(t) = [\sqrt{2}m(t) + n_c(t)]cos\omega_c t + n_s(t)sin\omega_c t$$

- When this signal is multiplied by $\sqrt{2}cos\omega_c t$ (synchronous demodulation) and then low-pass filtered, the terms $m(t)\cos 2\omega_c t$ and $m(t)\sin 2\omega_c t$ are suppressed. the resulting demodulator output $y_o(t)$ is

$$y_o(t) = m(t) + \frac{1}{\sqrt{2}}n_c(t)$$

- Hence,

$$S_o = \overline{m^2} = S_i \quad (5.6a)$$

$$N_o = \frac{1}{2}\overline{n_c^2(t)} \quad (5.6b)$$

- For white noise with power density N/2, we have

$$\overline{n_c^2(t)} = \overline{n^2(t)} = 2NB \quad (5.7)$$

- And

- Hence, from Eqs. (5.6a) and (5.7), we have

$$\frac{S_o}{N_o} = \frac{S_i}{NB} = \gamma$$

- Comparison of Eqs. (5.8) and (5.3) shows that for a fixed transmitted power (which also implies a fixed signal power at the demodulator input), the SNR at the demodulator output is the same for the baseband

and the DSB-SC systems. Moreover, quadrature multiplexing in DSB-SC can render its bandwidth requirement identical to that of baseband systems. Thus, theoretically, baseband and DSB-SC systems have identical capabilities

5.3.2 SSB-SC

- An SSB-SC system is shown in figure 5.4. The SSB signal $\varphi_{SSB}(t)$ can be expressed as

$$\varphi_{SSB}(t) = m(t)cos\omega_c t + m_h(t)sin\omega_c t \tag{5.9}$$

- The well-known spectrum of $\varphi_{SSB}(t)$ studied previously can be obtained (figure 5.4) by multiplying m(t) by 2 cos $\omega_c t$ and then suppressing the unwanted sideband. The power of the modulated signal 2m(t) cos $\omega_c t$ is $\overline{2m^2}$ [four times the power of m(t) cos $\omega_c t$]. Suppression of one sideband halves the power, Hence S_i, the power of $\varphi_{SSB}(t)$ is

$$S_i = \overline{m^2} \tag{5.10}$$

- Expressing the channel bandpass noise in terms of quadrature components the signal at the detector input, $y_i(t)$ is,

$$y_i(t) = [m(t) + n_c(t)].cos\omega_c t + [m_h(t) + n_s(t)].sin\omega_c t$$

- This signal is multiplied by 2cosct (synchronous demodulation) and then low-pass filtered to yield the demodulator output

$$y_o(t) = m(t) + n_c(t)$$

- Hence,

$$S_o = \overline{m^2} = S_i$$
$$N_o = \overline{n_c^2} \tag{5.11}$$

- We have already found $\overline{n_c^2}$ for the SSB channel noise (lower sideband) as

$$N_o = \overline{n_c^2} = NB$$

- Thus,

$$\frac{S_o}{N_o} = \frac{S_i}{NB} = \gamma \tag{5.12}$$

- This shows that baseband, DSB-SC and SSB-SC systems perform identically in terms of resource utilization. All of them yield the same output SNR for a given transmitted power and transmission bandwidth

Example 5.1

- In a DSB-SC system, the carrier frequency is f_c = 500 KHz, and the modulating signal m(t) has a uniform PSD band-limited to 4 KHz. The modulated signal is transmitted over a distortionless channel with a noise PSD $S_n(\omega) = 1/(\omega^2+a^2)$, where $a = 10^6\pi$. The useful signal power at the receiver input is 1 μW
- The received signal is bandpass filtered, multiplied by 2 cos $\omega_c t$ and then low-pass filtered to obtain the output $S_o(t) + n_o(t)$.
- Determine the output SNR

Solution

- If the received signal is Km(t) cos $\omega_c t$, the demodulator input is [Km(t)

+ $n_c(t)$] cos $\omega_c t$ + $n_s(t)$ sin $\omega_c t$. When this is multiplied by 2 cos $\omega_c t$ and low-pass filtered, the output is

$$S_o(t) + n_o(t) = Km(t) + n_c(t) \qquad (5.13)$$

- Hence,

$$S_o = K^2\overline{m^2} \qquad and \qquad N_o = \overline{n_c^2}$$

- But the power of the received signal Km(t) cos $\omega_c t$ is 1 μW. Hence,

$$\frac{K^2\overline{m^2}}{2} = 10^{-6}$$

- And

$$\mathrm{S_o} = \mathrm{K^2}\overline{\mathrm{m^2}} = 2\,\mathrm{x}10^{-6}$$

- To compute ,we use Eq. (5.11):

$$\overline{n_c^2} = \overline{n^2}$$

- Where is the power of the incoming bandpass noise of bandwidth 8 KHz centered at 500 KHz, that is,

$$\overline{n^2} = 2\int_{496,000}^{504,000} \frac{1}{\omega^2 + a^2} df \qquad ,, \qquad a = 10^6\pi$$

$$= \frac{1}{\pi}\int_{(2\pi)496,000}^{(2\pi)504,000} \frac{1}{\omega^2 + a^2} d\omega = \frac{1}{\pi a} tan^{-1}\frac{\omega}{a}$$

$$= 8.25\ x\ 10^{-10} = N_o$$

- Therefore,

$$\frac{S_o}{N_o} = \frac{2x10^{-6}}{8.25x10^{-10}} = 2.42x10^3 = 33.83 \text{ dB}$$

5.3.3 AM

- Am signal can be demodulated synchronously or by envelope detection. The former is of theoretical interest only. It is useful, however, for comparing the noise performance of the envelope detector. For this reason, we shall consider both of these methods

5.3.3.1 Coherent, or synchronous, Demodulation of AM

- Coherent AM is identical to DSB-SC in every respect except for the additional carrier. If the received signal $\sqrt{2}[A + m(t)]cos\omega_c t$ is multiplied by $\sqrt{2}cos\omega_c t$, the demodulator outputs is m(t). Hence,

$$S_o = \overline{m^2}$$

- The output noise will be exactly the same as that in DSB-SC [Eq. (5.7)]:

$$N_o = \overline{n_o^2} = NB$$

- The received signal is $\sqrt{2}[A + m(t)]cos\omega_c t$. Hence,

$$S_i = (\sqrt{2})^2 \frac{\overline{[A + m(t)]^2}}{2}$$
$$= \overline{[A + m(t)]^2}$$
$$= A^2 + \overline{m^2}(t) + 2A\overline{m(t)}$$

- Because m(t) is assumed to have zero mean,

$$S_i = A^2 + \overline{m^2(t)}$$

- And

$$\frac{S_o}{N_o} = \frac{\overline{m^2}}{NB}$$

$$= \frac{\overline{m^2}}{A^2 + \overline{m^2}} \frac{S_i}{NB}$$

$$= \frac{\overline{m^2}}{A^2+\overline{m^2}}\gamma \qquad (5.14)$$

- If $m(t)_{max} = m_p$, then $A \geq m_p$. For the maximum SNR, $A = m_p$, and

$$\left(\frac{S_o}{N_o}\right)_{max} = \frac{\overline{m^2}}{{m_p}^2 + \overline{m^2}}\gamma$$

$$= \frac{1}{(m_p^2/\overline{m^2}+1)}\gamma \qquad (5.15a)$$

- Because $(m_p^2/\overline{m^2}) \geq 1$,

$$\frac{S_o}{N_o} < \frac{\gamma}{2} \qquad (5.15b)$$

- It can be seen that the SNR in AM is at least 3 dB (and usually about 6 dB in practice) worse than that in DSB-SC and SSB-SC (depending on the modulation index and the signal waveform). For example, when m(t) is sinusoidal, $(m_p^2/\overline{m^2}) = 2$, and AM requires three times as much power (4.77 dB) as that needed for DSB-SC or SSB-SC

- In many communication systems the transmitter is limited by the peak power rather than the average power transmitted. In such a case, AM fares even worse. It can be shown that in tone modulation, for a fixed peak power transmitted, the output SNR of AM is 6 dB below that of DSB-SC and 9 dB below that of SSB-SC. These results are valid under conditions most favorable to AM, that is, with modulation index = 1. For <1, AM would be worse than this. For this reason, volume compression and peak limiting are generally used in AM transmission in order to have full modulation most of the time

5.3.3.2 AM Envelope detection

- Assuming the received signal to be $[A + m(t)] \cos \omega_c t$ the demodulator input is,

$$y_i(t) = [A + m(t)]cos\omega_c t + n(t)$$

- Using the quadrature component representation for n(t), we have,

$$y_i(t) = [A + m(t) + n_c(t)]cos\omega_c t + n_s(t)sin\omega_c t \quad (5.16a)$$

- The desired signal at the demodulator input is $[A + m(t)] \cos \omega_c t$, Hence, the signal power S_i is,

$$S_i = \frac{\overline{[A + m(t)]^2}}{2} = \frac{A^2 + \overline{m^2}}{2}$$

- To find the envelope of $y_i(t)$, we rewrite Eq. (5.16a) in polar form as,

$$y_i(t) = E_i(t)\cos[\omega_c t + \theta_i(t)] \quad (5.16b)$$

- Where the envelope $E_i(t)$ is,

$$E_i(t) = \sqrt{[A + m(t) + n_c(t)]^2 + n_s^2(t)} \qquad (5.16c)$$

- The envelope detector output is $E_i(t)$ [Eq. (5.16c)]. We shall consider two extreme cases : small noise and large noise

1. Small Noise Case: if [A+m(t)] >> n(t) for almost all t, then [A+m(t)] >> $n_c(t)$ and $n_s(t)$ for almost all t. in this case Ei(t) in Eq. (5.16c) can be approximated by [A+m(t)+$n_c(t)$],

$$E_i(t) \cong A + m(t) + n_c(t)$$

- The dc component A of the envelope detector output E_i is blocked by a capacitor, yielding m(t) as the useful signal and $n_c(t)$ as the noise. Hence,

$$S_o = \overline{m^2}$$

- And for noise

$$N_o = n_c^2(t) = 2NB$$

- And

$$\frac{S_o}{N_o} = \frac{\overline{m^2}}{2NB}$$

$$= \frac{\overline{m^2}}{A^2 + \overline{m^2}} \frac{S_i}{NB}$$

$$= \frac{\overline{m^2}}{A^2 + \overline{m^2}} \gamma$$

- Which is identical to the result for AM with synchronous demodulation [Eq. (5.14)]. Therefore for AM, when the noise is small compared to the signal, the performance of the envelope detector is identical to that of the synchronous detector

2. Large Noise Case: if n(t) >> [A+m(t)]. Hence, $n_c(t)$ and $n_s(t)$ >> [A+m(t)] for almost all t. Under this condition Eq. (5.16c) becomes,

$$E_i(t) \cong \sqrt{n_c^2(t) + n_s^2(t) + 2n_c(t)[A + m(t)]}$$
$$= E_n(t)\sqrt{1 + \frac{2[A + m(t)]}{E_n(t)} cos\theta_n(t)}$$

- Where $E_n(t)$ and $\theta_n(t)$, the enveloped and the phase of the noise n(t), are see in figure 5.5a

$$E_n(t) = \sqrt{n_c^2(t) + n_s^2(t)} \qquad (5.18a)$$

$$\theta_n(t) = tan^{-1}[\frac{n_s(t)}{n_c(t)}] \qquad (5.18b)$$

$$n_c(t) = E_n(t)cos\theta_n(t) \qquad (5.18c)$$

$$n_s(t) = E_n(t)sin\theta_n(t) \qquad (5.18d)$$

- Because $E_n(t)$ >> A + m(t), $E_i(t)$ may be further approximated as

$$E_i(t) \cong E_n(t)[1 + \frac{A + m(t)}{E_n(t)} cos\theta_n(t)$$
$$= E_n(t) + [A + m(t)]cos\theta_n(t) \qquad (5.19)$$

- A glance at Eq. (5.19) shows that the output contains no term proportional to m(t). the signal m(t) cos $\theta_n(t)$ represents m(t) multiplied by a time-varying function (actually a noise signal) cos $\theta_n(t)$ and, hence, is of no use in recovering m(t). in all previous cases, the output signal contained a term of the form am(t), where a was constant

- Furthermore, the output noise was additive (even for envelope detection with small noise). In Eq. (5.19), the noise is multiplicative. In this situation the useful signal quality at the output undergoes disproportionality rapid deterioration when the input noise increases beyond a certain level (i.e., when γ drops below a certain value)
- Calculation of the SNR for the intermediate case (transition region) is quite complex. Here we shall state the final results only:

$$\frac{S_o}{N_o} \cong 0.916 A^2 \overline{m^2 \gamma^2}$$

- Figure 5.5b shows the plot of So/No as a function of γ for AM with synchronous detection and AM with envelope detection. The threshold effect is Cleary seen from this figure. The threshold occurs when γ is on the order of 10 or less. For a reasonable-quality AM signal, γ should be on the order of 1000 (30 dB), and the threshold is rarely a limiting condition in practical cases

5.4 Angle-Modulated Systems

- A block diagram of an angle-modulated system is shown in figure 5.6. the angle-modulated (or exponentially modulated) carrier $\varphi_{EM}(t)$ can be written as,

$$\varphi_{EM}(t) = A\cos[\omega_c t + \psi(t) \qquad (5.120a)$$

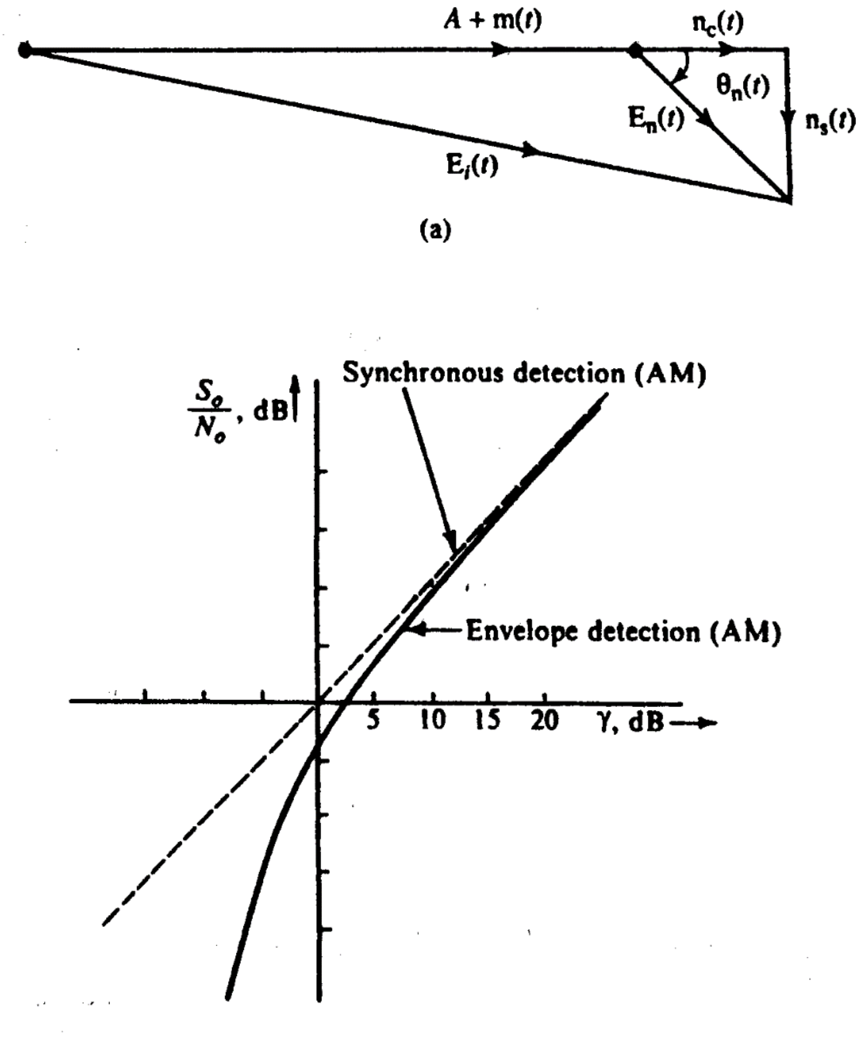
- Where,

$$\psi(t) = K_p m(t), \qquad for\ PM \qquad (5.20b)$$

$$= K_f \int_{-\infty}^{t} m(\alpha)d\alpha, \quad for\ FM \qquad (5.20c)$$

- And m(t) is the message signal. The channel noise n(t) at the demodulator input is a bandpass noise with PSD $S_n(\omega)$ and bandwidth $2(\Delta f + B)$. The noise n(t) can be expressed in terms of quadrature components as,

$$n(t) = n_c(t)cos\omega_c t + n_s(t)sin\omega_c t \qquad (5.21a)$$



(b)

Figure (5.5): Performance of AM (synchronous detection and enveloped detection)

[Refer to figure (5.5) in the text book. Page 195]

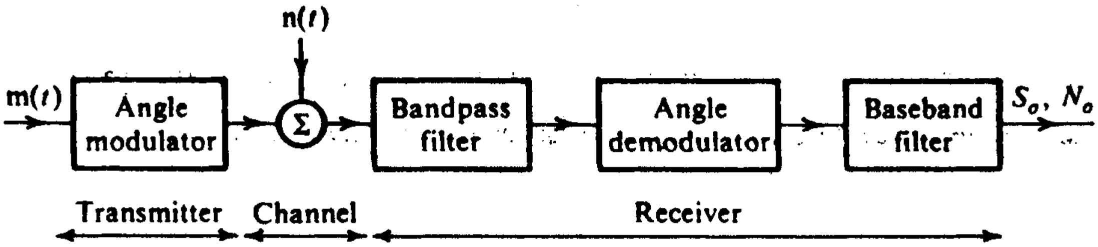


Figure (5.6): **Angle-modulated system**

[Refer to figure (5.6) in the text book. Page 196]

- Where $n_c(t)$ and $n_s(t)$ are low-pass signals of bandwidth Δf+B. The bandpass noise n(t) may also be expressed in terms of the enveloped $E_n(t)$ and phase $\varphi_n(t)$mas [see figure 5.5a and Eq. (5.18)]

$$n(t) = E_n(t)\cos[\omega_c t + \theta_n(t)] \qquad (5.21b)$$

- Angle modulation (and particularly wide-band angle modulation) is a nonlinear type of modulation. Hence, superposition does not apply. In AM, the signal output can be calculated by assuming the channel noise to be zero, and the noise output can be calculated by assuming the modulating signal to be zero

- This is a consequence of linearity. The signal and noise don't form intent nodulation components. Unfortunately, exponential modulation is nonlinear, and we cannot use superposition to calculate the output, as can be done in AM. We shall show that because of special circumstance, however, even in angle modulation the noise output can be calculated by assuming the modulating signal to be zero. To prove this we shall first consider the case of PM and then extend those results of FM

5.4.1 Phase Modulation

- Because narrow-band modulation is approximately linear, we need to consider only wideband modulation. The heart of the argument is that for wideband modulation, the signal m(t) changes very slowly relative to the noise n(t). the modulating signal bandwidth is B, and the noise bandwidth is $2(\Delta f + B)$, with $\Delta f >> B$

- Hence, the phase and frequency variations of the modulated carrier are much slower that are the variations of n(t)

- The modulated carrier appears to have constant frequency and phase over several cycles, and hence, the carrier appears to be unmodulated. We may therefore calculate the output noise by assuming m(t) to be zero (or constant). This is a qualitative justification for the linearity argument. A qualitative justification is given in the following development

- To calculate the signal and noise powers at the output, we shall first construct a phasor diagram of the signal $y_i(t)$ at the demodulator input, as shown in figure 5.7

$$y_i(t) = A\cos[\omega_c t + \psi(t)] + n(t)$$
$$= A\cos[\omega_c t + \psi(t)] + E_n(t)\cos[\omega_c t + \theta_n(t)]$$
$$= R(t)\cos[\omega_c t + \psi(t) + \Delta\psi(t)] \qquad (5.22)$$

- Where,

$$\psi(t) = K_p m(t) \,,,, \quad for\ PM \qquad (5.23)$$

- For the small-noise case, where $E_n(t) << A$ for "almost all t", $\Delta\Psi(t) <<$

π/2 for "almost all t", and

$$\Delta\psi(t) \cong \frac{E_n(t)}{a}\sin[\theta_n(t) - \psi(t)] \quad (5.24)$$

- The demodulator detects the phase of the input $y_i(t)$. hence, the demodulator output is

$$y_o(t) = \psi(t) + \Delta\psi(t) \quad (5.25a)$$

$$= K_p m(t) + \frac{E_n(t)}{A}\sin[\theta_n(t) - \psi(t)] \quad (5.25b)$$

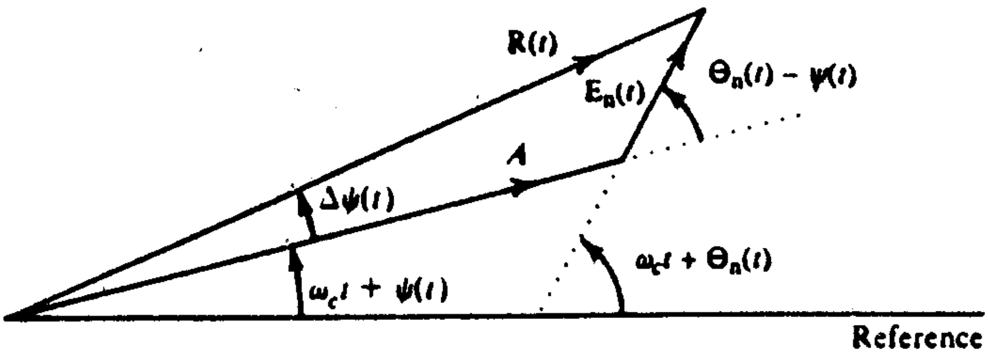


Figure (5.7): Phasor representation of signals in an angle-Modulated system

[Refer to figure (5.7) in the text book. Page 196]

- Note that the noise term ΔΨ(t) involves the signal Ψ(t) due to the nonlinear nature of angle modulation. Because Ψ(t) (baseband signal) varies much more slowly then $\theta_n(t)$ (wide-band noise), we can approximate Ψ(t) by a constant Ψ,

$$\Delta\psi(t) \cong \frac{E_n(t)}{A}\sin[\theta_n(t) - \psi]$$

$$= \frac{E_n(t)}{A} \sin\theta_n(t)\, cos\psi - \frac{E_n(t)}{A} \cos\theta_n(t)\, sin\psi$$

$$= \frac{n_s(t)}{A} cos\psi - \frac{n_c(t)}{A} sin\psi$$

- Also, because $n_c(t)$ and $n_s(t)$ are incoherent for white noise,

$$S_{\Delta\psi}(\omega) = \frac{cos^2\psi}{A^2} S_{ns}(\omega) + \frac{sin^2\psi}{A^2} S_{nc}(\omega)$$

$$= \frac{S_{ns}(\omega)}{A^2} \qquad ,, [because\ S_{nc}(\omega) = S_{ns}(\omega)] \qquad (5.26)$$

- For a white channel noise with PSD N/2

$$S_{\Delta\psi(\omega)} = \begin{cases} \frac{N}{A^2} & |f| \le \Delta f + B \\ 0 & otherwise \end{cases} \qquad (5.27a)$$

- Note that if we had assumed (t) = 0 (zero message signal) in Eq. (5.26), we should have obtained exactly the same result

- The demodulated noise bandwidth is f + B. But because the useful signal bandwidth is only B, the demodulator output is passed through a low-pass filter of bandwidth B to remove the out-of-band noise. Hence, the PSD of the low-pass filter output noise is,

$$S_{\mathrm{no}(\omega)} = \begin{cases} \frac{N}{A^2} & |\omega| \le 2\pi B \\ 0 & |\omega| \ge 2\pi B \end{cases} \qquad (5.27b)$$

- And

$$N_o = 2B\left(\frac{N}{A^2}\right) = \frac{2NB}{A^2} \qquad (5.28a)$$

- From Eq. (5.25b) we have,

$$S_o = K_p^2 \overline{m^2} \tag{5.28b}$$

- Thus,

$$\frac{S_o}{N_o} = (AK_p)^2 \frac{\overline{m^2}}{2NB} \tag{5.28c}$$

- These results are valid for small noise, and they apply to both WBPM and NBPM. We also have,

$$\gamma = \frac{S_i}{NB} = \frac{A^2/2}{NB} = \frac{A^2}{2NB}$$

- And

$$\frac{S_o}{N_o} = K_p^2 \overline{m^2} \gamma \tag{5.29}$$

- Also, for PM

$$\Delta\omega = K_p m_p' \qquad ,, \quad m_p' = [m(t)]_{max}$$

- Hence,

$$\frac{S_o}{N_o} = (\Delta\omega)^2 (\frac{\overline{m^2}}{m_p'^2}) \gamma \tag{5.30}$$

- Note that, the bandwidth of the angle-modulated waveform is about 2Δf (for the wide-band case). Thus, the output SNR increases with the square of the transmission bandwidth; that is, the output SNR increases by 6 dB for each doubling of the transmission bandwidth

- Remember, however, that this result is valid only when the noise power is much smaller than the carrier power. Hence, the output SNR cannot be increased indefinitely by increasing the transmission bandwidth because this also increases the noise power, and at some stage the small-noise assumption is violated

- When the noise power becomes comparable to the carrier power, the threshold appears as explained later, and a further increase in bandwidth actually reduces the output SNR instead of increasing it

- Let us apply Eq. (5.30) to tone modulation, where m(t) = $\alpha\cos\omega_m t$. for this case $\overline{m^2} = \alpha^2/2$, and $m_p' = |m(t)|_{max} = \alpha\omega_m$. Hence,

$$\frac{S_o}{N_o} = \frac{1}{2}\left(\frac{\Delta\omega}{\omega_m}\right)^2\gamma = \frac{1}{2}\left(\frac{\Delta f}{f_m}\right)^2\gamma \tag{5.31}$$

- Note that eq. (5.28c), (5.29), (5.30), and (5.31) are valid for both NBPM and WBPM

5.4.2 Frequency Modulation

- Frequency modulation may be considered as a special case of phase modulation, where the modulating signal is $\int_{-\infty}^{t} m(\alpha)d\alpha$ (figure 5.8)

- At the receiver, we can demodulate FM with a PM demodulator followed by a differentiator, as shown in figure 5.8. The PM demodulator output is $K_f\int_{-\infty}^{t} m(\alpha)d\alpha$. The subsequent differentiator yields the output $K_f m(t)$, so that,

$$S_o = K_f^2 \overline{m^2} \tag{5.32}$$

- The phase demodulator output noise will be identical to that calculated earlier, with PSD N/A^2 for white channel noise. This noise is passed through an ideal differentiator whose transfer function is $j\omega$. Hence, the PSD $S_{no}(\omega)$ of the output noise is $|j\omega|^2$ times the PSD in Eq. (5.27b)

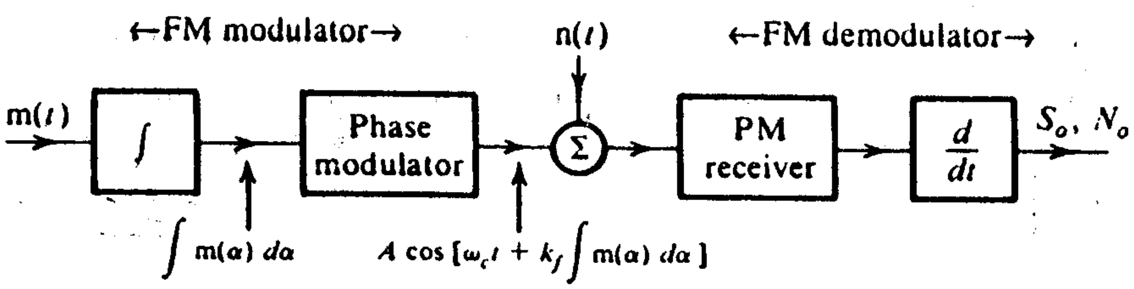


Figure (5.8): FM system as a special case of PM system

[Refer to figure (5.8) in the text book. Page 200]

$$S_{no(\omega)} = \begin{cases} \dfrac{N}{A^2}\omega^2 & |\omega| \le 2\pi B \\ 0 & |\omega| \ge 2\pi B \end{cases} \tag{5.33}$$

- The PSD of the output noise is parabolic (figure 5.9), and the output noise power is

$$N_o = 2\int_0^B \frac{N}{A^2}(2\pi f)^2 df$$
$$= \frac{8\pi^2 N B^3}{3A^2} \tag{5.34}$$

- Hence, the output SNR is

$$\frac{S_o}{N_o} = 3\left(\frac{K_f^2\overline{m^2}}{(2\pi B)^2}\right)\left(\frac{\frac{A^2}{2}}{NB}\right)$$

$$= 3\left[\frac{K_f^2\overline{m^2}}{(2\pi B)^2}\right]\gamma \quad (5.35)$$

- Because $\Delta\omega = K_f m_p$

$$\frac{S_o}{N_o} = 3\left(\frac{\Delta f}{B}\right)^2\left(\frac{\overline{m^2}}{m_p^2}\right)\gamma \quad (5.36)$$

$$= 3\beta^2\gamma\left(\frac{\overline{m^2}}{m_p^2}\right) \quad (5.37)$$

- Recall that the transmission bandwidth is about 2f. Hence, for each doubling of the bandwidth, the output SNR increases by 6 dB. Just as in the case of PM, the output SNR does not increase indefinitely because threshold appears as the increased bandwidth makes the channel noise power comparable to the carrier power

- For tone modulation, $\frac{\overline{m^2}}{m_p^2} = 0.5$ and

$$\frac{S_o}{N_o} = \frac{3}{2}\beta^2\gamma \quad (5.38)$$

- The output SNR S_o/N_o (in decibel) in Eq. (5.38) is plotted in figure 5.10 as a function of γ (in decibels) for various values of β. the dotted portion of the curves indicates the threshold region. Although the curves in figure 5.10 are valid for tone modulation only $\left(\frac{\overline{m^2}}{m_p^2} = 0.5\right)$, they can be used for any other modulating signal m(t) simply by shifting them vertically by a factor $(\frac{\overline{m^2}}{m_p^2}/0.5) = 2\frac{\overline{m^2}}{m_p^2}$,

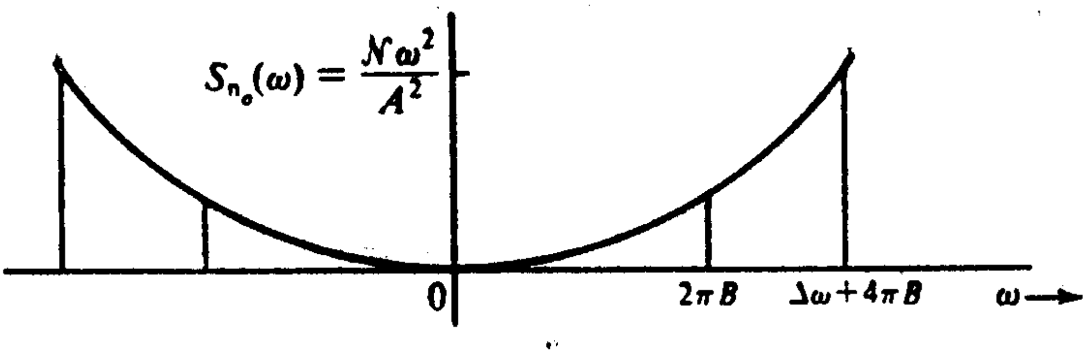


Figure (5.9): **PSD of output noise in FM receiver**

[Refer to figure (5.9) in the text book. Page 202]

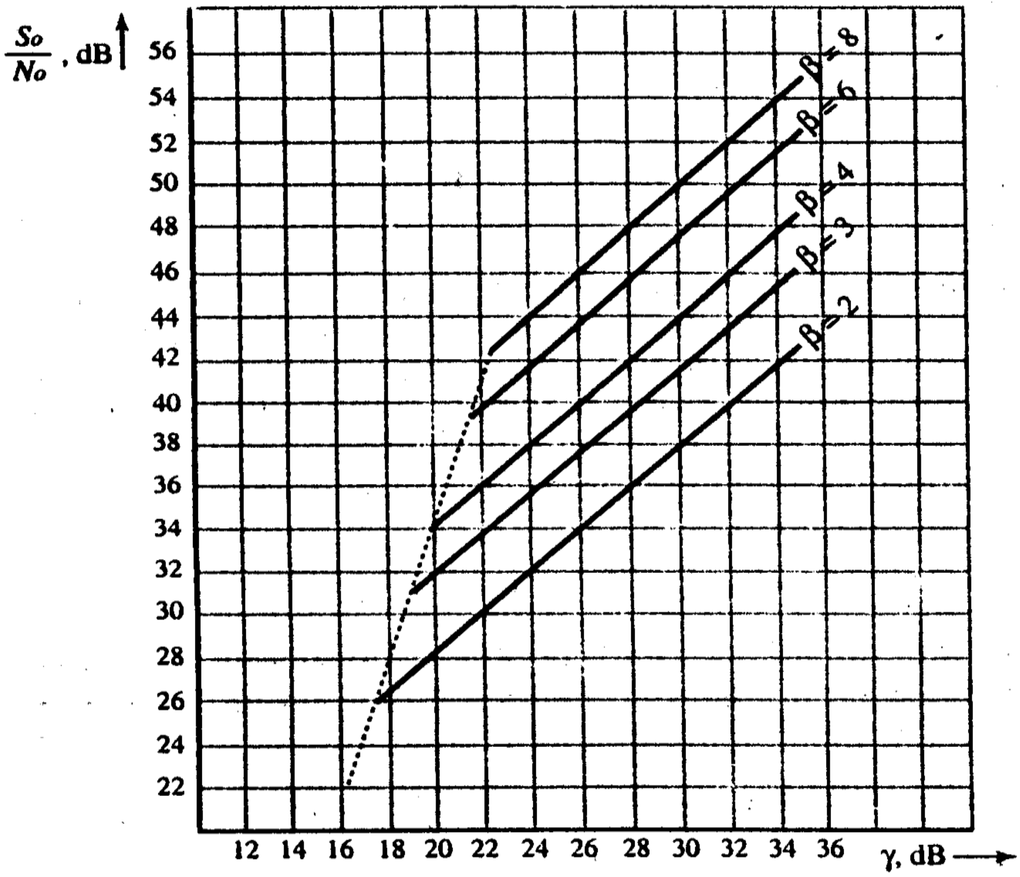


Figure (5.10): **Performance of FM system**

[Refer to figure (5.10) in the text book. Page 202]

- From Eq. (5.31) and (5.38), we observe that for tone modulation FM is superior to PM by a factor of 3. This does not mean that FM is superior to PM for other modulating signals as well. In fact, we shall see that PM is superior to FM for most of the practical signals. From Eq. (5.30) and (5.36), it can be seen that

$$\frac{(\frac{S_O}{N_O})_{PM}}{(\frac{S_O}{N_O})_{FM}} = \frac{(2\pi B)^2 m_p^2}{2m_p'^2} \qquad (5.39)$$

- Hence, if $(2\pi B)^2 m_p^2 > 3m_p'^2$, PM is superior to FM. If the PSD of m(t) is concentrated at lower frequencies, low-frequency components predominate in m(t), and $m_p'$is small. This favor PM; therefore, in general, PM is superior to FM when $S_m(\omega)$ is concentrated at lower frequencies and FM is superior to PM when $S_m(\omega)$ is concentrated at higher frequencies

- This explain why FM is superior to PM for tone modulation, where all the signal power is concentrated at the highest frequency in the band. But for most of the practical signals, the signal power is concentrated at lower frequencies, and PM proves superior to FM

- **Narrow-band modulation:** the equations thus far are valid for both narrowband and wideband modulation. We recall that narrow-band exponential modulation (NBEM) is approximately linear and is very similar to AM. In fact, the output SNRs for NBEM and AM are similar. To see this, consider the cases of NBPM and AM

$$\varphi_{AM}(t) = A\, cos\omega_c t + m(t) cos\omega_c t$$

$$\varphi_{NBPM}(t) = A\, cos\omega_c t - AK_p m(t) sin\omega_c t$$
$$= A\, cos\omega_c t - m(t)\sin\omega_c t$$

- Where $m_1(t) = AK_p m(t)$. Both φ_{AM} and φ_{NBPM} contain a carrier and a DSB term. In φ_{NBPM} the carrier and the DSB component are out of phase by π/2 rad, whereas in φ_{AM} they are in phase. But the π/2 rad phase difference has no effect on the power. Thus, m(t) in φ_{AM} analogous to $m_i(t)$ in φ_{NBPM}

- Now let us compare the output SNRs for AM and NBPM. For AM [Eq. (5.17)]

$$(\frac{S_o}{N_o})_{AM} = \frac{\overline{m^2}}{A^2 + \overline{m^2}}\gamma$$

- Whereas for NBPM with $m(t) = AK_p m(t)$. Eq. (5.29) can be expressed as,

$$(\frac{S_o}{N_o})_{PM} = K_p^2\overline{m^2}\gamma$$
$$= \frac{\overline{m_1^2}}{A^2}\gamma$$

- Note that for NBPM, we require that $|K_p m(t)| \ll 1$, that is, $m(t)/A << 1$, Hence,

$$A^2 \cong A^2 + \overline{m_1^2}$$

- And

$$(\frac{S_o}{N_o})_{PM} = \frac{\overline{m_1^2}}{A^2 + \overline{m_1^2}}\gamma$$

- Which is of the same form as $(So/No)_{AM}$. Hence, NBPM is very similar to AM. Under the best possible conditions, however, AM outperforms NBPM because for AM, we need only to satisfy the conditions

- [A + m(t)] > 0 which implies $[m(t)]_{max}$ A. Thus for tone modulation, the modulation index for AM can be nearly equal to unity. For NBPM, however, the narrow-band condition would be equivalent to requiring μ << 1. Hence, although AM and NBPM have identical performance for a given value of μ, AM has the edge over NBPM from the SNR viewpoint

- It is interesting to look for the line (in terms of Δf) that separates narrow-band and wideband FM. We may consider the dividing line to be that value of Δf for which the output SNR for FM given in Eq. (5.37) is equal to the maximum output SNR for FM. The maximum SNR for AM occurs when μ = 1, or when A = m_p. Hence, equating Eq. (5.37) with Eq. (5.17) [with A = m_p]

$$3\beta^2\gamma\left(\frac{\overline{m^2}}{m_p^2}\right) = \frac{\overline{m^2}}{m_p^2 + \overline{m^2}}\gamma$$

- Or

$$\beta^2 = \frac{1}{3}\left[\frac{1}{1+(\overline{m^2}/m_p^2)}\right] \tag{5.40}$$

- Because $\overline{m^2}/m_p^2 < 1$ for practical signals, and

$$\beta^2 < \frac{1}{3}$$

- Or

$$\beta < 0.6 \quad (5.41a)$$

- This gives

$$\Delta f = 0.6B \quad (5.41b)$$

5.5 Pulse-Modulated Systems

- Among pulse-modulated systems (PAM, PWM, PPM, and PCM), only PCM is of practical importance. The other systems are rarely used in practice. It can be shown that the performance of PAM is similar to that of AM-SC systems (i.e., the output SNR is equal to γ). The PWM and PPM systems are capable of exchanging the transmission bandwidth with output SNR, as in angle-modulated systems

- In PWM, the output SNR is proportional to the transmission bandwidth B_T. This performance is clearly inferior to that of angle-modulated systems, where the output SNR increases as B_T^2. In PPM systems under optimum conditions, the output SNR increases as B_T^2 but still inferior to FM by a factor of 6.

- Another PM system that deserves mention is the delta-modulation (DM) system. For speech signal, this system's performance is comparable to that of PCM for a bandwidth expansion ratio B_T/B of 7 to 8. For $B_T/B > 8$, PCM is superior to DM, and for $B_T/B < 8$, DM is superior to PCM

5.5.1 Noise in Pulse-Position Modulation

- In a PPM system, the transmitted information is contained in the relative position of the modulated pulses. The presence of additive noise affects the performance of such a system by falsifying the time at which the modulated pulses are judged to occur. Immunity to noise can be established by making the pulse build up so rapidly that the time interval during which noise can exert any perturbation is very short

- Indeed, additive noise would have no effects on the pulse positions if the received pulses were perfectly rectangular, because the presence of noise introduces only vertical perturbations. However, the reception of perfectly rectangular pulses implies an infinite channel bandwidth, which is of course impractical. Thus, with a finite channel bandwidth in practice, we find that the received pulses have a finite rise time, and so the performance of the PPM receiver is affected by noise

- As with a CW modulation system, the noise performance of a PPM system may be described in terms of the output signal-to-noise ratio. Also, to find the noise improvement produced by PPM over baseband transmission of a message signal, we may use the figure of merit defined as the output signal-to-noise ratio of the PPM system divided by the channel signal-to-noise ratio

5.5.2 Noise consideration in PCM systems

- In PCM, a baseband signal m(t) band-limited to B Hz and with amplitudes in the range of $-m_p$ to m_p is sampled at a rate of 2B samples per second. The sample amplitudes are quantized into L levels, each separated by $2m_p/L$. Each quantized sample is encoded into n binary digits ($2^n = L$). The binary signal is transmitted over a channel. The receiver detects the binary signal and reconstructs quantized samples (decoding). The quantized samples are then passed through a low-pass filter to obtain the desired signal m(t)

- There are two sources of error in PCM, 1) quantization or "rounding off" error and, 2) detection error. The latter is caused by error in the detection of the binary signal at the receiver

- The performance of a PCM system is influenced by two major sources of noise:
    1. Channel noise, which is introduced anywhere between the transmitter output and the receiver input. Channel noise is always present, once the equipment is switched on, which directly affect the detection error (bit error, probability of symbol error)

    2. Quantization noise, which is introduced in the transmitter and is carried all the way along to the receiver output. Unlike channel noise, quantization noise is signal-dependent in the sense that it disappears when the message signal is switched off

- Naturally, these two sources of noise appear simultaneously once the

PCM system is in operation. However, the traditional practice is to consider them separately, so that we may develop insight into their individual effects on the system performance

- The main effect of channel noise is to introduce bit error into the received signal, in case of a binary PCM system, the presence of a bit error causes symbol 1 to be mistaken for symbol 0, or vice versa. Clearly, the more frequently bit error occur, the more dissimilar the receiver output becomes compared to the original message signal

- The fidelity of information transmission by PCM in the presence of channel noise may be measured in terms of the average probability of symbol error, which is defined as the probability that the reconstructed symbol at the receiver output differs from the transmitted binary symbol, on the average

- The average probability of symbol error, also referred to as the error rate, assumes that all the bits in the received binary wave are of equal importance. When, however, there is more interest in reconstructing the analog waveform of the original message signal, different symbol errors may need to be weighted differently; for example, an error in the most significant bit in a code word (representing a quantized sample of the message signal) is more harmful than an error in the least significant bit

- To optimize system performance in the presence of channel noise, we need to minimize the average probability of symbol error. For this evaluation, it is customary to model the channel noise, originating at the front end of the receiver, as additive, white, and Gaussian

- The effect of channel noise can be made practically negotiable by ensuring the use of an adequate signal energy-to-noise density ratio through the provision of proper spacing between the regenerative repeaters in the PCM system. In such a situation, the performance of the PCM system is essentially limited by quantization noise acting alone

- From the discussion of quantization noise we recognize that quantization noise is essentially under the designer's control. It can be made negligibly small through the use of an adequate number of representation levels in the quantizer and the selection of a companding strategy matched to the characteristics of the type of message signal being transmitted

- We thus find that the use of PCM offers the possibility of building a communication system that is rugged with respect to channel noise on a scale that is beyond the capability of any CW modulation or analog pulse modulation system

5.5.3 Error Threshold

- The underlying theory of error rate calculation in a PCM system is discussed in details in literatures. For the present, it suffices to say that the average probability of symbol error in a binary encoded PCM receiver due to additive white Gaussian noise depends solely on E_b/N_o, the ratio of the transmitted signal energy per bit, E_b, to the noise spectral density, N_o

- Note that the ratio E_b/N_o is dimensionless even though the quantities E_b

and N_o have different physical meaning. In table 5.1 , we represent a summary of this dependence for the case of a binary PCM system using non return-to-zero signaling. The results represented in the last column of the table assume a bit rate of 10^5 b/s

Table 5.1. Influence of Eb/N0 on the Probability of Error

E_b/N_o	Probability of Error P_e	For a Bit Rate of 10^5 b/s This Is About One Error Every	
4.3 dB	10^{-2}	10^{-3}	second
8.4	10^{-4}	10^{-1}	second
10.6	10^{-6}	10	second
12.0	10^{-8}	20	minutes
13.0	10^{-10}	1	day
14.0	10^{-12}	3	month

- From table 5.1 it is clear that there is an error threshold (at about 11 dB). For E_b/N_o below the error threshold the receiver performance involves significant numbers of errors, and above it the effect of channel noise is practically negligible. In other words, provided that the ratio E_b/No exceeds the error threshold, channel noise has virtually no effect on the receiver performance, which is precisely the goal of PCM

- When, however, E_b/N_o drops below the error threshold, there is a sharp increase in the rate at which error occur in the receiver. Because decision errors result in the construction of incorrect code words, we find that when the errors are frequent, the reconstructed message at the receiver output bears little resemblance to the original message

- Comparing the figure of 11 dB for the error threshold in a PCM system using NRZ signaling with the 60-70 dB required for high-quality transmission of speech using amplitude modulation, we see that PCM requires much less power, even though the average noise power in the PCM system is increased by the R-fold increase in bandwidth, where R is the number of bits in a code word (i.e., bits per sample)

- In most transmission systems, the effects of noise and distortion from the individual links accumulate. For a given quality of overall transmission, the longer the physical separation between the transmitter and the receiver, the more serve are the requirements on each link in the system

- In a PCM system, however, because the signal can be regenerated as often as necessary, the effects of amplitude, phase, and nonlinear distortions in one link (if not too serve) have practically no effect on the regenerated input signal to the next link

- We have also seen that the effect of channel noise can be made practically negligible by using a ratio E_b/N_o above threshold. For all practical purposes, then , the transmission requirements for a PCM link are almost independent of the physical length of the communication channel

- Another important characteristic of a PCM system is its ruggedness to interference, caused by stray impulses or cross-talk. The combined presence of channel noise and interference cause the error threshold necessary for satisfactory operation of the PCM system to increase. If an adequate margin over the error threshold is provided in the first place, however, the system can withstand the [presence of relatively large amounts of interference. In other words, a PCM system is quite rugged

5.6 Noise Considerations in Digital Radio Systems

5.6.1 Probability of Error and Bit Error Rate

- Probability of error P(e) and bit error rate (BER) are often used interchangeably, although in practice they do have slightly different meanings. P(e) is a theoretical (mathematical) expectation of the bit error rate for a given system

- BER is an empirical (historical) record of a system's actual bit error performance. For example, if a system has a P(e) of 10^{-5}, this means that mathematically, you can expect one bit error in every 100,000 bits transmitted ($1/10^5 = 1/100{,}000$). If a system has a BER of 10^{-5}, this means that in past performance there was one bit error for every 100,000 bits transmitted. A bit error rate is measured, then compared to the expected probability of error to evaluate a system's performance

- Probability of error is a function of the carrier-to-noise power ratio (or, more specifically, the average energy per bit-to-noise power density ratio) and the number of possible encoding conditions used (M-ary).

- Carrier-to-noise power ratio is the ratio of the average carrier power (the combined power of the carrier and its associated sidebands) to the thermal noise power Carrier power can be stated in watts or dBm. where

$$C_{(dBm)} = 10 \log [C_{(watts)} / 0.001] \qquad (5.42)$$

- Thermal noise power is expressed mathematically as

$$N = KTB \text{ (watts)} \qquad (5.43a)$$

- Where
  - N = thermal noise power (watts)
  - K = Boltzmann's proportionality constant (1.38×10^{-23} joules per kelvin)
  - T= temperature (kelvin: 0 K=-273° C, room temperature = 290 K)
  - B = bandwidth (hertz)

Stated in dBm, $$N_{(dBm)} = 10 \log [KTB / 0.001] \qquad (5.43b)$$

- Mathematically, the carrier-to-noise power ratio is

$$C / N = C / KTB \text{ (unitless ratio)} \qquad (5.44a)$$

- Where
  - C = carrier power (watts)
  - N = noise power (watts)

Stated in dB,

$$C / N \text{ (dB)} = 10 \log [C / N]$$
$$= C_{(dBm)} - N_{(dBm)} \qquad (5.44b)$$

- Energy per bit is simply the energy of a single bit of information. Mathematically, energy per bit is

$$E_b = CT_b \text{ (J/bit)} \qquad (5.45a)$$

- Where
  - E_b = energy of a single bit (joules per bit)

– T_b = time of a single bit (seconds)

– C = carrier power (watts)

Stated in dBJ, $E_{b(dBJ)} = 10 \log E_b$ (5.45b)

- And because $T_b = 1/f_b$, where f_b is the bit rate in bits per second, E_b can be rewritten as

$$E_b = C / f_b \text{ (J/bit)} \quad (5.45c)$$

Stated in dBJ, $E_{b(dBJ)} = 10 \log C / f_b$ (5.45d)

$$= 10 \log C - 10 \log f_b \quad (5.45e)$$

- Noise power density is the thermal noise power normalized to a 1-Hz bandwidth (i.e., the noise power present in a 1-Hz bandwidth). Mathematically, noise power density is

$$N_o = N / B \text{ (W/Hz)} \quad (5.46a)$$

- Where

– N_o = noise power density (watts per hertz)

– N = thermal noise power (watts)

– B = bandwidth (hertz)

Stated in dBm,

$$N_{o(dBm)} = 10 \log (N/0.001) - 10 \log B \quad (5.46b)$$

$$= N_{(dBm)} - 10 \log B \quad (5.46c)$$

- Combining Equations 5.43a and 5.46a yields

$$N_o = KTB / B = KT \text{ (W/ Hz)} \quad (5.46d)$$

Stated in dBm,

$$N_{o(dBm)} = 10 \log (K/0.001) + 10 \log T \qquad (5.46e)$$

- Energy per bit-to-noise power density ratio is used to compare two or more digital modulation systems that use different transmission rates (bit rates), modulation schemes (FSK, PSK, QAM), or encoding techniques (M-ary). The energy per bit-to-noise power density ratio is simply the ratio of the energy of a single bit to the noise power present in 1 Hz of bandwidth. Thus E_b/N_o normalizes all multiphase modulation schemes to a common noise bandwidth allowing for a simpler and more accurate comparison of their error performance

- Mathematically, E_b/N_o is

$$E_b/N_o = (C/f_b) / (N/B) = CB/Nf_b \qquad (5.47a)$$

- Where E_b/N_o *is* the energy per bit-to-noise power density ratio. Rearranging Equation 5.47a yields the following expression:

$$E_b/N_o = (C/N) \times (B/f_b) \qquad (5.47b)$$

- Where
  - E_b/N_o = energy per bit-to-noise power density ratio
  - C/N = carrier-to-noise power ratio
  - B/f_b = noise bandwidth-to-bit rate ratio

Stated in dB,

$$E_b/N_o\ (dB) = 10 \log (C/N) + 10 \log (B/f_b) \qquad (5.47c)$$

$$= 10 \log E_b - 10 \log N_o \qquad (5.47d)$$

- From Eq. (5.47b) it can be seen that the E_b/N_o ratio is simply the product of the carrier-to-noise power ratio and the noise bandwidth-to-bit rate ratio. Also, from Eq. (5.47b) it can be seen that when the bandwidth equals the bit rate, $E_b/N_o = C/N$

- In general, the minimum carrier-to-noise power ratio required for QAM systems is less than that required for comparable PSK systems. Also, the higher the level of encoding used (the higher the value of M), the higher the minimum carrier-to-noise power ratio

Example 5.2

- For a QPSK system and the given parameters, determine

  a. Carrier power in dBm.

  b. Noise power in dBm.

  c. Noise power density in dBm.

  d. Energy per bit in dBJ.

  e. Carrier-to-noise power ratio in dB.

  f.. E_b/N_o ratio.

  – $C = 10^{-12}$ W , $F_b = 60$ kbps

  – $N = 1.2 \times 10^{-14}$ W , $B = 120$ kHz

Solution

a. The carrier power in dBm is determined by substituting into Equation 5.42:

$$C = 10 \log (10^{-12} / 0.001) = -90 \text{ dBm}$$

b. The noise power in dBm is determined by substituting into Equation 5.43b:

$$N = 10 \log [(1.2\text{x}10^{-14}) / 0.001] = -109.2 \text{ dBm}$$

c. The noise power density is determined by substituting into Equation 5.46e:

$$N_o = -109.2 \text{ dBm} - 10 \log 120 \text{ kHz} = -160 \text{ dBm}$$

d. The energy per bit is determined by substituting into equation 5.45d:

$$E_b = 10 \log (10^{-12} / 60 \text{ kbps}) = -167.8 \text{ dBJ}$$

e. The carrier-to-noise power ratio is determined by substituting into Equation 5.44b:

$$C / N = 10 \log (10^{-12} / 1.2 \text{ x } 10^{-14}) = 19.2 \text{ dB}$$

f. The energy per bit-to-noise density ratio is determined by substituting into Equation 5.47c:

$$E_b / N_o = 19.2 + 10 \log 120 \text{ kHz} / 60 \text{ kbps} = 22.2 \text{ dB}$$

5.6.2 PSK Error Performance

- The bit error performance is related to the distance between points on a signal state-space diagram.

- For example, on the signal state-space diagram for BPSK shown in Figure 5.11a, it can be seen that the two signal points (logic 1 and logic 0) have maximum separation (d) for a given power level (D). In fact, one BPSK signal state is the exact negative of the other

- The figure shows, a noise vector (V_N), when combined with the signal vector (V_s), effectively shifts the phase of the signaling element (V_{SE}) alpha degrees. If the phase shift exceeds ±90°, the signal element is shifted beyond the threshold points into the error region.

- For BPSK, it would require a noise vector of sufficient amplitude and phase to produce more than a ±90° phase shift in the signaling element to produce an error. For PSK systems, the general formula for the threshold points is

$$TP = \pm \pi / M \qquad (2.47)$$

- Where M is the number of signal states.

- The phase relationship between signaling elements for BPSK (i.e., 180^0 out of phase) is the optimum signaling format, referred to as antipodal signaling, and occurs only when two binary signal levels are allowed and when one signal is the exact negative of the other. Because no other bit-by-bit signaling scheme is any better, antipodal performance is often used as a reference for comparison.

- The error performance of the other multiphase PSK systems can be compared with that of BPSK simply by determining the relative decrease in error distance between points on a signal state-space diagram.

- For PSK, the general formula for the maximum distance between signaling points is given by

$$\text{Sin } \theta = \sin 360^0 / 2M = (d/2) / D \qquad (5.49)$$

- Where
  - d = error distance
  - M = number of phases
  - **D =** peak signal amplitude

- Rearranging equation 5.49 and solving for d yields

$$d = \left(2\sin\frac{180^0}{M}\right)xD \qquad (5.50)$$

- Figure 5.11b shows the signal state-space diagram for QPSK. From Figure 5.11b and Equation 3.2, it can be seen that QPSK can tolerate only a ± 45° phase shift.

- From Equation 5.48 the maximum phase shift for 8-PSK and 16-PSK is ±22.5° and ±11.25°, respectively.

- Consequently, the higher the level of modulation (i.e., the greater the value of M) require a greater energy per bit-to-noise power density ratio to reduce the effect of noise interference. Hence, the higher the level of modulation, the smaller the angular separation between signal points and the smaller the error distance.

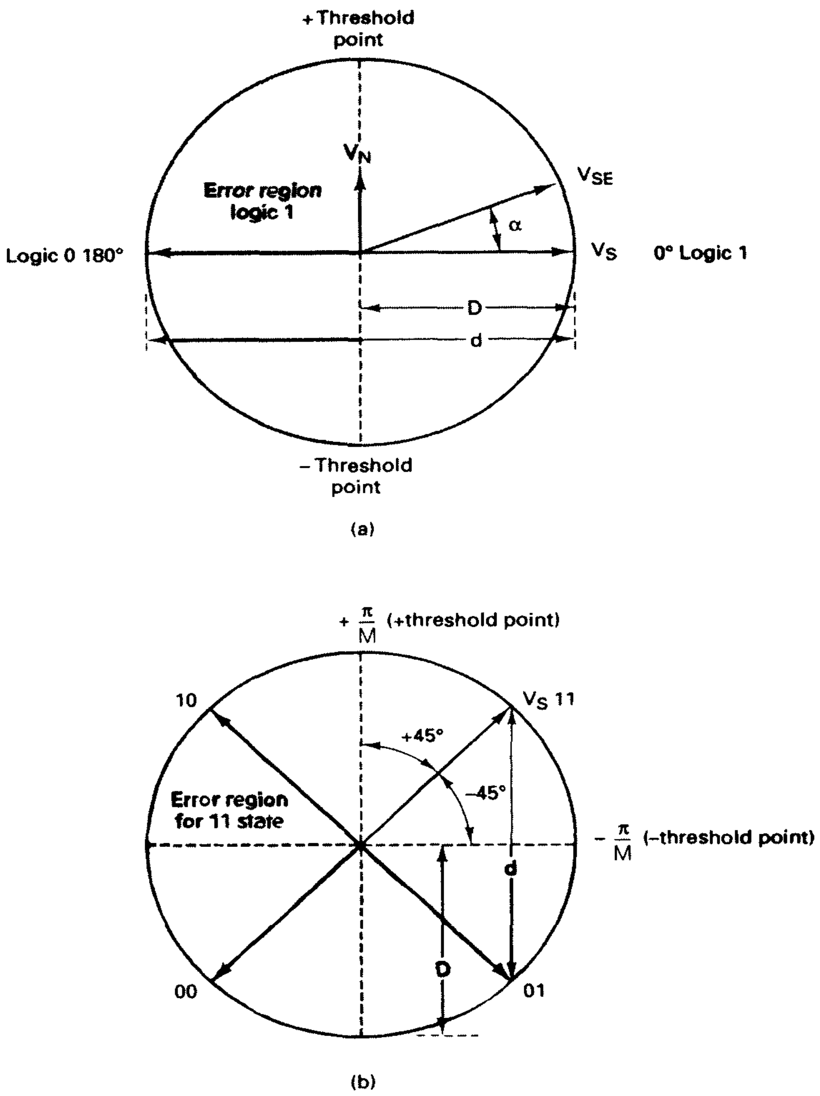


Figure (5.11): PSK error region: (a) BPSK; (b) QPSK

[Refer to figure (5.11) in the text book. Page 216]

- The general expression for the bit error probability of an M-phase PSK system is

$$P(e) = (1 / \log_2 M)\ \mathrm{erf}\ (z) \qquad (5.51)$$

- Where erf = error function

$$z = \sin(\pi/M)\ (\sqrt{\log_2 M})(\sqrt{E_b / N_0})$$

- By substituting into Equation 5.51, it can be shown that QPSK provides the same error performance as BPSK. This is because the 3-dB reduction in error distance for QPSK is offset by the 3-dB decrease in its bandwidth (in addition to the error distance, the relative widths of the noise bandwidths must also be considered).

- Thus, both systems provide optimum performance. Figure 5.12 shows the error performance for 2-, 4-, 8-, 16-, and 32-PS K systems as a function of E_b / N_o.

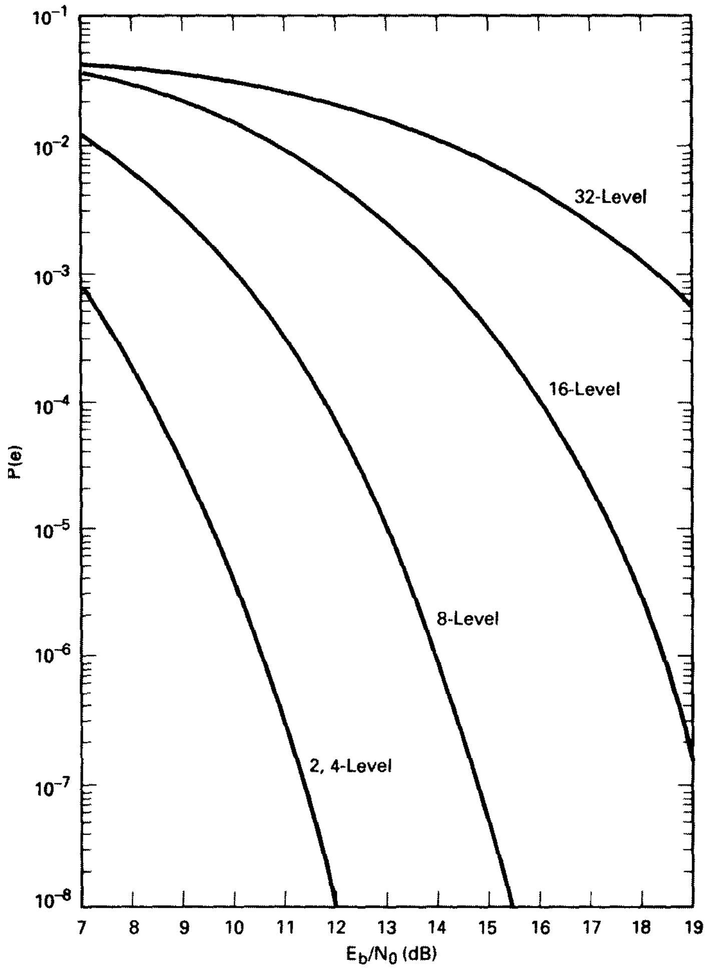


Figure (5.12): Error rates of PSK modulation systems

[Refer to figure (5.12) in the text book. Page 217]

Example 5.3

- Determine the minimum bandwidth required to achieve a P(e) of 10^{-7} for an 8-PSK system operating at 10 Mbps with a carrier-to-noise power ratio of 11.7 dB.

Solution

- From Figure 5.12, the minimum E_b / N_o ratio to achieve a P(e) of 10^{-7} for an 8-PSK system is 14.7 dB. The minimum bandwidth is found by rearranging Equation 5.47b:

  - $B / f_b = E_b / N_o = C / N$

    $= 14.7 \text{ dB} – 11.7 \text{ dB} = 3 \text{ dB}$

  - $B / f_b = \text{antilog } 3 = 2$; $B = 2 \times 10 \text{ Mbps} = 20 \text{ MHz}$

5.6.3 QAM Error Performance

- For a large number of signal points (i.e., M-ary systems greater than 4), QAM outperforms PSK. This is because the distance between signaling points in a PSK system is smaller than the distance between points in a comparable QAM system. The general expression for the distance between adjacent signaling points for a QAM system with L levels on each axis is

$$d = \frac{\sqrt{2}}{L-1} D \qquad (5.52)$$

- Where
  - d = error distance

– L = number of levels on each axis

– D = peak signal amplitude

- In comparing Equation 5.50 to Equation 5.52, it can be seen that QAM systems have an advantage over PSK systems with the same peak signal power level. The general expression for the bit error probability of an L-level QAM system is

$$P(e) = \frac{1}{\log_2 L}\left(\frac{L-1}{L}\right)erfc(z) \qquad (5.53)$$

- Where erfc(z) is the complementary error function

$$z = \frac{\sqrt{log_2 L}}{L-1}\sqrt{\frac{E_b}{N_o}}$$

- Figure 5.13 shows the error performance for 4-, 16-, 32-, and 64-QAM systems as a function of E_b/N_o.

- Table 5.2 lists the minimum carrier-to-noise power ratios and energy per bit-to-noise power density ratios required for a probability of error 10^{-6} for several PSK and QAM modulation schemes.

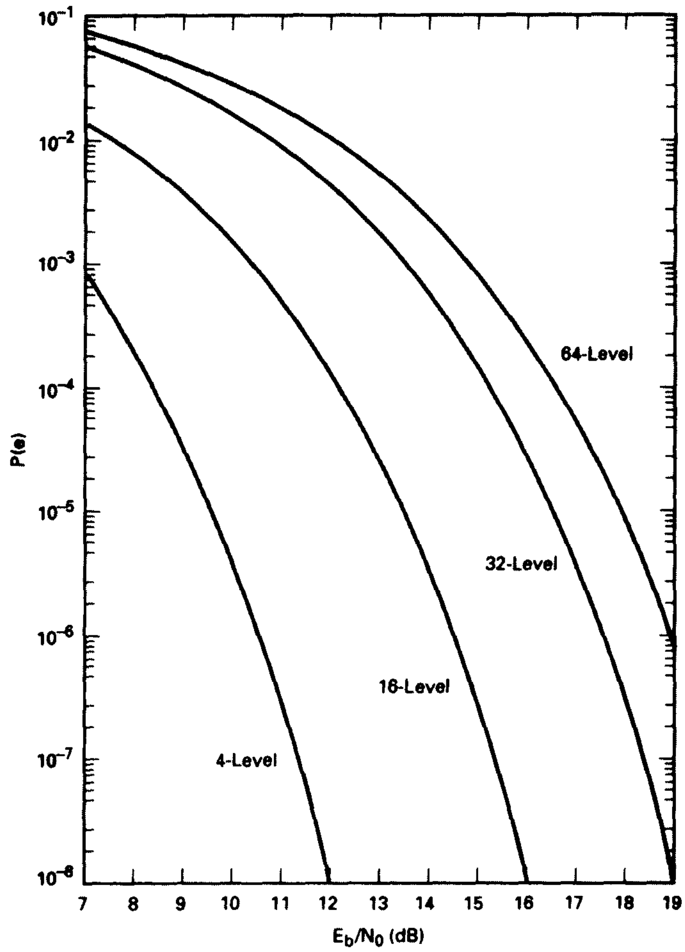


Figure (5.13): Error rates of QAM modulation systems

[Refer to figure (5.13) in the text book. Page 219]

Table 5.2 Performance Comparison of various digital modulation (BER = 10^{-6})

Modulation Technique	C/N Ratio (dB)	E_b/N_0 Ratio (dB)
BPSK	10.6	10.6
QPSK	13.6	10.6
4-QAM	13.6	10.6
8-QAM	17.6	10.6
8-PSK	18.5	14
16-PSK	24.3	18.3
16-QAM	20.5	14.5
32-QAM	24.4	17.4
64-QAM	26.6	18.8

Example 5.4

- Which system requires the highest E_b/N_o ratio for a probability of error of 10^{-6}, a four-level QAM system or an 8-PSK system?

Solution

- From Figure 5.13, the minimum E_b/N_O ratio required for a four-level QAM system is, 10.6 dB. From Figure 5.12, the minimum E_b/N_o ratio required for an 8-PSK system is 14 dB Therefore, to achieve a P(e) of 10^{-6}, a four-level QAM system would require 3.4 dB less E_b/N_O ratio.

5.6.4 FSK Error Performance

- The error probability for FSK systems is evaluated in a somewhat different manner than PSK and QAM. There are essentially only two types of FSK systems: noncoherent (asynchronous) and coherent (synchronous)

- With noncoherent FSK, the transmitter and receiver are not frequency or phase synchronized. With coherent FSK, local receiver reference signals are in frequency and phase lock with the transmitted signals. The probability of error for noncoherent FSK is

$$P(e) = 1/2 \exp\left(\frac{-E_b}{2N_o}\right) \qquad (5.54)$$

- The probability of error for coherent FSK is

$$P(e) = erfc\sqrt{\frac{E_b}{N_0}} \qquad (5.55)$$

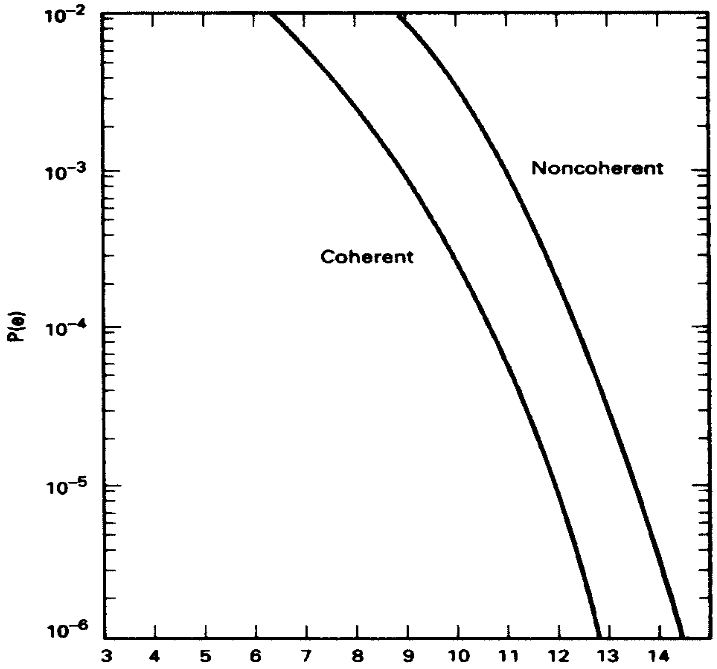


Figure (5.14): Error rates for FSK modulation systems

[Refer to figure (5.14) in the text book. Page 220]

- Figure 5.14 shows probability of error curves for both coherent and non-coherent FSK for several values of E_b/N_o. From Equations 5.54 and 5.55, it can be determined that the probability of error for non-coherent FSK is greater than that of coherent FSK for equal energy per bit-to-noise power density ratios.