GOVERNMENT OF TAMILNADU
DIRECTORATE OF TECHNICAL EDUCATION
CHENNAI – 600 025

STATE PROJECT COORDINATION UNIT

Diploma in Electronics and Communication Engineering

Course Code: 1040

M – Scheme

e-TEXTBOOK
on
DIGITAL COMMUNICATION
for
V Semester DECE

Convener for ECE Discipline:

Dr. M.JEGANMOHAN,M.E.,MBA.,Ph.D.,(Management).,Ph.D.,(Engg).,M.I.S.T.E.,
Principal,
Government Polytechnic College,
Uthappanaickanoor,
Usilampatti, Madurai – 625 536.

Team Members for Digital Communication

Mr. G.ATHIRAJAN, B.E., M.E.,
HOD / ECE,
K.L.Nagaswamy Memorial Polytechnic College,
Madurai – 625 009.

Mrs. RM.Shanthi, B.E., M.E.,
Senior Lecturer /ECE,
K.L.Nagaswamy Memorial Polytechnic College,
Madurai – 625 009.

Mr. P.Perumal, B.E., M.E.,
Senior Lecturer /ECE,
K.L.Nagaswamy Memorial Polytechnic College,
Madurai – 625 009.

Validated By

Dr. S.J. THIRUVENGADAM, M.E.,Ph.D.,
Professor / ECE,
Thiagarajar College of Engineering
Madurai – 625 015.

34071 DIGITAL COMMUNICATION

DETAILED SYLLABUS

UNIT	NAME OF THE TOPIC	HOURS
I	**BASICS OF DIGITAL COMMUNICATION** Digital Communication signal processing – Typical Block diagram and transformations – Advantages over analog communication – Channels for Digital Communication – Telephone, Optical fiber, Satellite. Classification of Signals – Deterministic and random signals – Periodic and Non-Periodic signals – Analog and Discrete Signals – Energy and Power Signals – Unit Impulse Function. Information Capacity (Definition only) – Shannon's limit for information capacity (Definition only) – Data transmission – Serial and parallel transmission – Synchronous and asynchronous transmission.	13
II	**FORMATTING AND BASE BAND MODDULATION** Base band system – The Sampling Theorem – Impulse sampling – Natural Sampling – Sample and Hold Operation – Spectra – Nyquist Theorem – Aliasing – Signal interface for a digital system – Sampling and quantizing effects – Quantization noise – Channel effects – channel noise – PCM – Uniform and Non-uniform Quantization. **BASEBAND TRANSMISSION** PCM waveform types – non return-to-zero (NRZ) – return-to-zero (RZ) – Phase encoded – Multilevel binary – Spectral attributes of PCM waveforms – Bits per PCM word and Bits per symbol – PCM word size – M-ary pulse modulation waveforms.	13
III	**BASEBAND CODING TECHNIQUES** Rationale for coding – Types of codes – Discrete memoryless channel – Error control coding methods – forward error correction – error detection with retransmission – types of errors – random error and burst error – Principles of linear block codes - Hamming code – Binary cyclic codes – Cyclic Redundancy Check code (CRC) – Convolution code.	13
IV	**DIGITAL MODULATION TECHNIQUES** Digital modulation Techniques – Listing of various types – Coherent binary modulation techniques – Coherent quadrature modulation techniques – Non Coherent binary modulation techniques – Minimum Shift Keying (MSK) – Block diagram of MSK transmitter and receiver – TDM – Frame Structure, ASCII Framing – E1 Framing, T1 Framing for Telephone. Detection of signals – Coherent detection of PSK – Sampled matched filter – Coherent detection of FSK – Non-coherent	12

	detection – Binary differential PSK.	
V	**<u>SPREAD SPECTRUM TECHNIQUES</u>** Spread spectrum communication – Beneficial attributes of spread spectrum systems – Pseudo noise sequences – Randomness properties – Balance property, Run property and Correlation property – Direct sequence spread spectrum systems – Processing gain and performance – Frequency hopping systems – Frequency hopping with diversity - fast hopping versus slow hopping – Synchronization – Jamming consideration – Commercial application – CDMA Digital Cellular System.	**12**
	REVISION & TEST	**12**

REFERENCE BOOKS:

Sl. No.	Title	Author	Publisher with Edition
1	Digital communications Fundamentals & Applications	Bernard Sklar & Pabitra Kumar Ray	Pearson – Second Edition – 2009
2	Digital Communications	Simon Haykin	John Wiley India edition – 2006
3	Digital Communication	Dr. J.S. Chitode	Technical Publications – Pune Second Edition, 2011.
4	Digital and analog communication system	B.P. Lathi, Zhi Ding	International 4^{th} Edition – OXFORD University Press
5	Digital Communication	P. Ramakrishna Rao	TMH, 2011
6	Principles of Communications System	Taub & Schilling	TMH, Third edition, 2008
7	Digital Communications	John G. Prokais	2011
8	Digital Communications	Dr. K.N. Hari Bhat Dr. D, Ganesh Rao	Saanguine technical Publisher, 2005

CONTENTS

Unit - I BASICS OF DIGITAL COMMUNICATION

OBJECTIVES

- To study about the basic elements or building blocks that constitutes a digital communication system.
- To study about various types of digital communication channels.
- To study about the various types of signals.
- To study about data transmission.

1.1 INTRODUCTION

The purpose of a communication system is to transmit information-bearing signals from a source, located at one location, to a user destination, located at another distant location. Based on the nature of signal processing applied to the information-bearing signal, communication systems may be broadly divided into two major systems. They are:

1) Analog Communication System

2) Digital Communication System

In an analog communication system, the information bearing analog signal is continuously varying in both amplitude and time. It is used directly to modify some characteristics of a high frequency sinusoidal carrier wave, such as amplitude, phase or frequency. Speech signal, video signal, temperature signal, pressure signal etc., are some examples of analog signal.

In digital communication system, the information bearing digital signal is processed such that it can be represented by a sequence of binary digits (discrete messages). Then it is used for ON/OFF keying of some characteristic of a high frequency sinusoidal carrier wave, such as amplitude, phase or frequency. If the input message signal is in analog form, then it is converted to digital form by the processes of sampling, quantizing and encoding. Computer data and telegraph signals are some examples of digital signal. The key feature of a digital communication system is that it deals with a finite set of discrete messages.

Digital communication systems are becoming increasingly attractive due to the ever-growing demand for data communication. Because digital transmission offers data processing options and flexibilities not available with analog transmission. Further, developments in digital techniques have led to more and more powerful microprocessors, larger and larger memory devices and a number of programmable logic devices. Availability of these devices has made the design of digital communication systems highly convenient.

1.2 DIGITAL COMMUNICATION SIGNAL PROCESSING

The transmission of information (voice, video, or data) over a path (channel) may consist of wires, waveguides, or space. The principle feature of a digital communication system is that during a finite interval of time, it sends a signal waveform from a finite set of possible waveforms. During propagation, the amplitude and shape of the signal waveform gets degraded. The objective of the receiver is to determine from a noise-perturbed signal which waveform from the finite set of waveforms was sent by the transmitter.

Figure 1.1 illustrates an ideal binary digital pulse propagating along a transmission line. The shape of the waveform is affected by two basic mechanisms.

1) Due to some non-ideal frequency transfer function of all transmission lines and circuits.
2) Unwanted electrical noise or other interference further distorts the pulse waveform.

Both of these mechanisms cause the pulse shape to degrade as a function of line length, as shown in the figure 1.1.

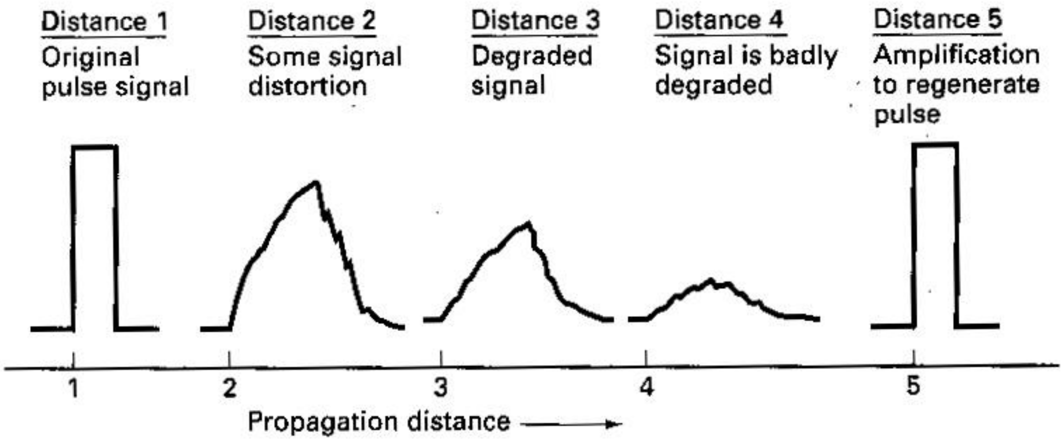


Figure 1.1 Pulse degradation and regeneration

During propagation, before the pulse is degraded to an ambiguous state, some corrective signal processing methods have to be done. This process is called as Regeneration.

In Regeneration, the pulse is amplified by a digital amplifier that recovers its original ideal shape. The pulse is thus “reborn” or regenerated. Circuits that perform this function at regular intervals along a transmission system are called Regenerative repeaters.

1.3 ADVANTAGES OF DIGITAL COMMUNICATION OVER ANALOG COMMUNICATION

1. The use of “Regenerative repeaters” generate strong error free signal at a good power level.
2. Digital circuits are less subject to distortion and interference than analog circuits.
3. With digital techniques, extremely low error rates producing high signal fidelity are possible through error detection and correction.
4. Digital circuits are more reliable and can be produced at a lower cost than analog circuits.
5. Digital hardware lends itself to more flexible implementation than analog circuits.
6. The combining of digital signals using Time Division Multiplexing (TDM) is simpler than the combining of analog signals using Frequency Division Multiplexing(FDM).
7. Different types of digital signals (data, telegraph, telephone, television) can be treated as identical signals in transmission and switching - a bit is a bit.
8. Digital techniques lend themselves naturally to signal processing functions that protect against interference and jamming or that provide encryption and privacy.
9. Also, much data communication is from computer to computer, or from digital instruments or terminal to computer. Such digital terminations are naturally best served by digital communication links.
10. Storage and retrieval of voice, data or video at intermediate points (in the transmission path) is easy and is inexpensive in terms of storage space.
11. Signal processing and image-processing operations like compression of voice and image signals, etc. can easily be carried out.
12. Adaptive equalization can be implemented.
13. Very powerful encryption and decryption algorithms are available for digital data so as to maintain a high level of secrecy of communication.
14. Availability of powerful microprocessors, larger memory devices, and number of programmable logic devices has made the design of digital communication systems highly convenient.

15. The mathematical theory of logic circuits called as switching theory is a very useful concept in digital communication.
16. The effect of noise, temperature and parameter variations is very small in digital circuits.

1.4 DISADVANTAGES OF DIGITAL COMMUNICATION OVER ANALOG COMMUNICATION

1. Digital systems tend to be very signal-processing intensive compared with analog systems.
2. Digital systems need to allocate a significant share of their resources to the task of synchronization at various levels.
3. Sometimes non-graceful degradation occurs in digital communication systems, ie., when the signal-to-noise ratio drops below a certain threshold, the quality of service can change suddenly from very good to very poor.
4. Digital communication systems generally need more bandwidth than analog communication systems.
5. Digital components generally consume more power as compared to analog components.

1.5 TYPICAL BLOCK DIAGRAM AND TRANSFORMATIONS

The principal feature of a Digital communication system is that during a finite interval of time, the transmitter sends a waveform from a finite set of possible waveforms. The receiver has to determine from a noise perturbed signal which waveform from the finite set of waveforms was sent by the transmitter.

The block diagram of a typical digital communication system with only the essential blocks is shown in the figure 1.2(a). The functions of encryption, multiplexing, spreading, multiple access and equalization are optional.

The upper blocks- Formatter, Source encoder, channel encoder, Baseband processor/ Band pass modulator- denote signal transformations from the source to the transmitter. The lower blocks-Baseband decoder/Bandpass demodulator, channel decoder, source decoder, Deformatter – denote signal transformations from the receiver to the sink. The lower blocks essentially reverse the signal processing steps performed by the upper blocks. We shall discuss the basic functions of each of these blocks.

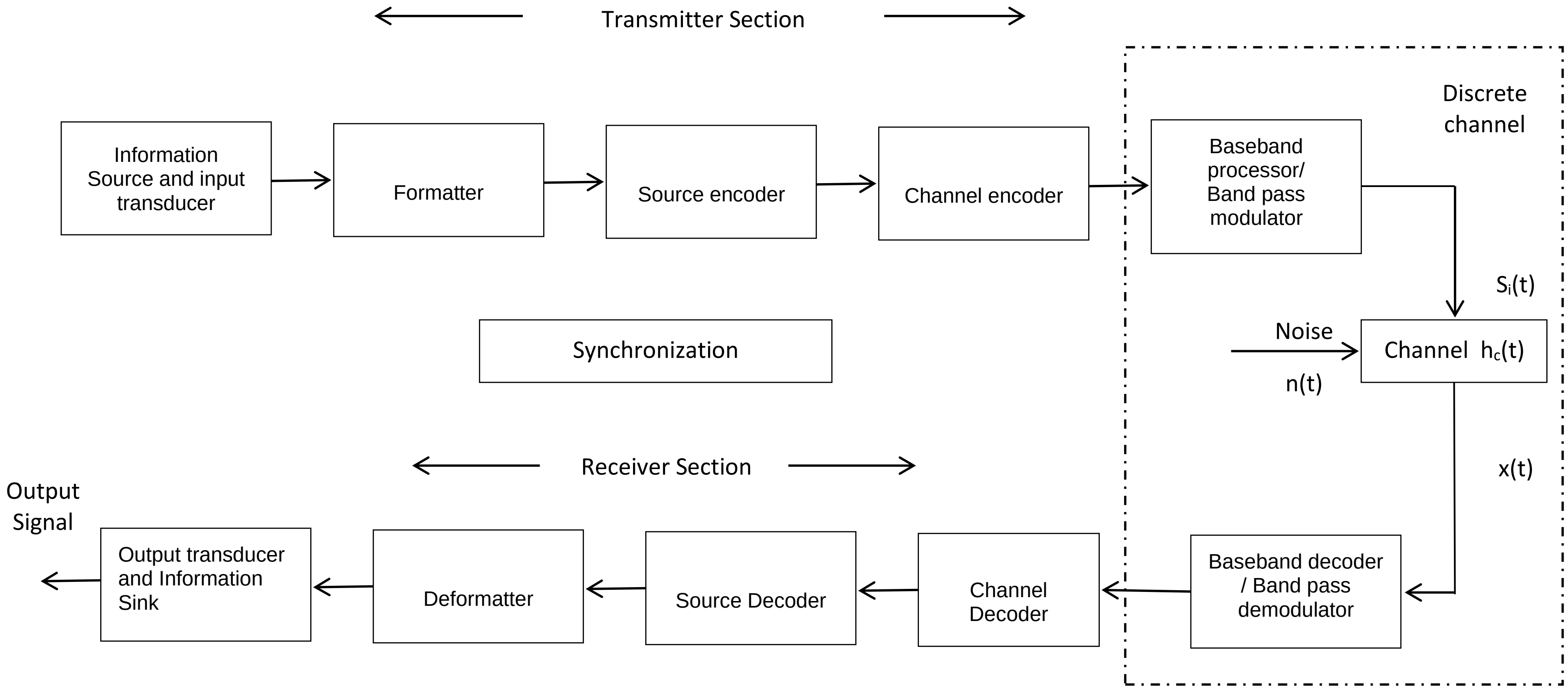


Figure 1.2(a): Block diagram of a Typical Digital Communication System

TRANSMITTER SECTION

1) Information source

The Source is where the information to be transmitted, originates. The information / message may be available in digital form (eg: computer data, tele-type data). If the information / message available is a non-electrical signal, (eg: video signal, voice signal) then it is first converted into a suitable electrical signal using an input transducer. Then the analog electrical signal is sampled and digitized using an analog to digital converter to make the final source output to be in digital form.

2) Formatter

Formatting transforms the source information into binary digits (bits). The bits are then grouped to form digital messages or message symbols. Each such symbol (m_i, where i = 1,2,3……M) can be regarded as a member of a finite alphabet set containing M members. Thus for M=2, the message symbol m_i is binary (it constitutes just a single bit). For M>2, such symbols are each made up of a sequence of two or more bits (M-ary)

3) Source encoder

The process of efficiently converting the output of either an analog or digital source into a sequence of binary digits is called source encoding or data compression. Source coding produces analog-to-digital (A/D) conversion for analog sources. It also removes redundant (unneeded) information. By reducing data redundancy, source codes can reduce a system's data rate (ie., reduced bandwidth).

Formatting and source coding are similar processes, in that they both involve data digitization. However, source coding involves data compression in addition to digitization. Hence, a typical digital communication system would either use formatter, (for digitizing alone) or source encoder (for both digitizing and compressing).

4) Channel encoder

The channel encoder introduces some redundancy in the binary information sequence, in a controlled manner. Such introduction of controlled redundancy can be used at the receiver to provide error correction capability to the data being transmitted. This minimises the effects of noise and interference encountered in the transmission of the signal through the channel. Hence channel coding increases the reliability of the received data and improves the fidelity of the received signal. Channel coding is used for reliable transmission of digital data.

5) Base band processor

For low speed wired transmission, each symbol to be transmitted is transformed from a binary representation (voltage levels representing binary ones and zeros) to a baseband waveform. The baseband refers to a signal whose frequency range extends from DC up to a few MHz. The baseband processor is a pulse modulation circuit. When pulse modulation is applied to binary symbols, the resulting binary waveform is called Pulse Code-Modulation (PCM) waveform. In telephone applications, the PCM waveforms are often called as Line codes. After pulse modulation, each message symbol takes the form of a baseband waveform, $g_i(t)$, where i=1,2…M.

6) Band pass Modulator

For transmission of high speed digital data (eg. Computer communication systems), the digital signal needs to be modulated. The primary purpose of the digital modulator is to map the binary information sequence into high frequency analog signal waveforms (carrier signals).The term band pass is used to indicate that the baseband waveform $g_i(t)$ is frequency translated by a carrier wave to a frequency that is much larger than the spectral content of $g_i(t)$. The digitally modulated signal is a band pass waveform $S_i(t)$, where i=1,2,…..M. The digital modulator may simply map the binary digit 0 into a waveform $S_1(t)$ and the binary digit 1 into a waveform $S_2(t)$. We call this as binary modulation (M=2).

Alternatively, the modulator may transmit K coded information bits at a time by using $M=2^K$ distinct waveforms $S_i(t)$, i=1,2,……..M, one waveform for each of the 2^K possible bit sequences. We call this as M-ary modulation (M>2). The band pass modulator is used for efficient transmission of digital data. The baseband processor block is not required, if the bandpass modulator block is present. Therefore, these two blocks are shown as mutually exclusive blocks.

CHANNEL

The communication channel is the physical medium that is used to send the signal from the transmitter to the receiver. In wireless transmission, the channel may be the atmosphere (free space). On the other hand, telephone channels usually employ a variety of physical media, including wirelines, optical fibre cables, and wireless (microwave radio).

The transmitted signal is corrupted in a random manner by a variety of possible mechanisms, such as additive thermal noise generated by electronic devices, man-made noise, eg., automobile ignition noise and atmospheric noise, eg., electrical lightning discharges during thunderstorms.

As the transmitted signal $S_i(t)$ propagates over the channel, it is impacted by the channel characteristics, which can be described in terms of the channel's impulse response $h_c(t)$. Also, at various points along the signal route, additive random noise n(t) distorts the signal. Hence the received signal x(t) must be termed as the corrupted version of the transmitted signal $S_i(t)$. The received signal x(t) can be expressed as

$$x(t)= S_i(t) * h_c(t) + n(t) \qquad i=1,2.....M$$

where * represents a convolution operation and n(t) represents a noise process.

RECEIVER SECTION

1. Baseband decoder

The baseband decoder block converts back the line coded pulse waveform to transmitted data sequence.

2. Band pass demodulator

The receiver front end and/or the demodulator provides frequency down conversion for each of the received band pass waveform x(t). Digital demodulation is defined as recovery of a waveform (base band pulse). The demodulator restores x(t) to an optimally shaped baseband pulse z(t) in preparation for detection. Detection is defined as decision-making regarding the digital meaning of that waveform.

Typically there are several filters associated with the receiver and demodulator

(i) Filtering to remove unwanted high frequency terms (in the Frequency down conversion of band pass waveforms.)

(ii) Filtering for pulse shaping.

(iii) Filtering option by equalisation to reverse any degrading effects on the signal caused by the poor impulse response of the channel.

Finally the detector transforms the shaped pulse to an estimate of the transmitted data symbols (binary or M-ary).

Demodulator is typically accomplished with the aid of reference waveforms. When the reference used is a measure of the entire signal attributes (particularly phase), the process is termed coherent. When phase information is not used, the process is termed non-coherent.

3. Channel decoder

The estimates of the transmitted data symbols are passed to the channel decoder. The channel decoder attempts to reconstruct the original information sequence from knowledge of the code used by the channel encoder and the redundancy contained in the received data. A measure of how well the demodulator and decoder perform is the frequency with which errors occur in the decoded sequence. This is the important measure of system performance called as Probability of bit error (P_e).

4. Source decoder

The source decoder accepts the output sequence from the channel decoder. From the knowledge of the source encoding method used, it attempts to reconstruct the original signal from the source. Because of channel decoding errors and possible distortion introduced by the source decoder, the signal at the output of the source decoder is an approximation to the original source output. The difference of this estimate and the original digital signal is the distortion introduced by the digital communication system.

5. Deformatter

If the original information source was not in digital data form and the output of the receiver needs to be in the original form of information, a deformatter block is needed. It converts back the digital data to either discrete form (like keyboard characters) or analog form (speech signal).

6. Information sink

If an analog output is needed in non-electrical form, the output transducer converts the estimate of digital signal to the required analog output. The information sink may be computer, data terminal equipment or an user.

7. Synchronization

Synchronization and its key element, a clock signal, is involved in the control of all signal processing within the digital communication system. It actually plays a role in regulating the operation of almost every block. Synchronization involves the estimation of both time and frequency. Coherent systems need to synchronize their frequency reference with the carrier in both frequency and Phase. For non-coherent systems, phase synchronization is not needed.

BASIC DIGITAL COMMUNICATION TRANSFORMATIONS

The basic signal processing functions which may be viewed as transformations can be classified into the following nine groups.

1. Formatting and Source Coding:

Formatting and source coding are similar processes, in that they both involve data digitization. Source coding also involves data compression in addition to digitization.

2. Baseband Signaling

Baseband signaling process involves generation of PCM waveforms or line codes.

3. Bandpass signaling

During demodulation, when the references used are a measure of all the signal attributes (particularly phase), the process is termed coherent. When phase information is not used, the process is termed non coherent.

4. Equalization

An equalization filter is needed for those systems where channel induced ISI (Intersymbol interference) can distort the signals.

5. Channel Coding

Waveform coding and structured sequences are the two methods of channel coding. Waveform coding involves the use of new waveforms. Structured sequences involve the use of redundant bits.

6. Multiplexing and multiple access

Multiplexing and multiple access both involve the idea of resource sharing. Multiplexing takes place locally and multiple access takes place remotely.

7. Spreading

Spreading is used in military applications for achieving interference protection and privacy. Signals can be spread in frequency, in time, or in both frequency and time.

8. Encryption

Encryption and decryption are the basic goals, which are communication privacy and authentication. Maintaining privacy means preventing unauthorized persons from extracting information (eavesdropping) from the channel. Establishing authentication means preventing unauthorized persons from injecting spurious signals (spoofing) into the channel.

9. Synchronization

Synchronization involves the estimation of both time and frequency. Coherent systems need to synchronize their frequency reference with the carrier in both frequency and phase. For non coherent systems, phase synchronization is not needed.

The figure 1.2(b) shows the basic digital communication transformations.

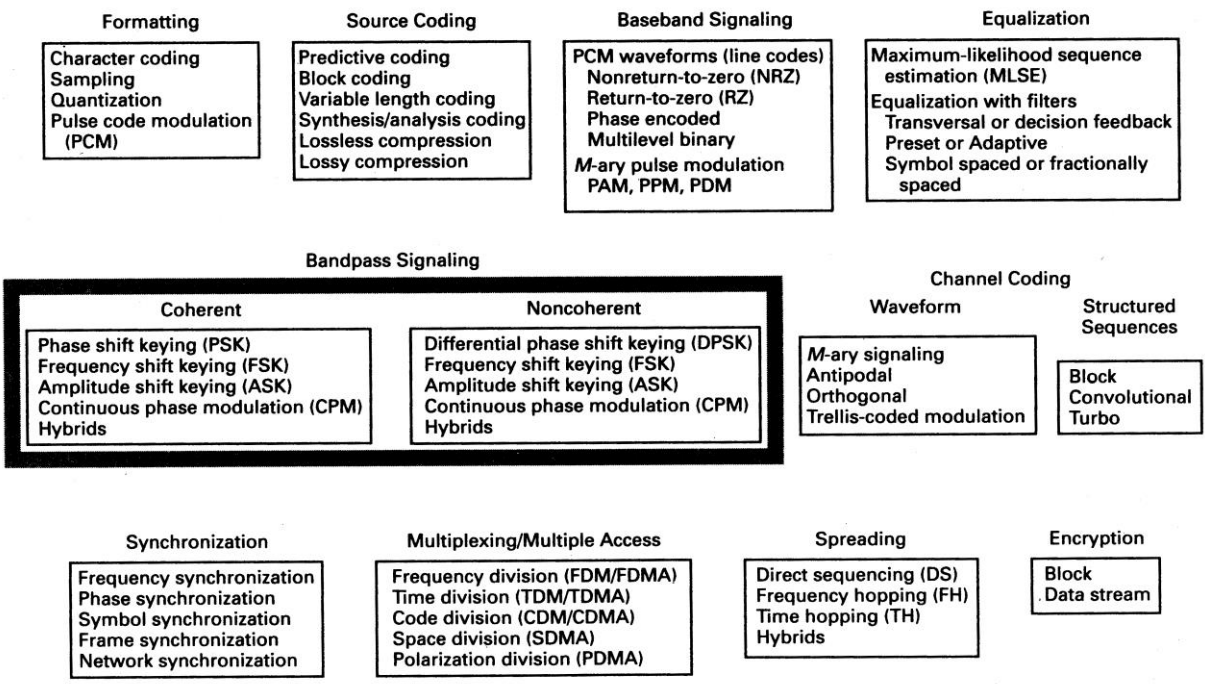


Fig. 1.2(b) Basic digital communication transformations

Performance criteria

A digital communication system transmits signals that represent digits. These digits form a finite set or alphabet, and the set is known a priori to the receiver. A figure of merit for digital communication systems is the probability of incorrectly detecting a digit or the probability of error (P_e).

1.6 CHANNELS FOR DIGITAL COMMUNICATION

The transmission of information across a communication network is accomplished in the physical layer by means of a communication channel. One common problem in signal transmission through any channel is additive noise.

Other types of signal degradation that may occur over the channel are signal attenuation, amplitude and phase distortion, and multipath distortion.

The two channel characteristics, Power and Bandwidth, constitute the primary communication resources available to the designer. The effects of noise may be minimized by increasing the power in the transmitted signal. However, equipment and other practical constraints limit the power level in the transmitted signal. There is limitation for available channel bandwidth also. This is due to the physical limitations of the medium and the electronic components used to implement the transmitter and receiver. These two limitations constrain the amount of data that can be transmitted reliably over any communication channel.

Depending on the mode of transmission used, we may distinguish two basic groups of communication channels.

I. Channels based on guided propagation - Telephone channels, coaxial cables, and optical fibres.

II. Channels based on free space (unguided) propagation – wireless broadcast channels, mobile radio channels, and satellite channels.

Here, we describe some of the important characteristics of three channels used for digital communication – Telephone, optical fibre and satellite channels.

1.6.1 Telephone channel

The telephone network makes extensive use of wirelines for voice signal transmission, as well as voice and data transmission. Earlier the telephone channel is built using twisted wire pairs for signal transmission. Twisted pair wirelines are basically guided electromagnetic channels that provide relatively modest bandwidths. Now, Telephone channel has evolved into a worldwide network that encompasses a variety of transmission media (open-wire lines, coaxial cables, optical fibres, microwave radio, and satellites) and a complex of switching systems. This makes the telephone channel an excellent choice for data communication over long distances.

The telephone channel has a bandpass characteristic occupying the frequency range 300Hz to 3400Hz. It has a high signal-to-noise ratio of about 30dB, and approximately linear response. The channel has a flat amplitude response over the pass-band. But data and image transmissions are strongly influenced by phase delay variations. Hence an equalizer is designed to maintain a flat amplitude response and a linear phase response. Transmission rates up to 16.8 kilobits per seconds (kbps) have been achieved over telephone lines. Telephone channels are naturally susceptible to electromagnetic interference (EMI), the effects of which are mitigated through twisting the wires.

1.6.2 Optical fibre channel

An optical fibre is a dielectric waveguide that transports light signals from one place to another. It consists of a central core within which the light signal is confined. It is surrounded by a cladding layer having a refractive index slightly lower than the core. The core and cladding are both made of pure silica glass. Optical fibre communication system uses a light source (LED or Laser) as the transmitter or modulator. At the receiver, the light intensity is detected by a photodiode, whose output is an electrical signal. Sources of noise in fibre optical channels are photodiodes and electronic amplifiers.

Optical fibres have unique characteristics that make them highly attractive as a transmission medium. They are

- Enormous potential bandwidth, resulting from the use of optical carrier frequencies around 2×10^{14}Hz.
- Low transmission losses, as low as 0.1dB/km.
- Immunity to electromagnetic interference and hence no cross talk.
- Small size and weight.
- Highly reliable photonic devices available for signal generation and signal detection.
- Ruggedness and flexibility.

These unique characteristics have resulted in a rapid deployment of optical fibre channels for telecommunication services including voice, data, facsimile and video.

1.6.3 Satellite channel

A satellite channel consists of a satellite in geostationary orbit, an uplink from a ground station, and a downlink to another ground station. Typically, the uplink and the downlink operate at microwave frequencies with the uplink frequency higher than the downlink frequency. The most popular frequency band for satellite communications is 6GHz for the uplink and 4GHz for the downlink. Another popular band is 14/12GHz. On board the satellite there is a low-power amplifier, which is usually operated in its non-linear mode for high efficiency.

Thus, the satellite channel may be viewed as a powerful repeater in the sky. It permits communication (from one ground station to another) over long distances at high bandwidths and relatively low cost. The non-linear nature of the channel restricts its use to constant envelope modulation techniques (i.e, Phase modulation, frequency modulation).

Communications satellites in geostationary orbit offer the following unique system capabilities:

- Broad area coverage.
- Reliable transmission links
- Wide transmission bandwidths.

In the 6/4GHz band, a typical satellite is assigned a 500MHz bandwidth. This bandwidth is divided among 12 transponders on board the satellite. Each transponder, using approximately 36MHz of the satellite bandwidth, can carry atleast one color TV signal, 1200 voice circuits, or digital data at a rate of 50Mbps.

We may classify communication channels in different ways.

A) A channel may be Linear (e.g., wireless radio channel) or non-linear(e.g., satellite channel).

B) A channel may be time invariant (e.g. Optical Fibre Channel) or time varying (e.g., mobile radio channel).

C) A channel may be bandwidth limited (e.g., telephone channel) or power limited (e.g., optical fibre channel and satellite channel).

1.7 CLASSIFICATION OF SIGNALS

A signal is defined as a single-valued function of time that conveys information. Consequently, for every instant of time there is a unique value of the function. Signals can be categorised in a number of different ways. 1) Deterministic and Random Signals, 2) Periodic and Non-periodic signals, 3) Analog and Discrete Signals and 4) Energy and Power Signals. They are discussed below.

1.7.1 Deterministic and Random signals

A signal is termed as deterministic, if there is no uncertainty with respect to its value at any time. Deterministic signals or waveforms are modeled by explicit mathematical expressions, such as $x(t)= 2 \sin 5t$.

A signal is termed as random, if there is some degree of uncertainty before the signal actually occurs. Random signals or waveforms cannot be modeled by explicit mathematical expressions. But, a random waveform, also referred to as a random process, may exhibit certain regularities, when examined over a long period. Hence it can be described in terms of probabilities and statistical averages. Such a model, in the form of a probabilistic description of the random process is particularly useful for characterising signals and noise in communication systems.

1.7.2 Periodic and Non-Periodic signals

A signal x(t) is called periodic in time if there exists a constant $T_0>0$ such that

$$x(t) = x(t + T_0) \qquad \text{for } -\infty < t < \infty \qquad (1.1)$$

where, t denotes time.

The smallest value of T_0 that satisfies this condition is called the period of x(t). The period T_0 defines the duration of one complete cycle of x(t). The figure 1.3 shows a periodic signal.

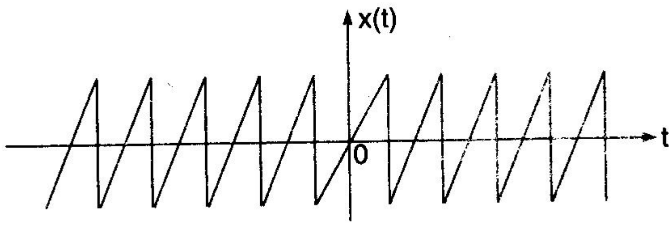


Figure 1.3 A Periodic Signal

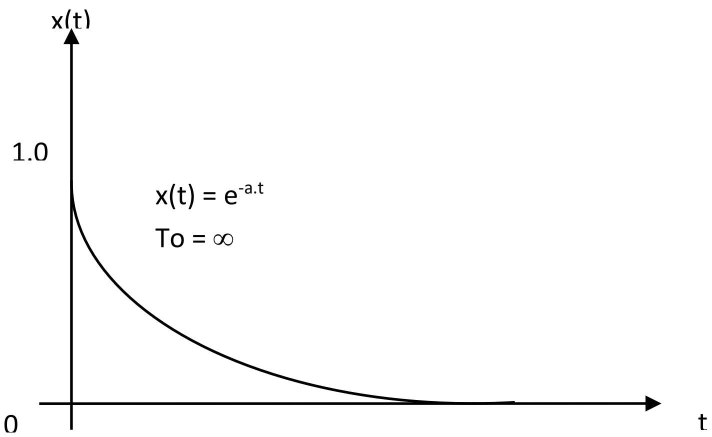


Figure 1.4 Non Periodic Signal

A signal x(t) for which there is no value of T_0 that satisfies the equation,

$$x(t) = x(t + T_0) \qquad \text{for } -\infty < t < \infty \qquad (1.2)$$

is called a non-periodic signal.

The figure 1.4 shows an aperiodic or non-periodic signal.

1.7.3 Analog and Discrete signals

An analog signal x(t) is a continuous function of time; that is, x(t) is uniquely defined for all values of t. When a physical waveform (e.g., speech) is converted into an electrical signal by means of a transducer, then an electrical analog signal is produced. The figure 1.5 shows an analog signal.

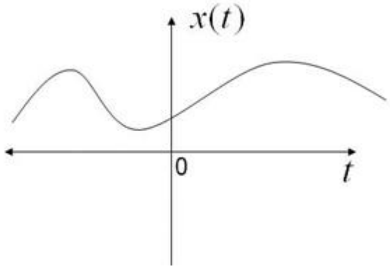


Figure 1.5 Analog Signal

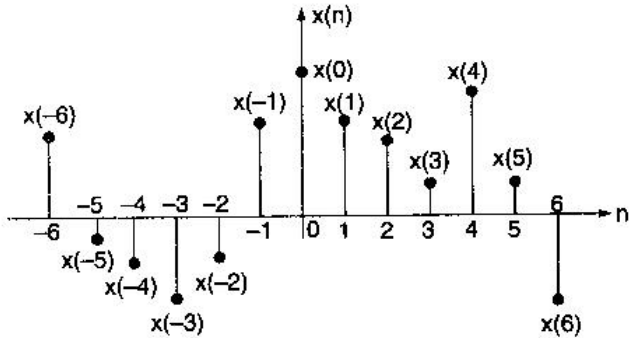


Figure 1.6 Discrete Signal

A discrete signal x(kT) is one that exists only at discrete times. It is characterised by a sequence of numbers defined for each time, kT, where k is an integer and T is a fixed time interval. The figure 1.6 shows a discrete signal.

1.7.4 Energy and Power signals

An electrical signal x(t) can be represented as a voltage v(t) or a current i(t). The instantaneous power p(t) across a resistor R is defined by

$$p(t) = \frac{V^2(t)}{R} \quad \text{(or)} \quad p(t) = i^2(t).R \tag{1.3}$$

In communication systems, power is often normalized by assuming R to be 1Ω. Therefore, regardless of whether the signal is a voltage or current waveform, we can express the instantaneous power as

$$\text{p(t)} = x^2(t) \tag{1.4}$$

The energy dissipated during the time interval ($-{}^{T}/_{2}, {}^{T}/_{2}$) by a real signal with instantaneous power can be written as

$$E_x^T = \int_{-T/2}^{T/2} x^2(t)dt \tag{1.5}$$

The average power dissipated by the signal during the interval is written as

$$P_x^T = \frac{1}{T}E_x^T = \frac{1}{T}\int_{-T/2}^{T/2} x^2(t)dt \tag{1.6}$$

The performance of a communication system depends on the received signal energy. The received energy does the work. Power is the rate at which energy is delivered. Therefore, in analysing communication signals, it is often desirable to deal with the waveform energy.

A signal x(t) is said to be an energy signal if, and only if, it has non zero but finite energy ($0 < E_x < \infty$) for all time, where

$$E_x = \lim_{T\to\infty} \int_{-T/2}^{T/2} x^2(t)dt = \int_{-\infty}^{\infty} x^2(t)dt \qquad (1.7)$$

An energy signal has finite energy but zero average power. Signals that are both deterministic and non-periodic are classified as energy signals.

A signal x(t) is defined as a power signal if, and only if, it has finite but non zero power ($0 < P_x < \infty$) for all time, where

$$P_x = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x^2(t)dt \qquad (1.8)$$

A power signal has finite average power but infinite energy. Periodic signals and random signals are classified as power signals.

1.7.5 The Unit Impulse function

A useful function in communication theory is the unit impulse or Dirac delta function δ(t). The impulse function is defined as an infinitely large amplitude pulse, with zero pulse width, and unity weight (area under the pulse), concentrated at the point where its argument is zero. The unit impulse is characterised by the following relationships:

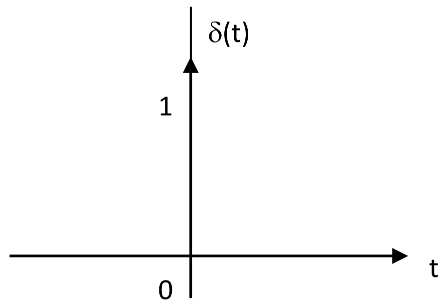


Figure 1.7 The Unit Impulse Function

$$\int_{-\infty}^{\infty} \delta(t)dt = 1 \qquad (1.9)$$

δ(t) = 0 for t≠0 (1.10)

δ(t) is unbounded at t=0 (1.11)

$$\int_{-\infty}^{\infty} x(t)\delta(t-t_o)dt = x(t_o) \qquad (1.12)$$

Equation (1.12) is known as shifting or sampling property of the unit impulse function. The unit impulse multiplier selects a sample of the function x(t) evaluated at $t=t_o$. The figure 1.7 shows an unit impulse function.

1.8 INFORMATION CAPACITY

The information capacity is defined as the maximum rate at which information can be transmitted across the channel without error. It is measured in bits per second.

A key issue in evaluating the performance of a digital communication system is that the maximum rate at which reliable communication can take place over the channel.

1.9 SHANNON'S LIMIT FOR INFORMATION CAPACITY

Shannon-Hartley capacity theorem:

The Shannon's channel capacity theorem defines the fundamental limit on the rate of error free transmission for a power-limited, band-limited Gaussian Channel. The information capacity C of a channel perturbed by additive white Gaussian Noise (AWGN) is a function of the average received signal power S, the average noise power N, and the bandwidth B. The information capacity relationship (Shannon-Hartley theorem) can be stated as

$$C = B\log_2\left(1+\frac{S}{N}\right), \text{ bits/s} \tag{1.13}$$

We can rewrite the noise power as $N=N_oB$, where N_o is the noise power spectral density. Hence, the theorem can be written as

$$C = B\log_2\left(1+\frac{S}{N_oB}\right), \text{ bits/s} \tag{1.14}$$

The significance of the channel capacity is as follows:

(i) If the information rate R from the source is less than or equal to channel capacity C ($R \leq C$), then it is possible to achieve reliable (error-free) transmission through the channel by appropriate coding.

(ii) If the information rate R from the source is greater than the channel capacity C ($R > C$), it is not possible to find a code that can achieve reliable (error-free) transmission through the channel.

Thus, Shannon established basic limits on communication of information and gave birth to a new field that is now called Information Theory.

Example problem 1.1: ***Calculate the capacity of a standard 4kHz telephone channel with a 32dB signal-to-noise ratio.***

Solution:

The standard telephone channels occupy the frequency range of 300Hz to 3400Hz. Hence, the bandwidth is B=3400-300=3100HZ

Signal-to-noise ratio (S/N) in decibels = 32dB

Hence, $10\log_{10}\left(\frac{S}{N}\right) = 32$

$$\log_{10}\left(\frac{S}{N}\right) = \frac{32}{10} = 3.2$$

$$\Rightarrow \frac{S}{N} = \text{anti}\log(3.2) = 1584.89$$

Therefore, $\frac{S}{N} = 1585$

Capacity of a channel, C = $B.\log_2\left(1+\frac{S}{N}\right)$

On substituting the values of B and $\frac{S}{N}$, we have

$$C = 3100 \times \log_2(1+1585)$$

$$= 3100 \times \log_2(1586)$$

$$= 3100 \times \frac{\log_{10} 1586}{\log_{10} 2} = 3100 \times \frac{3.2003}{0.3010}$$

$$= 3100 \times 10.63 = 32953$$

$\Rightarrow$capacity, C = 32953 bits per second

Exercise Problem 1.8.1: A system has bandwidth of 4kHz and a signal-to-noise ratio of 28dB at the input to the receiver. Calculate its information carrying capacity.

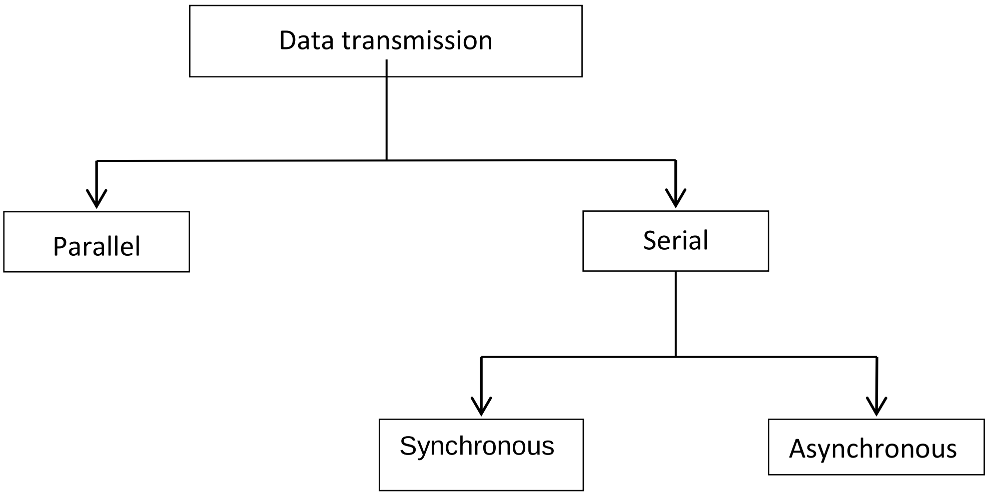
1.10 DATA TRANSMISSION

The data transmission involves the transmission of information such as digitized voice, digitized image and video, computer generated data, and so on. For data transmission and reception, a data network is established. A data network is a structure that enables a data user at one location to have access to some data processing function or service available at another location.

When we enter data into the computer via keyboard, each keyed element is encoded within the keyboard into an equivalent binary coded pattern. For this interchange of information, standard coding schemes are available. The most widely used codes that have been adopted for this function are the American Standard Code for Information Interchange codes (ASCII) (7 bits), and the Extended Binary Coded Decimal Interchange Codes (EBCDIC) (8 bits).

Data transmission refers to the movement of data in the form of bits between two or more digital devices. This transfer of data takes place via some form of transmission media (eg., twisted pair, coaxial cable, fibre optic cable etc.) In data transmission terminology, a station refers to a computer, terminal, telephone, or some other communication device. The channel or link connecting a pair of stations may be simplex, Half-duplex or full-duplex.

Types of Data transmission



Digital data can be transmitted in a number of ways from one station to another as shown in the types of data transmission.

1.11 SERIAL AND PARALLEL TRANSMISSION

1.11.1 Parallel transmission

In parallel transmission, we transfer a word or a byte at a time. Hence, all the bits of data are transmitted simultaneously on separate communication lines. In order to transmit 8 bits, we need 8 wires or lines. Thus each bit has its own line.

This is shown in Figure 1.8 where 8 wires are transmitting a byte (11001010) at a time.

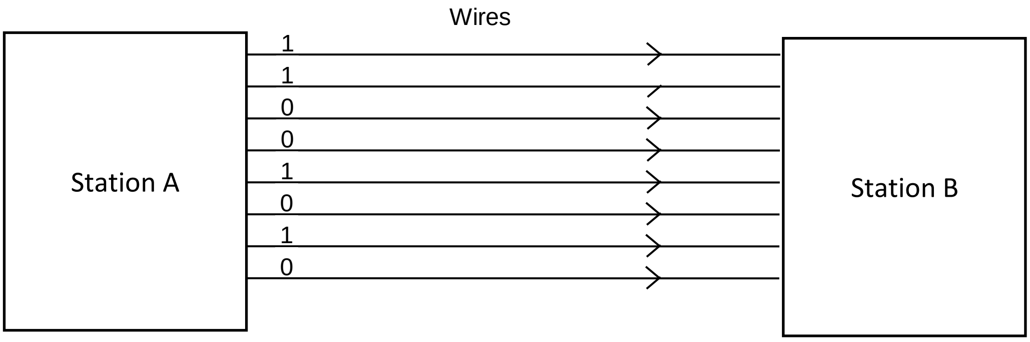


Figure 1.8 Parallel transmission

For efficient transmission of all 8 bits at a time, all the parallel wires have to be absolutely identical. Hence parallel transmission is used only for a very short distance. This method is used for data transmission within the computer system such as from the CPU registers to the memory or vice versa through the data bus. The data bus essentially implements the parallel transmission of data. During byte or word transfer, all 'n' bits of one group are transmitted with each clock pulse from one device to another ie., multiple bits are sent with a single clock pulse. Hence parallel transmission method is very fast.

Advantages of parallel transmission

- Very fast data transfer.
- Easy maintenance, as it is possible to isolate any one line/wire for testing or other purpose.
- Less complexity.

Disadvantages of parallel transmission

- Since separate wire connection is required for transmission of every bit, cost of transmission is higher.
- Parallel transmission is not suitable for long distance communication.

Applications

- Used for data transmission within the computer system such as from processor to memory and vice versa through the data bus.
- Printers use parallel transmission.
- Suitable for short distance communication.

1.11.2 Serial Transmission

In serial transmission, data is transmitted as a single bit at a time using a fixed time interval for each bit. Hence, only a single pair of wires or lines is needed for serial transmission. Since there is only one communication line, the various bits of data are transmitted serially one after the other. For each clock pulse only single bit is transmitted. This is shown in Figure 1.9 where an 8 bit data (11001010) is sent. Here, the LSB bit '0' will be transmitted first followed by other bits in a serial manner. The MSB bit '1' will be transmitted in the end.

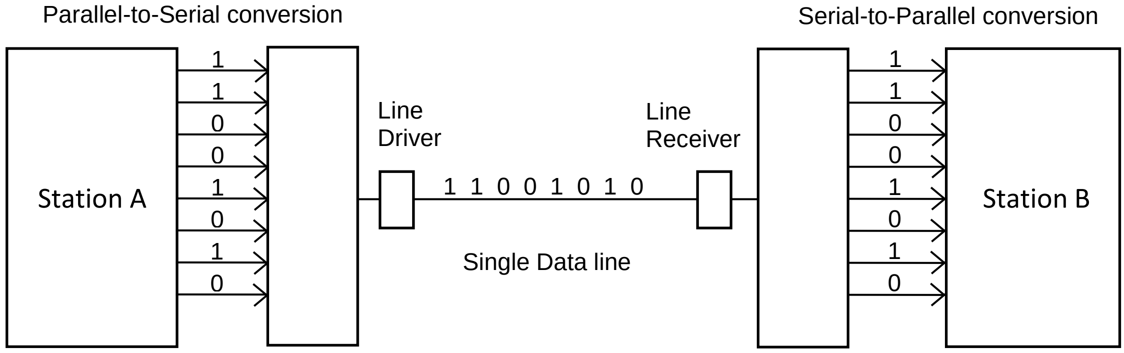


Figure 1.9 Serial Transmission

The conversion devices convert the parallel data into serial data at the sender side so that it can be transmitted over single line. On receiver side, serial data received is again converted to parallel form so that the internal circuitry of computer can accept the data. All the bits are collected, measured and put together as bytes in memory. Serial transmission is very convenient for long distance communication, since single data line is required. In an asynchronous communication system, the conversion between parallel and serial form is usually performed by an integrated circuit called a Universal Asynchronous Receiver-Transmitter (UART).

Advantages of serial transmission

- Single data line reduces the cost of the system.
- Most popular for long distance communication.

Disadvantages of serial transmission

- System becomes complex due to the use of conversion devices at sender and receiver.
- As bits are transmitted serially one after the other, this method is slower compared to parallel transmission.
- For each clock pulse, only single bit is transmitted.

Applications

- Serial transmission is used in most of the long distance communications.

1.11.3 Comparison

Table 1.1 shows the comparison between serial and parallel transmission.

Table 1.1: Comparison between Serial and Parallel transmission

Factor	Serial transmission	Parallel transmission
1) Number of bits transmitted at one clock pulse.	One bit	‘n’ bits
2) No of lines/wires required to transmit ‘n’ bits.	One wire	‘n’ wires
3) Speed of data transfer	Slow	Fast
4) Cost of transmission	Low as one line is required	Higher as ‘n’ lines are required.
5) Application	Long distance communication between two computers	Short distance communication like computer to printer.

1.12 Synchronous and Asynchronous transmission

There are two types of Serial transmission. They are designated as Asynchronous or Synchronous depending on how the timing and framing information is transmitted. Bit synchronization is a function that is required to determine when the beginning and end of the data transmission occurs.

1.12.1 Asynchronous transmission

In asynchronous transmission, the transmitter sends data with its own timing clock, which is not known to the receiver.

Asynchronous transmission sends only one character at a time. The character may be a letter of the alphabet or number or control character. It sends one byte of data at a time.

Bit synchronization between two devices is made possible using start bit and stop bit, as shown in the Figure 1.10

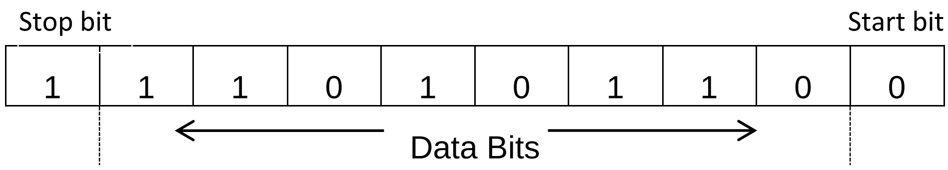


Figure 1.10 Data bytes

In an asynchronous system, the transmit and receive clocks are free-funning and are set to approximately the same speed. A 'Start' bit is transmitted at the beginning of each character. Start bit alerts the receiver to the arrival of new group of bits. A start bit usually '0' is added to the beginning of each byte.

A 'stop' bit usually '1' is sent at the end of the character. Stop bit indicates the end of data ie., to let the receiver know that byte is finished. One or more additional '1' bits are appended to the end of the byte.

In asynchronous transmission, the framing is set by the start bit. The timing remains sufficiently accurate throughout the limited duration of the character. There is idle time between the transmissions of different data bytes. This idle time is also known as Gap.

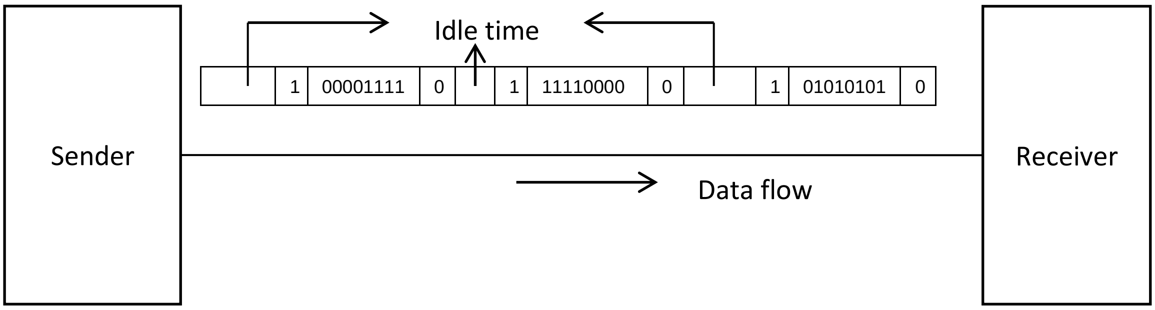


Figure 1.11 Asynchronous transmission

The gap or idle time can be of varying intervals, as shown in the Figure 1.11. This mechanism is called Asynchronous, because at byte level sender and receiver need not to be synchronized. But within each byte, receiver must be synchronized with the incoming bit stream.

Advantages of Asynchronous transmission

- The transmission can start as soon as data byte to be transmitted becomes available.
- It is possible to transmit signals from sources having different bit rates.
- This mode of data transmission is easy to implement.
- This method of data transmission is cheaper in cost. Eg. If lines are short, asynchronous transmission is better, because line cost would be low and idle time will not be expensive.

Disadvantages of asynchronous transmission

- This method is less efficient and slower than synchronous transmission due to the overhead of extra bits (start and stop) and insertion of gaps into bit stream.
- Successful transmission depends on the recognition of the start bits. These bits can be missed or corrupted.

Applications

- Asynchronous transmission is well suited for keyboard type-terminals and paper tape devices, because it does not require any local storage at the terminal or the computer.
- Asynchronous transmission is best suited to Internet traffic in which information is transmitted in short bursts. This type of transmission is used by modems.
- Used for communication between microcomputers.

1.12.2 Synchronous transmission

In synchronous transmission the transmitter and receiver are synchronized to the same clock frequency. This mode of transmission is more efficient than asynchronous because start and stop bits are not necessary. Here, data is sent in blocks that may contain multiple bytes. There is no gap or idle time between the various bytes in the data stream, as shown in the Figure 1.12.

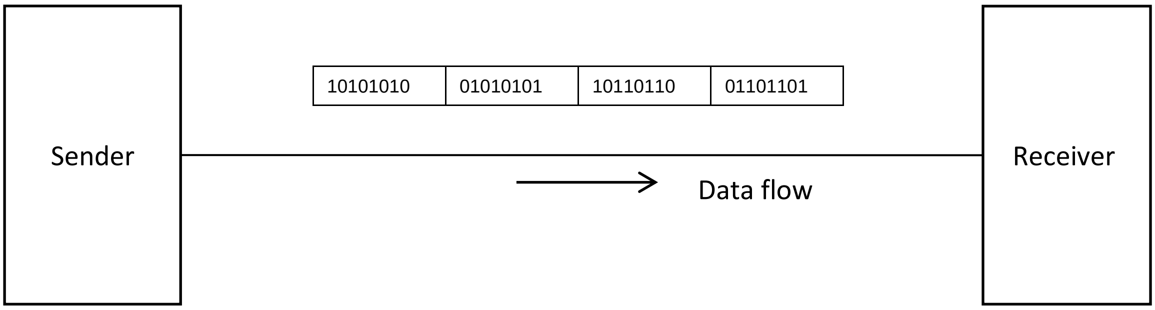


Figure 1.12 Synchronous Transmission

Bit synchronization is established between sender and receiver by “timing” the transmission of each bit. Since the various bytes are placed on the link without any gap, it is the responsibility of receiver to separate the bit stream into bytes so as to reconstruct the original information. The longer message blocks used in synchronous communication make it necessary to lock the sender and receiver clocks together exactly.

Framing

In synchronous communication, it is only necessary to locate the beginning of a block of data. As long as the transmitter and receiver clocks remain synchronized, the receiver can keep track of the remaining data without the need for stop and start bits. Thus synchronous systems can be much more efficient than asynchronous systems.

Synchronous Data-Link Protocols

Synchronous communication protocols are either character oriented or bit oriented. Character oriented protocol example is BISYNC, an IBM product. These protocols begin a block with at least two synchronizing (SYN) characters, followed by control and data characters, as shown in the Figure 1.13

SYN	SYN	SOH	Header	STX	DATA	ETX	BCC

SOH - Start of Header

STX - Start of Text

ETX - End of Text

BCC - Block Check Character

Figure 1.13 Character oriented Synchronous Protocol

A bit pattern of 00101101 may be reserved for the SYN character for indicating the start of the block, ie., for synchronization. The Block check character BCC is used for error control.

Examples of bit oriented protocols include High-level data link control (HDLC), an ISO standard, and synchronous data link control (SDLC), an IBM product. Each block of data is preceded by an eight-bit pattern called a flag, which signals the start of a frame. The flag consists of bit pattern 01111110. The data block can have either a fixed or variable length. Figure 1.14 illustrates a data block in HDLC.

Flag	Address Field	Control Field	Information Field (Data)	FCS	Flag

FCS- Frame Check Sequence

Figure 1.14 Bit oriented Synchronous protocol

After the data, there are frame check bits (FCS) that can be used for error control. If another frame follows immediately, the flag that ends one frame starts the next frame.

Advantages of Synchronous transmission

- Synchronous transmission is faster than asynchronous transmission.
- It is more efficient than asynchronous because start and stop bits are not necessary.
- There is no gap or idle time between the various bytes in the data stream.
- It is suitable for high speed communication between computers.

Disadvantages of synchronous transmission

- It has greater complexity in both hardware and software.
- It requires accurately synchronised clocks at both the sender and receiver. The system requires proper synchronization.
- It requires local buffer storage at the two ends of line to assemble data blocks.
- It is costly compared to asynchronous method.

Applications:

- Synchronous transmission systems are higher-speed communication typical of mainframe computers.
- These are also used for transmitting the digitized analog signals such as are found in the telephone system.

1.12.3 Comparison

Table 1.2 shows the comparison between asynchronous and synchronous transmission

Table 1.2: Comparison between Asynchronous and synchronous transmissions

Factor	Asynchronous	Synchronous
1. Data sent at one time.	Usually one byte	Multiple bytes
2. Start and stop bits.	Required	Not required
3. Gap between data units.	Present	Not present
4. Data transmission speed.	Slow	Fast
5. Cost.	Low	High
6. System complexity.	simple	Complex

SHORT QUESTIONS AND ANSWERS

1. What is Digital Communication?

In a digital communication system, the information bearing digital signal is processed such that it can be represented by a sequence of binary digits (discrete messages). Then it is used for ON/OFF keying of some characteristic of a high frequency sinusoidal carrier wave, such as amplitude, phase or frequency.

If the input message signal is in analog form, then it is converted to digital form by the processes of sampling, quantizing and encoding.

2. What is the principle feature of a digital communication system?

The principle feature of a digital communication system is that during a finite interval of time, it sends a signal waveform from a finite set of possible waveforms. The receiver has to determine from a received noise-perturbed signal which waveform from the finite set of waveforms was sent by the transmitter.

3. What is meant by Regeneration?

Regeneration is one of the corrective signal processing method. In Regeneration, the received pulse is amplified by a digital amplifier that recovers its original ideal shape. The pulse is thus reborn or regenerated. Circuits that perform this function at regular intervals along a transmission system are called Regenerative repeaters.

4. Mention some of the advantages of digital communication.

- Very powerful encryption and decryption algorithms are available for digital data so as to maintain a high level of secrecy of communication.
- With digital techniques, extremely low error rates producing high signal fidelity are possible through error detection and correction.
- The use of ‘Regenerative repeaters’ generate strong error free signal at a good power level.
- Digital circuits are less subject to distortion and interference than analog circuits.
- Signal processing and image processing operations like compression of voice and image signals can be easily carried out.

5. Mention some of the disadvantages of digital communication

- Digital communication systems generally need more bandwidth than analog communication systems.
- Digital systems need to allocate a significant share of their resources to the task of synchronization at various levels.
- When the signal-to-noise ratio drops below a certain threshold, the quality of service can change suddenly from very good to very poor.

6. List the available channels for digital communication.

I. Wireline channels using a physical media

1. Twisted wire pair (STP, UTP)
2. Coaxial Cable
3. Optical Fibre cable

II. Wireless channels using free space

1. Microwave radio channel
2. Satellite channel
3. Wireless broadcast channel
4. Wireless mobile channel

7. What is a signal? Write the classification of signals

A signal is defined as a single-valued function of time that conveys information. The signals are classified as

1. Deterministic and Random signals
2. Periodic and Non-periodic signals
3. Analog and Discrete signals
4. Energy and Power signals
5. The Unit Impulse Function

8. Define deterministic and random signals.

A signal is termed as deterministic, if there is no uncertainty with respect to its value at any time.

A signal is termed as random, if there is some degree of uncertainty before the signal actually occurs.

9. Define periodic and non-periodic signals.

A signal x(t) is called periodic in time if there exists a constant $T_o > 0$ such that

$$x(t) = x(t+T_o) \quad \text{for} \quad -\infty \leq t \leq \infty$$

where t denotes time.

A signal x(t) for which there is no value of T_o that satisfies the equation

$$x(t) = x(t+T_o) \quad \text{for} \quad -\infty \le t \le \infty$$

is called a non-periodic signal

10. Define analog and discrete signals

An analog signal x(t) is a continuous function of time; that is x(t) is uniquely defined for all values of t.

A discrete signal x(kT) is one that exists only at discrete times. It is characterised by a sequence of numbers defined for each time, kT, where k is an integer and t is a fixed time interval.

11. Define energy and power signal

A signal x(t) is said to be an energy signal if, and only if, it has non zero but finite energy ($0 < E_x < \infty$) for all time.

A signal x(t) is defined as a power signal if, and only if, it has finite but non zero power ($0 < p_x < \infty$) for all time.

12. Define Unit Impulse Function

The impulse function is defined as an infinitely large amplitude pulse, with zero pulse width, and unity weight (area under the pulse), concentrated at the point where its argument is zero. The unit impulse is characterised by the following relationships.

$$\int_{-\infty}^{\infty} \delta(t)dt = 1, \qquad \delta(t) = 0 \quad \text{for} \quad t \neq 0$$

13. Define information capacity. What is the Shannon's limit for information capacity

The information capacity is defined as the maximum rate at which information can be transmitted across the channel without error. The Shannon's channel capacity theorem defines the fundamental limit on the rate of error free transmission. It can be stated as

$$C = B \log_2\left(1 + \frac{S}{N}\right), \text{ bits/s}$$

where C→ Channel capacity

B → Bandwidth

$\frac{S}{N}$→ Signal to noise power ratio

14. What is meant by data transmission? What are its types?

The data transmission involves the transmission of information such as digitized voice, digitized image and video, computer generated data, and so on. There are two types of data transmission. They are 1) Parallel transmission and 2) Serial transmission. There are two methods in serial transmission. They are 1) Synchronous transmission and 2) Asynchronous transmission.

15. What is parallel transmission? What are its merits?

In parallel transmission, we transfer a word or a byte a time. Hence, all the bits of data are transmitted simultaneously on separate communication lines. The merits of parallel transmission are

1. Fast data transmission
2. Less complexity
3. Easy to isolate any data line

16. What is serial transmission? What are its merits?

In serial transmission, data is transmitted as a single bit at a time using a fixed time interval for each bit. Hence, only a single pair of wires or lines is needed for serial transmission. The merits of serial transmission are

1. Most popular for long distance communication.
2. Single data line reduces the cost of the system.

17. What are the applications of parallel and serial transmission?

Parallel transmission:

1. Printers use parallel transmission.
2. Used for data transmission within the computer system such as from processor to memory and vice versa through the data bus.

Serial transmission:

1. It is used in most of the long distance communications

18. Differentiate asynchronous and synchronous transmissions

In asynchronous transmission, the transmitter sends data with its own timing clock, which is not known to the receiver. It sends one byte of data at a time. In synchronous transmission, the transmitter and receiver are synchronized to the same clock frequency. Here, data is sent in blocks that may contain multiple bytes.

Unit - II FORMATTING AND BASEBAND MODULATION

OBJECTIVES

- To understand the baseband system.
- To learn about sampling and its types.
- To learn about quantization.
- To learn about PCM waveform types.

2.1 INTRODUCTION

Formatting is the first essential signal processing step in Digital communication. The purpose of formatting is to insure that the message or source signal is compatible with digital processing. Transmit formatting is a transformation from source information to digital symbols. When data compression in addition to formatting is employed, the process is termed source coding.

The digital messages are considered to be in the logical format of binary ones and zeros. These messages are transformed by baseband processor (pulse modulation) into baseband waveforms. Such waveforms can then be transmitted over a cable.

2.2 BASEBAND SYSTEMS

The functional diagram of formatting and transmission of baseband signals (Baseband Systems) is shown in the Figure 2.1

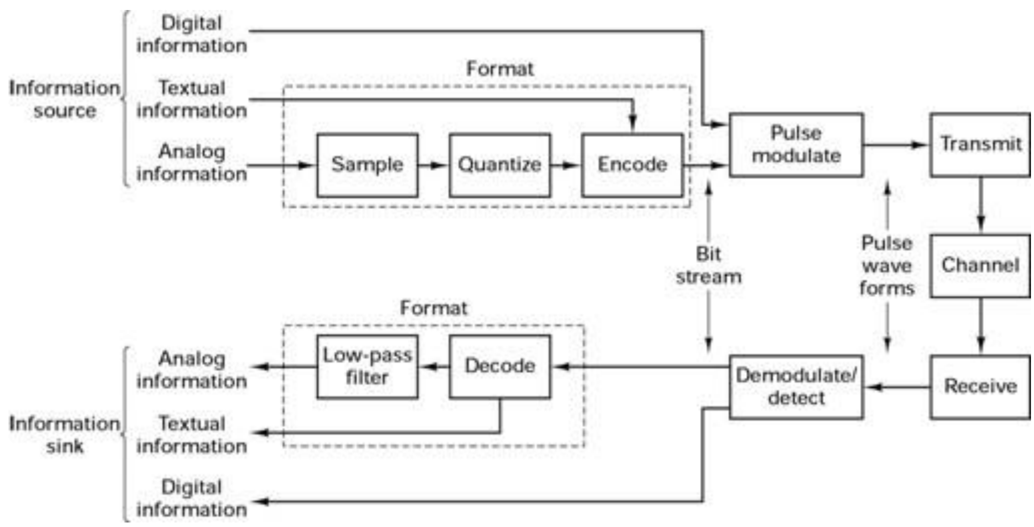


Figure 2.1 Formatting and Transmission of baseband signals (Baseband Systems)

Formatting transforms the source information into bits, thus assuring compatibility between the information and the signal processing steps within the digital communication system. The information remains in the form of a bit stream upto the pulse modulation block.

Information sources can be analog or discrete. Hence, the output of an information source may be digital information, textual information or an analog information. Data already in a digital format would bypass the formatting function. Textual information is transformed into binary digits by the use of coder. If the data is in the form of alphanumeric text, then it will be character encoded with one of several standard formats such as ASCII, EBCDIC, Baudot, and Hollerith.

Analog information is formatted using three separate processes: Sampling, quantization and coding. For all types of information sources, the formatting step results in a sequence of binary digits.

The pulse modulator converts the bit stream into a sequence of pulse waveforms. The characteristics of this sequence of pulses correspond to the digits being sent. These pulse waveforms are then transmitted through a baseband channel, such as pair of wires or a coaxial cable.

After transmission through the channel, the pulse waveforms, are recovered (demodulated) and detected to produce an estimate of the transmitted digits. The final step is the reverse formatting, which recovers an estimate of the source information.

2.3 FORMATTING ANALOG INFORMATION

Analog information sources can be transformed into digital sources through the use of sampling and quantization. We utilize sampling to convert a continuous time signal to a discrete time signal, process the discrete-time signal using a discrete time system and then convert back to continuous-time signals.

2.3.1 The Sampling theorem

Sampling of the signals is the fundamental operation in signal-processing. A continuous-time signal is first converted to discrete-time signal by sampling process. Sampling theorem gives the complete idea about the sampling of signals. The output of the sampling process is called pulse amplitude modulation (PAM). Because the successive output intervals can be described as a sequence of pulses with amplitudes derived from the input waveform samples. The analog waveform can be approximately retrieved from a PAM waveform by simple low-pass filtering.

The statement of sampling theorem can be given in two parts as below:

(i) A band limited signal of finite energy, which has no frequency components higher than f_m Hertz, is completely described by its sample values at uniform intervals less than or equal to $\frac{1}{2f_m}$ seconds apart.

(ii) A band limited signal of finite energy, which has no frequency components higher than f_m Hertz, may be completely recovered from the knowledge of its samples taken at the rate of $2f_m$ samples per second.

Combining the two parts, the uniform sampling theorem may be stated as follows:

"A continuous-time signal may be completely represented in its samples and recovered back if the sampling frequency is $f_s \geq 2f_m$".

Here f_s is the sampling frequency and f_m is the maximum frequency present in the signal.

2.3.2 Nyquist theorem

The Nyquist theorem provides a prescription for the nominal sampling interval required to avoid aliasing. It may be stated as follows:

"The sampling frequency (f_s) must be at the rate equal to or greater than twice the highest frequency component (f_m) present in the signal ie.,$f_s \geq 2f_m$, in order to recover the signal exactly."

(i) When the sampling rate becomes exactly equal to $2f_m$ samples per second, then it is called as Nyquist rate. Nyquist rate is also defined as the minimum sampling rate. It is given by

$$f_s = 2f_m \tag{2.1}$$

(ii) Similarly, maximum sampling interval is called as Nyquist interval. It is given by

$$\text{Nyquist interval, } T_s = \frac{1}{2f_m} \text{ seconds,} \tag{2.2}$$

where, $f_s = \frac{1}{T_s}$

(iii) The restriction of $f_s \geq 2f_m$, stated in terms of the sampling rate, is known as the Nyquist criterion. The Nyquist criterion is a theoretically sufficient condition to allow an analog signal to be reconstructed completely from a set of uniformly spaced discrete time samples.

2.4 SAMPLING TECHNIQUES

The sampling of a continuous-time signal is done in several ways. Basically, there are three types of sampling techniques. They are:

1. Impulse sampling
2. Natural sampling
3. Flat top sampling (Sample and hold operation)

2.4.1 Impulse sampling or Ideal sampling

If the sampling function is a train of impulses, then the method is called Impulse sampling or Ideal sampling. Figure 2.2(c) shows this sampling function. Figure 2.2(g) shows a circuit to produce this sampling. This circuit is known as the switching sampler.

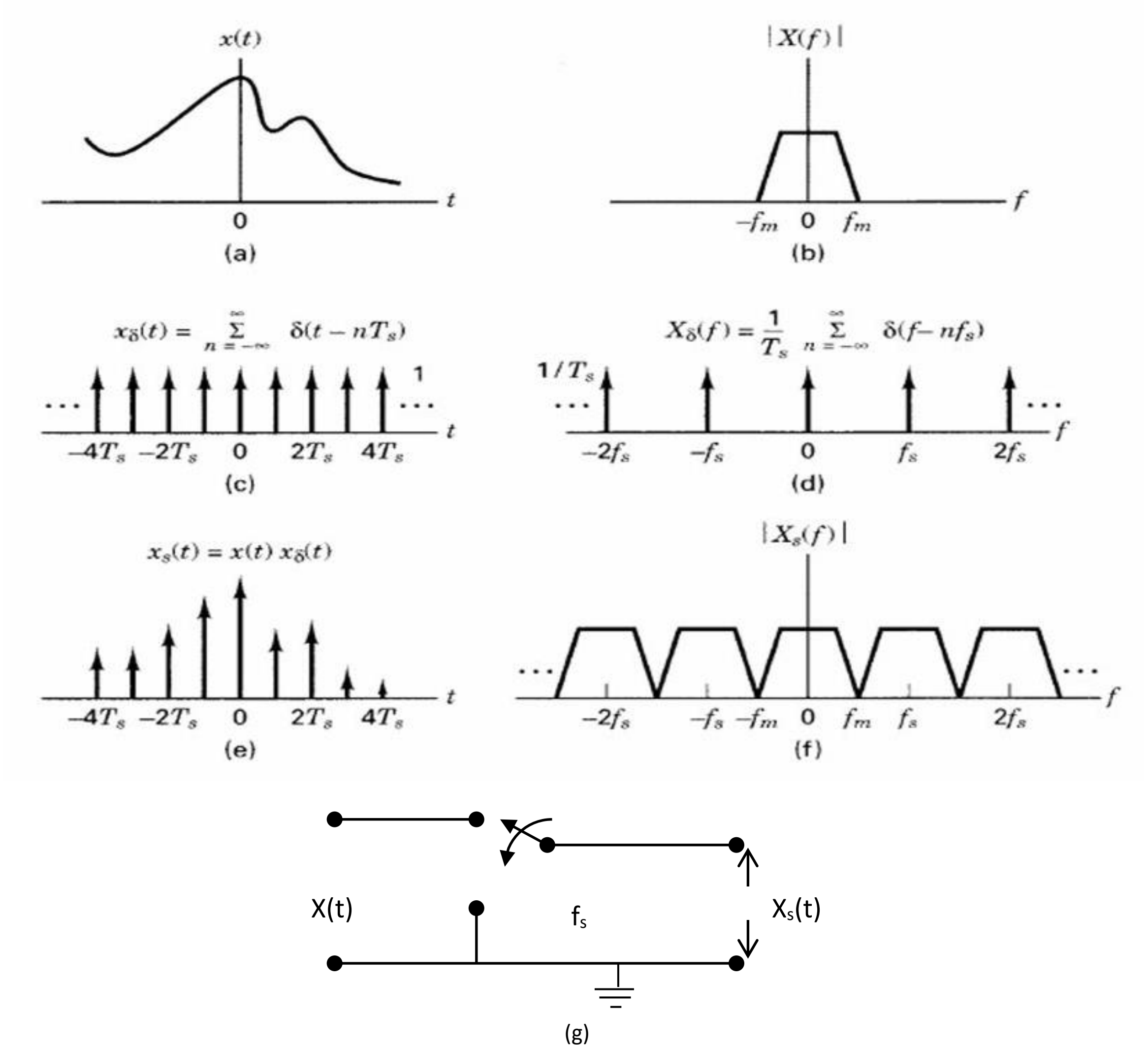


Figure 2.2 Impulse Sampling

Let us choose $T_s=\frac{1}{2f_m}$, so that the Nyquist criterion is just satisfied. The circuit simply consists of a switch. If we assume that the closing time 't' of the switch approaches zero, then the output $x_s(t)$ will contain only instantaneous value of the input signal x(t). This instantaneous sampling gives a train of impulses of height equal to the instantaneous value of the input signal x(t) at the sampling instant.

The train of impulses (sampling function) may be represented as

$$x_\delta(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \quad (2.3)$$

Where T_s is the sampling period and $\delta(t)$ is the unit impulse or Dirac delta function. The sampled signal $x_s(t)$ is expressed as the multiplication of x(t) and $x_\delta(t)$.

Thus, $$x_s(t) = x(t).x_\delta(t) \quad (2.4)$$

$$x_s(t) = x(t).\sum_{n=-\infty}^{\infty} \delta(t - nT_s) \quad (2.5)$$

$$\Rightarrow x_s(t) = \sum_{n=-\infty}^{\infty} x(nT_s).\,\delta(t - nT_s) \quad (2.6)$$

Demerits of Impulse sampling:

Impulse sampling results in the samples whose width T approaches zero. Due to this, the power content in the instantaneously sampled pulse is negligible. Thus, this method is not suitable for transmission purpose.

Spectrum:

The spectrum $X_\delta(f)$ of the sampled signal $x_s(t)$ is shown in the figure 2.2(f) for $f_s=2f_m$. Using the frequency convolution property of the Fourier transform, we can transform the time domain product $x(t).x_\delta(t)$ of equation(2.6) to the frequency domain convolution $X(f)*X_\delta(f)$.

Therefore, $$X_s(f) = X(f) * X_\delta(f)$$

$$= X(f) *\left[\frac{1}{T_s}\sum_{n=-\infty}^{\infty} \delta(f - nf_s)\right]$$

$$\Rightarrow X_s(f) = \frac{1}{T_s}\sum_{n=-\infty}^{\infty} X(f - nf_s) \quad (2.7)$$

(i) From the figure, we infer that, if the sampling rate is chosen such that $f_s = 2f_m$, then each spectral replicate is separated from each of its neighbours by a frequency band exactly equal to f_s Hertz. Therefore, the analog waveform can theoretically be completely recovered from the samples, by the use of filtering.

(ii) If the sampling rate is chosen such that $f_s > 2f_m$, the spectral replications will move farther apart in frequency, as shown in Figure 2.3(a),

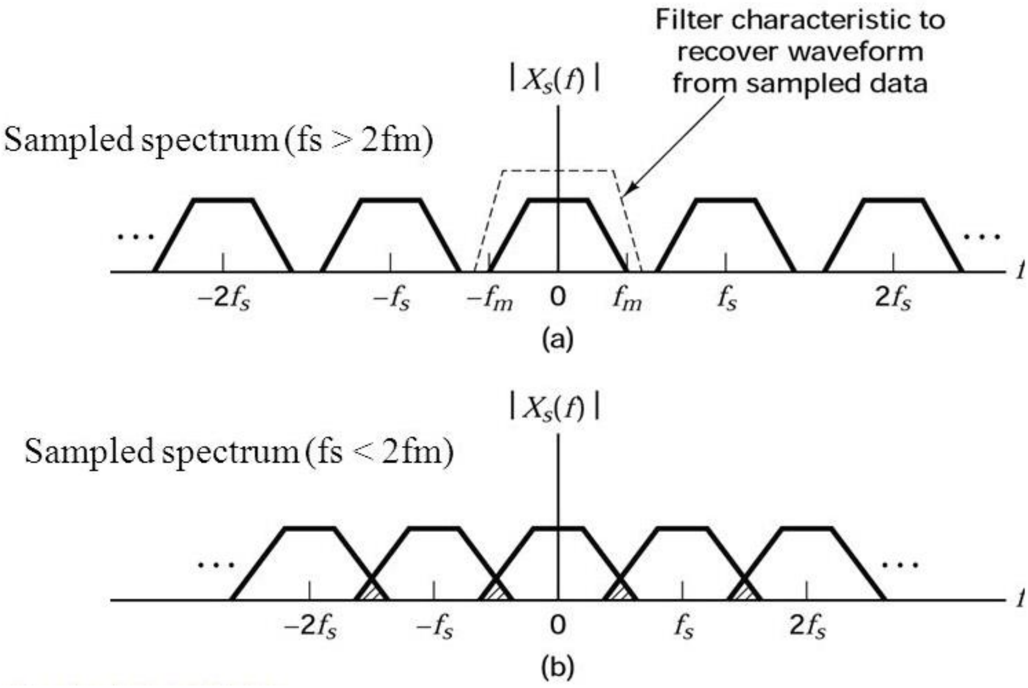


Figure 2.3 Spectra for various sampling rates

Now, it is easier to perform the filtering operation. A typical low-pass filter characteristic that might be used to separate the baseband spectrum from those at higher frequencies is also shown in the figure.

(iii) When the sampling rate is reduced, such that $f_s < 2f_m$, the spectral replications will overlap, as shown in the figure 2.3(b). Therefore, some information will be lost. This phenomenon is called "aliasing" which results from under sampling (sampling at too low a rate).

Conclusion

- The Nyquist rate, $f_s = 2f_m$ is the sampling rate below which aliasing occurs.
- To avoid aliasing, the Nyquist criterion, $f_s \geq 2f_m$ must be satisfied.

2.4.2 Natural Sampling

Natural sampling is a practical method. Here the sampling function is a pulse train or switching waveform $x_p(t)$. Figure 2.4(c) shows this sampling function. Figure 2.4(g) shows a functional diagram of a natural sampler.

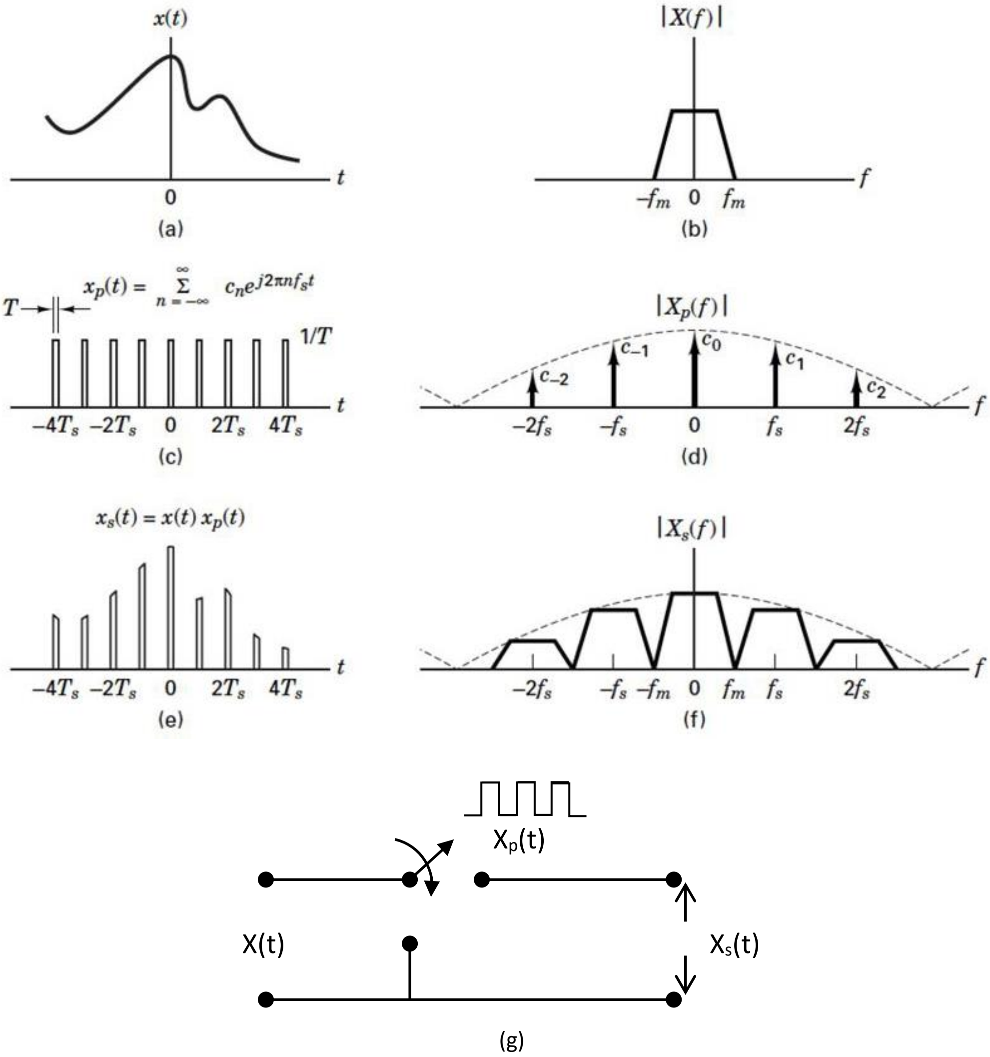


Figure 2.4 Natural Sampling

Here also, we choose $T_s=\frac{1}{2f_m}$, so that the Nyquist criterion is just satisfied. The circuit simply consists of a switch. The pulse train $x_p(t)$ is applied to the switch. Each pulse in $x_p(t)$ has width T and amplitude $\frac{1}{T}$. The multiplication of input analog signal x(t) by the pulse train $x_p(t)$ can be viewed as the opening and closing of the switch. The resulting sampled data sequence, $x_s(t)$ is shown in figure 2.4(e). It can be represented as

$$x_s(t) \quad = \quad x(t).x_p(t) \qquad (2.8)$$

This process is called natural sampling, since the top of each pulse in the sampled data sequence retains the shape of its corresponding analog segment during the pulse interval. We can express the periodic pulse train as a Fourier series in the form

$x_p(t) \quad = \quad \sum_{n=-\infty}^{\infty} C_n e^{j2\pi n f_s t}$ (2.9)

where $C_n \quad = \quad \frac{1}{T_s}\,\mathrm{sinc}\left(\frac{nT}{T_s}\right)$, T is the pulse width and $\frac{1}{T}$ is the pulse amplitude.

Hence, the sampled data sequence is given by

$x_s(t) \quad = \quad x(t)\sum_{n=-\infty}^{\infty} C_n e^{j2\pi n f_s t}$ (2.10)

Disadvantages

Each pulse in the sampled data sequence has varying top according to signal variation. During transmission, noise interferes the top of pulses. Then it becomes difficult to determine the shape of top of the pulse at the receiver.

Spectrum

The spectrum of the naturally sampled signal is shown in Figure 2.4(f). The transform $X_s(f)$ of the sampled waveform is found as follows:

$X_s(f) \quad = \quad \mathcal{F}\{x(t).\sum_{n=-\infty}^{\infty} C_n\, e^{j2\pi n f_s t}\}$ (2.11)

We can solve for $X_s(f)$ as below,

$X_s(f) \quad = \quad \sum_{n=-\infty}^{\infty} C_n\, X(f - nf_s)$ (2.12)

Equation (2.12) and Figure 2.4(f) illustrate that $X_s(f)$ is a replication of X(f), periodically repeated in frequency every f_s Hertz. However, we see that $X_s(f)$ is weighted by the Fourier series coefficients (C_n) of the pulse train, compared with a constant value in the impulse sampling.

2.4.3 Flat Top Sampling or Sample-and-Hold operation

In case of natural sampling, the pulse has varying top according to the signal variation. Therefore, amplitude detection of the pulse is not exact and errors are introduced in the signal. This problem will be solved by having flat top pulses. A sample and hold circuit is used to generate flat top pulses.

Figure 2.5 shows the functional diagram of a sample and hold circuit. The circuit consists of two field effect transistor (FET) switches and a capacitor.

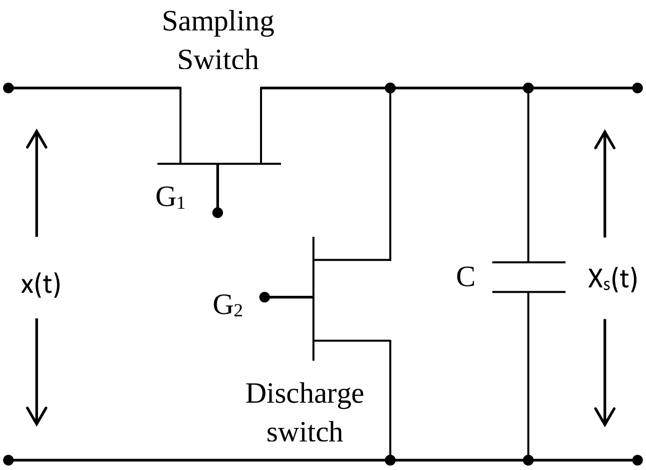


Figure 2.5 Sample and hold circuit for flat top sampling

By applying a short pulse to the gate G_1, the sampling switch is closed for a very small period. During this period, the capacitor 'C' is quickly charged upto a voltage equal to the instantaneous sample value of the incoming signal x(t). The sampling switch is now opened and the capacitor holds the charge. The discharge switch is then closed by a pulse applied to the gate G_2 to discharge capacitor to zero volts. The discharge switch is then opened and thus capacitor has no voltage. After the period of T_s, sampling switch is closed to take new sample. This periodic gating of sample and hold circuit generates a sequence of flat top samples as shown in the figure 2.6.

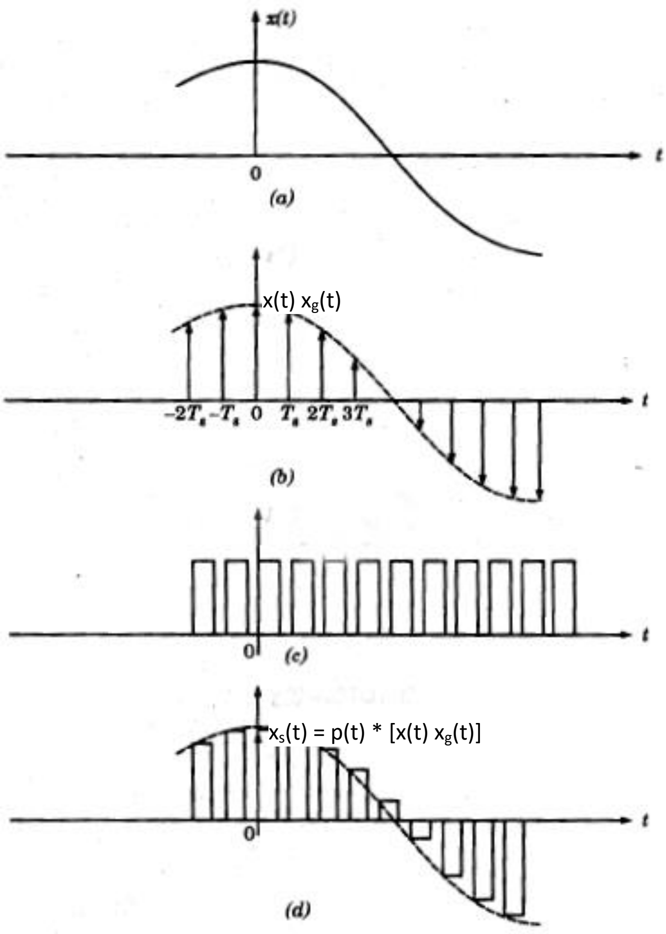


Figure 2.6 Flat Top Sampling

It may be noted that only starting edge of the pulse represents instantaneous value of the baseband signal x(t). Sample and hold can be described by the convolution of the sampled pulse train, $[x(t)x_\delta(t)]$, with a unity amplitude rectangular pulse P(t) of pulse width T_s. Hence, convolution results in the flat top sampled sequence.

$$x_s(t) = P(t) * [x(t)x_\delta(t)]$$

$$= P(t) * [x(t)\sum_{n=-\infty}^{\infty} \delta(t - nT_s)] \quad (2.13)$$

Spectrum

The Fourier transform, $x_s(f)$, of the time convolution in equation (2.13) is the frequency-domain product of the transform P(f) of the rectangular pulse and the periodic spectrum of the impulse-sampled data. Therefore,

$$X_s(f) = P(f)\, \mathcal{F}\{x(t) \sum_{n=-\infty}^{\infty} \delta(t - nT_s)\}$$

$$= P(f)\left\{X(f) * \left[\frac{1}{T_s}\sum_{n=-\infty}^{\infty} \delta(f - nf_s)\right]\right\}$$

$$\Rightarrow X_s(f) = P(f).\frac{1}{T_s}\sum_{n=-\infty}^{\infty} X(f - nf_s) \quad (2.14)$$

The effect of this product operation results in a spectrum similar in appearance to the naturally sampled signal shown in the figure 2.4(f). The main effect of the hold operation is the significant attenuation of the higher frequency spectral replicates, which is a desired effect.

Example 2.1: A continuous-time signal is given as x(t) = 8cos 200πt. Determine the minimum sampling rate ie., Nyquist rate required to avoid aliasing.

Solution:

The continuous time signal,

$$x(t) = 8\cos 200\pi t$$

We have,

$$x(t) = A\cos(2\pi f)t = A\cos\omega t$$

$$\Rightarrow A\cos(2\pi f)t = 8\cos 200\pi t$$

On comparing the values,

$$2f = 200$$

$$\Rightarrow f = \frac{200}{2} = 100Hz$$

Hence, the highest frequency component of the given continuous time signal is f_m=100Hz. Therefore, minimum sampling rate required to avoid aliasing is the Nyguist rate given by

$$f_s = 2f_m = 2 \times 100 = 200Hz.$$

2.5 ALIASING

When a continuous-time band-limited signal is sampled at a rate lower than Nyquist rate, $f_s<2f_m$, it is termed as undersampling. The spectrum of the sampled signal is shown in the figure 2.7 and 2.8.

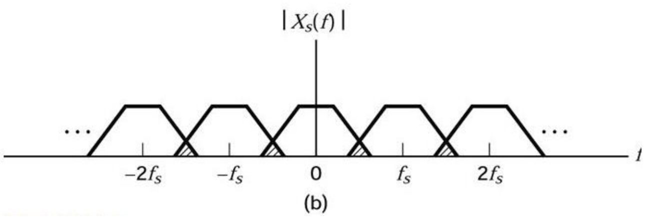


Figure 2.7 Sampled Spectrum for fs< 2 fm

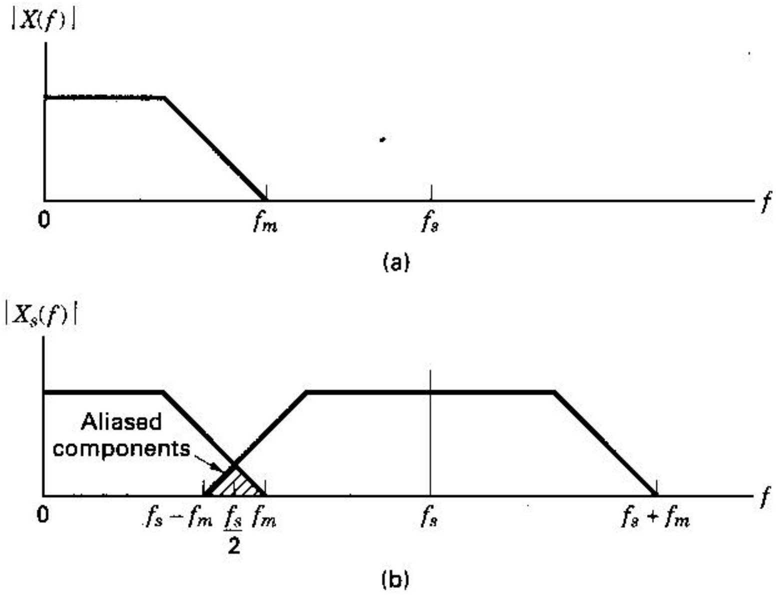


Figure 2.8 Aliasing in the frequency domain

Aliasing is a phenomenon which results from the effect of undersampling. The aliased spectral components represent ambiguous data that appear in the frequency band between (f_s-f_m) and f_m. Aliasing may be defined as the phenomenon in which a high frequency component in the frequency-spectrum of the signal takes identity of a lower-frequency component in the spectrum of the sampled signal.

Let the frequencies above half the sampling frequency ($>\frac{f_s}{2}$) be called as the fold over frequencies. From the figure, we see some overlapping in the periodic replications. This overlapping of successive periods of the spectrum causes the fold over frequencies in the original signal to appear as frequencies below half the sampling frequency ($<\frac{f_s}{2}$), in the sampled signal. This will cause distortion in the reconstructed signal. This phenomenon is called Aliasing.

Methods to prevent aliasing:

There are two ways of eliminating aliasing using antialiasing filters.

(i) The analog signal is prefiltered using a low pass filter. The bandwidth of the filter is less than or equal to half the sampling frequency (ie.,$f'_m \leq \frac{f_s}{2}$). Thus, there are no aliased components seen in the sampled signal spectrum as shown in the Figure 2.9.

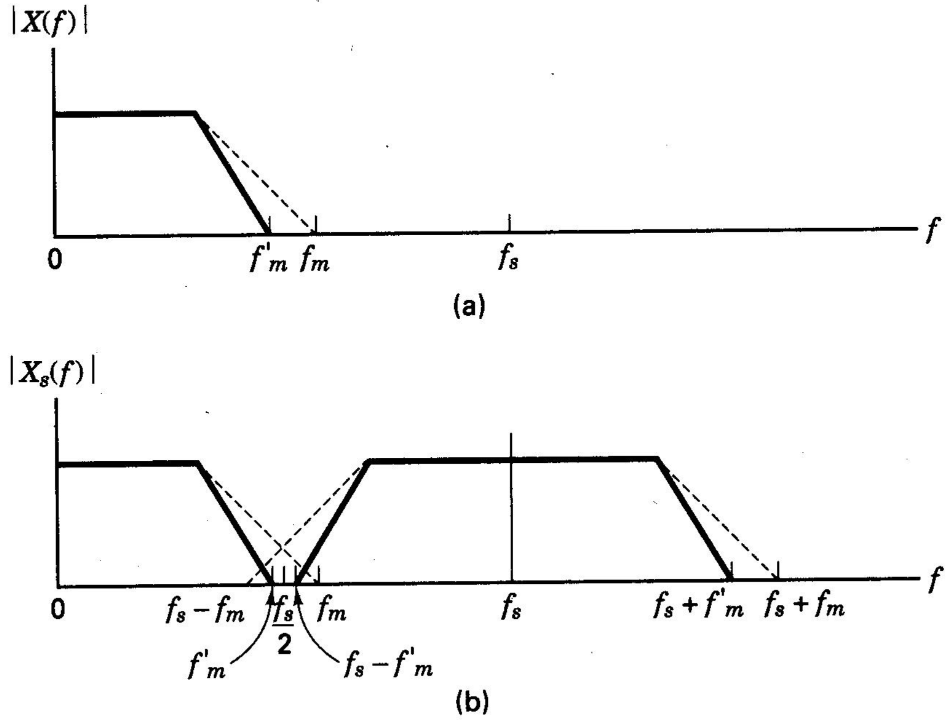


Figure 2.9 Prefiltering

Eliminating the aliasing terms prior to sampling is good engineering practice.

(ii) When the signal structure is well known, the aliased terms can be eliminated after sampling. Here, the low pass filter operates on the sampled data. The figure 2.10 shows how the aliased components are removed by post filtering after sampling.

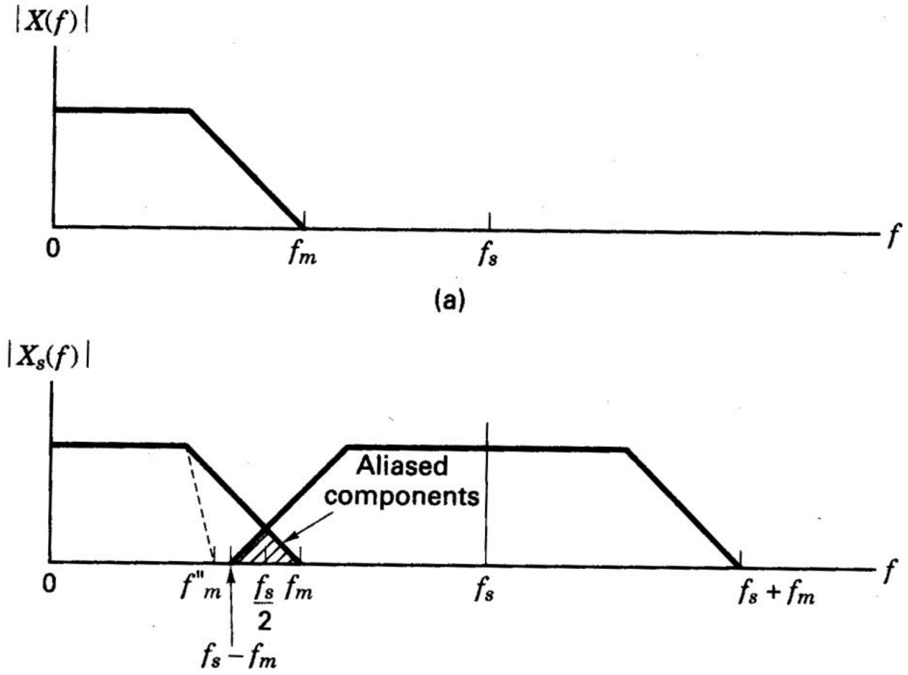


Figure 2.10 Post Filtering

Table 2.1: Performance comparison of three sampling techniques

Parameter	Ideal sampling	Natural Sampling	Flat Top Sampling
1) Sampling Principle	It uses multiplication.	It uses chopping principle.	It uses sample and hold circuit.
2) Generation circuit	Figure 2.2g	Figure 2.4g	Figure 2.5
3) Sampling rate	Sampling rate tends to be infinite.	Sampling rate satisfies Nyquist criteria.	Sampling rate satisfies Nyquist criteria.
4) Noise interference	Noise interference is maximum.	Noise interference is minimum.	Noise interference is maximum
5) Feasibility	This is not a practically possible method.	This method can be used practically.	This method is used practically.

2.6 SIGNAL INTERFACE FOR A DIGITAL SYSTEM

We know that a digital system deals with a finite number of values. An analog signal, such as voice, has a continuous range of amplitudes. When it is sampled, the samples also cover a continuous amplitude range. Because, within the finite amplitude range of the signal we find an infinite number of amplitude levels. Hence, the sampled data are not compatible with a digital system.

Figure 2.11 illustrates four ways in which analog source information can be described.

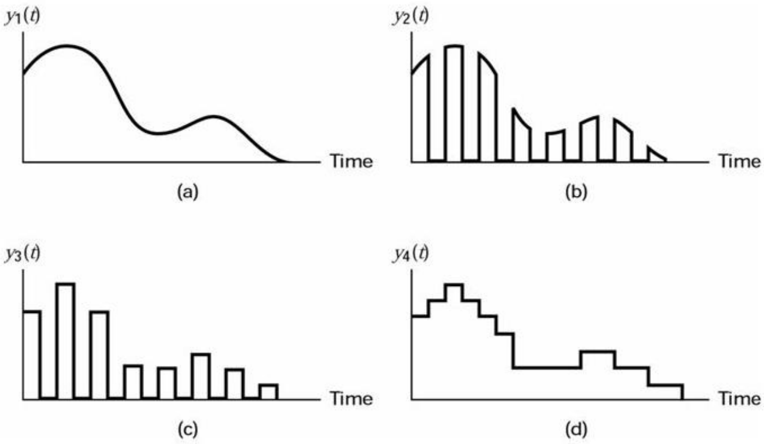


Figure 2.11 Source Data

(i) Figure 2.11(a) shows the original analog waveform.

(ii) Figure 2.11(b) represents natural-sampled version of the original analog waveform. Here, the amplitude of each natural sample still has an infinite number of possible values. Hence, it is not compatible with a digital system.

(iii) Figure 2.11(c) illustrates the original waveform represented by discrete pulses. Here the pulses have flat tops and the pulse amplitude values are limited to a finite set. Each pulse is expressed as a level from a finite number of predetermined levels. These pulses are referred to as quantized samples. Such a format is the best choice for interfacing with a digital system.

(iv) The format in Figure 2.11(d) may be viewed as the output of a sample and hold circuit. When the sample values are quantized to a finite set, this format can also interface with a digital system.

Therefore, the existence of a finite number of discrete amplitude levels is a basic condition for interfacing with a digital system. The conversion of analog (continuous) sample of the signal into a digital (discrete) form is called the quantizing process.

2.7 QUANTIZATION

The process of representing a large (possibly infinite) set of values with a much smaller set of values is called quantization.

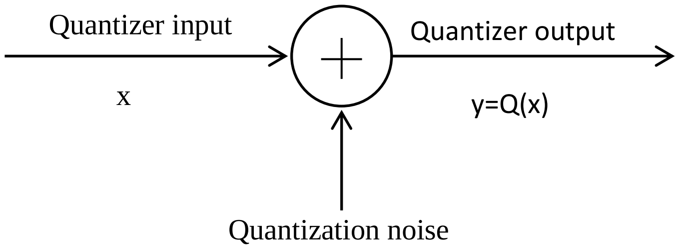


Figure 2.12 Quantizer

In a linear analog system, the transfer characteristic representing the relation between the input and the output is a straight line. For a quantizer, the transfer characteristic is staircase like in appearance

The quantizing process has a two-fold effect:

1) The peak-to-peak range of input sample values is subdivided into a finite set of decision levels or decision threshold that are aligned with the “risers” of the staircase, and
2) The output is assigned a discrete value selected from a finite set of representation levels or reconstruction values that are aligned with the “treads” of the staircase.

The combination of sampler and quantizer is called Analog-to-Digital(A/D) converter or digitizer.

2.8 SOURCES OF CORRUPTION

The analog signal recovered from the sampled, quantized, and transmitted pulses will contain corruption from several sources. The sources of corruption are related to

1) Sampling and quantizing effects (Quantization noise, Quantizer saturation and Timing jitter)
2) Channel effects (channel noise and Intersymbol interference)

2.8.1 Sampling and quantizing effects

2.8.1.1 Quantization noise

The distortion inherent in quantization is a round-off or truncation error. The sample values of an analog baseband signal are rounded-off to the nearest permissible representation levels of the quantizer. This rounding-off or approximation involves discarding some of the original analog information. The distortion introduced by the need to approximate the analog waveform with quantized samples, is referred to as quantization noise. The amount of such noise is inversely proportional to the number of levels employed in the quantization process.

2.8.1.2 Quantizer saturation

The quantizer (or analog-to-digital converter) allocates L levels to the task of approximating the continuous range of inputs with a finite set of outputs. The range of inputs for which the difference between the input and output is small is called the operating range of the converter.

If the input exceeds this range, the difference between the input and the output becomes large. At this condition, the converter is operating in saturation. Generally, saturation is avoided by the use of Automatic Gain Control (AGC), which effectively extends the operating range of the converter.

2.8.1.3 Timing jitter

We know that the samples of the analog signal are uniformly spaced. If there is a slight jitter in the position of the sample, the sampling is no longer uniform. The jitter is usually a random process and thus the sample positions are not accurately known.

The effect of the jitter is equivalent to frequency modulation (FM) of the baseband signal.

(i) If the jitter is random, a low-level wideband spectral contribution is induced. The properties are very close to those of quantizing noise.

(ii) If the jitter exhibits periodic components, the periodic FM will induce low-level spectral lines in the data.

Timing jitter can be controlled with very good power supply isolation and stable clock references.

2.8.2 Channel effects

2.8.2.1 Channel noise

Channel noise may be introduced anywhere along the transmission path. The channel noise is the combined effect of thermal noise, interference from other users, and interference from circuit switching transients. Channel noise may be modeled as

Additive White Gaussian Noise (AWGN) with zero mean and Power spectral density $\frac{N_o}{2}$.

The effect of channel noise is to introduce errors in detecting the pulses carrying the digitized samples. Channel-induced errors can degrade the reconstructed signal quality quite quickly. This rapid degradation of output signal quality with channel induced errors is called a threshold effect.

If the channel noise is small, it does not corrupt the reconstruct signals. The only noise present in the reconstruction is the quantization noise. On the other hand, if the channel noise is large enough to affect our ability to detect the waveforms, then there will be reconstruction errors. A large difference in behaviour can occur for very small changes in channel noise level.

2.8.2.2 Intersymbol interference

The channel is always band limited. A band limited channel disperses or spreads a pulse waveform passing through it.

(i) When the channel bandwidth is much greater than the pulse band width, the spreading of the pulse will be slight.

(ii) When the channel bandwidth is close to the signal bandwidth, the spreading will exceed symbol duration and cause signal pulses to overlap. This overlapping is called Intersymbol interference (ISI).

ISI causes system degradation (higher error rates). We may use an adaptive equaliser to correct the channel induced degradations. Also, if we transmit a sinc pulse instead of a rectangular pulse, then the ISI can be reduced to zero. This is known as Nyquist Pulse Shaping.

2.9 Pulse Code Modulation (PCM)

Pulse Code Modulation (PCM) refers to the class of baseband signals obtained from the quantized PAM signals by encoding each quantized sample into a digital word. The essential operations in the transmitter of a PCM system are sampling, quantizing and encoding as shown in the Figure 2.13.

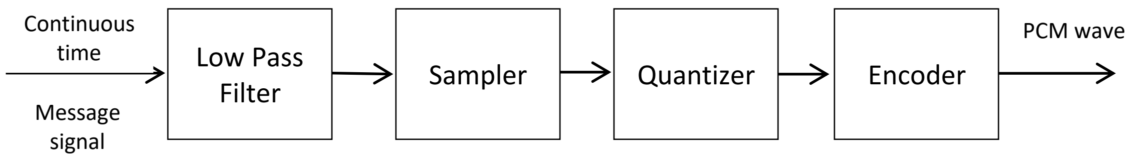


Figure 2.13 PCM transmitter

The source information is sampled and quantized to one of 'L' levels. Then each quantized sample is digitally encoded into an l-bit codeword, where $l = \log_2 L$. For baseband transmission, the codeword bits will then be transformed to pulse waveforms. The essential features of binary PCM are shown in the Figure 2.14.

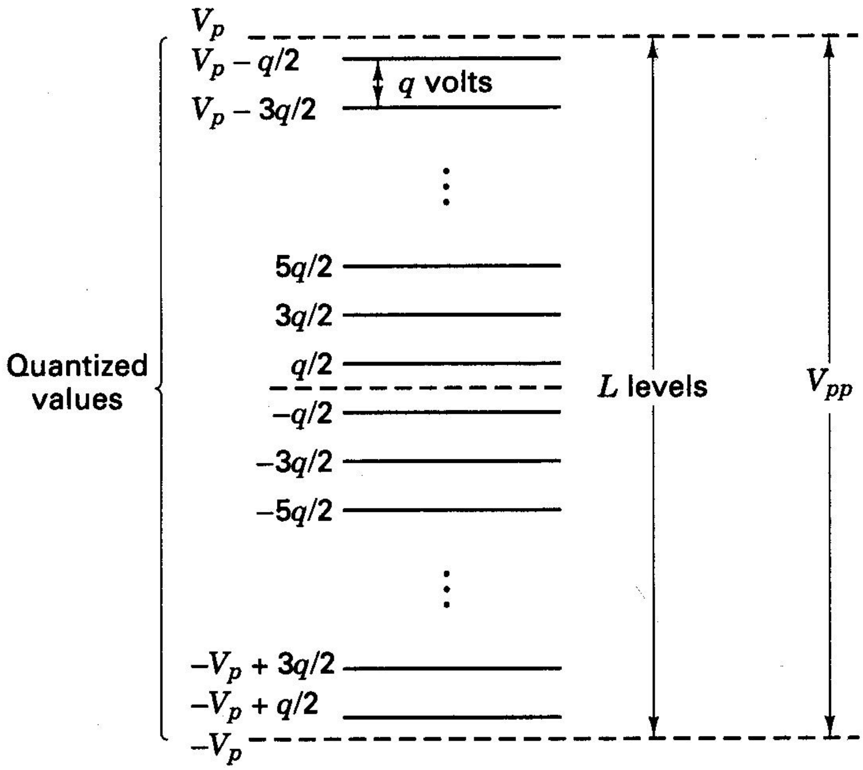


Figure 2.14 (a) Quantization Levels

The Figure 2.14(a) illustrates an L-level linear quantizer for an analog signal with a peak-to peak voltage range of $V_{pp} = V_p - (-V_p) = 2\ V_p$ volts. The quantized pulses assume positive and negative values. The stepsize between quantization levels, called the quantile interval, is denoted by q volts. When the quantization levels are uniformly distributed over the full range, the quantizer is called a uniform or linear quantizer. Each sample value of the analog waveform is approximated with a quantized pulse. The degradation of the signal due to quantization is therefore limited to half a quanitle interval, $\pm\frac{q}{2}$volts.

Figure 2.14(b) shows an analog signal x(t) limited in its excursions to the range -4 to +4V. The stepsize between quantization levels has been set at 1V. Thus, eight quantization levels are employed. These are located at -3.5, -2.5,……+3.5V. Assign the code number 0 to the level at -3.5V, code number 1 to the level at -2.5V, and so on, until the level at 3.5V, which is assigned the code number 7.

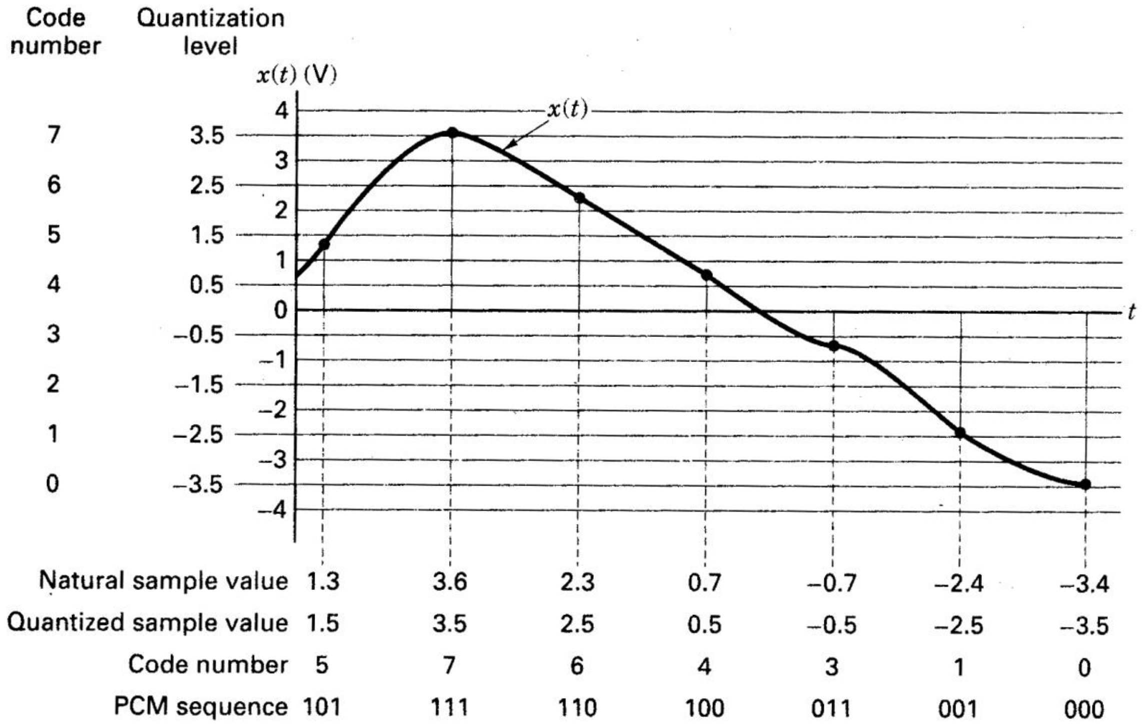


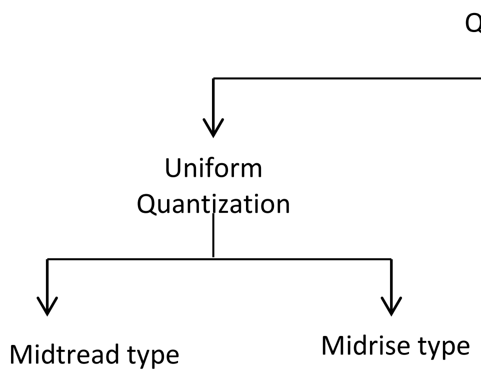
Figure 2.14 (b) Pulse Code Modulation

Each code number has its representation in binary arithmetic, ranging from 000 for code number 0 to 111 for code number 7. The quantile intervals between the levels should be equal. The ordinate in Figure 2.14(b) is labeled with quantization levels and their code numbers. Each sample of the analog signal is assigned to the quantization level closest to the value of the sample. There are four representations of x(t) as follows: the natural sample values, the quantized sample values, the code numbers, and the PCM sequence.

Here, each sample is assigned to one of eight levels or a three-bit PCM sequence. Increasing the number of levels will reduce the quantization noise. If we double the number of levels to 16, each analog sample will be represented as a four-bit PCM sequence. But when there are more bits per sample, the data rate is increased, and the cost is a greater transmission bandwidth. Thus, we can obtain better fidelity at the cost of more transmission bandwidth.

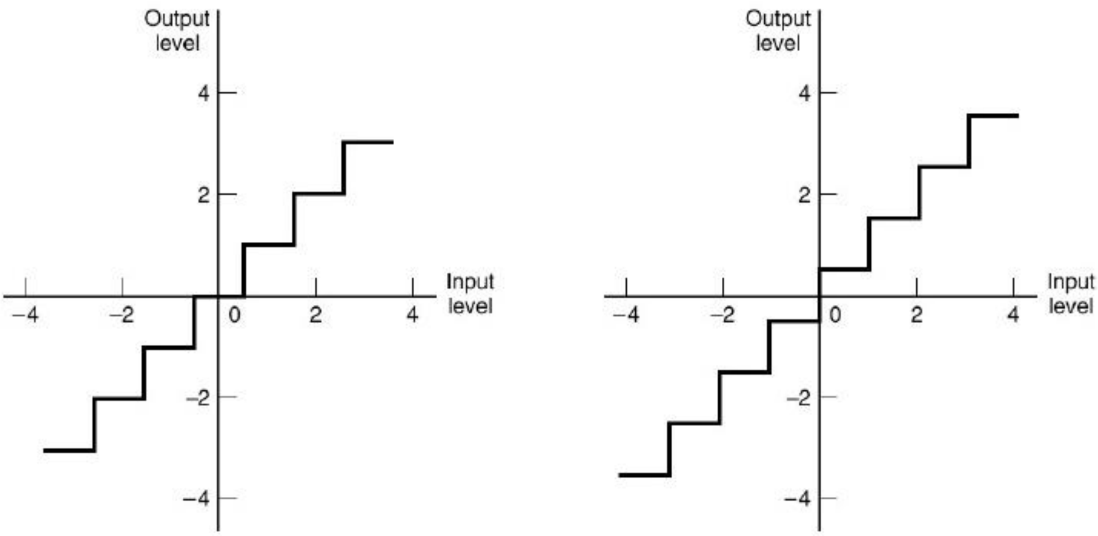
2.10 UNIFORM AND NON-UNIFORM QUANTIZATION

In pulse code modulation both the parameters time and amplitude are expressed in discrete form. The sampling process converts the continuous time values of the analog signal into discrete time values. The quantization process converts the continuous amplitude values into a finite (discrete) set of allowable values. This process is called “discretization” in time and amplitude. Here, we shall study about the quantization process. Basically, quantization process may be classified as follows:



2.10.1 Uniform quantization

When the quantization levels are uniformly distributed over the full amplitude range of the input signal, the quantizer is called an uniform or linear quantizer. In uniform quantization, the stepsize between quantization levels remains the same throughout the input range. The quantizer characteristic can also be midtread or midrise type, as shown in the Figure 2.15.



a) Midtread Type **b) Midrise Type**

Figure 2.15 Two types of Uniform Quantization

(a) For the uniform quantizer of midtread type, the origin lies in the middle of a tread of the staircase like graph.

(b) For the Uniform quantizer of midrise type, the origin lies in the middle of a rising part of the staircase like graph.

Both the midtread and midrise types of uniform quantizers are symmertric about the origin. Hence they are also called as symmetric quantizer.

2.10.2 Non-uniform quantization

If the quantizer characteristic is nonlinear, then the quantization is known as non-uniform quantization. In non-uniform quantization, the step size is not constant. The step size is variable, depending on the amplitude of input signal.

2.10.2.1 Companding

The non-uniform quantization is practically achieved through a process called companding. Figure 2.16 shows a companding model. The compressor amplifies weak signals and attenuates strong signals.

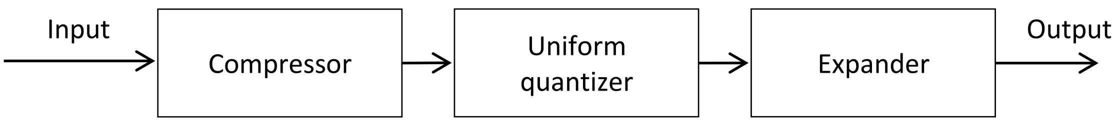
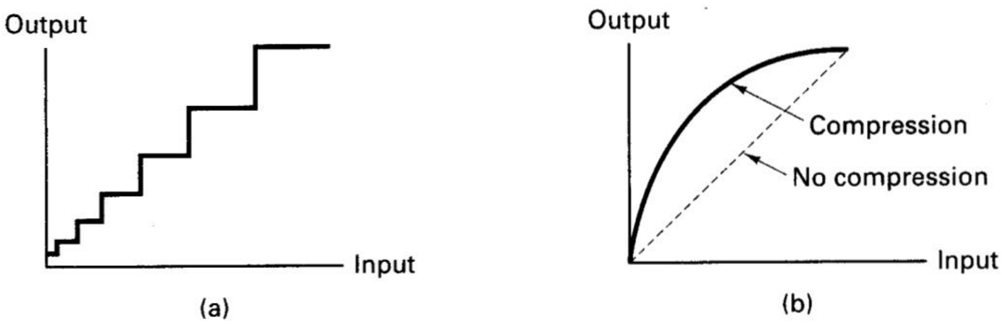


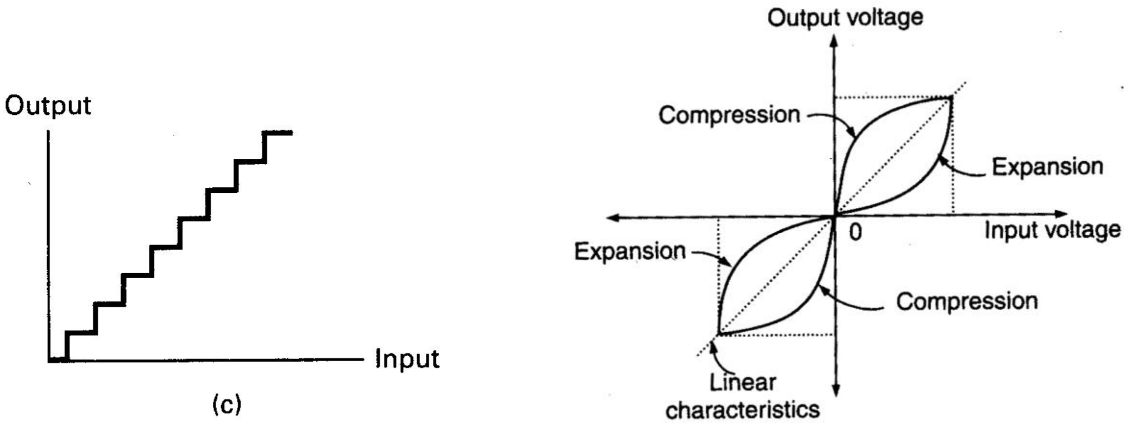
Figure 2.16 A companding model

This process is called compression. At the receiver, the expander does the opposite function of compression. Thus the expander provides expansion. Therefore, the compression of the signal at the transmitter and the expansion at the receiver is combined to be called as companding.

Companding = Compressing + Expanding

The non-uniform quantizer characteristic is shown in the figure 2.17





(c)

(d)

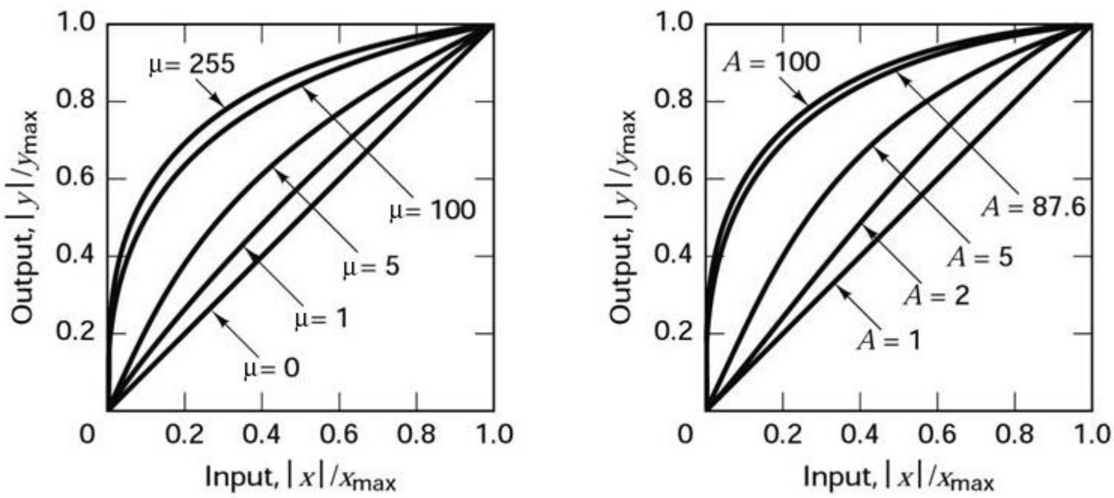
Figure 2.17 Non uniform Quantizer Characteristic

2.10.2.2 Companding Characteristics

We need linear compressor characteristics for small amplitudes of the input signal and a logarithmic characteristic elsewhere. In practice, this is achieved by using following two methods

(i) μ-law companding and (ii) A-law companding

The figure 2.18 shows the compression characteristics.



a) μ-law **b) A-law**

Figure 2.18 Compression Characteristics

(i) In North America, a μ-law compression characteristic is used. It is expressed mathematically as

$$Y = Y_{max}\frac{\log_e\left[1+\mu\left(|x|/x_{max}\right)\right]}{\log_e(1+\mu)} Sgn\ x \tag{2.15}$$

where

x	-	Input Voltage
Y	-	Output Voltage
X_{max}, Y_{max}	-	maximum positive excursions
μ	-	Positive constant

$$\text{Sgn } x = \begin{cases} +1 & for\ x \geq 0 \\ -1 & for\ x < 0 \end{cases}$$

In North America, the standard value for μ is 255. μ = 0 corresponds to linear amplification (uniform quantization).

(ii) In Europe, A-law compression characteristic is used. It is expressed mathematically as

$$Y = \begin{cases} Y_{max} \dfrac{A(|x|/x_{max})}{1+\log_e A} Sgn\ x, & 0 < \dfrac{|x|}{x_{max}} \leq \dfrac{1}{A} \\ Y_{max} \dfrac{1+\log_e[A(|x|/x_{max})]}{1+\log_e A} Sgn\ x, & \dfrac{1}{A} < \dfrac{|x|}{x_{max}} < 1 \end{cases} \tag{2.16}$$

where A is a positive constant. The standard value for A is 87.6. Here A = 1 corresponds to linear amplification (uniform quantization).

2.10.3 Performance comparison for speech signal

Human speech is characterised by unique statistical properties. For most voice communication channels, very low speech volumes predominate (for about 50% of time). Large amplitude values are relatively rare (only 15% of time).

1) Suppose that uniform quantization is applied to the speech signal. Since the step size is uniform over the full range of the input signal, many of the quantizing steps would rarely be used. The quantization noise is same for all signal magnitudes. Hence the signal to noise ratio (SNR) is worse for low level signals than for high level signals. Therefore, uniform quantization would be wasteful for speech signals.
2) Suppose that non uniform quantization is applied to the speech signal. Non uniform quantization can provide fine quantization of the weak signals and coarse quantization of the strong signals. Hence, noise can be made proportional to signal size. The effect is to improve the overall SNR by reducing the noise for the predominant weak signals.

Figure 2.19 compares the quantisation of a strong signal versus a weak signal for uniform and non-uniform quantization.

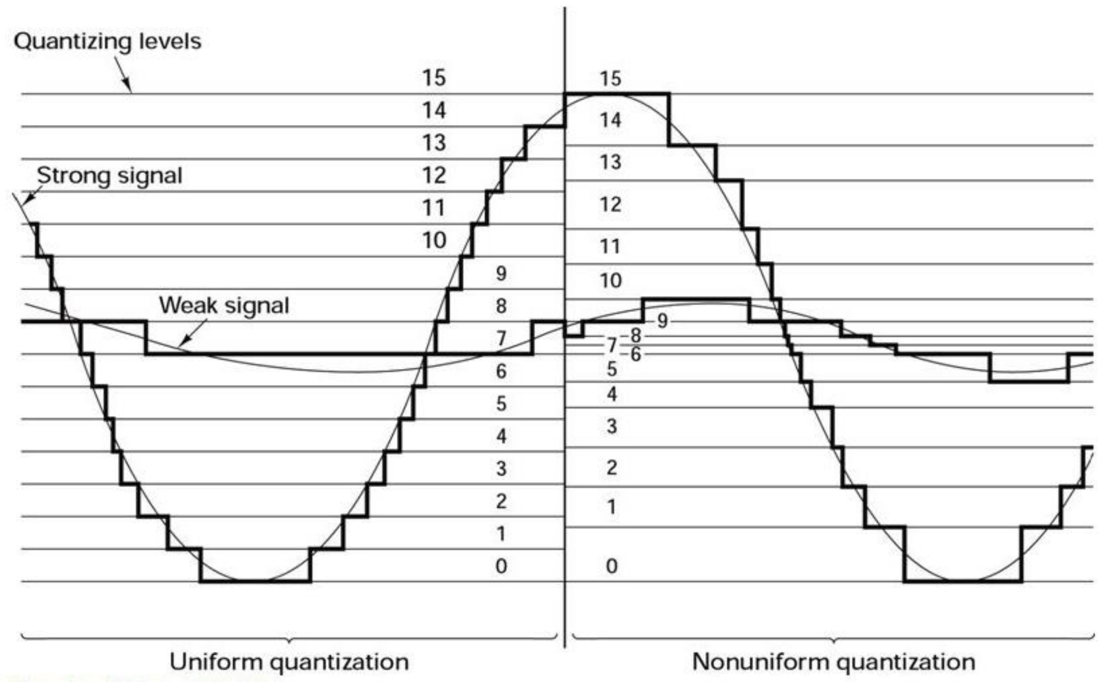


Figure 2.19 Performance Comparison

2.11 BASEBAND TRANSMISSION

A baseband digital communication system is made up of several elements.

1) Source, 2) Multiplexer 3) Line Coder and 4) Regenerative repeater

Source

If the information source is analog in nature, we use sampling, quantizing and encoding. We transform the analog waveforms into binary digits by the use of Pulse Code Modulation (PCM). If the information source is digital in nature, the data is already in the form of sequence of bits.

Multiplexer

We can combine several such sources through a digital multiplexer using the process of interleaving. Thus a channel is time-shared by several messages simultaneously.

Line Coder

In order to transmit these multiplexed messages which are in the form of sequence of bits through a baseband channel, the binary digits are coded into electrical pulses or waveforms. This process is known as Line Coding or transmission coding. There are several possible ways of assigning waveforms (ie., Pulses) to the digital data. One representation is shown in the Figure 2.20.

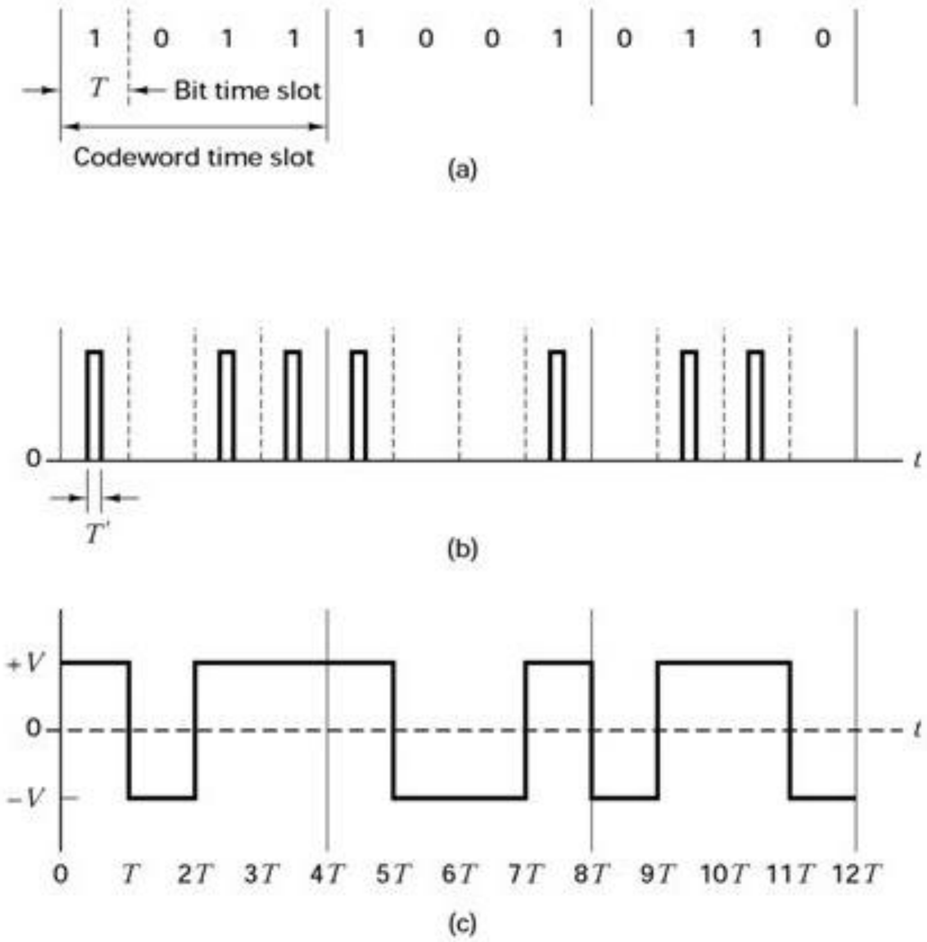


Figure 2.20 Example of waveform representation of binary digits

Thus a sequence of electrical pulses having the pattern shown in Figure 2.20 (b) or (c) can be used to transmit the information in the PCM bit stream, and hence the information in the quantized samples of a message.

Regenerative Repeater

Regenerative repeaters are used at regularly spaced intervals along a digital transmission line. They detect the incoming digital signal and regenerate new clean pulses for further transmission along the line.

2.12 PCM Wave form types

When pulse modulation is applied to a binary symbol, the resulting binary waveform is called a Pulse Code Modulation (PCM) wave form.

In telephony applications, these PCM waveforms are often called as Line Codes. When pulse modulation is applied to a non-binary symbol, the resulting waveform is called M-ary pulse modulation waveform.

Several types of PCM waveforms are illustrated in Figure 2.21.

The PCM waveforms can be classified into the following four groups.

1. Non return to Zero (NRZ)
2. Return to Zero (RZ)
3. Phase encoded
4. Multilevel binary

2.12.1 Non-Return to Zero (NRZ)

The NRZ group is probably the most commonly used PCM waveform. If the waveform stays at any non-zero level for the whole bit interval T, then it is called Non Return to Zero (NRZ) waveform. It can be subdivided into the following subgroups, NRZ – L (L for level), NRZ – M (M for Mark) and NRZ – S (S for Space).

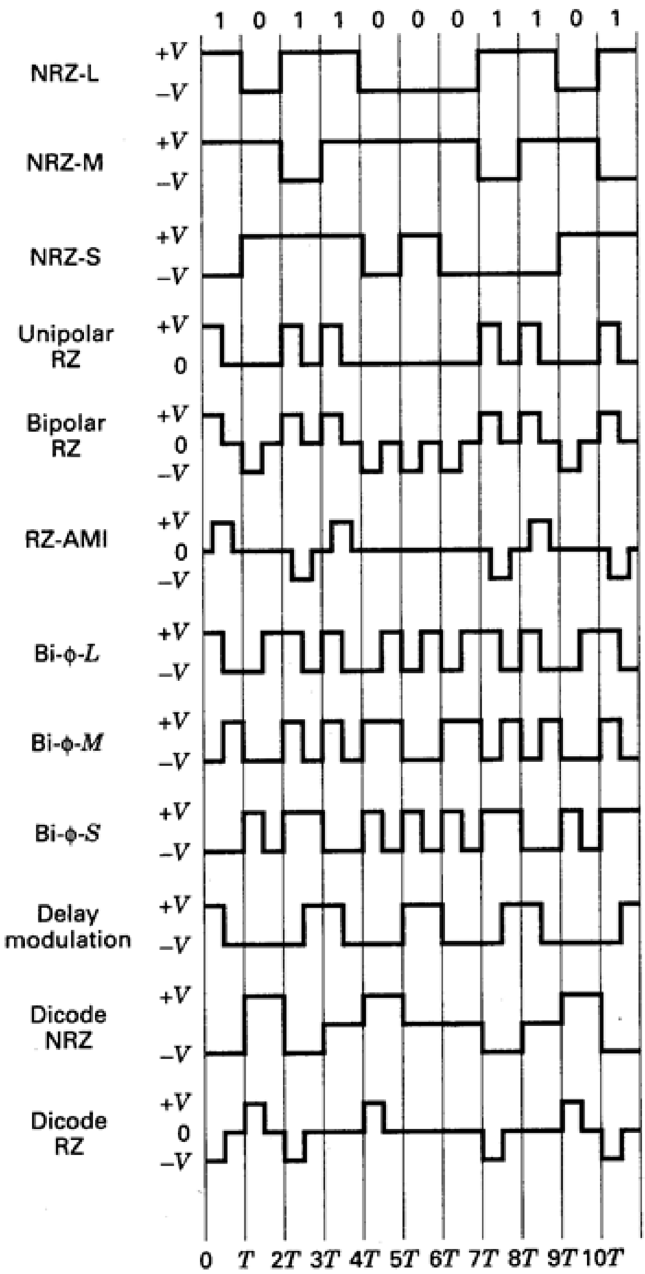


Figure 2.21 Various PCM Waveforms

1. NRZ – L is used extensively in digital logic circuits. A binary one is represented by one voltage level and a binary zero is represented by another voltage level. There is a change in level whenever the data change from a one to a zero or from a zero to a one.

2. With NRZ – M, the one, or mark, is represented by a change in level, and the zero, or space, is represented by no change in level. This is often referred to as differential encoding. NRZ – M is used primarily in magnetic tape recording.
3. NRZ – S is the complement of NRZ – M. A one is represented by no change in level, and a zero is represented by a change in level.

2.12.2 Return to Zero (RZ)

If the waveform comes back to zero level after a portion of bit interval T, then it is called RZ waveform. RZ group can be subdivided into the following subgroups; Unipolor RZ, Bipolor RZ, and RZ-AMI.

These codes find application in baseband data transmission and in magnetic recording.

1) With unipolar RZ, a one is presented by a half-bit-wide pulse, and a zero is represented by the absence of a pulse.
2) With bipolar-RZ, the ones and zeros are represented by opposite level pulses that are one-half bit wide. There is a pulse present in each bit interval.
3) RZ-AMI (AMI for "Alternate Mark Inversion") is a signaling scheme used in telephone systems. The ones are represented by equal amplitude alternating pulses. The zeros are represented by the absence of pulses.

2.12.3 Phase encoded

In phase encoded scheme, the time position of the occurrence or transition of a pulse waveform is utilized to distinguish between different logic levels. The phase encoded group consists of bi-ϕ-L (bi-phase-level), better known as Manchester coding; bi-ϕ-M (bi-phase-Mark); bi-ϕ-S (bi-phase-space); and delay modulation (DM) or Miller coding. The phase-encoding schemes are used in magnetic recording systems and optical communications and in some satellite telemetry links.

1) With bi-ϕ-L, a one is represented by a half-bit –wide pulse positioned during the first half of the bit interval. A zero is represented by a half-bit-wide pulse positioned during the second half of the bit interval.
2) With bi-ϕ-M, a transition occurs at the beginning of every bit interval. A one is represented by a second transition one-half bit interval later. A zero is represented by no second transition.

3) With bi-ϕ-S, a transition also occurs at the beginning of every bit interval. A one is represented by no second transition. A zero is represented by a second transition one-half bit interval later.
4) With delay modulation, a one is represented by a transition at the midpoint of the bit interval. A zero is represented by no transition, unless it is followed by another zero. In this case, a transition is placed at the end of the bit interval of the first zero.

2.12.4 Multilevel Binary

The binary waveforms which use three levels to encode the binary data, instead of two levels, are referred as multilevel binary waveforms. Bipolar RZ and RZ-AMI schemes belong to this group. This group also contains formats called dicode and duobinary.

1) With dicode NRZ, the one-to-zero, or zero-to-one data transition changes the pulse polarity. Without a data transition, the zero level is sent.
2) With dicode-RZ, the one-to-zero, or zero-to-one transition produces a half duration polarity change. Otherwise, a zero level is sent.
3) Duobinary signalling: Duobinary signalling is also referred to as correlative coding and partial response signaling. This technique introduced some controlled amount of ISI into the data stream. This improves bandwidth efficiency at the expense of an increase in power.

2.13 SELECTION OF A PCM WAVEFORM

There are a variety of PCM waveform formats to represent binary digits. We now consider the selection of a particular waveform format for the transmission of baseband signals through the channel. In choosing a PCM waveform for a particular application, some of the parameters to be considered are given below.

1) **DC component:** Eliminating the dc energy from the signal's power spectrum enables the system to be ac coupled.
2) **Self Clocking (Self Synchronization):** Symbol or bit synchronization is required for any digital communication system. The Manchester code has a transition in the middle of every bit interval whether a one or a zero is being sent. This guaranteed transition provides a clocking signal.
3) **Error detection:** Duobinary signaling scheme provides the means of detecting data errors without introducing additional error detection bits into the data sequence.

4) **Bandwidth Compression:** The Multilevel codes increase the efficiency of bandwidth utilization by allowing a reduction in required bandwidth for a given data rate.
5) **Differential Encoding:** This technique is useful because it allows the polarity of differentially encoded waveforms to be inverted without affecting the data detection.
6) **Noise Immunity:** The various PCM waveform types can be further characterized by probability of bit error versus signal-to-noise ratio. The NRZ waveforms have better error performance than does the unipolar RZ waveform.
7) **Spectral compatibility with Channel:** The transmission bandwidth of the code must be sufficiently small compared to the channel bandwidth so that ISI is not a problem. When channel characteristic varies over different frequency bands, a PCM waveform with a similar power spectral density would be preferred.

2.14 SPECTRAL ATTRIBUTES OF PCM WAVEFORMS

The most common criteria used for comparing PCM waveforms and for selecting a particular waveform type are

- Spectral characteristics
- Bit synchronization capabilities
- Error detecting capabilities
- Interference and noise immunity
- Cost and complexity of implementation

The figure 2.22 shows the spectral characteristics of some of the most popular PCM waveforms. It is important to choose a waveform which is optimum in terms of both its spectral occupancy and detection error performance.

The figure plots power spectral density in Watts / Hertz versus normalized bandwidth, BT, where B is bandwidth and T is the duration of the pulse. We often refer BT as the time-bandwidth product of the signal. Since the pulse or symbol rate R_s is the reciprocal of T, normalized bandwidth can also be expressed as B/R_s. Hence, the units of normalized bandwidth are Hertz / (Pulse/s) or Hertz / (Symbol/s). It describes how efficiently the transmission bandwidth is being utilized for each waveform of interest.

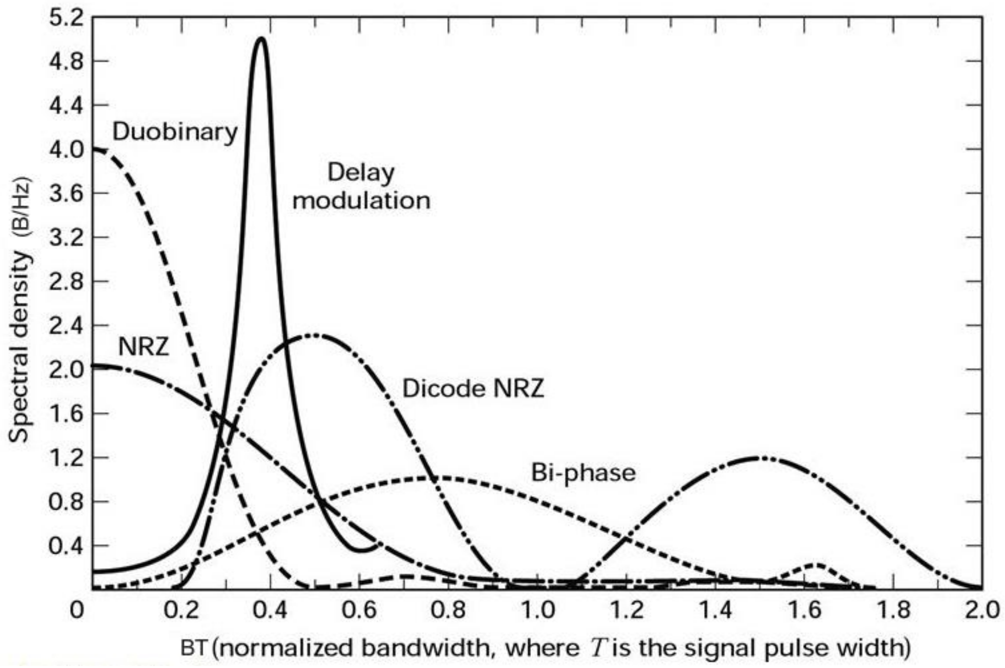


Figure 2.20 Spectral densities of various PCM waveforms

(i) Any waveform type that requires less than 1.0 Hertz for sending 1 symbol/s is relatively bandwidth efficient. Example waveforms are delay modulation and duobinary.

(ii) Any waveform type that requires more than 1.0 Hertz for sending 1 symbol/s is relatively bandwidth inefficient. Example waveform is bi-phase (Manchester) signaling.

(iii) On comparing spectral concentration of signaling energy, the NRZ and duobinary schemes have large spectral components at dc and low frequency. But bi-phase has no energy at dc.

(iv) An important parameter for measuring bandwidth efficiency is R/B having units of bits/hertz. For a given signaling scheme, R/B describes how much data throughput can be transmitted for each Hertz of available bandwidth.

2.15 BITS PER PCM WORD AND BITS PER SYMBOL

The formatting process involves conversion of analog information into a bit stream by means of sampling, quantization and coding. Each analog sample is transformed into a PCM word made up of group of bits. The PCM word size can be described by the number of quantization levels allowed for each sample. This is identical to the number of values that the PCM word can assume.

The number of quantization levels is based on the number of encoding bits required to identify that set of levels. We know that the relationship between the size of a set of message symbols and the number of bits needed to represent the symbol is

$$M = 2^k \tag{2.17}$$

Similarly for PCM, the relationship between the number of levels per sample and the number of bits needed to represent those levels is given by

$$L = 2^l \tag{2.18}$$

Where L → number of quantization levels in the PCM word

l→ number of bits needed to represent those levels

For example, if each sample has 256 quantization levels, then the bits per PCM word is given by, 2^l = 256 ⇒ l = $log_2 256$ = 8.

Hence the number of bits needed to represent 256 quantization levels is l = 8.

2.16 PCM WORD SIZE

The required number of bits per analog sample is called as the PCM word size. For digital telephone channels, each speech sample is PCM encoded using 8 bits, resulting in 2^8 or 256 levels per sample. The choice of PCM word size or bits per sample, depends on how much quantization distortion is tolerable. We may develop a general relationship between the PCM word size and the allowable quantization distortion.

Let the magnitude of the quantization distortion error$|e|$, be specified as a fraction P of the peak-to-peak analog voltage Vpp as follows.

$$|e| \leq PV_{pp} \tag{2.19}$$

The quantization error cannot be larger than $\frac{q}{2}$, where q is the quantile interval. Hence,

$$|e|_{max} = \frac{q}{2} = \frac{\text{Vpp}}{2L} \tag{2.20}$$

Where L is the number of quantization levels.

From equations (2.19) and (2.20), we can write,

$$\frac{\text{Vpp}}{2L} \leq PV_{pp}$$

$$\Rightarrow \frac{1}{2L} \leq P$$

$$\Rightarrow \quad \frac{1}{2P} \leq L$$

$$\Rightarrow \quad L \geq \frac{1}{2P}$$

Since $L = 2^l$,

$$2^l \geq \frac{1}{2P}$$

$$\Rightarrow \quad l \geq log_2\left(\frac{1}{2P}\right) \text{ bits} \qquad (2.21)$$

2.17 M-ARY PULSE MODULATION WAVEFORMS

When information samples without any quantization are modulated on to pulses, the resulting pulse modulation can be called analog pulse modulation. There are three basic ways to modulate information on to a sequence of pulses. We can vary the pulse's amplitude (Pulse Amplitude Modulation, PAM), position (Pulse – Position Modulation, PPM) and duration (Pulse Duration Modulation or Pulse Width Modulation, PDM or PWM).

When information samples are quantized and then modulated onto pulses, the resulting pulse modulation is called digital pulse modulation. Pulse Code Modulation (PCM), Differential Pulse Code Modulation (DPCM), Delta Modulation (DM) and Adaptive Delta Modulation (ADM) are the four ways of digital pulse modulation techniques.

If the output of the pulse generator consists of binary pulses, ie. pulses with one of the two possible amplitude levels, then the resulting waveform is binary pulse modulation waveform.

If the output of the pulse generator consists of pulses with M possible amplitude levels, the resulting waveform is M-ary pulse modulation. The value of M is typically an integer power of 2.

(i) In the case of M-ary PAM, one of the M allowable amplitude levels are assigned to each of the M possible symbol values. The binary PCM waveforms having two amplitude values (eg. NRZ, RZ) represent the special case (M=2) of the general M-ary PAM that requires M levels. PAM is similar to analog amplitude modulation.

(ii) In the case of M-ary PPM waveforms, modulation is effected by delaying a pulse occurrence, by an amount that corresponds to the value of the information symbols. PPM is similar to analog phase modulation. Here the pulse amplitude is held constant.

(iii) In the case of M-ary PDM, modulation is effected by varying the pulse width by an amount that corresponds to the value of the symbols. Here also, the pulse amplitude is held constant. PDM is similar to analog frequency modulation.

Need for M-ary scheme (or) Advantages / Merits of M-ary Scheme

We shall discuss about M-ary PAM waveforms as they are comparable to PCM waveforms.

1) The transmission bandwidth required for binary digital waveforms such as PCM may be very large. To reduce the required bandwidth, we may use M-ary scheme or multilevel signaling.

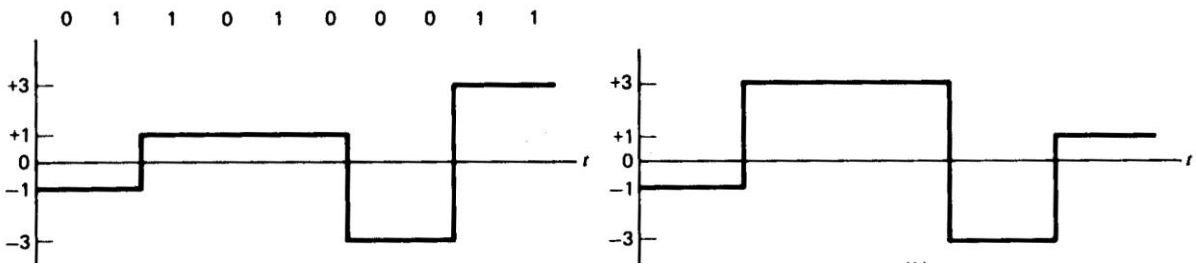
Consider a bit stream with data rate, R bits per second. Instead of transmitting a pulse waveform for each bit, we may partition the data into k-bit groups, and then use (M = 2^k) – level pulses for transmission. With such multilevel signaling, or M-ary PAM, each pulse waveform can now represent a k-bit symbol in a symbol stream moving at the rate of R / K symbols per second. Thus for a given data rate, multilevel signaling (where M > 2), can be used to reduce the number of symbols transmitted per second. Hence, the transmission bandwidth requirements of the channel is reduced.

2) M-ary PAM is well suited for the transmission of digital data over channels that offer a limited bandwidth and a high signal-to-noise ratio.

3) M-ary PAM enables the transmission of data through the channel at a rate that is log_2M faster than the corresponding binary scheme.

Example of M-ary Scheme

The utilization of bandwidth can be made more efficient by adopting an M-ary format for the representation of the input binary data. An example of this representation is the polar quaternary NRZ format as shown in Figure 2.21.



(a) Natural-encoded **(b) Gray-encoded**

Figure 2.21 Polar Quaternary NRZ format

A quaternary code has four distinct symbols, referred to as dibits (pairs of bits). Each dibit is assigned a level in accordance with the natural code described in the below table 2.2.

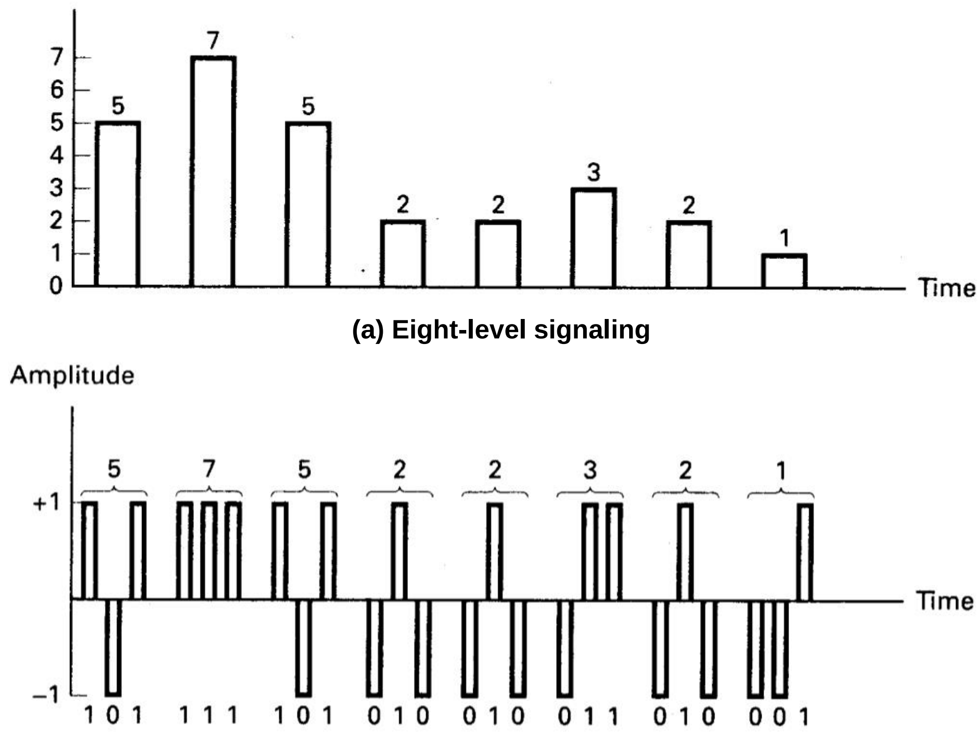
Table 2.2 Natural and Gray Codes

Level	Natural Code	Gray Code
-3	0 0	0 0
-1	0 1	0 1
+1	1 0	1 1
+3	1 1	1 0

To obtain the waveform of Figure 2.21(a) the input binary sequence 0110100011 is viewed as a new sequence of dibits{ 01, 10, 10, 00, 11 }. The figure 2.21(b) shows a second coding scheme, called the Gray Code. In Gray Code, the adjacent bits are arranged in such a way that they differ by only one bit.

Disadvantages / Demerits of M-ary scheme

Consider an M-ary format of an octal pulse as shown in Figure 2.22.



(a) Two-level signaling

Figure 2.22 PCM signaling (M-ary)

It is not easy for the receiver to distinguish among the eight possible levels of each octal pulse. The transmission of an 8 level (compared with a 2 level) pulse requires a greater amount of energy for equivalent detection performance. For equal average power in the binary and octal pulses, it is easier to detect the binary pulses. Because the detector has more signal energy per level for making a binary decision than an 8 level decision. But for equivalent binary transmission, the octal pulse must be replaced with three binary pulses as shown in the figure.

Table 2.3 Performance comparison of various line codes:

Sl. No.	Performance Parameters	Polar RZ	Polar NRZ	AMI	Manchester	Polar quarternary NRZ
1.	Transmission of DC component	Yes	Yes	No	No	Possible
2.	Signaling rate	$\frac{1}{T}$	$\frac{1}{T}$	$\frac{1}{T}$	$\frac{1}{T}$	$\frac{1}{2T}$
3.	Noise immunity	Low	Low	High	High	High
4.	Synchronizing capability	Poor	Poor	Very Good	Very Good	Poor
5.	Bandwidth requirement	$\frac{1}{T}$	$\frac{1}{2T}$	$\frac{1}{2T}$	$\frac{1}{T}$	$\frac{1}{2T}$
6.	Cross talk	High	High	Low	Low	Low

SHORT QUESTIONS AND ANSWERS

1. What is meant by Formatting?

Formatting is the first essential signal processing step in digital communication. Formatting transforms the source information into bits, thus assuring compatibility between the information and the signal processing steps within the digital communication system.

i. Data already in digital format would bypass the formatting function.
ii. Textual information is transformed into binary digits by the use of coder.
iii. Analog information is formatted using three separate processes of sampling, quantizing and coding.

2. Define sampling theorem.

Sampling theorem can be stated in two parts as below.

(i) A band-limited signal of finite energy, which has no frequency components higher than f_m Hertz, is completely described by its sample values at uniform intervals less than or equal to $\frac{1}{2f_m}$.

(ii) A band-limited signal of finite energy, which has no frequency components higher than f_m Hertz, may be completely recovered from the knowledge of its samples taken at the rate of $2f_m$ samples per second.

3. Define Nyquist Theorem.

The Nyquist theorem can be stated as follows. "The sampling frequency (f_s) must be at the rate equal to or greater than twice the highest frequency component (f_m) present in the signal ie. $f_s \geq 2f_m$, in order to recover the signal exactly."

4. What is meant by sampling? Name the different methods of sampling.

The process of converting continuous-time signal into discrete-time signal is called sampling. Basically, there are three types of sampling techniques.

1. Impulse sampling or Ideal sampling.
2. Natural sampling
3. Sample and hold operation or flat top sampling.

5. What is impulse sampling? What is its disadvantage?

If the sampling function is a train of impulses, then the method is called as impulse or ideal sampling. This method results in the samples whose width approaches zero. Due to this, the power content in the instantaneously sampled pulse is negligible. Hence, this method is not suitable for transmission purpose.

6. What is natural sampling? What are its disadvantages?

If the sampling function is a pulse train or switching waveform, then the method is called as natural sampling. In this method, each pulse in the sampled data sequence has varying top according to signal variation. During transmission, noise interferes the top of pulses. Then it becomes difficult to determine the shape of top of the pulse at the receiver.

7. What is sample and hold operation?

In natural sampling, the pulse has varying top according to the signal variation. Therefore, amplitude detection of the pulse is not exact and errors are introduced in the signal. This problem will be solved by having flat top pulses. A sample and hold circuit is used to generate flat top pulses.

The sample and hold operation can be described by the convolution of the sampled pulse train with a unity amplitude rectangular pulse. The convolution process results in the flat top sampled sequence.

8. What is aliasing?

Aliasing may be defined as the phenomenon in which a high frequency component in the frequency spectrum of the signal takes identity of a lower frequency component in the spectrum of the sampled signal. Aliasing results from the effect of undersampling ($f_s < 2f_m$).

9. How we can prevent aliasing?

There are two ways of eliminating aliasing using antialiasing filters.

(i) The analog signal is prefiltered using a low pass filter. The bandwidth of the filter is less than or equal to half the sampling frequency (ie.,$f_m < \frac{f_s}{2}$).

(ii) The analog signal is postfiltered after sampling using a lowpass filter. When the signal structure is well known, the aliased terms can be eliminated after sampling.

10. What is a quantization noise?

The sample values of an analog baseband signal are rounded-off to the nearest permissible representation levels of the quantizer. The distortion introduced by this approximation of the quantized samples is referred to as quantization noise. The amount of such noise is inversely proportional to the number of levels employed in the quantization process.

11. What do you mean by channel noise?

The channel noise is the combined effect of thermal noise, interference from other users, and interference from circuit switching transients. It may be introduced anywhere along the transmission path. It may be modelled as Additive White Gaussian Noise (AWGN) with zero mean and power spectral density, $\frac{N_o}{2}$.

12. What is InterSymbol Interference (ISI)?

A band limited channel disperses or spreads a pulse waveform passing through it. When the channel bandwidth is close to the signal bandwidth, the spreading will exceed symbol duration and cause signal pulses to overlap. This overlapping is called InterSymbol Interference (ISI).

13. What is PCM?

Pulse Code Modulation (PCM) refers to the class of baseband signals obtained from the quantized PAM signals by encoding each quantized sample into a digital word. The essential operations in the transmitter of a PCM system are sampling, quantizing and encoding.

14. What is quantization? Mention its types.

The quantization process converts the continuous amplitude values into a finite (discrete) set of allowable values. This process is called “discretization” in time and amplitude. Basically, quantization process may be classified as below.

I. Uniform quantization
   1. Midtread type
   2. Midrise type

II. Non-Uniform quantization

15. What is Uniform Quantization?

When the quantization levels are uniformly distributed over the full amplitude range of the input signal, then the quantizer is called as an Uniform or Linear quantizer. Here, the stepsize between quantization levels remains the same throughout the input range.

16. What is non-uniform quantization?

If the quantization levels are not uniformly distributed over the full amplitude range of the input signal, then the quantizer is called as non-uniform or non-linear quantizer. Here the stepsize between quantization levels varies according to certain law.

17. What is companding?

Companding is the process of signal compression and expansion. At the transmitter, the compressor amplifies weak signals and attenuates strong signals. At the receiver, the expander does the opposite function of compression. To overcome the problems of quantization noise, companding process is used.

18. What is meant by Line Coding?

In order to transmit the multiplexed message signals which are in the form of sequence of bits through a baseband channel, the binary digits are coded into electrical pulses or waveforms. This process is known as Line coding or transmission coding.

19. What is meant by PCM waveform? What are its types?

When pulse modulation is applied to binary symbols, the resulting binary waveform is called a Pulse Code Modulation (PCM) waveform. In telephone applications, these PCM waveforms are often called as Line Codes. The PCM waveforms can be classified into the following four groups.

1. Non Return to Zero (NRZ)
2. Return to Zero (RZ)
3. Phase encoded
4. Multilevel Binary

20. What is Non-Return to Zero (NRZ) waveform?

If the PCM waveform stays at any non zero level for the whole bit interval T, then it is called as Non-Return to Zero (NRZ) waveform. It can be subdivided into the following subgroups.

(i) NRZ – L (L for level)
(ii) NRZ – M (M for mark)
(iii) NRZ – S (S for space)

21. What is Return to Zero (RZ) waveform?

If the PCM waveform comes back to Zero level after a portion of bit interval T, then it is called as Return to Zero (RZ) waveform. It can be subdivided into the following subgroups.

(i) Unipolar RZ
(ii) Bipolar RZ
(iii) RZ-AMI (AMI for Alternate Mark Inversion)

22. What is phase encoded waveform?

In phase encoded scheme, the time position of the occurrence or transition of a pulse waveform is utilized to distinguish between different logic levels. It can be subdivided into the following subgroups.

(i) Bi-ϕ-L (Bi-phase-Level) or Manchester coding
(ii) Bi-ϕ-M (Bi-phase-Mark)
(iii) Bi-ϕ-S (Bi-phase-Space)
(iv) Delay Modulation (DM) or Miller Coding

23. What is Multilevel binary waveform?

The binary waveforms which use three levels to encode the binary data, instead of two levels, are referred as multilevel binary waveforms. The different formats used in multilevel binary are given below.

(i) Dicode NRZ
(ii) Dicode RZ
(iii) Duobinary signalling
(iv) Bipolar RZ
(v) RZ – AMI

24. Mention the criteria used for selecting or choosing a particular PCM waveform type.

The most common criteria used for comparing PCM waveforms and for selecting a particular waveform type are

1. Spectral attributes or characteristics
2. Bit synchronization capabilities
3. Error detection performance
4. Interference and noise immunity
5. Cost and complexity or implementation

25. Define PCM wordsize.

The number of bits required to represent the quantization levels of analog sample is called as the PCM wordsize. For digital telephone channels, each speech sample has 256 quantization levels. Hence the PCM word size is 8 bits ($2^8 = 256$).

26. State the types of Pulse Modulation.

Analog Pulse Modulation	Digital Pulse Modulation
1. Pulse Amplitude Modulation (PAM)	1. Pulse Code Modulation (PCM)
2. Pulse Width Modulation (PWM)	2. Differential Pulse Code Modulation (DPCM)
3. Pulse Position Modulation (PPM)	3. Delta Modulation (DM)
	4. Adaptive Delta Modulation (ADM)

27. Define M-ary Pulse Modulation Waveforms.

In the pulse modulation schemes, if the output of the pulse generator consists of pulses with M possible amplitude levels, the resulting waveform is called M-ary Pulse Modulation Waveform. The value of M is typically an integer power of 2.

28. What are the advantages and disadvantages of M-ary Pulse Modulation schemes?

Advantages

1. M-ary scheme or multilevel signaling reduces the required transmission bandwidth.
2. High signal to noise ratio.
3. M-ary scheme is $\log_2 M$ times faster than the corresponding binary scheme.

Disadvantages

1. M-ary scheme requires a greater amount of energy for equivalent detection performance compared to binary scheme.

Unit - III BASEBAND CODING TECHNIQUES

OBJECTIVES

- To learn about the need for baseband coding techniques.
- To learn about the types of codes, and types of error control coding methods.
- To learn in detail about Block codes and convolution codes.

3.0 INTRODUCTION

In any digital communication system, the two important desirable features are the higher transmission rate and good reliability (ie., low probability of error). To achieve these features baseband coding techniques are used. There are two types of baseband coding. They are source coding and channel coding.

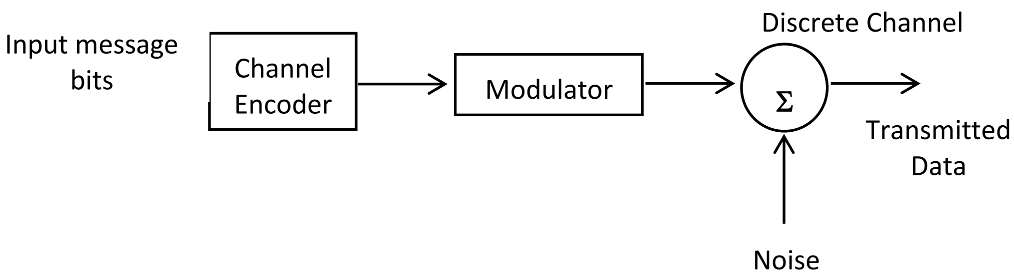
Source coding is used for an efficient representation of data generated by a discrete source. Source coding minimizes the average bit rate required for representation of the source by reducing the redundancy of the information source.

Channel coding is used for the reliable transmission of digital information over the channel. Channel coding improves communication performance by enabling the transmitted signals to better withstand the effects of various channel impairments, such as noise, interference and fading. Channel coding methods introduce controlled redundancy in order to provide error detecting and correcting capability to the data being transmitted. Hence channel coding is also called as error control coding. Here we shall see in detail about error control coding.

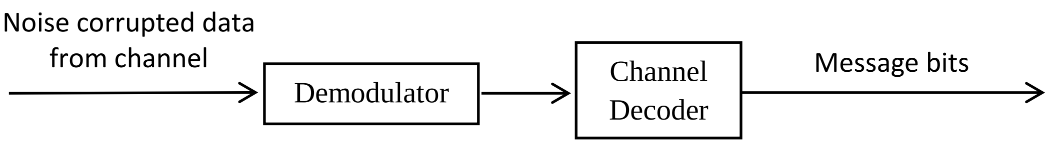
3.1 RATIONALE FOR CODING (NEED FOR CODING)

- The primary communication resources are the transmitted signal power and channel bandwidth.
- These two parameters, together with the power spectral density of receiver noise, determine the signal energy per bit-to-noise power density ratio, E_b/N_o.
- This ratio E_b/N_o uniquely determines the probability of error (P_e) or bit error rate (BER), for a particular modulation scheme.
- The channel induced noise can introduce errors in the transmitted binary data. ie., a bit 0 may change to bit 1 or a bit 1 may change to bit 0. The reliability of data transmission gets severely affected because of these errors.
- Accordingly, in practice with the available modulation schemes, it is not possible to provide acceptable data quality of low error performance. Also, there is a limitation on the achieved maximum value of the ratio E_b/N_o.

- Therefore, for a fixed E_b/N_o, the only practical option available for changing data quality from problematic to acceptable level is to use coding.
- Another practical requirement for the use of coding is to reduce the required E_b/N_o for a fixed bit error rate. The reduction in E_b/N_o will reduce the required transmitted power.
- This in turn, reduces the hardware costs by requiring a smaller antenna size.
- The figure 3.1 shows the digital communication system which uses channel coding.



(a) Transmitter



(b) Receiver

Figure 3.1 Digital Communication System with Channel Coding

The channel encoder adds extra bits (redundancy) to the message bits in a controlled manner. The encoded signal is modulated and then transmitted over the noisy channel.

- After demodulation, the channel decoder identifies the redundant bits and uses them to detect and correct the errors in the message bits. Thus the errors introduced due to channel noise are minimized by channel coding.

Drawbacks of channel coding

- The addition of redundant bits in the coded messages, increases the required transmission bandwidth.
- The use of coding adds complexity to the communication system.

Therefore, the design trade-offs in the use of error control coding to achieve acceptable error performance must include consideration of bandwidth and system complexity.

3.2 TYPES OF CODES:

We can use many different codes for error control coding. The classification of codes is given below.

I. Based on Methodology / Architecture

1. Block Codes:

To generate an (n, k) block code, the channel encoder accepts information in successive k-bit blocks. For each block, it adds (n-k) redundant bits, thereby producing an overall encoded block of n bits. Block codes do not need memory.

2. Convolutional Codes:

In a convolutional code, the encoding operation is the discrete-time convolution of the input sequence with the impulse response of the encoder. The convolutional encoder accepts the message bits continuously and generates the encoded sequence continuously. Convolutional codes need memory. Convolutional codes are a subclass of Tree codes.

II. Based on hardware mechanization required to generate

1. Linear Code:

A linear code has the property that any two codewords of a linear code can be added in modulo-2 arithmetic to produce a third codeword in the code. The codes used in practical applications are almost always linear codes.

2. Nonlinear code:

Modulo-2 addition of the nonlinear codewords doesnot necessarily produces third codeword. The nonlinear codes have been less important.

III. Based on structure of Architecture (Tree / Block)

This classification refers to the way in which redundant bits (check bits) are added to the message (information) bits in block codes.

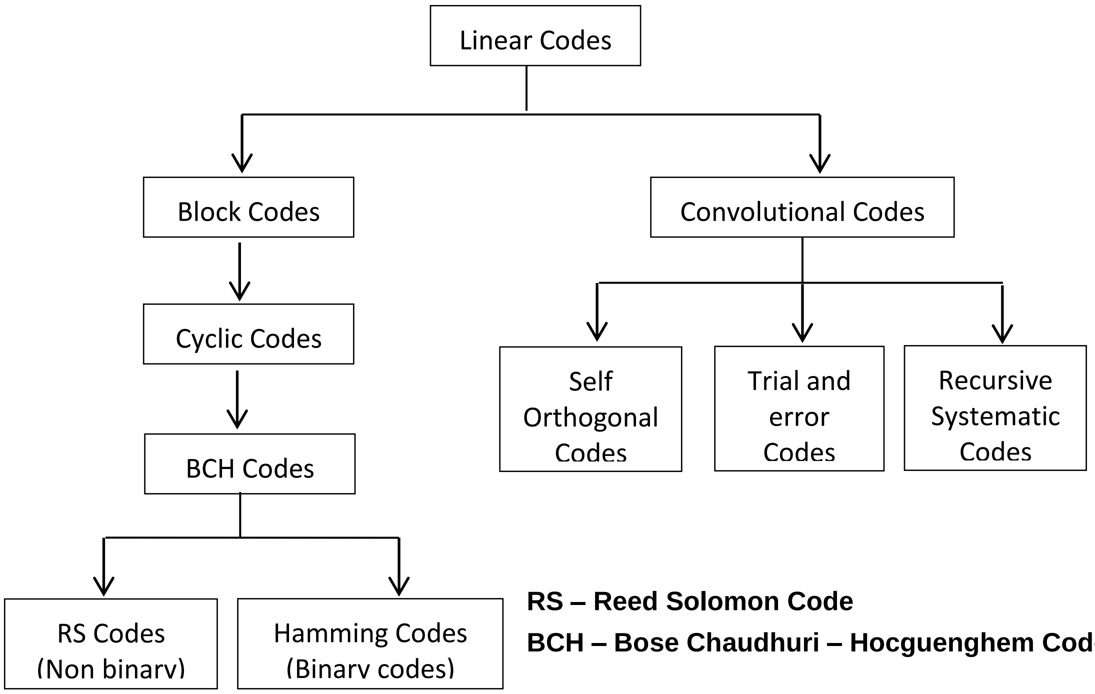
1. Systematic Code:

In a systematic block code, the redundant (check) bits are added in such a way that the message bits appear first and then redundant (check) bits.

2. Non-Systematic Code:

In a Non-systematic block code, it is not possible to identify message bits and redundant (check) bits. They are mixed in the block.

The various error correction / detection codes in the class of linear codes are listed below.



RS – Reed Solomon Code

BCH – Bose Chaudhuri – Hocguenghem Code

3.3 DISCRETE MEMORYLESS CHANNEL

The propagating medium or electromagnetic path connecting the transmitter and receiver is called the channel. The characteristics of a channel can be studied by representing the channel as a statistical model. The Discrete Memoryless Channel (DMC) is one of such channel models.

A discrete memoryless channel is a statistical model with an input X and an output Y as shown in the Figure 3.2. The output Y is a noisy version of X. Both X and Y are random variables.

During each unit of the time (signaling interval), the channel accepts an input symbol X selected from an alphabet $\mathcal{X}$. In response, it emits an output symbol Y from an alphabet $\mathcal{Y}$. The channel is said to be "discrete" when both of the alphabets $\mathcal{X}$ and $\mathcal{Y}$ have finite sizes. Also, the channel is said to be "memoryless" when the current output symbol depends only on the current input symbol and not on any of the previous inputs.

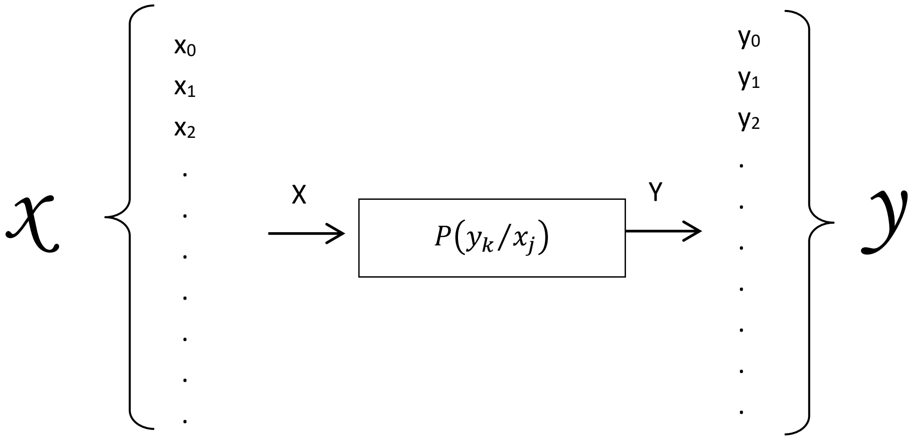


Figure 3.2 Discrete Memoryless channel

The channel is described in terms of an input alphabet $\mathcal{X} = \{x_0, x_1, x_2, \ldots. x_{j-1}\}$ and an output alphabet $\mathcal{Y} = \{y_0, y_1, y_2, \ldots. y_{k-1}\}$. The transition probability is given by $P(y_k/x_j) = P(Y = y_k/X = x_j)$ for all j and k. (3.1)

We have, $0 \leq P(y_k/x_j) \leq 1$ for all j and k.

- The transition probability $P(y_k/x_j)$is the conditional probability that the channel output Y = y_k, given that the channel input X = x_j. There is a possibility of errors arising from the process of information transmission through the channel.
- Thus, when k = j, the transition probability $P(y_k/x_j)$represents a conditional probability of correct reception.
- When k ≠ j, it represents a conditional probability of error.

Channel Matrix

- The discrete memoryless channel can be described by the complete set of transition probabilities.
- The various transition probabilities of the channel are arranged in the form of a matrix as below. This matrix is called as channel matrix.

$$P = \begin{bmatrix} p(y_0/x_0) & p(y_1/x_0) & \cdots\cdots & p(y_{k-1}/x_0) \\ p(y_0/x_1) & p(y_1/x_1) & \cdots\cdots & p(y_{k-1}/x_1) \\ \vdots & & & \\ p(y_0/x_{j-1}) & p(y_1/x_{j-1}) & \cdots\cdots & p(y_{k-1}/x_{j-1}) \end{bmatrix}$$

Probability of error

- The average probability of symbol error, P_e, is defined as the probability that the output random variable Y_k is different from the input random variable X_j, averaged over all $k \neq j$.
- The difference $(1 – P_e)$ is the average probability of correct reception.

Binary Symmetric Channel

The Binary Symmetric Channel (BSC) is a special case of the discrete memoryless channel with j = k = 2. The transition probability diagram of this channel is shown in the figure 3.3

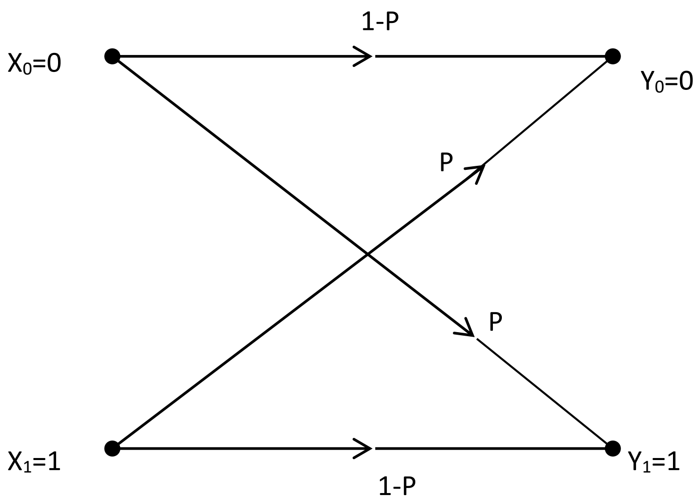


Figure 3.3 Binary Symmetric Channel

This channel has two input symbols ($x_0 = 0$, $x_1 = 1$) and two output symbols ($y_0 = 0$, $y_1 = 1$). The channel is symmetric because the probability of receiving a 1 if a 0 is sent is the same as the probability of receiving a 0 if a 1 is sent. Its channel matrix is given by $\begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$, where p refers to common transition probability or conditional probability of error.

3.4 ERROR CONTROL CODING METHODS

Error control coding techniques involve systematic addition of redundant bits to the transmitted information to achieve error detection and correction at the receiver. Thus the controlled redundancy in the transmitted message reduces probability of error at the receiver.

- The redundant bits added to the message are called check bits. Errors can be detected and corrected with the help of these bits.

- The check bits reduce the data rate through the channel.
- It is not possible to detect and correct all the error in the received message. Errors upto certain limit can only be detected and corrected.
- There are two main methods used for error control coding. They are
  1) Forward acting Error Correction (FEC)
  2) Error detection with retransmission or Automatic Repeat Request (ARQ)

  Sometimes, a hybrid system employing both FEC and ARQ may be used.

3.4.1 Forward Error Correction (FEC) Method

Forward Error Correction (FEC) method consists of a channel encoder at the transmitter and a channel decoder at the receiver as shown in the figure 3.4. If the channel decoder detects any error, it corrects the error using error correction codes.

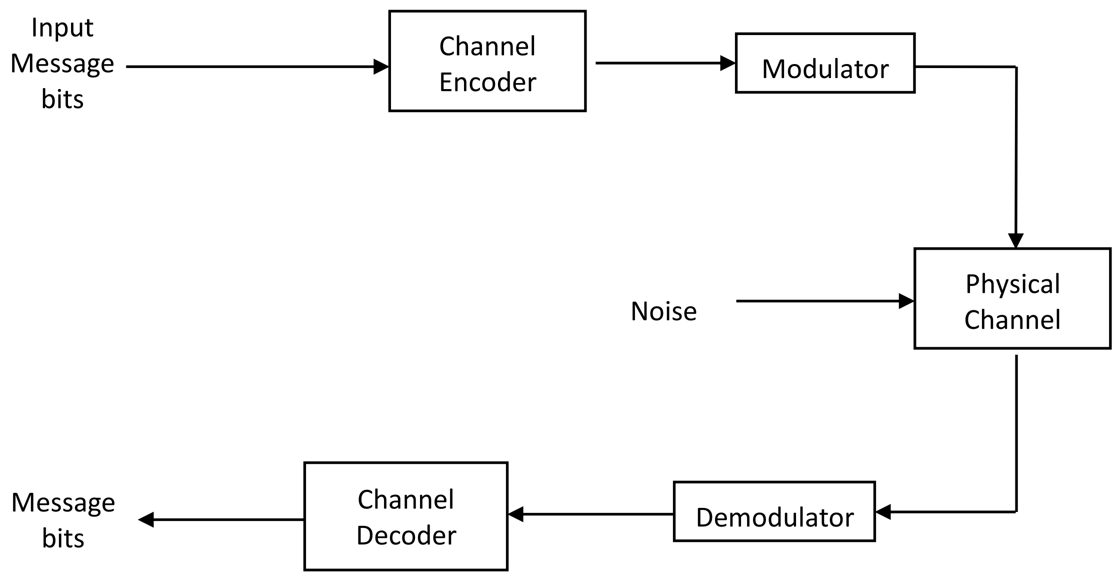


Figure 3.4 System employing FEC

The channel encoder will add redundant bits (check bits) in a controlled manner. Thus the encoded data is produced at a higher rate. The check bits are used by the channel decoder to detect and correct errors. The error detection and correction capability of the receiver depends upon the number of check bits in the transmitted message.

Advantages of using FEC

- The feedback channel or return path is not needed.

- A constant overall delay is obtained. Hence the system is faster.
- The information throughput efficiency is constant in FEC systems.

Disadvantages of using FEC

- The overall probability of errors is higher, because some of the errors cannot be corrected.
- When very high reliability is needed, selection of an appropriate error correcting code and implementing its decoding algorithm may be difficult.
- Only a relatively moderate information throughput is obtained.

Codes used in FEC method

- There are several error correcting codes. They can be classified broadly into 1) Block Codes and 2) Convolutional Codes
- A subclass of linear block codes called cyclic codes is popularly used in FEC method. Cyclic code examples are BCH codes, Hamming codes, Golay codes and Reed Solomon Codes.

3.4.2 Error detection with retransmission or Automatic Repeat Request(ARQ)

In this method, if the channel decoder detects any error, it discards that part of the data sequence and requests the transmitter for retransmission. The block diagram of a basic ARQ system is show in the figure 3.5.

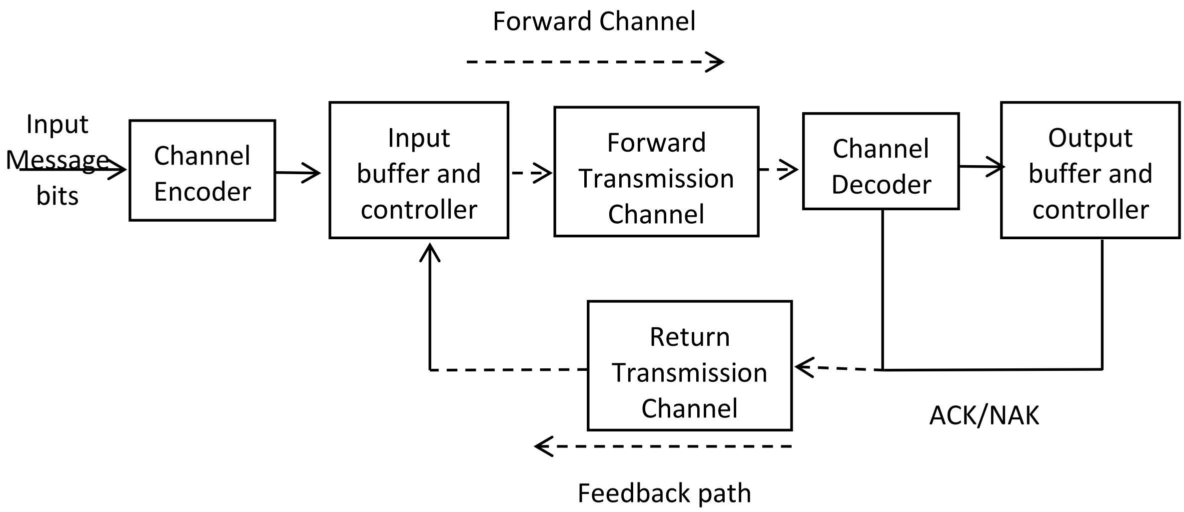


Figure 3.5 : Block diagram of basic ARQ system

The channel encoder produces codewords for each message signal at its input. Each codeword is temporarily stored in the buffer and transmitted over the forward transmission channel. At the receiver, the channel decoder decodes the codewords and look for errors.

- If no error is detected, the decoder sends a positive acknowledgement (ACK) through the return transmission channel.
- If errors are detected, the decoder sends a negative acknowledgement (NAK). The transmitter then once again transmits that part of the codeword sequence in which error was detected.

Here, the decoder does not correct the errors. It just detects the errors and request the transmitter for retransmission. Hence this method is called as Automatic Repeat Request (ARQ).

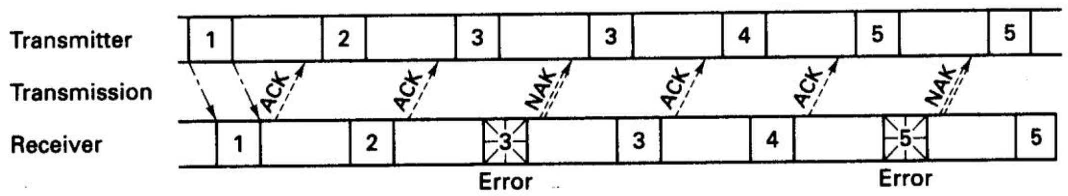
Types of ARQ System

Basically there are two types of ARQ. They are 1) Stop-and-wait ARQ 2) Continuous ARQ

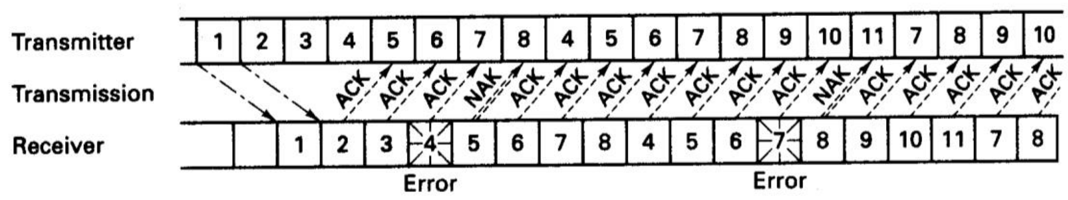
1) Stop-and-wait ARQ

In this method, the transmitter transmits a codeword and then waits. The decoder checks the received codeword for error. if no error is detected, the receiver sends an acknowledgement (ACK) signal through the return path. On receipt of the (ACK) signal, the transmitter transmits the next codeword.

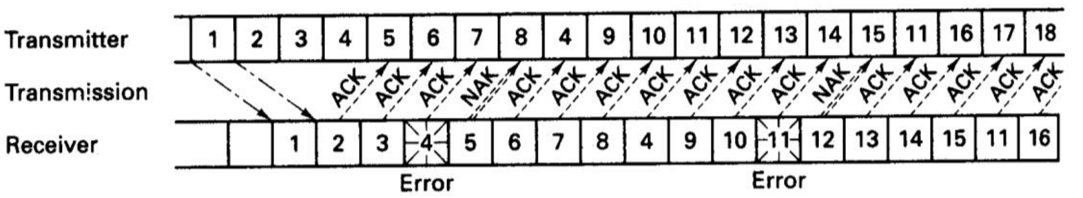
If any error is detected in the received codeword, the decoder sends a Negative Acknowledgement (NAK) to the transmitter. On receipt of the (NAK) signal, the transmitter retransmits the same codeword. The Figure 3.6 (a) illustrates this procedure.



(a) Stop-and-wait ARQ



(b) Go back-N ARQ



(c) Selective repeat ARQ

Figure 3.6 Automatic Repeat Request (ARQ) Procedures

2. Continuous ARQ

Continuous ARQ systems are of two types.

a) Go back–N ARQ Systems.
b) Selective repeat ARQ systems.

a) Go back-N ARQ System

In this procedure, the transmitter sends the message continuously without waiting for an ACK signal from the receiver. Suppose that the decoder detects an error in the k^{th} codeword. Then an NAK signal is sent to the transmitter indicating that the k^{th} codeword is in error.

The transmitter on receiving the NAK signal, goes back to the k^{th} code word and starts transmitting all the codewords from the k^{th} onwards. The figure 3.6(b) illustrates this procedure.

b) Selective repeat ARQ

In this procedure, the transmitter goes on sending the messages one after the other without waiting for an ACK. Suppose that the decoder detects an error in the k^{th}codeword. Then an NAK signal is sent to the transmitter indicating that the k^{th}codeword is in error.

The transmitter on receiving the NAK signal, retransmit only the k^{th} codeword. Then it resumes transmission of the messages in a sequential order starting from where it broke the sequence. The figure 3.6(c) illustrates this procedure.

Advantages of using ARQ

- This method has lower probability of error.
- The stop-and-wait ARQ systems are very simple. They are used on terrestrial microwave links as the round-trip delay is very small.

- Go back N ARQ is useful in satellite links in which the round trip delay is large.
- Selective ARQ provides the best throughput efficiency.
- The major advantage of ARQ over FEC is that error detection requires much simpler decoding equipment and much less redundancy.
- ARQ is an adaptive method, since information is retransmitted only when errors occur.

Disadvantages of using ARQ

- The ARQ system is slow, because of large overall delay.
- Expensive input and output buffers are required.
- The implementation cost is high.

Codes used in ARQ method

In ARQ method, we need a code for the purpose of error detection only. The Cyclic Redundancy Check (CRC) codes are well-suited for error detection. Simple parity check codes are also used for error detection.

3.5 TYPES OF ERRORS

Depending upon the nature of the noise, the codewords transmitted through the channel is affected differently. Hence, there is a possibility that bit 0 transmitted may be received as bit 1 or vice versa. This is called as the error introduced by the noise in the transmitted code word. There are mainly two types of errors introduced during data transmission. They are 1) Random error and 2) Burst error. Both random error and burst errors occur in the contents of a message. Hence they may also be referred to as “content errors”.

Alternatively, it is possible that a data block may be lost in the network as it has been delivered to a wrong destination. It is referred as the “Flow Integrity error”.

3.5.1 Random Error

Random errors are caused by Additive White Gaussian Noise (AWGN) in the channel. Noise affects the transmitted symbols independently. Hence, the errors introduced in the particular interval does not affect the performance of the system in subsequent intervals. The errors are totally uncorrelated. Therefore, they are also called as independent errors. The channels that are mostly influenced by white

Gaussian noise are satellite and deep-space communication links. The use of forward-error correcting codes is well suited for these channels.

3.5.2 Burst Error

Burst errors are caused by Impulse noise in the channel. Impulse noise affects several consecutive bits and errors tend to occur in clusters. Hence the burst errors are dependent on each other in successive message intervals. The channels that are mostly influenced by impulse noise are telephone channels and radio channels.

In telephone channels, burst of errors result from impulse noise on circuits due to lightning, and transients in central office switching equipment. In radio channels, bursts of errors are produced by atmospherics, multi-path fading, and interferences from other users of the frequency band. An effective method for error protection over these channels is based on ARQ method.

3.5.3 Compound Error

In many of the real communication channels, there is a possibility that both the white Gaussian noise and impulse noise will affect the channel. Hence the errors introduced will be of both random (independent) and burst errors. Therefore, if there is a mixture of random and burst errors, then such errors are called as compound errors.

3.6 APPLICATIONS OF ERROR CONTROL CODING TECHNIQUES

1) Coding for White Gaussian Noise (AWGN) Channels:

- Forward Error Correcting (FEC) codes are used for error correction in the channels that can be modeled as White Gaussian noise channels.
- Typical applications include line-of-sight radio links such as satellite and deep space communication links.

2) Coding for compound-error channels:

- Many real communication channels exhibit a mixture of random (independent) and burst errors. We refer to such channels as compound-error channels. Examples are telephone channels and radio channels.
- An effective method for error protection over compound-error channels is based on Automatic Repeat Request (ARQ).

3) Block codes for error Control in data storage

- Block codes are widely used to provide error control for magnetic tapes, mass storage system, magnetic disks, and other data storage systems.

4) Trellis-Coded Modulation (TCM) for efficient utilization of Bandwidth and power

- We obtain the advantages of error correction coding with the penalty of increased bandwidth.
- In communication, we have to efficiently utilize the primary resources of bandwidth and power.
- Trellis-Coded Modulation (TCM) is one way of using error correction coding and efficiently utilizing bandwidth.
- This technique combines convolutional coding and modulation into a single function.
- TCM is applied in the new generation of modems being developed for telephone channel.

3.7 IMPORTANT TERMS USED IN ERROR CONTROL CODING

Codeword:

The encoded block of 'n' bits is called a codeword. It contains message bits (k) and redundant check bits (q).

Block length: The number of bits 'n' after coding is called the block length of the code.

Code rate: The code rate 'r' is defined as the ratio of message bits (k) and the encoder output bits (n). Hence,

$$\text{Code rate, } r = \frac{k}{n} \tag{3.2}$$

where $0 < r < 1$

Code Vector: An 'n' bit code word can be visualized in an N-dimensional space as a vector whose elements or co-ordinates are the bits in the code word.

Code Efficiency: The code efficiency is the ratio of the message bits to the transmitted bits for that block by the encoder. Hence, Code efficiency $= \dfrac{\text{Message bits}}{\text{Transmitted bits}} = \left[\dfrac{k}{n}\text{x}100\right]\%$

Weight of the code: The number of non-zero elements in the transmitted code vector is called code vector weight.

Hamming distance: The hamming distance (d) between the two code vectors is equal to the number of elements in which they differ. Eg. Let X = 101 and Y = 110. Then hamming distance (d) between X and Y code vectors is 2.

Minimum hamming distance:

The smallest hamming distance between the valid code vectors is termed as the minimum hamming distance (d_{min}).

Modulo-2 arithmetic

(i) Addition

Modulo-2 addition is an exclusive OR operation.

$$0 \oplus 0 = 0$$
$$0 \oplus 1 = 1$$
$$1 \oplus 0 = 1$$
$$1 \oplus 1 = 0$$

(ii) Multiplication

Multiplication of binary digits follows AND logic

$$0 \,.\, 0 = 0$$
$$0 \,.\, 1 = 0$$
$$1 \,.\, 0 = 0$$
$$1 \,.\, 1 = 1$$

3.8 LINEAR BLOCK CODES

Block codes are a class of parity check codes that can be characterized by the (n, k) notation. Block codes are used as forward error correction codes. If the block codes exhibit linearity property, then they are called as Linear Block Codes. In a linear code, modulo-2 addition of any two code words will produce another valid codeword. Typical examples of Linear Block codes are

1. Hamming Codes
2. Cyclic Codes
3. BCH Codes
4. Golay Codes
5. Reed-Solomon codes

- In linear block codes, if the redundant bits are added in such a way that the message bits appear first and then redundant bits, then they are called as systematic linear block codes.
- In linear block codes, if the redundant bits are added in such a way that, it is not possible to identify message bits and redundant bits, then they are called as Non-systematic linear block codes.

3.8.1 Principles of linear block codes

The figure 3.6 shows the functional diagram of channel encoder, for generating systematic linear block codes

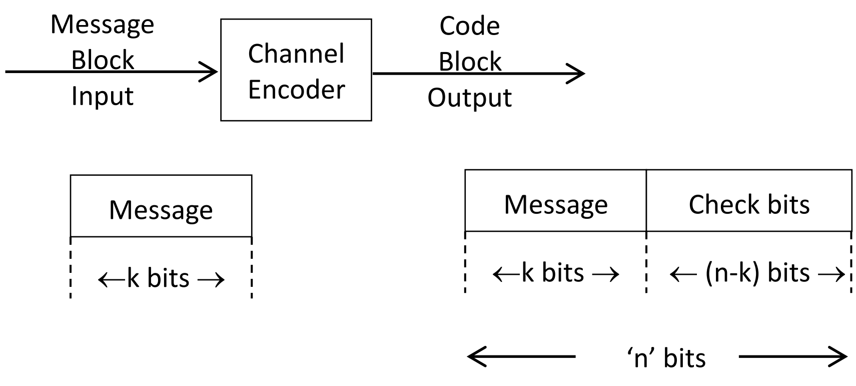


Figure 3.6: Functional diagram of block coder

- The input binary data sequence is divided into block of 'k' message bits. Hence, each block of message can be considered as a message vector 'M' of

$$M = (m_1\ m_2\ \ldots\ldots\ m_k) \tag{3.3}$$

- The channel encoder adds (n-k) check bits (redundant bits) to each block of message input. Let n – k = q. The 'q' check bits are computed from the message bits, according to the prescribed encoding rule. Then the check bits vector 'c' is

$$C = (c_1\ c_2\ \ldots\ c_q) \tag{3.4}$$

- Thus, for each block of 'k' message bits, 'q' check bits are added in the encoder. Hence the code word generated at the output of the encoder has total bits of k + q = n. Therefore, such codes are characterized as (n, k) linear block codes.

- The code vector 'X' at the output of the encoder is

$$X = (m_1\ m_2\ \ldots\ m_k \quad c_1\ c_2\ \ldots\ c_q) = (M \vdots C) \tag{3.5}$$

  Here M is of order 1 x k

  C is of order 1 x q, where q = n – k

  and X is of order 1 x n

- Let G be the generator matrix of order k x n, used by the encoder to generate check bits. Then the output codevector is given by

$$X = M\,.\,G$$

  This can be represented in matrix form as

$$[X]_{1\,\mathrm{x}\,n} = [M]_{1\,\mathrm{x}\,k} \cdot [G]_{k\,\mathrm{x}\,n} \tag{3.6}$$

- The generator matrix depends on the type of linear block code used. It is generally represented as

$$[G]_{k\,\mathrm{x}\,n} = \left[I_{k\,\mathrm{x}\,k}\ \vdots\ P_{k\,\mathrm{x}\,q}\right]_{k\,\mathrm{x}\,n} \tag{3.7}$$

  Where $I_{k\,\mathrm{x}\,k}$ = k x k Identity matrix given by $\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & 1 \end{bmatrix}$

  and $P_{k\,\mathrm{x}\,q}$ = k x q submatrix or coefficient matrix.

$$P_{k\,\mathrm{x}\,q} \text{ is given by } P_{k\,\mathrm{x}\,q} = \begin{bmatrix} P_{11} P_{12} & \cdots & P_{1q} \\ P_{21} P_{22} & \cdots & P_{2q} \\ \vdots \quad \vdots & & \vdots \\ P_{k1} P_{k2} & \cdots & P_{kq} \end{bmatrix} \tag{3.8}$$

- From this submatrix, we can obtain the check bits vector.
  The check bits vector, $[C]_{1 \times q} = [M]_{1 \times k} \cdot [P]_{k \times q}$ (3.9)

- On substituting the matrix form we obtain

$$[c_1 c_2 \cdots c_q]_{1 \times q} = [m_1 m_2 \cdots m_k]_{1 \times k} \begin{bmatrix} P_{11} P_{12} & \cdots & P_{1q} \\ P_{21} P_{22} & \cdots & P_{2q} \\ \vdots \quad \vdots & & \vdots \\ P_{k1} P_{k2} & \cdots & P_{kq} \end{bmatrix}_{k \times q} \quad (3.10)$$

- By solving the above matrix equation, we can obtain the check bits vector. The check bits are

$$\begin{aligned} c_1 &= m_1 p_{11} \oplus m_2\, p_{21} \oplus \ldots\ldots \oplus m_k\, p_{k1} \\ c_2 &= m_1 p_{12} \oplus m_2\, p_{22} \oplus \ldots\ldots \oplus m_k\, p_{k2} \qquad (3.11) \\ &\vdots \\ c_q &= m_1 p_{1q} \oplus m_2\, p_{2q} \oplus \ldots\ldots \oplus m_k p_{kq} \end{aligned}$$

  Here all the additions are mod-2 addition.

Example 3.1

The generator matrix for a block code is given below. Find all code vectors of this code.

$$G = \begin{bmatrix} 1 & 0 & 0 & \vdots & 1 & 0 & 0 \\ 0 & 1 & 0 & \vdots & 0 & 1 & 1 \\ 0 & 0 & 1 & \vdots & 1 & 0 & 1 \end{bmatrix}$$

Solution:

The Generator matrix G is generally represented as

$$[G]_{k \times n} = \begin{bmatrix} I_{k \times k} & \vdots & P_{k \times q} \end{bmatrix}_{k \times n}$$

Hence

$$\begin{bmatrix} I_{k \times k} & \vdots & P_{k \times q} \end{bmatrix}_{k \times n} = \begin{bmatrix} 1 & 0 & 0 & \vdots & 1 & 0 & 0 \\ 0 & 1 & 0 & \vdots & 0 & 1 & 1 \\ 0 & 0 & 1 & \vdots & 1 & 0 & 1 \end{bmatrix}$$

On comparing,

The number of message bits, $k = 3$

The number of code word bits, $n = 6$

The number of check bits, $q = n – k = 6 – 3 = 3$

Hence, the code is a (6, 3) systematic linear block code. From the generator matrix, we have

$$\text{Identity matrix,} \quad I_{k \times k} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\text{The coefficient or submatrix,} \quad P_{k \times q} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

Therefore, the check bits vector is given by,

$$[C]_{1 \times q} = [M]_{1 \times k}[P]_{k \times q}$$

On substituting the matrix form,

$$[c_1\ c_2\ c_3] = [m_1\ m_2\ m_3] \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

From the matrix multiplication, we have

$$c_1 = (1 \times m_1) \oplus (0 \times m_2) \oplus (1 \times m_3)$$

$$c_2 = (0 \times m_1) \oplus (1 \times m_2) \oplus (0 \times m_3)$$

$$c_3 = (0 \times m_1) \oplus (1 \times m_2) \oplus (1 \times m_3)$$

On simplifying, we obtain

$$c_1 = m_1 \oplus m_3$$

$$c_2 = m_2$$

$$c_3 = m_2 \oplus m_3$$

Hence the check bits (c_1 c_2 c_3) for each block of (m_1 m_2 m_3) message bits can be determined.

(i) For the message block of (m_1 m_2 m_3) = (0 0 0), we have

$$c_1 = m_1 \oplus m_3 = 0 \oplus 0 = 0$$
$$c_2 = m_2 = 0$$
$$c_3 = m_2 \oplus m_3 = 0 \oplus 0 = 0$$

(ii) For the message block of (m_1 m_2 m_3) = (0 0 1), we have

$$c_1 = m_1 \oplus m_3 = 0 \oplus 1 = 1$$
$$c_2 = m_2 = 0$$
$$c_3 = m_2 \oplus m_3 = 0 \oplus 1 = 1$$

(iii) For the message block of (m_1 m_2 m_3) = (0 1 0), we have

$$c_1 = m_1 \oplus m_3 = 0 \oplus 0 = 0$$
$$c_2 = m_2 = 1$$
$$c_3 = m_2 \oplus m_3 = 1 \oplus 0 = 1$$

Similarly, we can obtain check bits for other message blocks. The table 3.1 lists all the message bits, their check bits and code word vectors.

Table 3.1 Code Vectors of Example 3.1

Message Vector in one block			**Check bits**			**Code word Vector**					
m_1	**m_2**	**m_3**	**$c_1 = m_1 \oplus m_3$**	**$c_2 = m_2$**	**$c_3 = m_2 \oplus m_3$**	**m_1**	**m_2**	**m_3**	**c_1**	**c_2**	**c_3**
0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	1	0	1	0	0	1	1	0	1
0	1	0	0	1	1	0	1	0	0	1	1
0	1	1	1	1	0	0	1	1	1	1	0
1	0	0	1	0	0	1	0	0	1	0	0
1	0	1	0	0	1	1	0	1	0	0	1
1	1	0	1	1	1	1	1	0	1	1	1
1	1	1	0	1	0	1	1	1	0	1	0

Error detection and correction capability of block codes

The error detection and correction capabilities of a coding technique depend on the minimum hamming distance d_{min} as shown in Table 3.2.

Table 3.2 Error control capabilities

S.No.	Name of errors detected / corrected	Distance requirement
1.	Detect upto 's' errors per code word	$d_{min} \geq s + 1$
2.	Correct upto 't' errors per code word	$d_{min} \geq 2t + 1$
3.	Correct upto 't' errors and detect s > t errors per code word	$d_{min} \geq t + s + 1$

3.8.2 Hamming Codes

Hamming codes are (n, k) linear block codes. They can be generated either systematically or non-systematically.

3.8.2.1 Systematic form of Hamming codes

The systematic form of Hamming codes satisfy the following conditions.

1) Number of check bits, $q \geq 3$
2) Code word length, $n = 2^q - 1$
3) Number of message bits, $k = n - q$
4) Minimum hamming distance, $d_{min} = 3$
5) Since $d_{min} = 3$,

   The error detection capability is

   $$d_{min} \geq s + 1 \quad \Rightarrow 3 \geq s + 1 \quad \Rightarrow s \leq 3 - 1 \quad \Rightarrow s \leq 2$$

   The error correction capability is

   $$d_{min} \geq 2t + 1 \quad \Rightarrow 3 \geq 2t + 1 \quad \Rightarrow 2t \leq 2 \quad \Rightarrow t \leq 1$$

Hence by using Hamming code, we can correct single bit errors and detect errors in two bits.

Parity check matrix (H)

A parity check matrix (H) is used in Hamming codes for encoding and decoding. For every block code there is a q x n parity check matrix (H). It is defined as

$$[H]_{q \times n} \quad = \quad [P^T \vdots I]_{q \times n} \tag{3.12}$$

Where P^T is the Transpose of submatrix

I is the Identity matrix

From equation (3.8), we have the submatrix as

$$P_{k \times q} \quad = \quad \begin{bmatrix} P_{11} P_{12} & \cdots & P_{1q} \\ P_{21} P_{22} & \cdots & P_{2q} \\ \vdots \quad \vdots & & \vdots \\ P_{k1} P_{k2} & \cdots & P_{kq} \end{bmatrix}$$

By changing rows to columns, we can get the transpose of the submatrix.

$$\left[\mathrm{P}_{k \times q}\right]^T \quad = \quad [P^T]_{q \times k}$$

$$[P^T]_{q \times k} \quad = \quad \begin{bmatrix} P_{11} & P_{21} & \cdots & P_{k1} \\ P_{12} & P_{22} & \cdots & P_{k2} \\ \vdots & \vdots & & \vdots \\ P_{1q} & P_{2q} & \cdots & P_{kq} \end{bmatrix} \tag{3.13}$$

On substituting in equation (3.12), the parity check matrix (H) is

$$[H]_{q \times n} \quad = \quad \begin{bmatrix} P_{11} & P_{21} & \cdots & P_{k1} & \vdots & 1 & 0 & 0 & \cdots & 0 \\ P_{12} & P_{22} & \cdots & P_{k2} & \vdots & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ P_{1q} & P_{2q} & \cdots & P_{kq} & \vdots & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}_{q \times n} \tag{3.14}$$

Example 3.2

The parity check matrix of a (7, 4) linear block code is given by

$$H \quad = \quad \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

i. Find the generator matrix (G).
ii. List all the code vectors.
iii. How many errors can be detected?
iv. How many errors can be corrected?
v. Draw the encoder circuit.

Solution:

For a(7,4) linear block code, we have n = 7, k = 4 and q = n – k = 7 – 4 = 3.

Thus n = 2^q – 1 = 2^3 – 1 = 7.

Hence, the given code is Hamming code.

(i) To find the generator matrix:

We know that $[H]_{q \text{ x } n} = [P^T \vdots I]_{q \text{ x } n}$

Given that, $[H]_{3 \text{ x } 7} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$

Then the transpose of submatrix is

$$[P^T]_{q \text{ x } k} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}_{3 \text{ x } 4}$$

The submatrix is given by changing rows to columns

$$[P]_{k \text{ x } q} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}_{4 \text{ x } 3}$$

Therefore, the Generator matrix is given by

$$[G]_{k \text{ x } n} = \left[I_{k \text{ x } k} \;\vdots\; P_{k \text{ x } q} \right]_{k \text{ x } n}$$

$$\Rightarrow \quad [G]_{4 \text{ x } 7} = \begin{bmatrix} 1 & 0 & 0 & 0 & \vdots & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & \vdots & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & \vdots & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & \vdots & 0 & 1 & 1 \end{bmatrix}_{4 \text{ x } 7}$$

(ii) To find all the code words

The check bits vector is given by

$$[C]_{1 \text{ x } q} = [M]_{1 \text{ x } k}[P]_{k \text{ x } q}$$

Hence, $[c_1\ c_2\ c_3] = [m_1\ m_2\ m_3\ m_4] \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$

From the matrix multiplication, we have

$c_1 = m_1 \oplus m_2 \oplus m_3$

$c_2 = m_1 \oplus m_2 \oplus m_4$

$c_3 = m_1 \oplus m_3 \oplus m_4$

We can now determine the check bits (c_1 c_2 c_3) for each block of (m_1 m_2 m_3 m_4) message bits. The generated code words are given in Table 3.3.

Table 3.3: Code vectors of Example 3.2

Sl. No.	Message Vector				Check bits			Code word Vector X							Weight of Code Vector w(X)
	m_1	m_2	m_3	m_4	$c_1= m_1 \oplus m_2 \oplus m_3$	$c_2= m_1 \oplus m_2 \oplus m_4$	$c_3= m_1 \oplus m_3 \oplus m_4$	m_1	m_2	m_3	m_4	c_1	c_2	c_3	
1.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2.	0	0	0	1	0	1	1	0	0	0	1	0	1	1	3
3.	0	0	1	0	1	0	1	0	0	1	0	1	0	1	3
4.	0	0	1	1	1	1	0	0	0	1	1	1	1	0	4
5.	0	1	0	0	1	1	0	0	1	0	0	1	1	0	3
6.	0	1	0	1	1	0	1	0	1	0	1	1	0	1	4
7.	0	1	1	0	0	1	1	0	1	1	0	0	1	1	4
8.	0	1	1	1	0	0	0	0	1	1	1	0	0	0	3
9.	1	0	0	0	1	1	1	1	0	0	0	1	1	1	4
10.	1	0	0	1	1	0	0	1	0	0	1	1	0	0	3
11.	1	0	1	0	0	1	0	1	0	1	0	0	1	0	3
12.	1	0	1	1	0	0	1	1	0	1	1	0	0	1	4
13.	1	1	0	0	0	0	1	1	1	0	0	0	0	1	3
14.	1	1	0	1	0	1	0	1	1	0	1	0	1	0	4
15.	1	1	1	0	1	0	0	1	1	1	0	1	0	0	4
16.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	7

The smallest weight of any non-zero code vector is 3. Hence the minimum hamming distance is $d_{min} = 3$.

(iii) Error detection capability

The error detection upto 's' errors per word is

$$d_{min} \geq s + 1$$

The minimum distance is $d_{min} = 3$

Therefore, $3 \geq s + 1$

$s \leq 3 - 1$

$\Rightarrow \quad s \leq 2$

Thus error in two bits will be detected.

(iv) Error correction capability

The error correction upto 't' errors per word is

$$d_{min} \geq 2t + 1$$

Hence, $3 \geq 2t + 1$

$2t \leq 3 - 1$

$2t \leq 2$

$\Rightarrow \quad t \leq 1$

Thus error in one bit will be corrected.

(v) Encoder circuit

Figure 3.7 shows the encoder circuit of the given (7, 4) Hamming code.

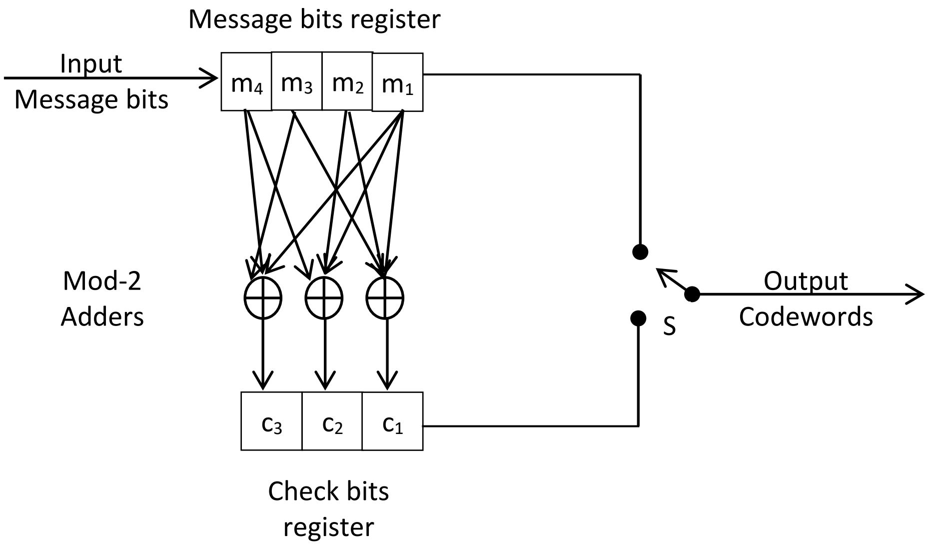


Figure 3.7 Encoder of (7, 4) Hamming code

3.8.2.2 Non-Systematic form of Hamming Code:

In a non-systematic block code, it is not possible to identify message bits and check bits. They are mixed in the block. The error detection and correction capability of non-systematic hamming code can be explained by an example.

Example 3.3

Consider a data (message) block of 1 1 0 1. The hamming code adds three parity bits to the message bits in such a way that both message bits and check bits get mixed. The check bit locations are as shown below.

1	2	3	4	5	6	7	→ Bit location
p_1	p_2	D	p_3	D	D	D	

- Here p_1, p_2 and $p_3$represent the parity check bits to be added. D represents the data (message) bits. Then we have

1	2	3	4	5	6	7
p_1	p_2	1	p_3	1	0	1

- The first parity bit, p_1 provides even parity from a check of bit locations 3, 5 and 7. Here they are 1, 1 and 1 respectively. Hence p_1 will therefore be 1 to achieve even parity.
- The second parity bit, p_2 checks locations 3, 6 and 7. Here they are 1, 0 and 1 respectively. Hence p_2 will be 0 to achieve even parity.
- The third parity bit p_3, checks locations 5, 6 and 7. Here they are 1, 0 and 1 respectively. Hence p_3 will be 0 to achieve even parity.
- The resulting 7-bit code word generated is as below.

1	2	3	4	5	6	7	
p_1	p_2	D	p_3	D	D	D	
1	0	1	0	1	0	1	→Code word transmitted

- Suppose that this code word is altered during transmission. Assume that location 5 changes from 1 to 0. Hence the received code word with error is given below.

1	2	3	4	5	6	7
p_1	p_2	D	p_3	D	D	D
1	0	1	0	0	0	1

- At the decoder, we have to evaluate the parity bits to determine where error occurs. This is accomplished by assigning a 1 to any parity bit which is incorrect and a 0 to the parity bit which is correct.
- We check p_1 for locations 3, 5 and 7. Here they are 1, 0 and 1. For even parity p_1 should be 0, but we have received p_1 as 1, which is incorrect. We assign a 1 to p_1.
- We check p_2 for locations 3, 6 and 7. Here they are 1, 0 and 1 respectively. For even parity p_2 should be 0 and we have also received p_2 as 0, which is correct. We assign a 0 to p_2.
- We check p_3 for locations 5, 6 and 7. Here they are 0, 0 and 1 respectively. For even parity p_3 should be 1, but we have received p_3 as 0, which is incorrect. We assign a 1 to p_3.

- The three assigned values result in the binary form of 1 0 1, which has a decimal value of 5. This means that the bit location containing the error is 5. The decoder then change the 5th location bit from 0 to 1.
- The hamming code is therefore capable of locating a single error. But it fails if multiple errors occur in one data block.

3.8.3 Binary Cyclic Codes

Binary cyclic codes are a subclass of the linear block codes. They have very good features which make them extremely useful. Cyclic codes can correct errors caused by bursts of noise that affect several successive bits. The very good block codes like the Hamming codes, BCH codes and Golay codes are also cyclic codes.

A linear block code is called as cyclic code if every cyclic shift of the code vector produces another code vector.

A cyclic code exhibits the following two properties.

(i) **Linearity Property:** A code is said to be linear if modulo-2 addition of any two code words will produce another valid codeword.

(ii) **Cyclic Property:** A code is said to be cyclic if every cyclic shift of a code word produces another valid code word. For example, consider the n-bit code word, $X = (x_{n-1}, x_{n-2}, \ldots\ldots x_1, x_0)$.

If we shift the above code word cyclically to left side by one bit, then the resultant code word is

$$X' = (x_{n-2}, x_{n-3}, \ldots\ldots x_1, x_0, x_{n-1})$$

Here X' is also a valid code word. One more cyclic left shift produces another valid code vector X''.

$$X'' = (x_{n-3}, x_{n-4}, \ldots\ldots x_1, x_0, x_{n-1}, x_{n-2})$$

Representation of codewords by a polynomial

- The cyclic property suggests that we may treat the elements of a code word of length n as the coefficients of a polynomial of degree (n-1).
- Consider the n-bit code word,

$$X = (x_{n-1}, x_{n-2}, \ldots\ldots x_1, x_0) \tag{3.13}$$

This code word can be represented in the form of a code word polynomial as below:

$$X(p) = x_{n-1}p^{n-1} + x_{n-2}p^{n-2} + \ldots\ldots + x_1 p + x_0 \tag{3.14}$$

where X (p) is the polynomial of degree (n-1).

p is an arbitrary real variable. For binary codes, the coefficients are 1s or 0s.

- The power of 'p' represents the positions of the code word bits. i.e., p^{n-1} represents MSB and p^0 represents LSB.
- Each power of p in the polynomial X(p) represents a one-bit cyclic shift in time. Hence, multiplication of the polynomial X(p) by p may be viewed as a cyclic shift or rotation to the right, subject to the constraint that $p^n = 1$.
- We represent cyclic codes by polynomial representation because of the following reasons.
  1. These are algebraic codes. Hence algebraic operations such as addition, subtraction, multiplication, division, etc. becomes very simple.
  2. Positions of the bits are represented with the help of powers of p in a polynomial.

3.8.3.1 Generation of code vectors in non-systematic form of cyclic codes

- Let M = (m_{k-1}, m_{k-2}, …… m_1, m_0) be 'k' bits of message vector. Then it can be represented by the polynomial as,

$$M(p) = m_{k-1}p^{k-1} + m_{k-2}p^{k-2} + \ldots\ldots + m_1p + m_0 \tag{3.15}$$

- The codeword polynomial X(P) is given as

$$X(P) = M(P) \cdot G(P) \tag{3.16}$$

where G(P) is called as the generating polynomial of degree 'q' (parity or check bits q = n – k). The generating polynomial is given as

$$G(P) = p^q + g_{q-1}p^{q-1} + \ldots\ldots.. + g_1p + 1 \tag{3.17}$$

Here g_{q-1}, g_{q-2} ….. g_1 are the parity bits.

- If M_1, M_2, M_3, …… etc. are the other message vectors, then the corresponding code vectors can be calculated as,

$$X_1(P) = M_1(P)\, G(P) \tag{3.18}$$

$$X_2(P) = M_2(P)\, G(P)$$

$$X_3(P) = M_3(P)\, G(P) \text{ and so on}$$

- All the above code vectors X_1, X_2, X_3 ….. are in non-systematic form and they satisfy cyclic property.

Example 3.3

The generator polynomial of a (7, 4) cyclic code is $G(p) = p^3 + p + 1$. Find all the code vectors for the code in non-systematic form.

Solution:

Here n = 7, k = 4

Therefore, q = n – k = 7 – 4 = 3

Since k = 4, there will be a total of $2^k = 2^4 = 16$ message vectors (From 0 0 0 0 to 1 1 1 1). Each can be coded in to a 7 bits codeword.

(i) Consider any message vector as

$$M = (m_3\ m_2\ m_1\ m_0) = (1\ 0\ 0\ 1)$$

The general message polynomial is

$$M(p) = m_3p^3 + m_2p^2 + m_1p + m_0, \text{ for } k = 4$$

For the message vector (1 0 0 1), the polynomial is

$$M(p) = 1 . p^3 + 0 . p^2 + 0 . p + 1$$

$$\Rightarrow \quad M(p) = P^3 + 1$$

The given generator polynomial is $G(p) = p^3 + p + 1$

In non-systematic form, the codeword polynomial is

$$X(p) = M(p) . G(p)$$

On substituting,

$$X(p) = (p^3 + 1) . (p^3 + p + 1)$$

$$= p^6 + p^4 + p^3 + p^3 + p + 1$$

$$= p^6 + p^4 + (1 \oplus 1)\, p^3 + p + 1 \qquad = \qquad p^6 + p^4 + p + 1$$

$$= 1.p^6 + 0.p^5 + 1.p^4 + 0.p^3 + 0.p^2 + 1.p + 1$$

The code vector corresponding to this polynomial is

$$X = (x_6\ x_5\ x_4\ x_3\ x_2\ x_1\ x_0)$$

$$\Rightarrow \quad X = (\ 1\ \ 0\ \ 1\ \ 0\ \ 0\ \ 1\ \ 1)$$

(ii) Consider another message vector as

$$M = (m_3\ m_2\ m_1\ m_0)$$

$$= (\ 0\ \ 1\ \ 1\ \ 0\ )$$

The polynomial is $M(p) = 0.p^3 + 1.p^2 + 1.p^1 + 0.1$

$\Rightarrow \quad M(p) \quad = \quad p^2 + p$

The codeword polynomial is X(p) = M(p) . G(p)

$\Rightarrow \quad X(p) \quad = \quad (p^2 + p) . (p^3 + p + 1)$

$= \quad p^5 + p^3 + p^2 + p^4 + p^2 + p$

$= \quad p^5 + p^4 + p^3 + (1 \oplus 1)\, p^2 + p$

$= \quad p^5 + p^4 + p^3 + p$

$= \quad 0.p^6 + 1.p^5 + 1.p^4 + 1.p^3 + 0.p^2 + 1.p + 0.1$

The codevector, X = (0 1 1 1 0 1 0)

Similarly, we can find code vector for other message vectors also, using the same procedure.

3.8.3.2 Generation of code vectors in systematic form of cyclic codes

- The code word for the systematic form of cyclic codes is given by

$X \quad = \quad (\text{ k message bits} \vdots \text{q check bits})$

$$\Rightarrow \quad X \quad = \quad (m_{k-1}\, m_{k-2} \ldots\ldots\, m_1\, m_0 \vdots\, c_{q-1}\, c_{q-2} \ldots\, c_1\, c_0) \qquad (3.19)$$

- In polynomial form, the check bits vector can be written as

$$C(p) \quad = \quad c_{q-1}p^{q-1} + c_{q-2}p^{q-2} + \ldots\ldots + c_1 p + c_0 \qquad (3.20)$$

- In systematic form, the check bits vector polynomial is obtained by

$$C(p) \quad = \quad \text{rem}\left[\frac{p^q.M(p)}{G(p)}\right] \qquad (3.21)$$

where M(p) is message polynomial

G(p) is generating polynomial

'rem' is remainder of the division

Example 3.4

The generator polynomial of a (7, 4) cyclic code is $G(p) = p^3 + p + 1$. Find all the codevectors for the code in systematic form.

Solution:

Here n = 7, k = 4

Therefore, q = n - k = 7 – 4 = 3

Since k = 4, there will be a total of $2^k = 2^4 = 16$ message vectors (From 0 0 0 0 to 1 1 1 1). Each can be coded into a 7 bits codeword.

(i) Consider any message vector as

M = $(m_3\ m_2\ m_1\ m_0)$ = (1 1 1 0)

By message polynomial, $M(p) = m_3p^3 + m_2p^2 + m_1p + m_0$, for k = 4.

For the message vector (1 1 1 0), the polynomial is

$$M(p) = 1.p^3 + 1.p^2 + 1.p + 0.1$$

$$\Rightarrow \quad M(p) = p^3 + p^2 + p$$

The given generator polynomial is

$$G(p) = p^3 + p + 1$$

The check bits vector polynomial is

$$C(p) = \text{rem}\left[\frac{p^q.M(p)}{G(p)}\right]$$

$$= \text{rem}\left[\frac{p^3(p^3+p^2+p)}{(p^3+p+1)}\right]$$

$$= \text{rem}\left[\frac{p^6+p^5+p^4}{p^3+p+1}\right]$$

We perform division as per the following method.

$$\begin{array}{r|l}
 & p^3 + p^2 \\
\hline
(p^3 + p + 1) & p^6 + p^5 + p^4 + 0.p^3 + 0.p^2 + 0.p + 0.1 \\
 & p^6 + 0.p^5 + p^4 + p^3 \\
\hline
\text{(Mod-2 addition)} & 0.p^6 + p^5 + 0.p^4 + p^3 + 0.p^2 + 0.p + 0.1 \\
 & \quad p^5 + 0.p^4 + p^3 + p^2 \\
\hline
 & \quad 0.p^5 + 0.p^4 + 0.p^3 \underbrace{+ p^2 + 0.p + 0.1}_{\text{remainder}}
\end{array}$$

Thus the remainder polynomial is $p^2 + 0.p + 0.1$.

This is the check bits polynomial C(p)

$$\therefore c(p) = p^2 + 0.p + 0.1$$

The check bits are c = (1 0 0)

Hence the code vector for the message vector (1 1 1 0) in systematic form is

$$X = (m_3\ m_2\ m_1\ m_0 \vdots\ c_2\ c_1\ c_0) = (\ 1\ 1\ 1\ 0\ 1\ 0\ 0\ )$$

(ii) Consider another message vector as

M = $(m_3\ m_2\ m_1\ m_0)$ = (1 0 1 0)

The message polynomial is $M(p) = p^3 + p$

Then the check bits vector polynomial is

$$C(p) = \text{rem}\left[\frac{p^q.M(p)}{G(p)}\right]$$

$$= \text{rem}\left[\frac{p^3(p^3+p)}{(p^3+p+1)}\right]$$

$$= \text{rem}\left[\frac{p^6+p^4}{p^3+p+1}\right]$$

The division is performed as below.

$$
\begin{array}{r|l}
 & p^3 + 1 \\ \hline
(p^3 + p + 1) & p^6 + 0.p^5 + p^4 + 0.p^3 + 0.p^2 + 0.p + 0.1 \\
 & p^6 + 0.p^5 + p^4 + \quad p^3 \\ \hline
\text{(Mod-2 addition)} & 0.p^6 + 0.p^5 + 0.p^4 + p^3 + 0.p^2 + 0.p + 0.1 \\
 & \qquad\qquad\qquad p^3 + 0.p^2 + \quad p + 1 \\ \hline
 & \qquad\qquad\quad 0.p^3 + \underbrace{0.p^2 + \quad p + 1}_{\text{remainder}}
\end{array}
$$

Thus the check bits polynomial is

$C(p) = 0.p^2 + 1.p + 1.1$

The check bits are $C = (0\ 1\ 1)$

Hence the codevector is $X = (1\ 0\ 1\ 0\ 0\ 1\ 1)$

Similarly, we can find code vector for other message vectors also, using the same procedure.

Advantages of Cyclic Codes

- Cyclic codes can correct burst errors that span many successive bits.
- They have an excellent mathematical structure. This makes the design of error correcting codes with multiple-error correction capability relatively easier.
- The encoding and decoding circuits for cyclic codes can be easily implemented using shift registers.
- The error correcting and decoding methods of cyclic codes are simpler and easy to implement. These methods eliminate the storage (large memories) needed for lookup table decoding. Therefore the codes become powerful and efficient.

Disadvantages of cyclic codes

- Even though the error detection is simpler, the error correction is slightly more complicated. This is due to the complexity of the combinational logic circuit used for error correction.

3.8.4 Cyclic Redundancy Check Code (CRC)

Cyclic codes are extremely well-suited for error detection. Because they can be designed to detect many combinations of likely errors. Also, the implementation of both encoding and error detecting circuits is practical. For these reasons, all the error detecting codes used in practice are of cyclic code type. Cyclic Redundancy Check (CRC) code is the most important cyclic code used for error detection in data networks & storage systems. CRC code is basically a systematic form of cyclic code.

CRC Generation (encoder)

The CRC generation procedure is shown in the figure 3.8.

- First we append a string of 'q' number of 0s to the data sequence. For example, to generate CRC-6 code, we append 6 number of 0s to the data.
- We select a generator polynomial of (q+1) bits long to act as a divisor. The generator polynomials of three CRC codes have become international standards. They are

(i)	CRC – 12 code	: $p^{12} + p^{11} + p^3 + p^2 + p + 1$
(ii)	CRC – 16 code	: $p^{16} + p^{15} + p^2 + 1$
(iii)	CRC – CCITT Code	: $p^{16} + p^{12} + p^5 + 1$

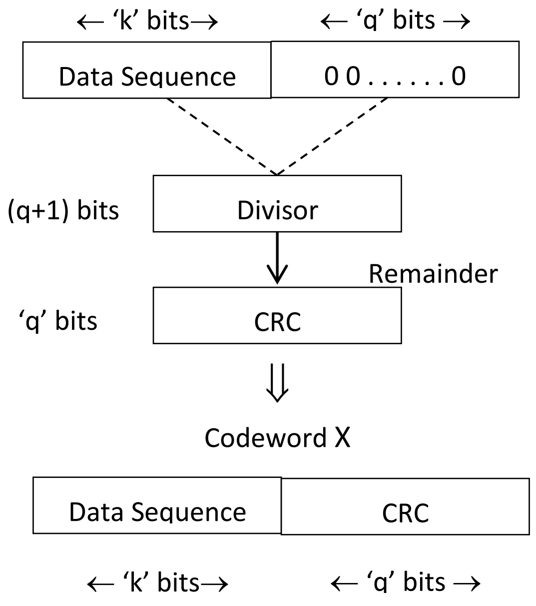


Figure 3.8 CRC Generation

- We divide the data sequence appended with 0s by the divisor. This is a binary division.
- The remainder obtained after the division is the 'q' bit CRC. Then, this 'q' bit CRC is appended to the data sequence. Actually CRC is a sequence of redundant bits.
- The code word generated is now transmitted.

CRC checker

The CRC checking procedure is shown in the figure 3.9

- The same generator polynomial (divisor) used at the transmitter is also used at the receiver.

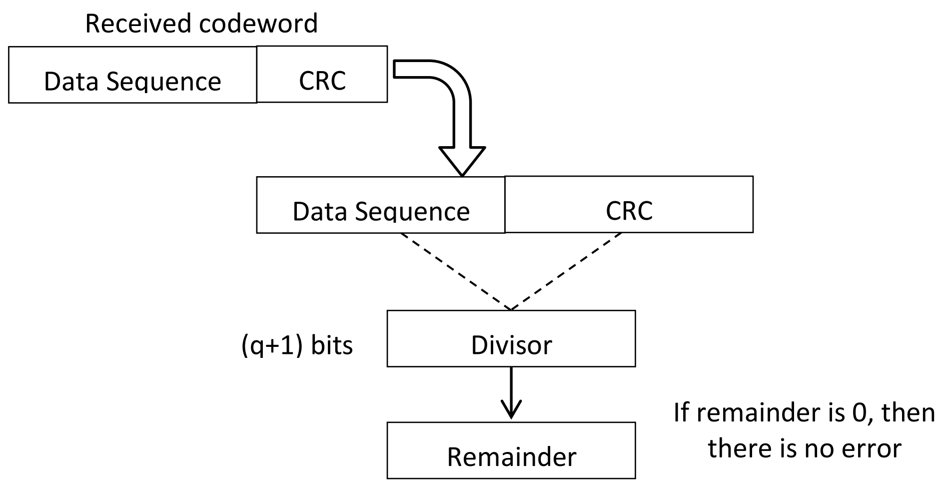


Figure 3.9 CRC checker

- We divide the received code word by the divisor. This is also a binary division.
- If the remainder is all 0s, then there are no errors in the received codeword, and hence must be accepted.
- If we have a non-zero remainder, then we infer that error has occurred in the received code word. Then this received code word is rejected by the receiver and an ARQ signalling is done to the transmitter.

Example 3.5

Generate the CRC code for the data word of 1 1 1 0. The divisor polynomial is $p^3 + p + 1$

Solution

Data Word (Message bits) = 1 1 1 0

Generator Polynomial (divisor) = $p^3 + p + 1$

Divisor in binary form = 1 0 1 1

The divisor will be of (q + 1) bits long.

Here the divisor is of 4 bits long.

Hence q = 3. We append three 0s to the data word.

Now the data sequence is 1 1 1 0 0 0 0. We divide this data by the divisor of 1 0 1 1. Binary division is followed.

```
          1 1 0 0
        ___________________
1 0 1 1 | 1 1 1 0 0 0 0
          1 0 1 1
          -----------------
          0 1 0 1 0
            1 0 1 1
          -----------------
            0 0 0 1 0 0
                  Remainder
```

The remainder obtained from division is 100. Then the transmitted codeword is 1 1 1 0 1 0 0.

Example 3.6

A codeword is received as 1 1 1 0 1 0 0. The generator (divisor) polynomial is $p^3 + p + 1$. Check whether there is error in the received codeword.

Solution

Received Codeword = 1 1 1 0 1 0 0

Divisor in binary form = 1 0 1 1

We divide the received codeword by the divisor.

```
          1 1 0 0
        ___________________
1 0 1 1 | 1 1 1 0 1 0 0
          1 0 1 1
          -----------------
          0 1 0 1 1
            1 0 1 1
          -----------------
            0 0 0 0 0 0
                  Remainder
```

The remainder obtained from division is zero. Hence there is no error in the received codeword.

Applications of CRC codes:

1. CRC codes are mainly used in ARQ system for error detection.
2. They are used in data networks, digital subscriber lines, storage systems etc.

i. CRC – 6
   CRC – 8 – WCDMA
   CRC – 10– CDMA 2000 } Mobile networks
   CRC – 16– CDMA 2000

ii. CRC – 8 → Digital Video Broadcasting

iii. CRC – CCITT → Bluetooth, XMODEM, X-25, V-41, HDLC FCS

iv. CRC – 32 → ANSI x3.66, ITU.T V.42, Ethernet, HDLC, MPEG-2, PNG

3.9 CONVOLUTIONAL CODES

In block coding, the encoder accepts a k-bit message block and generates an n-bit code word. Thus code words are produced on a block-by-block basis. Therefore, a buffer is required in the encoder to place the message block.

A subclass of Tree codes is convolutional codes. The convolutional encoder accepts the message bits continuously and generates the encoded codeword sequence continuously. Hence there is no need for buffer. But in convolutional codes, memory is required to implement the encoder.

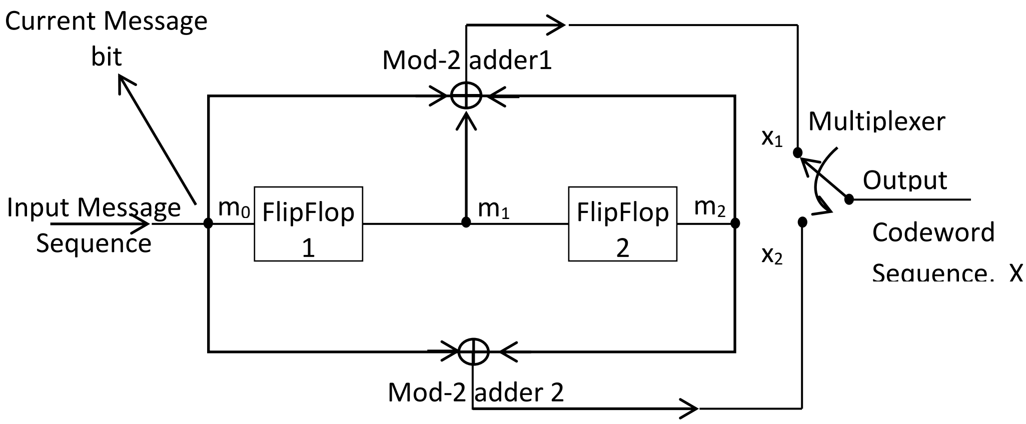


Figure 3.10(a) Convolutional Encoder

In a convolutional code, the encoding operation is the discrete-time convolution of the input data sequence with the impulse response of the encoder. The input message bits are stored in the fixed length shift register and they are combined with the help of mod-2 adders. This operation is equivalent to binary convolution and hence it is called convolutional coding. The figure 3.10 shows the connection diagram for an example convolutional encoder.

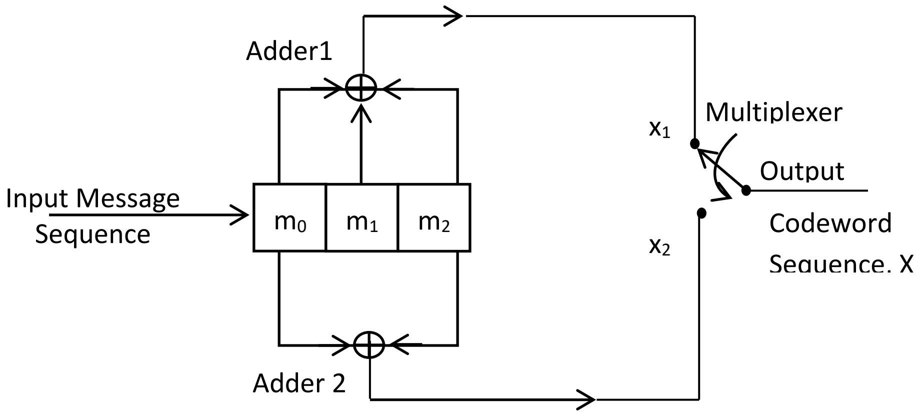


Figure 3.10(b) Convolutional Encoder redrawn alternatively

The encoder of a binary convolution code may be viewed as a finite-state machine. It consists of M-stage shift register with prescribed connections to modulo-2 adders. A multiplexer serializes the outputs of the adders. The convolutional codes generated by these encoders of Figure 3.10 are non-systematic form.

Consider that the current message bit is shifted to position m_0. Then m_1 and m_2 store the previous two message bits. Now, by mod-2 adders 1 and 2 we get the new values of X_1 and X_2. We can write

$$X_1 = m_0 \oplus m_1 \oplus m_2$$

and $$X_2 = m_0 \oplus m_2$$

The multiplexer switch first samples X_1 and then X_2. Then next input bit is taken and stored in m_0. The shift register then shifts the bit already in m_0 to m_1. The bit already in m_1 is shifted to m_2. The bit already in m_2 is discarded. Again X_1 and X_2 are generated according to this new combination of m_0, m_1 and m_2. This process is repeated for each input message bit. Thus the output bit stream for successive input bits will be,

$$X = x_1\, x_2\, x_1\, x_2\, x_1\, x_2 \ldots\ldots \text{ and so on}$$

In this convolutional encoder, for every input message bit, two encoded output bits X1 and X2 are transmitted. Hence number of message bits, k = 1. The number of encoded output bits for one message bit, n = 2.

Code rate:

The code rate of this convolutional encoder is given by

$$\text{Code rate, r} = \frac{Message\ bits\ (k)}{encoder\ output\ bits\ (n)} = \frac{k}{n} = \frac{1}{2}$$

where $0 < r < 1$

Constraint Length:

The constraint length (K) of a convolution code is defined as the number of shifts over which a single message bit can influence the encoder output. It is expressed in terms of message bits.

For the encoder of Figure 3.10, constraint length is K = 3 bits. Because, a single message bit influences encoder output for three successive shifts. At the fourth shift, the message bit is lost and it has no effect on the output.

For the encoder of Figure 3.10, whenever a particular message bit enters the shift register, it remains in the shift register for three shifts ie.,

First Shift → Message bit is entered in position m_0.

Second Shift → Message bit is shifted in position m_1.

Third Shift → Message bit is shifted in position m_2.

Constraint length 'K' is also equal to one plus the number of shift registers required to implement the encoder.

Dimension of the Code

The code dimension of a convolutional code depends on the number of message bits 'k', the number of encoder output bits, 'n' and its constraint length 'K'. The code dimension is therefore represented by (n, k, K).

For the encoder shown in figure 3.10, the code dimension is given by (2, 1, 3) where n = 2, k = 1 and constraint length K = 3.

Graphical representation of convolutional codes

Convolutional code structure is generally presented in graphical form by the following three equivalent ways.

1. By means of the state diagram
2. By drawing the code trellis
3. By drawing the code tree

These methods can be better explained by using an example.

Example 3.7

For the convolutional encoder given below in Figure 3.11, determine the following.

a) Code rate b) Constraint length c) Dimension of the code d) Represent the encoder in graphical form.

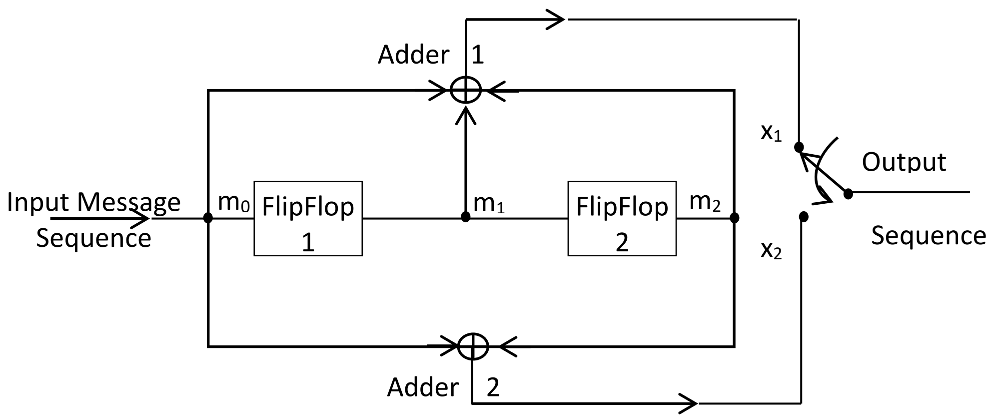


Figure 3.11 Convolutional Encoder Example

Solution:

a) Code rate:

The code rate, r = $\frac{k}{n}$

The number of message bits, k = 1.

The number of encoder output bits, n = 2.

Hence code rate, r = $\frac{1}{2}$

b) Constraint length:

Constraint length, k = 1 + number of shift registers.

Hence k = 1 + 2 = 3

c) Code dimension:

Code dimension = (n, k,K) = (2, 1, 3)

Hence the given encoder is of ½ rate convolutional encoder of dimension (2, 1, 3).

d) Graphical form representation

The encoder output is X = (x_1 x_2 x_1 x_2 x_1 x_2 ….. and so on)

The Mod-2 adder 1 output is $x_1 = m_0 \oplus m_1 \oplus m_2$

The Mod-2 adder 2 output is $x_2 = m_0 \oplus m_2$.

We can represent the encoder output for possible input message bits in the form of a Logic table.

Table 3.4 Logic Table

	Input Message bit	Present State		Next State		Encoder Output	
	m_0	m_1	m_2	m_1	m_2	x_1	x_2
A	0	0	0	0	0	0	0
	1	0	0	1	0	1	1
B	0	1	0	0	1	1	0
	1	1	0	1	1	0	1
C	0	0	1	0	0	1	1
	1	0	1	1	0	0	0
D	0	1	1	0	1	0	1
	1	1	1	1	1	1	0

Output $x_1 = m_0 \oplus m_1 \oplus m_2$ and $x_2 = m_0 \oplus m_2$

- The encoder output depends on the current input message bit and the contents in the shift register ie., the previous two bits.
- The present condition of the previous two bits in the shift register may be in four combinations. Let these combinations 00, 10, 01 and 11 be corresponds to the states A, B, C and D respectively.
- For each input message bit, the present state of the m_1 and m_2 bits will decide the encoded output.
- The logic table presents the encoded output x_1 and x_2 for the possible '0' or '1' bit input if the present state is A or B or C or D.

1. State diagram representation

- The state of a convolutional encoder is defined by the contents of the shift register. The number of states is given by $2^{k-1} = 2^{3-1} = 2^2 = 4$. Here K represents the constraint length.
- Let the four states be A = 00, B = 10, C = 01 and D = 11 as per the logic table. A state diagram as shown in the figure 3.12 illustrates the functioning of the encoder.

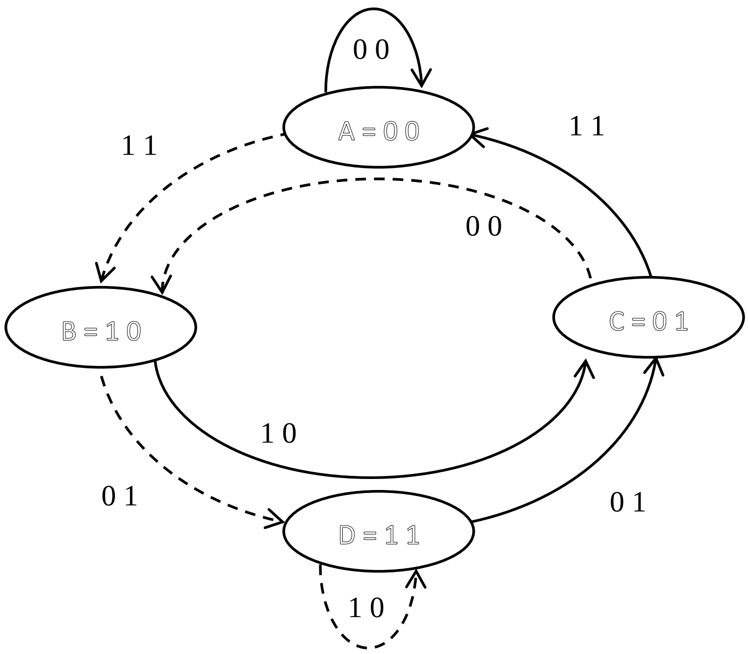


Figure 3.12 State diagram

- Suppose that the contents of the shift register is in the state A = 00. At this state, if the incoming message bit is 0, the encoder output is X = (x_1 x_2) = 00. Then if the m_0, m_1 and m_2 bits are shifted, the contents of the shift register will also be in state A = 00. This is represented by a solid line path starting from A and ending at A itself.
- At the 'A' state, if the incoming message bit is 1, then the encoder output is X = 11. Now if the m_0, m_1 and m_2 bits are shifted, the contents of the shift register will become the state B = 10. This is represented by a dashed line path starting from A and ending at B.
- Similarly we can draw line paths for all other states, as shown in the Figure 3.12.

2. Code tree representation

- The code tree diagram is a simple way of describing the encoding procedure. By traversing the diagram from left to right, each tree branch depicts the encoder output codeword.

- Figure 3.13 shows the code representation for this encoder.

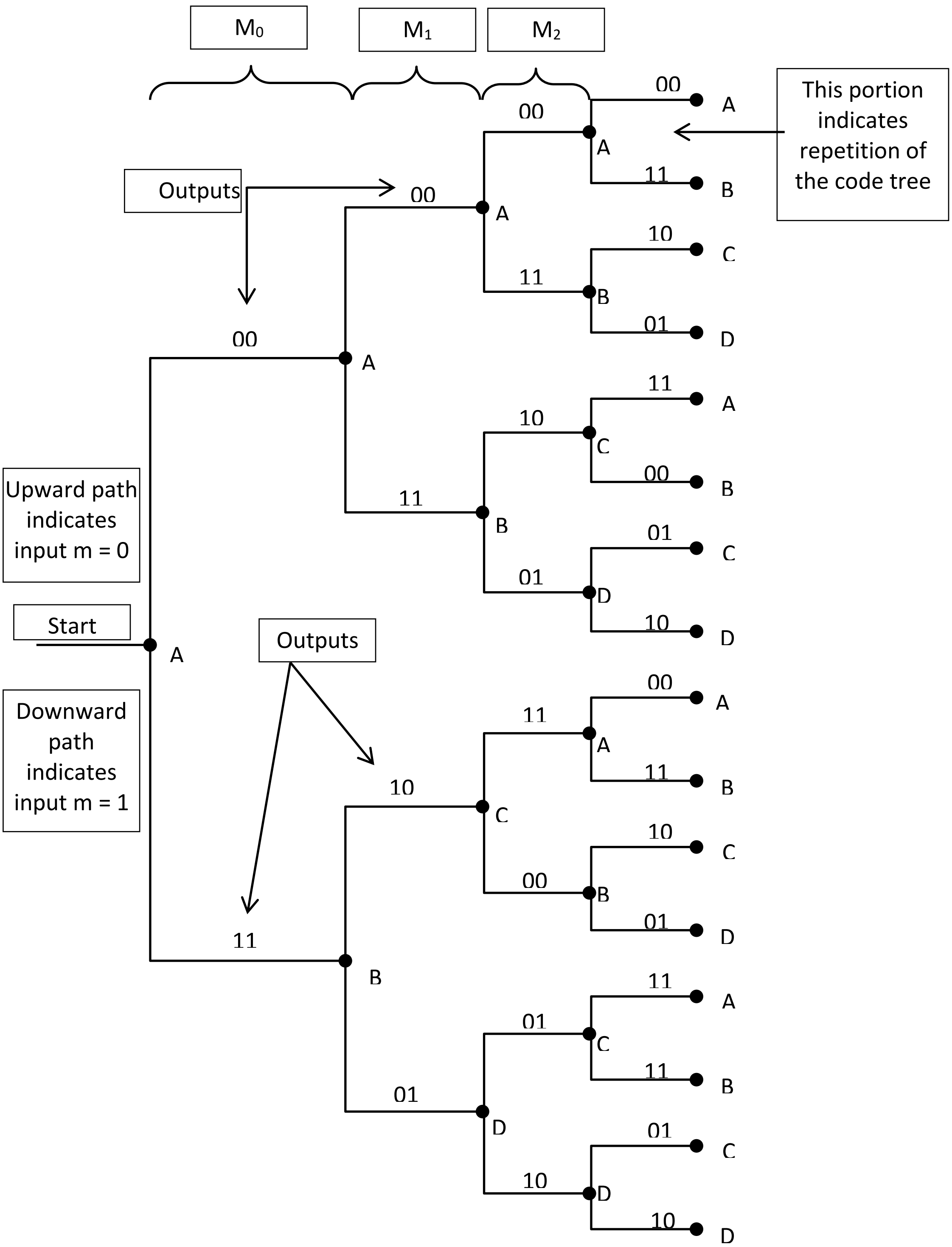


Figure 3.13 Code tree representation

- The code tree diagram starts at state A =00. Each state now represents the node of a tree. If the input message bit is m_0 = 0 at node A, then path of the tree goes upward towards node A and the encoder output is 00. Otherwise if the input message bit is m_0 = 1, the path of the tree goes down towards node B and the encoder output is 11.
- Similarly depending upon the input message bit, the path of the tree goes upward or downward. On the path between two nodes the outputs are shown.
- In the code tree, the branch pattern begins to repeat after third bit, since particular message bit is stored in the shift registers of the encoder for three shifts.

3. Code trellis representation

- Code trellis is the more compact representation of the code tree. In the code tree there are four states (or nodes). Every state goes to some other state depending upon the input message bit.
- Code trellis represents the single and unique diagram for such steady state transitions. The Figure 3.14 shows the code trellis diagram.

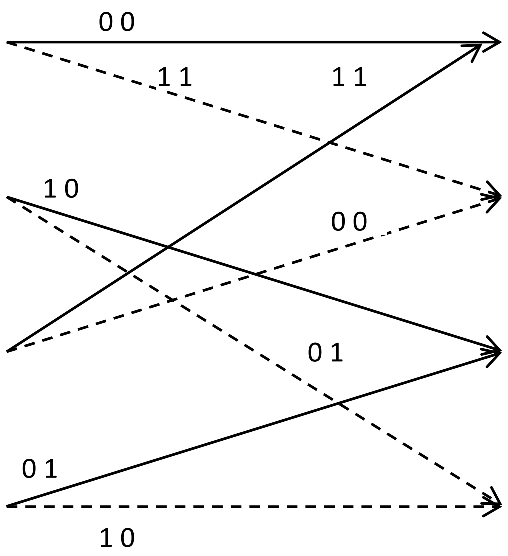
Present (Current) State	Output	Next State
0 0 = A		A = 0 0
1 0 = B		B = 1 0
0 1 = C		C = 0 1
1 1 = D		D = 1 1

Fig.3.14 Code Trellis

- The nodes on the left denote four possible current states and those on the right represents next state. The solid transition line represents for input message m_0 = 0 and dashed line represents input message m_0 = 1.
- Along with each transition line, the encoder output x_1 x_2 is represented during that transition.

Advantages of convolutional codes

- The convolutional codes operate on smaller blocks of data. Hence decoding delay is small.
- The storage hardware required is less.

Disadvantages of convolutional codes

- Due to complexity, the convolutional codes are difficult to analyse.
- These codes are not developed much as compared to block codes.

Comparison between Linear Block codes and Convolutional codes

Table 3.5 Comparison

Sl. No.	Linear Block Codes	Convolutional Codes
1.	Block codes are generated by X = MG (or) X(p) = M(p) .G(p)	Convolutional codes are generated by convolution between message sequencing and generating sequence.
2.	For a block of message bits, encoded block (code vector) is generated	Each message bit is encoded separately. For every message bit, two or more encoded bits are generated.
3.	Coding is block by block.	Coding is bit by bit.
4.	Syndrome decoding is used for most liklihood decoding.	Viterbi decoding is used for most liklihood decoding.
5.	Generator matrices, parity check matrices and syndrome vectors are used for analysis.	Code tree, code trellis and state diagrams are used for analysis.
6.	Generating polynomial and generator matrix are used to get code vectors.	Generating sequences are used to get code vectors.
7.	Error correction and detection capability depends upon minimum distance d_{min}.	Error correction and detection capability depends upon free distance d_{min}.

Exercise Problems:

1. The generator matrix for a (6, 3) block code is given below. Find all code vectors of this code.

$$G = \begin{bmatrix} 1 & 0 & 0 & \vdots & 0 & 1 & 1 \\ 0 & 1 & 0 & \vdots & 1 & 0 & 1 \\ 0 & 0 & 1 & \vdots & 1 & 1 & 0 \end{bmatrix}$$

2. The parity check matrix of a (7, 4) Hamming code is given below

$$H = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

   i. Find the code words generated
   ii. Find the error detection and correction capability
   iii. Draw the encoder circuit.

3. The generator polynomial of a (7, 4) cyclic code is $G(p) = p^3 + p^2 + 1$. Find all the code vectors for the code in both systematic and non-systematic form.

4. Determine CRC for the data sequence of 1 1 0 1 0 1 0 1 1, if the generator polynomial is $p^4 + p + 1$.

5. A code word is received as 1 0 1 1 1 1 0 1 0 0 0 1. The generator polynomial is p^3+1. Check whether there is error in the received codeword.

6. A rate ½, K = 3 binary convolutional encoder is shown in Figure 3.15.

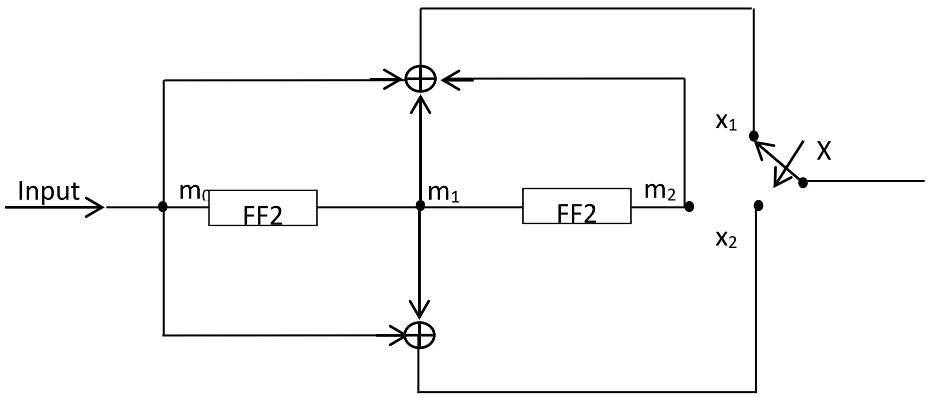


Figure 3.15

   a. Draw the tree diagram, trellis diagram and state diagram for this encoder.
   b. If the received signal at the decoder is (0 0 0 1 1 0 0 0), trace the decision on the code tree and find the input message bit sequence.

SHORT QUESTIONS AND ANSWERS

1. What is the need for coding? (Rationale for coding)

- The reliability of data transmission gets severely affected due to channel induced noise.
- With the available modulation schemes, it is not possible to provide acceptable data quality of low error performance.
- There is also a limitation on the achieved maximum value of E_b/N_o.
- Therefore, for a fixed E_b/N_o, the only practical option available for changing data quality from problematic to acceptable level is to use coding.

2. List the types of codes.

We can use many different codes for error control coding. They can be classified as below.

i. Based on methodology/Architecture
   1. Block Codes
   2. Convolutional Codes

ii. Based on hardware mechanization required to generate
   1. Linear Codes
   2. Mon-linear Codes

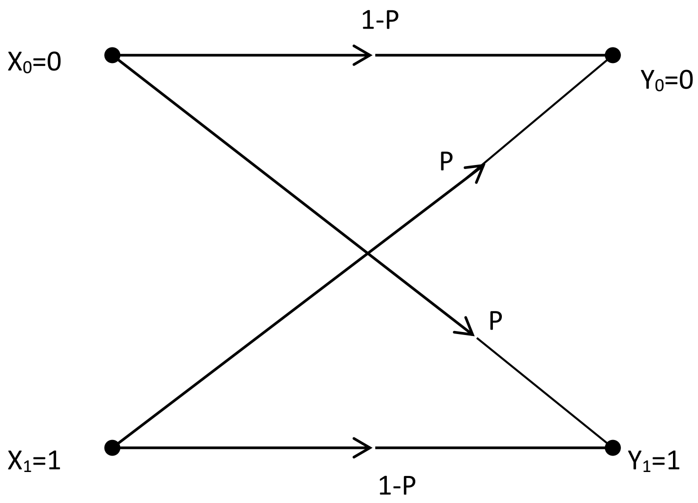
iii. Based on structure of architecture (Tree/Block)
   1. Systematic Codes
   2. Non-Systematic Codes

3. What is discrete memoryless channel?

The characteristics of a channel can be studied by representing the channel as a statistical model. The discrete memoryless channel is one of such channel models. The channel is said to be "discrete" when both the input X and output Y of the channel have finite sizes. The channel is said to be "memoryless" when the current output symbol depends only on the current input symbol and not on any of the previous inputs.

4. What is binary symmetric channel?

The Binary Symmetric Channel (BSC) is a special case of the discrete memoryless channel. The transition probability diagram of this channel is shown in the figure.



This channel is said to be binary because it has two input symbols ($X_0 = 0$ and $X_1 = 1$) and two output symbols ($Y_0 = 0$ and $Y_1 = 1$). The channel is symmetric because the probability of receiving a 1 if a 0 is sent is the same as the probability of receiving a 0 if a 1 is sent.

5. What is meant by error control coding? Name the methods of error control coding.

Error Control Coding techniques involve systematic addition of redundant bits to the transmitted information to achieve error detection and correction at the receiver. This controlled redundancy added in the transmitted message reduces probability of error at the receiver. There are two main methods used for error control coding. They are

1. Forward Error Correction (FEC)
2. Error detection with retransmission or Automatic Repeat Request (ARQ)

6. Briefly explain the Forward Error Correction Method.

The Forward Error Correction (FEC) method consists of a channel encoder at the transmitter and a channel decoder at the receiver. The channel encoder add redundant bits (check bits) to the message bits in a controlled manner. The channel decoder uses the check bits to detect and correct errors. The error detection and correction capability depends upon the number of check bits added.

7. What are the advantages and disadvantages of using forward error correction (FEC) method?

Advantages

- A constant overall delay is obtained. Hence the system is faster.
- The information throughput efficiency is constant in FEC systems.
- The feedback channel or return path is not needed.

Disadvantages

- Only a relatively moderate information throughput is obtained.
- The overall probability of errors is high, because some of the errors cannot be corrected.
- When very high reliability is needed, selection of an appropriate error correcting code and implementing its decoding algorithm, may be difficult.

8. Briefly explain the error detection with retransmission method.

The error detection with retransmission method is also called as Automatic Repeat Request (ARQ). It consists of a channel encoder along with an input buffer at the transmitter and a channel decoder along with an output buffer at the receiver. There is also a feedback path or return channel. Each codeword produced by the channel encoder is temporarily stored in the buffer and then transmitted. At the receiver, the channel decoder decodes the codewords and look for errors.

If no error is detected, the decoder sends a positive acknowledgement (ACK) through the return transmission channel. If any error is detected, it discards that part of the data sequence and sends a negative acknowledgement (NAK). The transmitter then once again transmits that part of the codeword sequence in which error was detected.

9. State the types of ARQ system

Basically, there are two types of ARQ. They are

1. Stop-and-Wait ARQ
2. Continuous ARQ
    a. Go back-N ARQ
    b. Selective Repeat ARQ

10. What are the advantages and disadvantages of using error detection with retransmission method or ARQ system?

Advantages

- This method has lower probability of error.
- Selective repeat ARQ provides the best throughput efficiency.
- It is an adaptive method, since information is retransmitted only when errors occur.

Disadvantages

- The ARQ system is slow, because of large overall delay.
- Expansive input and output buffers are required.
- The implementation cost is high.

11. List the error detection codes and error correction codes.

I. Error detection Codes
   1. Constant ratio Codes
   2. Redundant Codes
   3. Parity check Codes
   4. Cyclic Redundancy Check (CRC) Codes

II. Error Correction Codes
   A. Linear Block Codes
      1. Hamming Codes
      2. Cyclic Codes
      3. Bose-Chauduri-Hocguenghem (BCH) Codes
      4. Reed-Solomon (RS) Codes
   B. Convolutional Codes
      1. Self Orthogonal Codes
      2. Trial and Error Codes
      3. Recursive Systematic Codes

12. State the types of errors

There are mainly two types of errors introduced during data transmission.

1. Random Error: Random errors are caused by Additive White Gaussian Noise (AWGN) in the channel. Here noise affects the transmitted symbols independently.

2. Burst Error: Burst errors are caused by impulse noise in the channel. Impulse noise affects several consecutive bits and errors tend to occur in clusters.

There is a possibility that both the Gaussian noise and impulse noise will affect the channel. Therefore, if there is a mixture of random and burst errors, then such errors are called as compound errors.

13. Mention some applications of error control coding techniques

- For AWGN Channels, forward error correction (FEC) codes are employed. Typical applications include line-of-sight radio links such as satellite and deep space communication links.
- For compound-error channels, Automatic Repeat Request (ARQ) methods are employed. Typical applications include telephone channels and radio channels.
- Block codes are widely used to provide error control for magnetic tapes, mass storage systems, magnetic disks, and other data storage systems.
- Trellis-coded Modulation (TCM) technique combines convolutional coding and modulation into a single function. TCM is applied in the new generation of modems being developed for telephone channel.

14. Define Code Rate.

The code rate 'r' is defined as the ratio of the message bits (k) and the encoder output (codeword) bits (n).

$$\text{Code rate, } r = \frac{message\ bits}{Transmitted\ codeword\ bits} = \frac{k}{n}$$

where $0 < r < 1$

15. What is hamming distance?

The hamming distance (d) between the two code vectors is equal to the number of elements in which they differ. Eg, Let X = 101 and Y = 110. Then hamming distance between X and Y code vectors is 2.

The smallest hamming distance between the valid codevectors is termed as the minimum hamming distance (d_{min}).

16. Define codeword and codevector.

The encoded block of 'n' bits is called a codeword. It contains 'k' message bits and 'q' check bits.

An ‘n’ bit code word can be visualized in an N-dimensional space as a vector whose elements or co-ordinates are the bits in the codeword.

17. What are linear and non-linear codes?

In a linear code, modulo-2 sum of any two codevectors produces another valid code vector. The codes used in practical applications are almost always linear codes.

In a non-linear code, modulo-2-sum of any two codevectors does not necessarily produces another valid code vector.

18. What are systematic codes and non-systematic codes?

In a systematic code, the check bits are added in such a way that the message bits appear first and then check bits.

In a non-systematic code, it is not possible to identify message bits and check bits. They are mixed in the block.

19. What are linear block codes?

The input binary data sequence is divided into block of ‘k’ message bits. For each block of ‘k’ message bits, (n-k) check bits are added to produce ‘n’ bits codeword. Such codes are called (n, k) block codes.

Message bits k	Check bits q

← ‘n’ bits code word →

$(n = k + q)$

If the block codes satisfy linearity property, then they are called as linear block codes.

20. What are Hamming Codes?

Hamming codes are (n, k) linear block codes. They can be generated either systematically or non-systematically. The systematic form of Hamming codes satisfy the following conditions.

- Number of check bits, $q \geq 3$
- Codeword length, $n = 2^q - 1$
- Number of message bits, $k = n - q$
- Minimum hamming distance, $d_{min} = 3$

21. State the error detection and correction capability of linear block code or Hamming code.

- Detect upto 's' errors per codeword, $d_{min} \geq s+1$.
- Correct upto 't' error per codeword, $d_{min} \geq 2t + 1$.
- For Hamming code, $d_{min} = 3$, $s = 2$, $t = 1$

22. What is retransmission?

At the receiver, the channel decoder decodes the received codewords and look for errors. If no error is detected, the decoder sends a positive acknowledgement (ACK) through the return transmission channel. If any error is detected, it discards that part of the data sequence and sends a negative acknowledgement (NAK). The transmitter then once again transmits that part of the codeword sequence in which error was detected. This process is called retransmission.

23. What are binary cyclic codes?

Binary cyclic codes are a subclass of the linear block codes. A linear block code is called as cyclic code if every cyclic shift of the code vector produces another code vector. A cyclic code exhibits both the linearity property and cyclic property.

24. What are the advantages and disadvantages of cyclic codes

Advantages

- Cyclic codes can correct burst errors that span many successive bits.
- They have an excellent mathematical structure. This makes the design of error correcting codes with multiple-error correction capability relatively easier.
- The encoding and decoding circuits can be easily implemented using shift registers.
- The error correcting and decoding methods eliminate the storage (large memories) needed for lookup table decoding. Therefore the cyclic codes become powerful and efficient.

Disadvantages

- Even though the error detection is simpler, the error correction is slightly more complicated. This is due to the complexity of the combinational logic circuit used for error correction.

25. Define Cyclic Redundancy Check (CRC) code.

Cyclic Redundancy Check (CRC) code is an most important cyclic code used for error detection in data networks and storage systems. CRC code is basically a systematic form of cyclic code.

26. State the applications of CRC codes

1. CRC codes are mainly used in ARQ system for error detection.
2. They are used in data networks, digital subscriber lines, storage systems etc.

i) CRC – 6
CRC – 8 – WCDMA
CRC – 10 – CDMA 2000
CRC – 16 – CDMA 2000
} Mobile Networks

ii) CRC – 8 → Digital Video broadcasting

iii) CRC – CCITT → Bluetooth, XMODEM, X-25, V-41, HDLC FCS

iv) CRC – 32 → ANSI X3.66, ITU.T V.42, Ethernet, HDLC, MPEG-2, PNG

27. Define convolutional Coding

In convolution codes, the encoding operation is the discrete-time convolution of the input data sequence with the impulse response of the encoder.

The input message bits are stored in the fixed length shift register and they are combined with the help of mod-2 adders. This operation is equivalent to binary convolution and hence it is called convolutional coding.

28. Define constraint length in convolutional codes.

The constraint length 'K' of a convolution code is defined as the number of shifts over which a single message bit can influence the encoder output. It is expressed in terms of message bits. The code dimension of a convolutional code depends on the number of message bits 'k', the number of encoder output bits 'n' and its constraint length 'K'. The code dimension is therefore represented by (n, k, K).

29. What are the advantages and disadvantages of convolutional codes?

Advantages

- The convolutional codes operate on smaller blocks of data. Hence decoding delay is small.
- The storage hardware required is less.

Disadvantages

- Due to complexity, the convolutional codes are difficult to analyse.
- These codes are not developed much as compared to block codes.

Unit – IV DIGITAL MODULATION TECHNIQUES

OBJECTIVES

- To know the Digital Modulation techniques
- To study about Coherent and Non-Coherent modulation schemes
- To learn about TDM frame structure
- To study about Coherent and Non-Coherent detection schemes

4.0 INTRODUCTION

We have discussed Baseband pulse transmission in Unit II. In baseband pulse transmission, the input data is represented in the form of a discrete PAM signal (Line codes). The baseband signals have an adequately large power at low frequencies. So they can be transmitted over a pair of wires or coaxial cables.

But, it is not possible to transmit the baseband signals over radio links or satellites, since impractically large antennas would be required. Hence, the spectrum of the message signal has to be shifted to higher frequencies. This is achieved by using the baseband digital signal to modulate a high frequency sinusoidal carrier. The modulated signals are transmitted over a band pass channel, such as microwave radio link, satellite channel, optical fibre link etc. This process is called as digital carrier modulation or digital passband communication.

4.1 DIGITAL MODULATION:

Digital modulation may be defined as mapping a sequence of input binary digits into a set of corresponding high frequency signal waveforms. These modulated waveforms may differ in either amplitude or frequency or phase or some combination of two signal parameters (Amplitude and phase or frequency and phase).

4.1.1 Digital Modulation Techniques

The digital modulation techniques may be classified into two categories.

1. Coherent digital modulation techniques
2. Non-Coherent digital modulation techniques

1. Coherent Digital Modulation Techniques

Coherent digital modulation techniques employ coherent detection. In coherent detection, the local carrier generated at the receiver is phase locked with the carrier at the transmitter. Thus detection is done by correlating received noisy signal and locally generated carrier. The coherent detection is a synchronous detection. Coherent detection techniques are complex but provide better performance.

2. ***Non-Coherent Digital Modulation Techniques***

These techniques employ Non-Coherent detection. The detection process does not need receiver carrier to be phase locked with the transmitter carrier. Non-Coherent detection techniques are less complex. However the probability of error is high compared to Coherent detection.

4.1.2 Listing of various types:

Based on the mapping techniques, we can broadly classify the digital modulation methods.

I. Binary Scheme / M-ary Scheme:

In binary scheme, we send any one of the two possible signals during each signaling interval of duration T_b. Examples are

1. Amplitude Shift Keying (ASK), 2. Frequency Shift Keying (FSK) and 3. Phase Shift Keying (PSK)

M-ary Scheme:

In M-ary scheme, we can send any one of the M possible signals during each signaling interval of duration T_b. Examples are

1. M-ary ASK
2. M-ary FSK
3. M-ary PSK
4. Minimum shift keying (MSK) is a special form of continuous phase frequency shift keying (CPFSK).
5. Quadriphase shift keying (QPSK) is an example of M-ary PSK with M=4. Both MSK and QPSK are examples of quadrature carrier multiplexing system.
6. M-ary Quadrature Amplitude Modulation (M-ary QAM)

We may combine discrete changes in both the amplitude and phase of a carrier to produce M-ary Amplitude-Phase Keying (APK). M-ary QAM is a special form of this hybrid modulation.

II. Based on the performance of the modulation scheme and properties of modulated signal.

1. Power efficient scheme / Bandwidth efficient scheme
2. Continuous phase (CP) modulation / In phase Quadrature phase (IQ) modulation
3. Constant envelope modulation / Non-Constant envelope modulation
4. Linear modulation / Non-linear modulation
5. Modulation scheme with memory / modulation scheme without memory.

4.1.3 Design Goals of Digital Communication System

There are so many modulation/detection schemes available to the designer of a digital communication system. Each scheme offers system trade-offs of its own. The selection of a particular modulation/detection scheme is determined by the usage of available primary communication resources, transmitted power and channel bandwidth. In particular, the choice is based on achieving as many of the following design goals as possible.

1. Maximum data rate
2. Minimum possibility of symbol error
3. Minimum transmitted power
4. Minimum channel bandwidth
5. Maximum resistance to interfering signals
6. Minimum circuit complexity.

4.1.4 Gram-Schmidt Orthogonalization Procedure

The task of transforming an incoming message m_i, where i = 1, 2, …. M, into a modulated wave $S_i(t)$ may be divided into separate discrete time and continuous time operations. The Gram-Schmidt orthogonalization procedure permits the representation of any set of M energy signals, $\{S_i(t)\}$, as linear combinations of N orthonormal basis functions. Hence we may represent the given set of real-valued energy signals $S_i(t)$, $S_2(t)$….. $S_m(t)$, each of duration T seconds, in the form

$$S_i(t) = \sum_{j=1}^{N} S_{ij}\phi_j(t), \quad \begin{matrix} i = 1,2, \dots . M \\ o \le t \le T \end{matrix} \tag{4.1}$$

The real valued basis functions $\phi_1(t)$, $\phi_2(t)$, ……$\phi_N(t)$ are orthonormal. Hence we have

$$\int_0^T \phi_i(t)\phi_j(t)dt = \begin{cases} 1 & if\ i = j \\ 0 & if\ i \neq j \end{cases} \tag{4.2}$$

The first condition states that each basis function is normalized to have unit energy. The second condition states that the basis functions $\phi_1(t)$, $\phi_2(t)$, ……$\phi_N(t)$ are orthogonal with respect to each other over the interval $0 \le t \le T$.

In equation (4.1), the coefficients of the expansion are defined by

$$S_{ij} = \int_0^T S_i(t)\phi_j(t)dt \quad \begin{matrix} i = 1, 2, \dots . M \\ j = 1, 2, \dots . N \end{matrix} \tag{4.3}$$

Modulator Design:

Let the set of coefficients $\{S_{ij}\}$, j = 1, 2, … N operating as input. Then we may use the scheme shown in Figure 4.1 to generate the signal $S_i(t)$, i = 1, 2, … M as per equation (4.1).

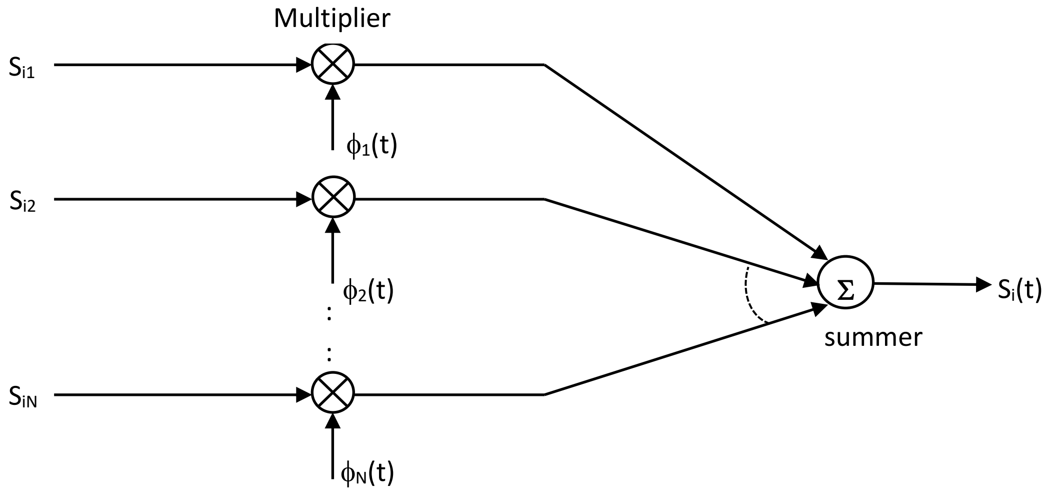


Figure 4.1 Scheme for generating the signal S_i (t)

It consists of a bank of N multipliers, with each multiplier supplied with its own basis function, followed by a summer. This scheme is performing a similar role to that of modulator in the transmitter.

Detector Design:

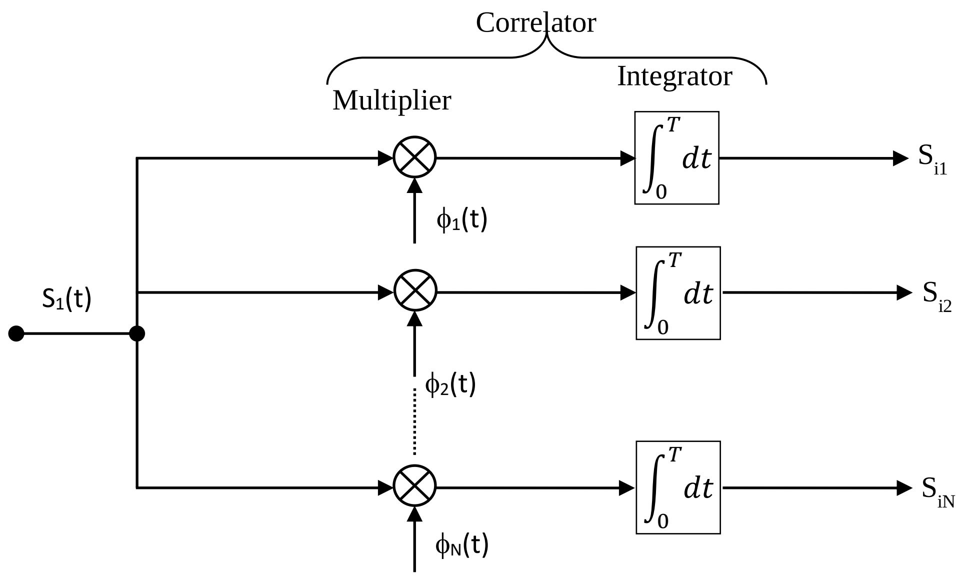


Figure 4.2 Scheme for generating the set of coefficients $\{S_{ij}\}$

Let the set of signals $\{S_i(t)\}, i = 1, 2, \ldots M,$ operating as input. We may use the scheme shown in figure 4.2 to calculate the set of coefficients $\{S_{ij}\}$, j = 1, 2, ….N as per equation (4.3). This scheme consists of a bank of N product integrators or correlators with a common input. Each multiplier is supplied with its own basis function. This scheme is performing a similar role to that of detector in the receiver.

4.2 COHERENT BINARY MODULATION TECHNIQUES

We know that binary modulation scheme has three basic forms.

1. Binary Amplitude Shift Keying (BASK)
2. Binary Frequency Shift Keying (BFSK)
3. Binary Phase Shift Keying (BPSK)

When these modulation schemes employ coherent detection at the receiver, then they are called as coherent binary modulation techniques. In coherent detection, the local carrier generated at the receiver is phase locked with the carrier at the transmitter.

4.2.1 Coherent Binary Phase Shift Keying (BPSK)

In phase shift keying, the modulation process involves switching or keying the phase of the carrier signal in accordance with the incoming data. The Figure 4.3(a) shows the block diagram for binary PSK transmitter.

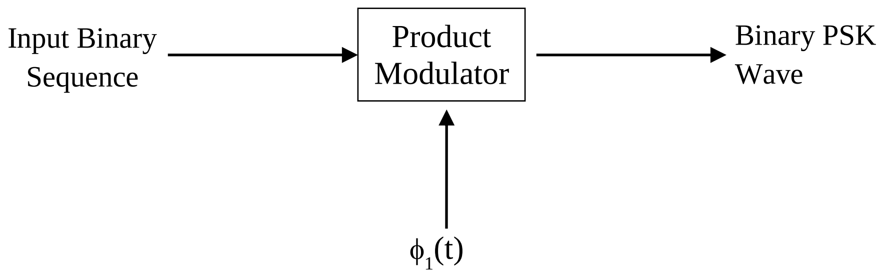


Figure 4.3(a) Binary PSK transmitter

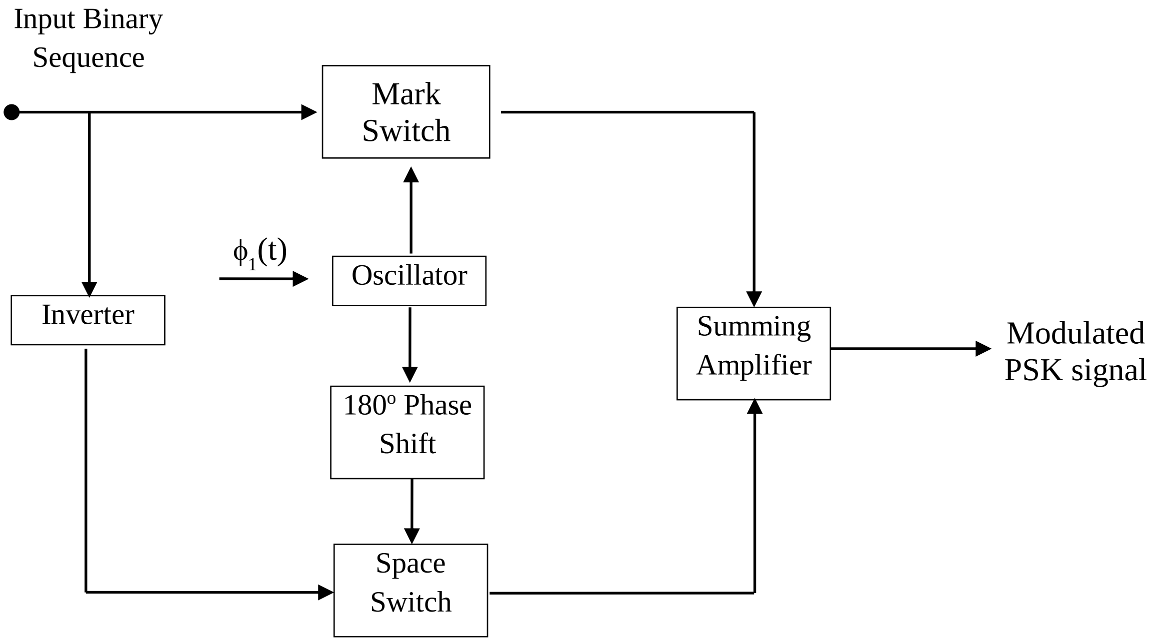


Figure:4.3 (b) Binary PSK modulator

In a coherent binary PSK system, the pair of signals, $S_1(t)$ and $S_2(t)$ are used to represent binary symbols 1 and 0 respectively. They are defined by

$$S_1(t) = \sqrt{\frac{2E_b}{T_b}}\, cos(2\pi f_c t) \qquad (4.4)$$

$$S_2(t) = \sqrt{\frac{2E_b}{T_b}}\, cos(2\pi f_c t + \pi)$$

$$= -\sqrt{\frac{2E_b}{T_b}}\, cos(2\pi f_c t) \qquad (4.5)$$

where $E_b \rightarrow$ transmitted signal energy per bit

$T_b \rightarrow$ bit duration, $0 \leq t \leq T_b$

We require only one basis function of unit energy.

$$\phi_1(t) = \sqrt{\frac{2}{T_b}}\, cos(2\pi f_c t),\ 0 \leq t \leq T_b \qquad (4.6)$$

We have to represent the input binary sequence in polar form with symbols 1 and 0 by constant amplitude levels of $\sqrt{E_b}$ and $\sqrt{-E_b}$, respectively. This binary wave and a sinusoidal carrier $\phi_1(t)$ are applied to a product modulator. The desired PSK wave is obtained at the modulator output. An alternate method of generating binary PSK is shown in Figure 4.3(b). In this method we use two balanced modulators as mark and space switch. The input binary data is applied directly to mark switch and after inverting to the space switch. The carrier signal $\phi_1(t)$ is fed directly to mark switch and 180° phase shifted to space switch. For binary input 1, the mark switch is closed and PSK wave is generated. For binary input 0, the space switch is closed and PSK wave is generated. The summing amplifier combines the output from mark and space switches.

Wave forms:

The Figure 4.4 shows the waveforms for coherent binary PSK modulation.

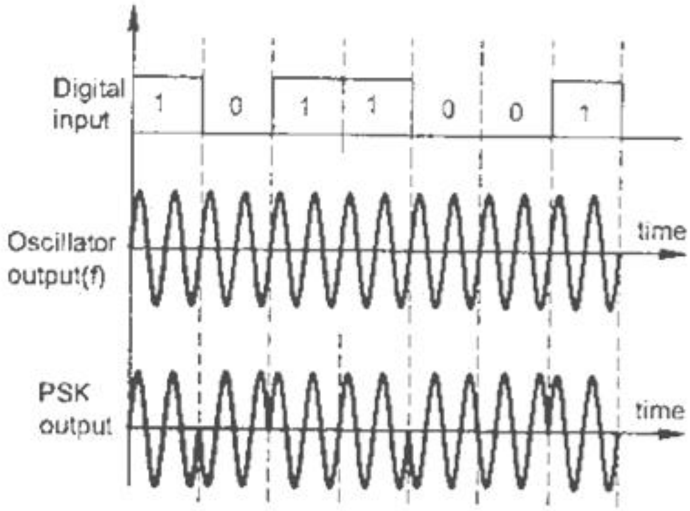


Figure 4.4 - Waveforms for BPSK

Merits of BPSK:

- BPSK requires lower bandwidth than BFSK
- BPSK has the minimum value of probability of error. Hence it provides best performance compared to BFSK and BASK schemes.
- It has very good noise immunity.

Demerits of BPSK:

In PSK, the information lies in the phase, and hence, it cannot be detected non-coherently.

4.2.2 Coherent Binary Frequency Shift Keying (BFSK):

In Frequency shift keying, the modulation process involves switching or keying the frequency of the carrier signal in accordance with the incoming data. The Figure 4.5 shows the block diagram of binary FSK transmitter.

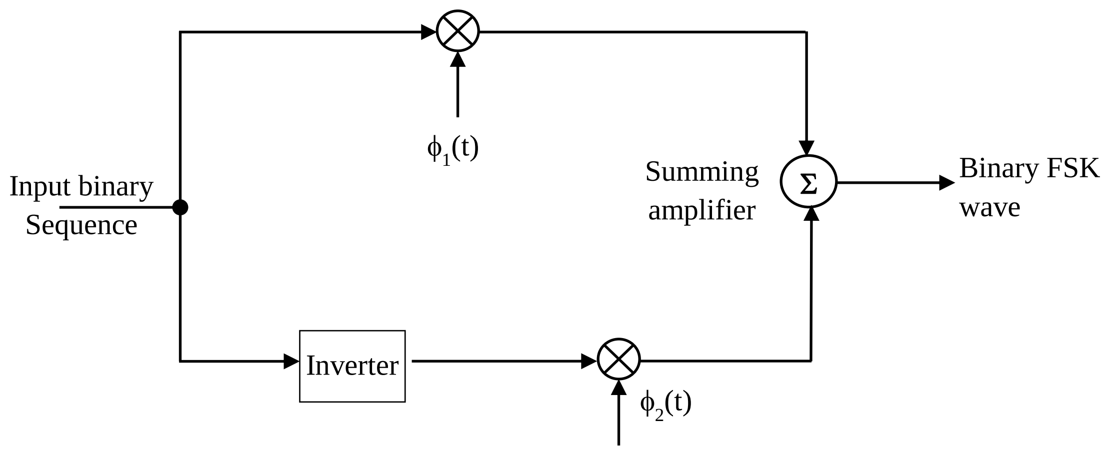


Figure 4.5 (a) BFSK Modulator

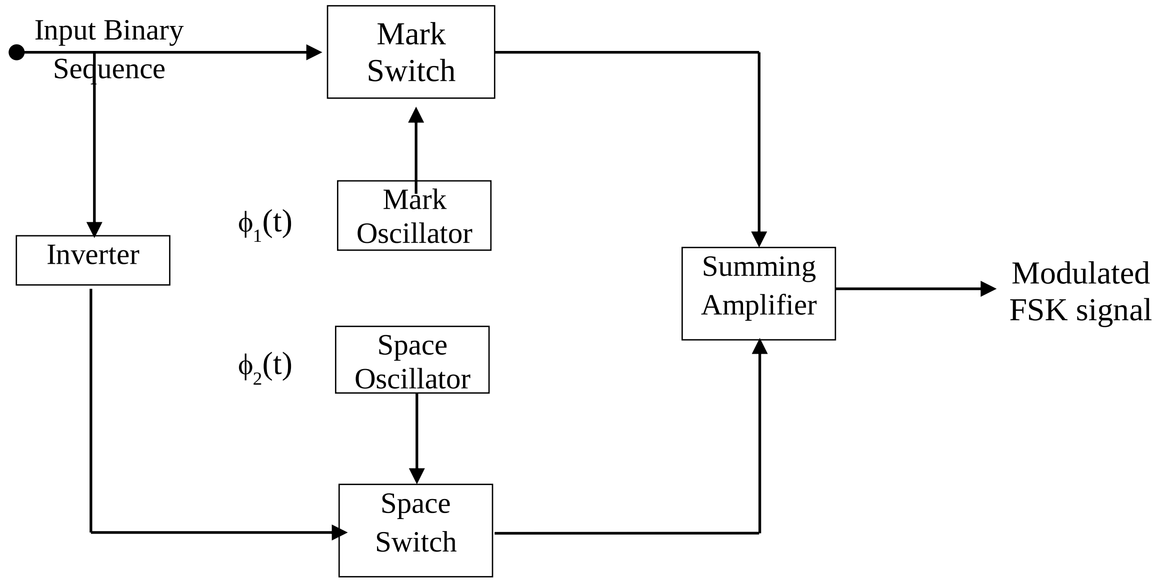


Figure 4.5 (b) BFSK Modulator

In a coherent binary FSK system, the pair of signals, $S_1(t)$ and $S_2(t)$ are used to represent binary symbols 1 and 0 respectively. They are defined by

$$S_1(t) = \sqrt{\frac{2E_b}{T_b}}\, cos(2\pi f_1 t) \quad (4.7)$$

$$S_2(t) = \sqrt{\frac{2E_b}{T_b}}\, cos(2\pi f_2 t) \quad (4.8)$$

Here we require two basic functions of unit energy.

$$\phi_1(t) = \sqrt{\frac{2}{T_b}}\, cos(2\pi f_1 t) \quad (4.9)$$

$$\phi_2(t) = \sqrt{\frac{2}{T_b}}\, cos(2\pi f_2 t) \quad (4.10)$$

Here $\phi_1(t)$ is applied to the upper product modulator (Also referred as Mark Switch). $\phi_2(t)$ is applied to lower product modulator (Also referred as space switch). The input binary data is applied directly to the mark switch and through an inverter to the space switch. For binary input 1, the mark switch is closed and FSK wave $S_1(t)$ is generated. For binary input 0, the space switch is closed and FSK wave $S_2(t)$ is generated. The summing amplifier combines the output from Mark and Space switches. In BFSK, the frequency of the modulated wave is shifted with a continuous phase, in accordance with the input binary wave. Hence phase continuity is always maintained including the inter-bit switching times. Therefore BFSK is also referred as continuous phase frequency shift keying (CPFSK).

Waveforms:

The figure 4.6 shows the waveforms for coherent binary FSK modulation.

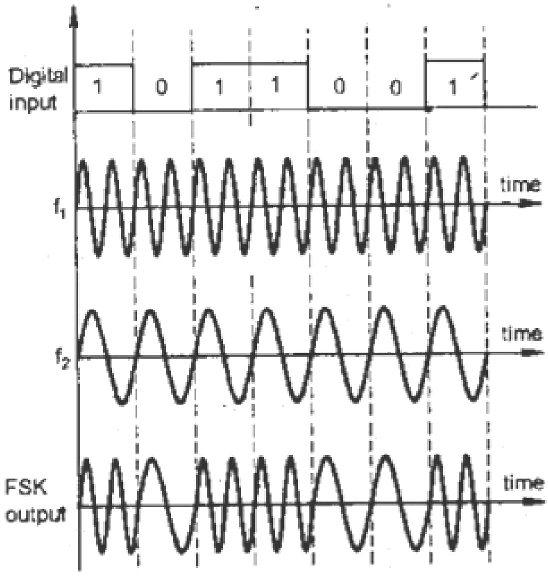


Figure 4.6 Waveforms for BFSK

Merits of BFSK

- It is relatively easy to implement.
- It has better noise immunity than ASK.

Demerits of BFSK

- BFSK requires high bandwidth compared to BPSK and BASK.

4.2.3 Coherent Binary Amplitude Shift Keying (BASK)

In Amplitude shift keying, the modulation process involves switching or keying the amplitude of the carrier signal in accordance with the incoming data. The Figure 4.7 shows the block diagram of binary ASK transmitter.

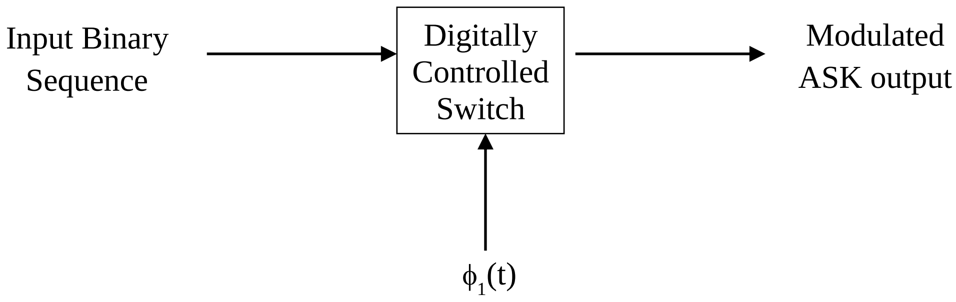


Figure 4.7 BASK Modulator

In a coherent binary ASK system, the pair of signals $S_1(t)$ and $S_2(t)$ are used to represent binary symbols 1 and 0 respectively. They are defined by

$$S_1(t) = \sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_c t) \tag{4.11}$$

$$S_2(t) = 0 \tag{4.12}$$

We require only one basis function of unit energy.

$$\phi_1(t) = \sqrt{\frac{2}{T_b}}\cos(2\pi f_c t) \tag{4.13}$$

The binary wave and the sinusoidal carrier $\phi_1(t)$ are applied to a product modulator. The product modulator acts like a digitally controlled switch. For binary input 1, the switch is closed and the carrier signal $\phi_1(t)$ is obtained as output signal. For binary input 0, the switch is open and hence there is no output signal. The resulting output will be the ASK waveform. The modulator simply does the on-off function. Hence BASK is also called as On-Off Keying (OOK).

Wave forms:

The Figure 4.8 shows the wave forms for coherent binary ASK modulation.

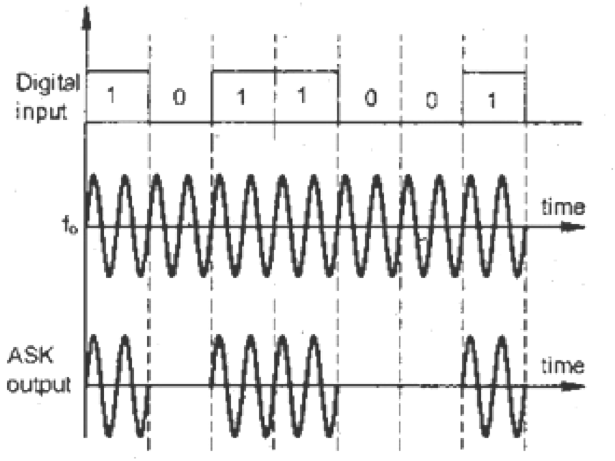


Figure 4.8: Wave forms for BASK

Merits of BASK:

- BASK is easy to generate and detect

Demerits of BASK:

- Bask is very sensitive to noise

4.2.4 Performance Comparison:

Table 4.1 shows the performance comparison of three basic digital modulation techniques.

Sl. No.	Parameters	BASK	BFSK	BPSK
1.	Switching or keying of	Amplitude	Frequency	Phase
2.	Bandwidth	$2f_b$	$4f_b$	$2f_b$
3.	Noise immunity	Low	High	High
4.	Probability of error	High	Low	Low
5.	Performance in presence of noise	Poor	Better than ASK	Best of three schemes
6.	System complexity	Simple	Moderately complex	Very Complex
7.	Bit rate or data rate	Suitable upto 100 bits / sec	Suitable upto 1200 bits / sec	Suitable upto high bit rates
8.	Demodulation method	Envelope detection	Envelope detection	Coherent detection

4.3 NON-COHERENT BINARY MODULATION TECHNIQUES:

The modulation scheme in which the detection process does not need receiver carrier to be phase locked with the transmitter carrier is said to be Non-Coherent modulation technique. The Non-Coherent binary modulation techniques are

1. Differential Phase Shift Keying (DPSK)
2. Binary Amplitude Shift Keying (BASK)
3. Binary Frequency Shift Keying (BFSK).

For BASK and BFSK, the modulator sections are the same for both coherent and non-coherent modulation techniques. We have already explained the modulator sections of BASK and BFSK. Now we shall see about Differential PSK.

The non-coherent binary FSK and DPSK schemes are treated as special cases of non-coherent orthogonal modulation.

Differential Phase Shift Keying (DPSK):

In the binary phase shift keying, we cannot have "Non-Coherent PSK", because detection without phase information is not possible. Hence, there is a "Pseudo PSK" technique called Differential Phase Shift Keying (DPSK). DPSK may be viewed as the non- coherent form of PSK.

DPSK eliminates the need for a coherent reference signal at the receiver by combining two basic operations at the transmitter.

1. Differential encoding of the input binary wave
2. Phase shift keying

To send symbol 1, we leave the phase of the current signal waveform unchanged. To send symbol 0, we phase advance the current signal waveform by 180°. The receiver is equipped with a storage capability, so that it can measure the relative phase difference between the waveforms received during two successive bit intervals. The Figure 4.8 shows the block diagram of a DPSK transmitter.

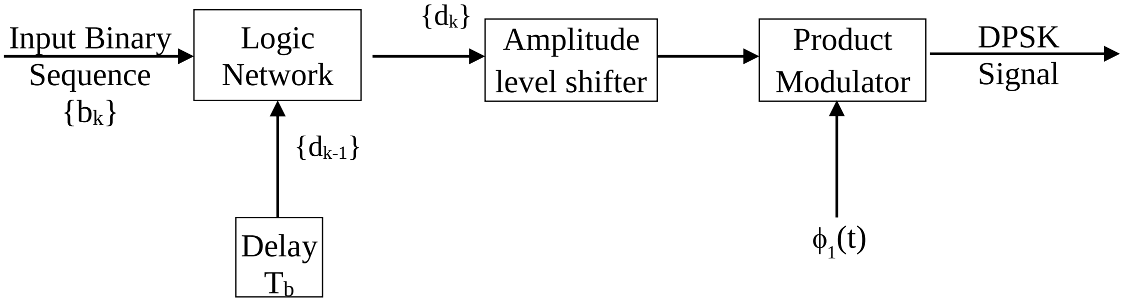


Figure 4.8 DPSK transmitter

It consists of a logic network and a one-bit delay element interconnected so as to convert the raw binary sequence {b_k} into a differentially encoded sequence {d_k}. This sequence is amplitude level encoded and then used to modulate a carrier wave $\left[\phi_1(t) = \sqrt{\frac{2}{T_b}}\, cos(2\pi f_c t)\right]$, thereby producing the desired DPSK signal.

The differential encoding process starts with an arbitrary first bit, serving as reference. Let {d_k} denote the differentially encoded sequence with this added reference bit.

(i) If the incoming binary symbol b_k is 1, leave the symbol d_k unchanged with respect to the previous bit.

(ii) If the incoming binary symbol b_k is 0, change the symbol d_k with respect to the previous bit.

The differentially encoded sequence {d_k} thus generated is used to phase-shift a carrier with phase angles 0 and π radians representing symbols 1 and 0, respectively. Table 4.2 illustrates the differential phase encoding process. Here, d_k is the complement of the modulo-2 sum of b_k and d_{k-1}.

Table 4.2 Illustrating the generation of DPSK signal

{b_k}		1	0	0	1	0	0	1	1
{d_{k-1}}		1	1	0	1	1	0	1	1
Differentially encoded sequence {d_k}	1	1	0	1	1	0	1	1	1
Transmitted phase (radians)	0	0	π	0	0	π	0	0	0

Merits of DPSK:

- DPSK scheme does not need carrier at the receiver end. Hence it has reduced system complexity.
- The bandwidth required is less than that required for BPSK.

Demerits of DPSK:

- It has higher value of probability of error than that of BPSK.
- Noise interference is more.
- In DPSK, previous bit is used to detect next bit. Hence, there is possibility of errors appearing in pairs.

4.4 COHERENT QUADRATURE MODULATION TECHNIQUES

One important goal in the design of a digital communication system is the efficient utilization of channel bandwidth. There are two bandwidth conserving Quadrature-Modulation schemes for the transmission of binary data.

They are:

1. Quadriphase-Shift Keying (QPSK)
2. Minimum Shift Keying (MSK)

These two schemes are both examples of the quadrature-carrier multiplexing system. They produce a modulated wave described as

$$S(t) = S_I(t) \cos(2\pi f_c t) - S_Q(t) \sin(2\pi f_c t) \tag{4.14}$$

where $S_I(t)$ is the in-phase component of the modulated wave, and $S_Q(t)$ is the quadrature component. QPSK is a quadrature-carrier signaling technique, which is an extension of binary PSK. MSK is a special form of continuous Phase Frequency Shift Keying (CPFSK).

4.4.1 Quadri Phase-Shift Keying (QPSK):

In QPSK, as with binary PSK, information carried by the transmitted signal is contained in the phase. The mapping or assignment of k information bits to the $M=2^k$ possible phases may be done in a number of ways. The preferred assignment is Gray encoding. For QPSK, we have k=2, and hence $M=2^2=4$. Therefore, the number of bits per symbol is two bits. Then the information bits 1 0, 0 0, 0 1, and 1 1 (Gray encoding) represent the phase values $\frac{\pi}{4}$, $3\frac{\pi}{4}$, $5\frac{\pi}{4}$ and $7\frac{\pi}{4}$ (45° , 135° , 225° , and 315°). For this set of values we may define the transmitted signal as

$$S_i(t) = \begin{cases} \sqrt{\frac{2E}{T}} cos\left[2\pi f_c t + (2i-1)\frac{\pi}{4}\right], & 0 \le t \le T \\ 0, & elsewhere \end{cases} \tag{4.15}$$

where i = 1, 2, 3, 4; E is the transmitted signal energy per symbol ($E = 2E_b$) and T is the symbol duration ($T = 2T_b$).

We can rewrite the equation (4.15) as

$$S_i(t) = \sqrt{\frac{2E}{T}} \cos(2\pi f_c t) \,.\, cos\left[(2i-1)\frac{\pi}{4}\right] - \sqrt{\frac{2E}{T}} \sin(2\pi f_c t) \,.\, sin\left[(2i-1)\frac{\pi}{4}\right] \tag{4.16}$$

Where $0 \le t \le T$

There are two orthonormal basis functions $\phi_1(t)$ and $\phi_2(t)$.

$$\phi_1(t) = \sqrt{\frac{2}{T}}\cos(2\pi f_c t) \quad \text{and} \tag{4.17}$$

$$\phi_2(t) = \sqrt{\frac{2}{T}}\sin(2\pi f_c t) \tag{4.18}$$

Transmitter

The figure 4.9 shows the block diagram of QPSK transmitter.

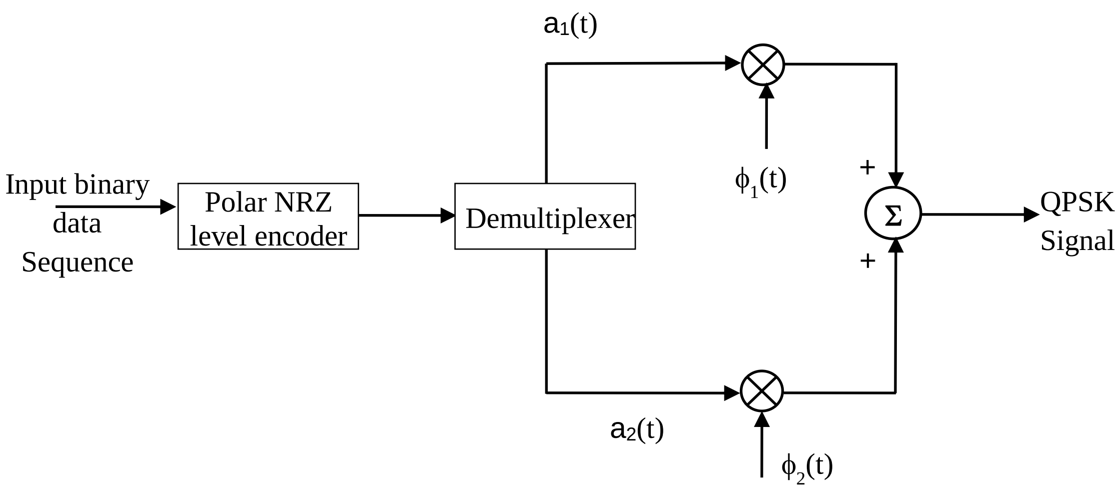


Figure 4.9 QPSK Transmitter

The incoming binary data sequence is first transformed into polar form by a non-return-to-zero (NRZ) level encoder. This binary wave is next divided by means of a demultiplexer into two separate binary waves consisting of the odd-and even-numbered input bits. These two binary waves are denoted by $a_1(t)$ and $a_2(t)$.

These two binary waves $a_1(t)$ and $a_2(t)$ are used to modulate a pair of quadrature carriers $\phi_1(t)$ and $\phi_2(t)$ respectively. The result is a pair of binary PSK signals. Finally, the two binary PSK signals are added to produce the desired QPSK signals.

Wave forms:

The figure 4.10 illustrates the sequences and waveforms involved in the generation of a QPSK signal.

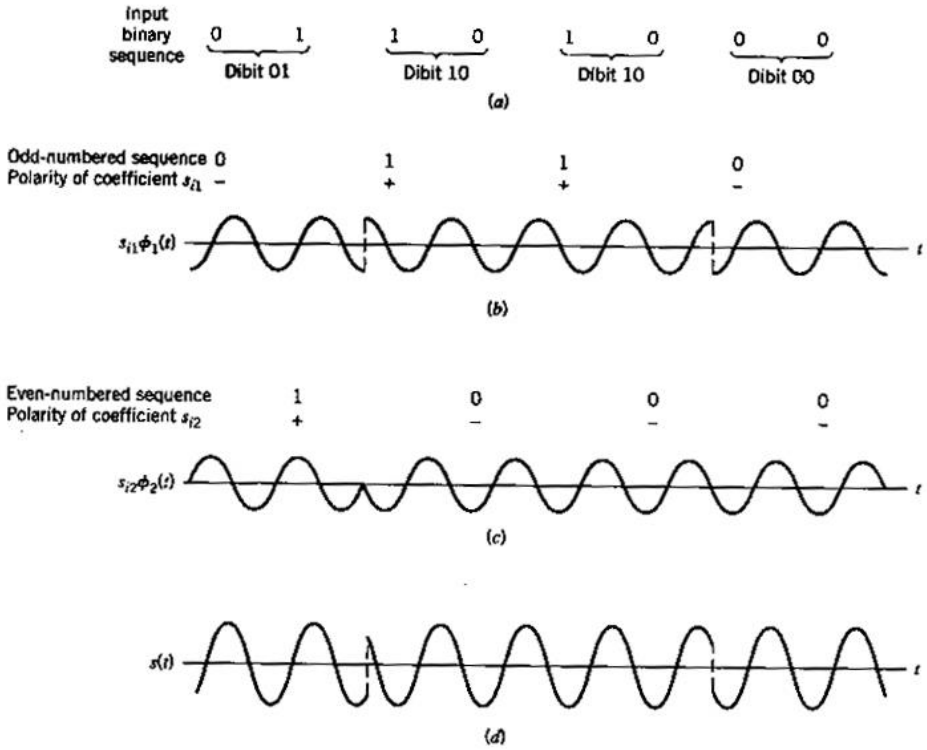


Figure 4.10 QPSK waveforms

Receiver:

The Figure 4.11 shows the block diagram of coherent QPSK receiver.

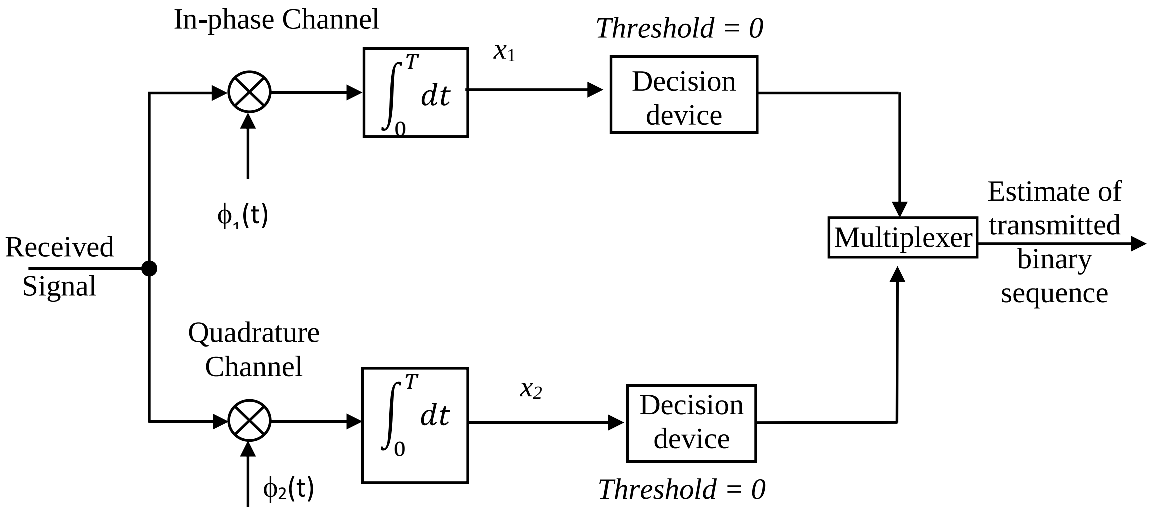


Figure 4.11 Coherent QPSK receiver

The received signal x(t) is applied to a pair of product integrators or correlators. The multiplier is supplied with locally generated coherent carrier signals $\phi_1(t)$ in the In-phase channel and $\phi_2(t)$ in the quadrature channel. The correlator outputs of x_1 and x_2 are produced in response to the received signal x(t). They are each compared with a threshold of Zero.

For the In phase channel, if $x_1>0$, a decision is made in favour of symbol 1, and if $x_1< 0$, a decision is made in favour of symbol 0. Similarly, for the quadrature channel, if $x_2> 0$, a decision is made in favour of symbol 1 and if $x_2< 0$, a decision is made in favour of symbol 0. Finally, these two binary sequences at the in-phase and quadrature channel outputs are combined in a multiplexer. This will reproduce the original binary sequence at the transmitter input. The minimum average probability of symbol error for QPSK is given by P_e = erfc$\left[\sqrt{\frac{E_b}{N_o}}\right]$.

Signal space diagram

For any modulation scheme, the analysis is based on the signal space diagram assuming an Additive White Gaussian Noise (AWGN) model. Signal-space approach is a plotting of possible message points. Such a set of possible message points is also referred to as a “Signal Constellation”.

In QPSK, there are four message points. The associated signal vectors are defined by

$$S_i = \begin{bmatrix} \sqrt{E}\,cos\left[(2i-1)\frac{\pi}{4}\right] \\ -\sqrt{E}sin\left[(2i-1)\frac{\pi}{4}\right] \end{bmatrix}, \quad i = 1, 2, 3, 4 \qquad (4.19)$$

The elements of the signal vectors, namely, s_{i1} and s_{i2} have their values shown in Table 4.3.

Table 4.3 Signal Space Characterization of QPSK

Input dibit $0 \le t \le T$	Phase of QPSK signal (radians)	Coordinates of message points	
		S_{i1}	S_{i2}
1 0	$\frac{\pi}{4}$	$+\sqrt{\frac{E}{2}}$	$-\sqrt{\frac{E}{2}}$
0 0	$3\frac{\pi}{4}$	$-\sqrt{\frac{E}{2}}$	$-\sqrt{\frac{E}{2}}$
0 1	$5\frac{\pi}{4}$	$-\sqrt{\frac{E}{2}}$	$+\sqrt{\frac{E}{2}}$
1 1	$7\frac{\pi}{4}$	$+\sqrt{\frac{E}{2}}$	$+\sqrt{\frac{E}{2}}$

Accordingly, a QPSK signal is characterized by having a two dimensional signal constellation (ie. N = 2) and four message points (ie., M = 4), as illustrated in Figure 4.12.

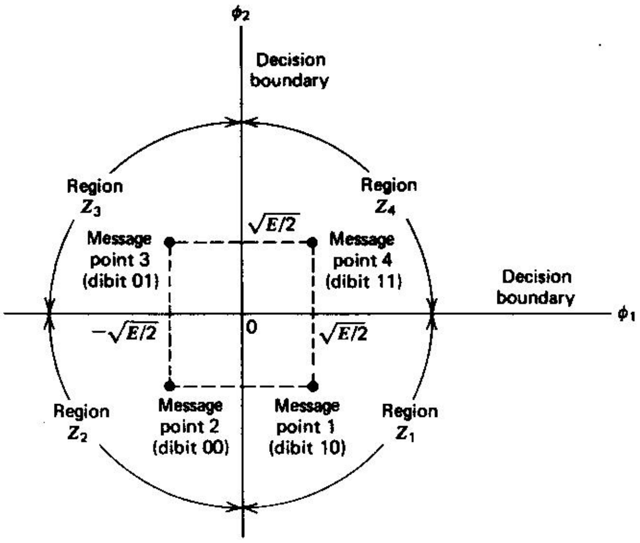


Figure 4.12 Signal space diagram for coherent QPSK system

Merits of QPSK:

- QPSK has very good noise immunity.
- More effective utilization of the available bandwidth of the transmission channel.
- It has low error probability

Demerits of QPSK:

- The generation and detection of QPSK Is complex.

How QPSK is better than BPSK:

- Due to multilevel modulation used in QPSK, it is possible to increase the bit rate to double the bit rate of BPSK without increasing the bandwidth.
- Available channel bandwidth is utilized in a better way by the QPSK system than the BPSK system.
- The noise immunity of QPSK is same as that of BPSK system.

4.4.2 Minimum Shift Keying (MSK):

Minimum shift keying (MSK) is a special form of binary CPFSK signal. A Continuous Phase Frequency Shift Keying (CPFSK) signal with a deviation ratio of $h = \frac{1}{2}$ is referred to as MSK. Using MSK, it is possible to improve the noise performance of the receiver significantly, by the proper use of the phase information. This improvement is achieved at the expense of increased receiver complexity.

Consider a CPFSK signal, defined for the interval $0 \leq t \leq T_b$ as below:

$$S(t) = \begin{cases} \sqrt{\frac{2E_b}{T_b}}\, cos[2\pi f_1 t + \theta(0)] \text{for symbol 1} \\ \sqrt{\frac{2E_b}{T_b}}\, cos[2\pi f_2 t + \theta(0)] \text{for symbol 0} \end{cases} \qquad (4.20)$$

where $E_b \rightarrow$ transmitted signal energy per bit and $T_b \rightarrow$ bit duration.

The phase $\theta(0)$ denotes the value of phase at time t=0. The frequencies f_1 and f_2 are sent in response to binary symbols 1 and 0 appearing at the modulator input respectively.

Another useful way of expressing the CPFSK signal S(t) is to represent it in the form of an angle modulated signal as follows:

$$S(t) = \sqrt{\frac{2E_b}{T_b}}\, cos[2\pi f_C t + \theta(t)] \qquad (4.21)$$

where $\theta(t)$ is the phase of S(t). The phase $\theta(t)$ of a CPFSK signal increases or decreases linearly with time during each bit duration of T_b seconds, as shown by

$$\theta(t) = \theta(0) \pm \frac{\pi h}{T_b} t\,, \quad 0 \le t \le T_b \qquad (4.22)$$

ie., $\theta(t) = \theta(0) + \frac{\pi h}{T_b} t \rightarrow$ for sending symbol 1

and $\theta(t) = \theta(0) - \frac{\pi h}{T_b} t \rightarrow$ for sending symbol 0

Substituting equation (4.22) into equation (4.21), and then comparing the angle of the cosine function with that of equation (4.20), we deduce the following pair of relations:

$$f_c + \frac{h}{2T_b} = f_1 \qquad (4.23)$$

$$f_c - \frac{h}{2T_b} = f_2 \qquad (4.24)$$

Solving equations (4.23) and (4.24) for f_c and h, we get

$$f_c = \frac{1}{2}(f_1 + f_2) \qquad (4.25)$$

and $$h = T_b (f_1 - f_2) \qquad (4.26)$$

The nominal carrier frequency f_c is therefore the arithmetic mean of the frequencies f_1 and f_2.

Deviation Ratio (h):

The difference between the frequencies f_1 and f_2, normalized with respect to the bit rate $1/T_b$ defines the dimensionless parameter h, which is referred to as the deviation ratio.

Phase Trellis:

From equation (4.22), for sending symbol 1, we have

$$\theta(t) = \theta(0) + \frac{\pi h}{T_b} t$$

At time t = T_b

$$\theta(T_b) = \theta(0) + \frac{\pi h}{T_b} . T_b \quad \Rightarrow \quad \theta(T_b) - \theta(0) = \pi h$$

For sending symbol 0, we have

$$\theta(t) = \theta(0) - \frac{\pi h}{T_b} . t$$

At time t = T_b

$$\theta(T_b) = \theta(0) - \frac{\pi h}{T_b} . T_b \quad \Rightarrow \quad \theta(T_b) - \theta(0) = - \pi h$$

Hence we may write

$$\theta(T_b) - \theta(0) = \begin{cases} \pi h & \text{for symbol 1} \\ -\pi h & \text{for symbol 0} \end{cases} \qquad (4.27)$$

Therefore, sending of symbol 1 increases the phase of a CPFSK signal S(t) by πh radians. Sending of symbol 0 decreases the phase by πh radians. We can plot the variation of phase θ(t) with respect to time t. Such a plot is called as a phase tree as shown in the Figure 4.13.

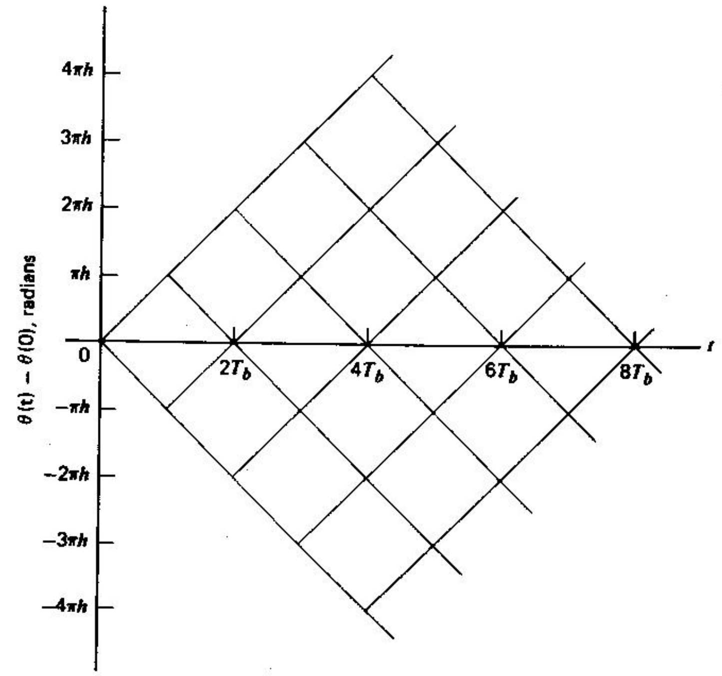


Figure 4.13 Phase Tree

The plot of phase tree follows a path consisting of a sequence of straight lines, the slope of which represents frequency changes.

The phase tree described in Figure 4.13 is a manifestation of phase continuity, which is an inherent characteristic of a CPFSK signal. In BFSK, which is a CPFSK scheme, the deviation ratio h is exactly unity. Hence the phase change over one bit interval is $\pm\pi$ radians.

For MSK scheme, the deviation ratio h is assigned the special value of $\frac{1}{2}$. Then the phase can take on only the two values $\pm\frac{\pi}{2}$ at odd multiples of T_b, and only the two values 0 and π at even multiples of T_b as shown in the Figure 4.14.

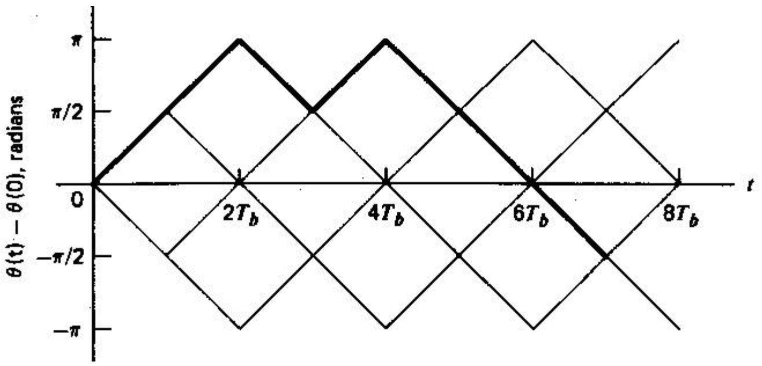


Figure 4.14 phase trellis (for the sequence 1101000)

This plot is called as a phase trellis, since a “trellis” is a tree like structure with remerging branches. Each path from left to right through the trellis corresponds to a specific binary sequence input. The path shown in boldface in the figure 4.14 corresponds to the binary sequence 1101000 with $\theta(0) = 0$.

Why the name MSK?

From equation (4.26), we have

$$h = T_b (f_1 - f_2)$$

On substituting $h = \frac{1}{2}$

$$\frac{1}{2} = T_b (f_1 - f_2) \quad \Rightarrow \quad f_1 - f_2 = \frac{1}{2T_b}$$

Since bit rate $R_b = \frac{1}{T_b}$, we can write

$$f_1 - f_2 = \frac{R_b}{2} \tag{4.28}$$

Hence the frequency deviation ($f_1 - f_2$) equals half the bit rate. This is the minimum frequency spacing that allows the two FSK signals representing symbols 1 and 0 as in equation 4.20 to be coherently orthogonal ie., they do not interfere with one another in the process of detection. It is for this reason, a CPFSK signal with a deviation ratio of $h = \frac{1}{2}$ is referred to as Minimum shift keying (MSK). MSK is also referred to as fast FSK.

4.4.2.1 MSK Transmitter

The Figure 4.15 shows the block diagram of MSK transmitter.

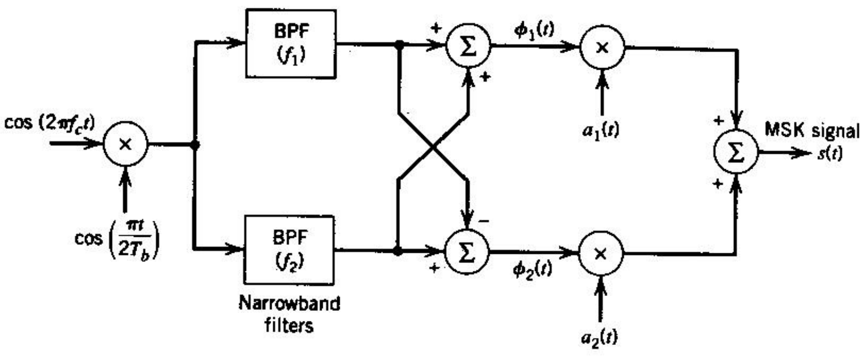


Figure 4.15 Block diagram of MSK transmitter communication system

The advantage of this method of generating MSK signals is that the signal coherence and deviation ratio are largely unaffected by variation in the input data rate. Two input sinusoidal waves, one of frequency $f_c = \frac{n_c}{4T_b}$ for some fixed integer n_c, and the other of frequency $\frac{1}{4T_b}$ are first applied to a product modulator.

This produces two phase-coherent sinusoidal waves at frequencies f_1 and f_2. They are related to the carrier frequency f_c and the bit rate $\frac{1}{T_b}$ such that

$$f_c + \frac{h}{2T_b} = f_1$$

$$f_c - \frac{h}{2T_b} = f_2, \qquad \text{where } h = \frac{1}{2}$$

These two sinusoidal waves are separated from each other by two narrow band filters, one centered at f_1 and the other at f_2. The resulting filter outputs are linearly combined to produce the pair of quadrature carriers. The orthonormal basis functions used as quadrature carriers are

$$\phi_1(t) = \sqrt{\frac{2}{T_b}}\cos\left(\frac{\pi}{2T_b}t\right) cos(2\pi f_c t),\ 0 \le t \le T_b \qquad (4.29)$$

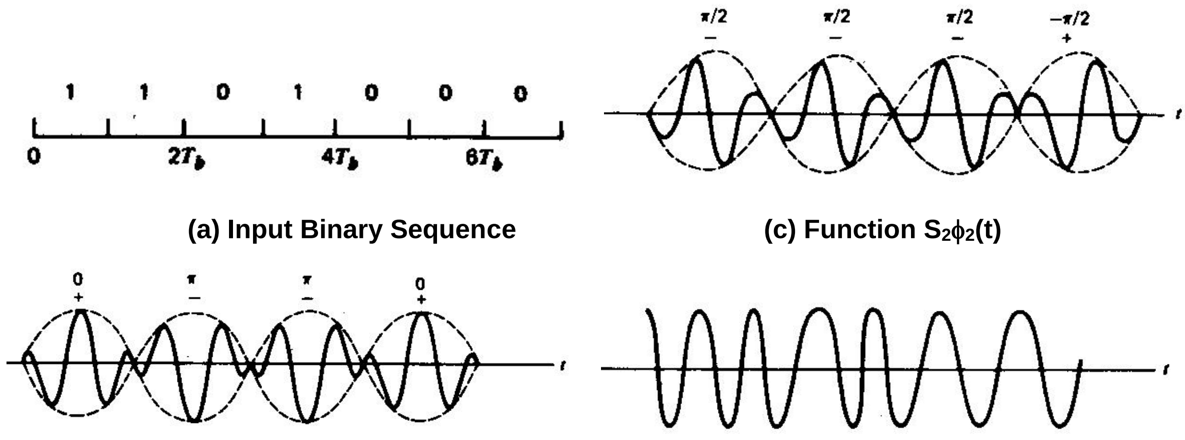
$$\phi_2(t) = \sqrt{\frac{2}{T_b}}\sin\left(\frac{\pi}{2T_b}t\right) sin(2\pi f_c t),\ 0 \le t \le T_b \qquad (4.30)$$

Finally, $\phi_1(t)$ and $\phi_2(t)$ are multiplied with two binary waves $a_1(t)$ and $a_2(t)$ having a bit rate equal to $\frac{1}{2Tb}$. The two binary waves $a_1(t)$ and $a_2(t)$ are extracted from the incoming binary sequence. The two multiplier outputs are summed to get the MSK signal output. We may express the MSK signal in the form of

$$s(t) = s_1\phi_1(t) + s_2\phi_2(t),\ 0 \le t \le T_b \qquad (4.31)$$

WAVE FORMS:

The Figure 4.16 shows the sequences and waveforms involved in the generation of MSK signal for the binary sequence 1101000



(a) Input Binary Sequence **(c) Function $S_2\phi_2(t)$**

(b) Function $S_1\phi_1(t)$ **(d) MSK Signal**

Figure 4.16 sequence and waveforms for MSK signal

4.4.2.2 MSK Receiver:

The Figure 4.17 shows the block diagram of coherent MSK receiver.

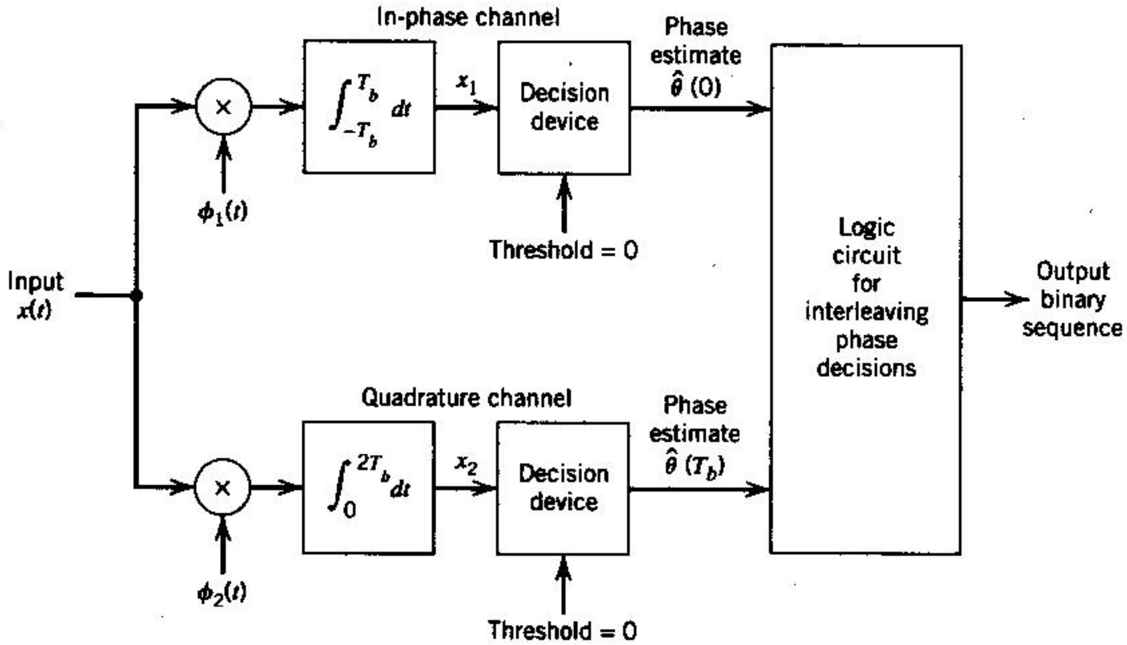


Figure 4.17 Block diagram of MSK receiver

The received signal x(t) is correlated with locally generated replicas of the coherent reference signals $\phi_1(t)$ and $\phi_2(t)$. In both cases the integration interval is $2T_b$ seconds. Also, the integration in the quadrature channel is delayed by T_b seconds with respect to that in the in-phase channel.

The resulting in-phase and quadrature channel correlator outputs, x_1 and x_2, are each compared with a threshold of zero.

- For the in-phase channel, if $x_1>0$, then choose the phase estimate $\hat{\theta}(0) = 0$. If $x_1<0$, then choose the estimate $\hat{\theta}(0) = \pi$.
- For the quadrature channel, if $x_2>0$, then choose the phase estimate $\hat{\theta}(T_b) = -\frac{\pi}{2}$. If $x_2<0$, then choose the estimate $\hat{\theta}(\mathrm{Tb}) = \frac{\pi}{2}$

Finally, these phase decisions are interleaved so as to reconstruct the original input binary sequence. The minimum average probability of symbol error in an AWGN channel, for MSK is given by $p_e=\mathrm{erfc}\left[\sqrt{E_b/N_o}\right]$

Signal Space Diagram:

The signal constellation for an MSK signal is two-dimensional (ie., N = 2), with four message points (ie., M = 4), as shown in the Figure 4.18.

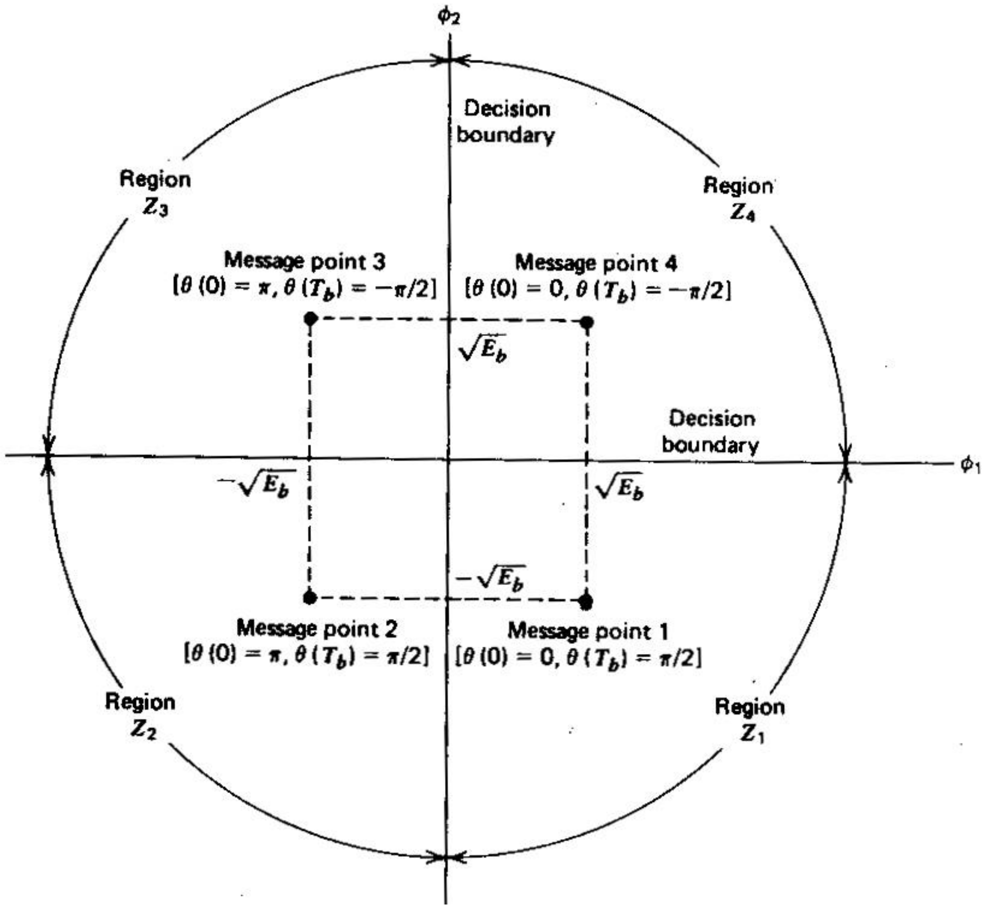


Figure 4.18 signal space diagram for MSK system

The coordinates of message points are given as

$S_1 \quad = \quad \sqrt{E_b}\cos[\theta(0)]$ and (4.32)

$S_2 \quad = \quad -\sqrt{E_b}\sin[\theta(T_b)]$ (4.33)

Table 4.4 presents a summary of the values of $\theta(0)$ and $\theta(T_b)$, as well as the corresponding values of S_1 and S_2.

Table 4.4: Signal-Space characterization of MSK

Transmitted binary symbol $0 \le t \le T_b$	**Phase states (radians)**		**Coordinates of message points**	
	θ(0)	**θ(T_b)**	**S_1**	**S_2**
1	0	$+\frac{\pi}{2}$	$+\sqrt{E_b}$	$-\sqrt{E_b}$
0	π	$+\frac{\pi}{2}$	$-\sqrt{E_b}$	$-\sqrt{E_b}$
1	π	$-\frac{\pi}{2}$	$-\sqrt{E_b}$	$+\sqrt{E_b}$
0	0	$-\frac{\pi}{2}$	$+\sqrt{E_b}$	$+\sqrt{E_b}$

Merits of MSK:

- MSK scheme has constant envelope (ie., there are no amplitude variations).
- It has coherent detection performance equivalent to that of QPSK.
- The MSK signal has a continuous phase (ie., there are no phase changes in the MSK signal)
- Filters to suppress the sidelobes which causes interchannel interference are not required.

Demerits of MSK:

- The generation and detection of MSK signal is more complicated.
- For a wireless communication system using MSK, the adjacent channel interference is not low enough to satisfy the practical requirements of such a multiuser communications environment.

4.4.3 Performance Comparison:

Table 4.5 shows the performance comparison of the quadrature modulation schemes of QPSK and MSK.

Table 4.5 Performance comparison

Sl. No.	Parameter	QPSK	MSK
1.	Switching or keying of	Phase	Frequency
2.	Bandwidth	f_b	1.5 f_b
3.	Bits per symbol	Two	Two
4.	System complexity	Moderately complex	Very complex
5.	Demodulation method	Coherent detection	Coherent detection
6.	Noise immunity	High	High
7.	Probability of symbol error	Low	Low
8.	Carrier signal	A pair of quadrature carriers	A pair of sinusoidally modulated quadrature carriers

4.5 DETECTION OF SIGNALS:

In previous sections, we described various types of modulation methods that may be used to transmit digital information through a communication channel. The modulator at the transmitter performs the function of mapping the digital binary input sequence into corresponding signal waveforms. Here, we shall study the performance characteristics of receivers for the various modulation methods. The Figure 4.19 shows the two basic steps in a digital receiver.

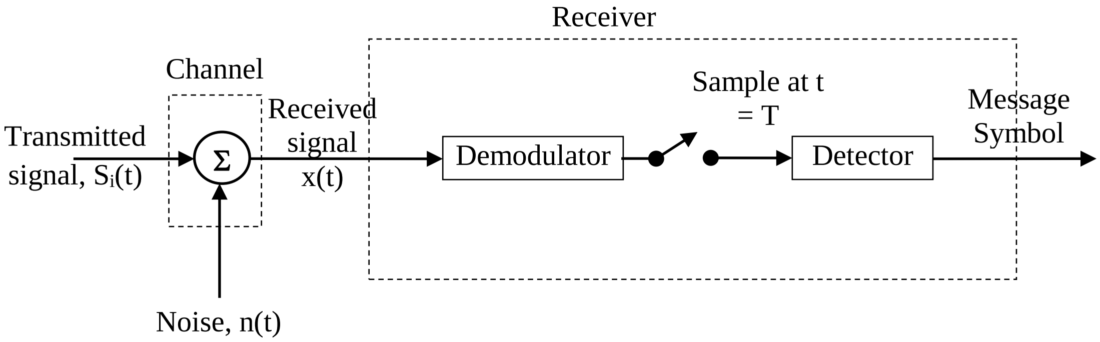


Figure 4.19 Two basic steps in digital receiver

The transmitted signal $S_i(t)$ is degraded by noise n(t) and impulse response of the channel $h_c(t)$. Hence the received signal is given by

$$x(t) = S_i(t) * h_c(t) + n(t) \quad (4.34)$$

On receiving the signal x(t), the digital receiver performs two basic functions of demodulation and detection.

1. Demodulator

The demodulator is a frequency down conversion block. The function of the signal demodulator is to convert the received waveform x(t) in to an N-dimensional vector $x=[x_1, x_2, \dots x_N]$ where N is the dimension of the transmitted signal waveforms. Signal demodulator can be realized in two ways. They are

A) Based on the use of signal correlators (product integrators)
B) Based on the use of matched filters.

2. Detector:

The function of the detector is to decide which of the M possible signal waveforms was transmitted based on the vector x. The optimum detector is designed to minimize the probability of error.

4.5.1 Correlation Receiver:

The basic function of a correlator is to product integrate the received noisy signal with each of the reference carrier signals. It decomposes the received signal into N-dimensional vectors $(x_1, x_2, \dots x_N)$. The Figure 4.20 shows the block diagram of correlation type receiver, using a bank of N correlators.

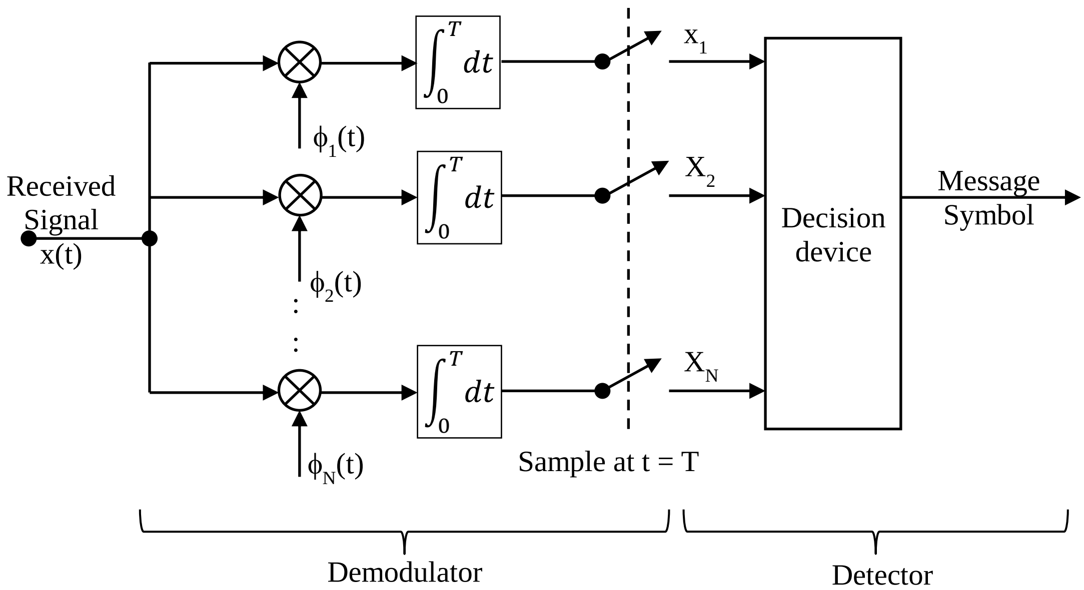


Figure 4.20 Correlator receiver

Here the demodulators imply the use of analog hardware (multipliers and integrators) and continuous signals. The mathematical operation of a correlator is correlation; a signal is correlated with a replica of itself. The demodulator outputs are sampled at the rate t=T to obtain the vector $x=x_1, x_2, \ldots x_N$.

A decision device is used as a detector. The function of the detector is to decide which of the symbols was actually transmitted. The decision rule for the detector is to choose a symbol based on location of received vector x in the particular decision regions of the signal space.

4.5.2 Matched Filter Receiver:

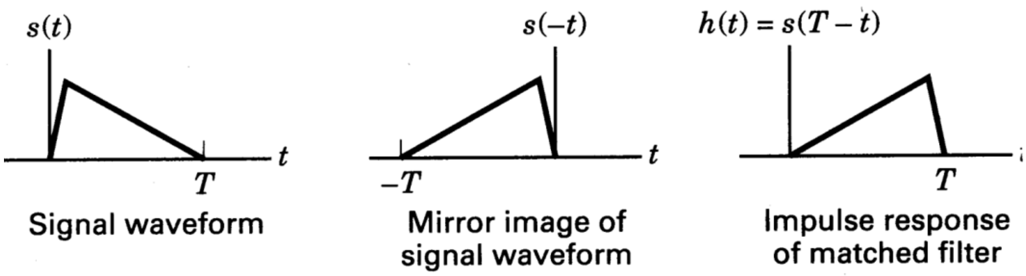


Figure 4.21 Matched filter characteristics

The mathematical operation of a matched filter is convolution; a signal is convolved with the impulse response of a filter. A matched filter is a linear filter designed to provide the maximum signal-to- noise power ratio at its output for a given transmitted symbol waveform. Also, the impulse response of the filter is a delayed version of the mirror image (rotated on the t = 0 axis) of the signal waveform. Therefore, if the signal waveform is S(t), its mirror image is S(-t), and the mirror image delayed by T seconds is S(T-t), as shown in the Figure 4.21.

Thus a matched filter can be implemented using digital hardware and sampled waveforms. The figure 4.22 shows the block diagram of matched filter receiver. Here we use a bank of N linear filters followed by envelope detectors. Envelope detectors are used to avoid poor sampling that arises in the absence of prior information about the phase θ.The demodulator outputs are sampled at the rate t=T to obtain the vectors $x=x_1, x_2, \ldots x_N$. The decision device follows the decision rule to decide which of the symbols was actually transmitted.

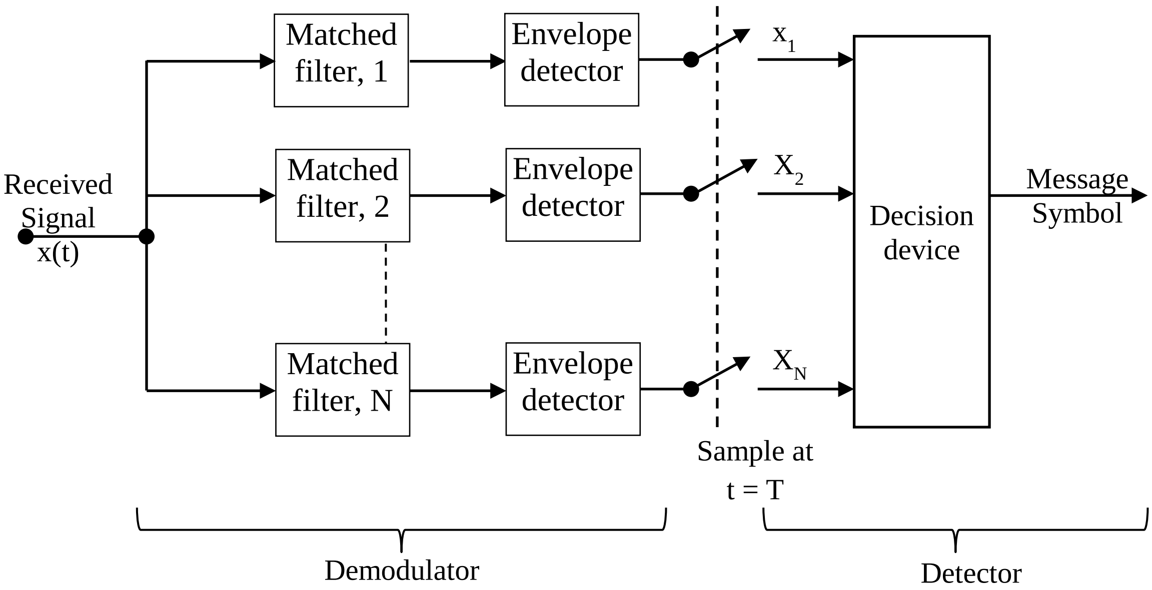


Figure 4.22 Matched filter receiver

4.6 COHERENT DETECTION OF PSK

The figure 4.23 shows the block diagram of coherent binary PSK receiver.

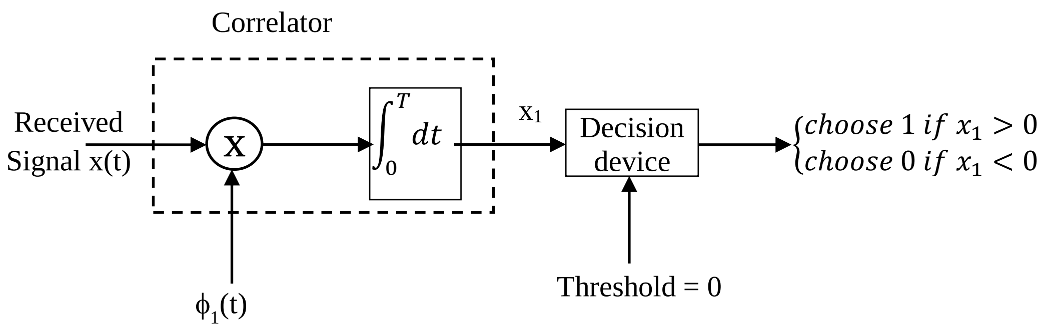


Figure 4.23 coherent binary PSK receiver

The noisy BPSK signal x(t) received from the channel is applied to a correlator. The correlator is also supplied with a locally generated coherent reference signal $\phi_1(t)$. The correlator output, x_1, is compared with a threshold of zero volts. If $x_1>0$, the receiver decides in favour of symbol 1. If $x_1<0$, it decides in favour of symbol 0. If x_1 is exactly zero, the receiver makes a random guess in favour of 0 or 1. The average probability of symbol error or, equivalently, the bit error rate for coherent BPSK is

$$P_e = \frac{1}{2}\,\text{erfc}\left[\sqrt{\frac{E_b}{N_0}}\right] \qquad (4.35)$$

Signal - Space Diagram:

A coherent BPSK system is characterized by having a signal space that is one-dimensional (ie., N=1), and with two message points (ie.,M=2), as shown in figure 4.24

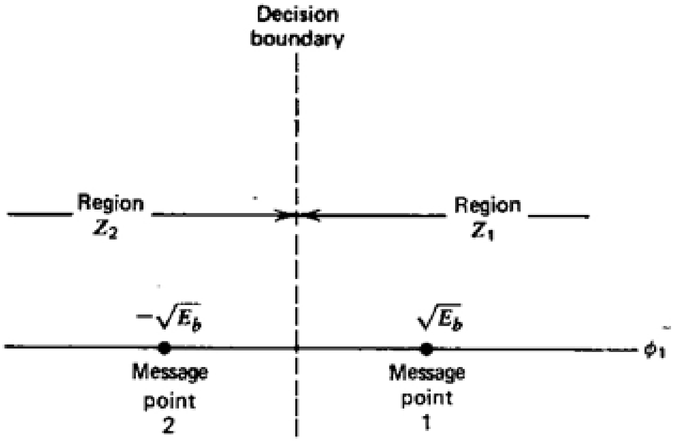


Figure 4.24 Signal-Space diagram for coherent BPSK

The coordinates of message points are given by

$S_{11} = +\sqrt{E_b}$ and

$S_{21} = -\sqrt{E_b}$

Hence the message point corresponding to $S_1(t)$ is located at $S_{11} = +\sqrt{E_b}$. The message point corresponding to $S_2(t)$ is located at $S_{21} = -\sqrt{E_b}$. Now the decision rule is

- If the received signal point falls in region Z_1, guess that symbol 1 was transmitted.
- If the received signal point falls in region Z_2, guess that symbol 0 was transmitted.

4.7 SAMPLED MATCHED FILTER

A matched filter is a linear filter designed to provide the maximum signal to noise power ratio. The impulse response of the matched filter is a delayed version of the mirror image (rotated on the t = 0 axis) of the input signal waveform. Therefore, if the signal waveform is S(t), its mirror image is S(-t), and the mirror image delayed by T seconds is S(T-t). Thus the impulse response h(t) of a filter matched to S(t) is described by

$$h(t) = \begin{cases} S(T-t) & 0 \le t \le T \\ 0 & elsewhere \end{cases} \qquad (4.36)$$

Matched filter can be implemented using digital techniques and sampled waveforms. The figure 4.25 shows how a matched filter can be implemented using digital hardware.

The input signal X(t) comprises a prototype signal $S_i(t)$ plus noise n(t). The bandwidth of the signal is $W = \frac{1}{2T}$, where T is the symbol time. Thus, the minimum Nyquist sampling rate is $f_s = 2W = \frac{1}{T}$. The sampling time T_b needs to be equal to or less than the symbol time. Therefore, there must be atleast one sample per symbol. In real systems, sampling is usually performed at a rate that exceeds the Nyquist minimum by a factor of 4.

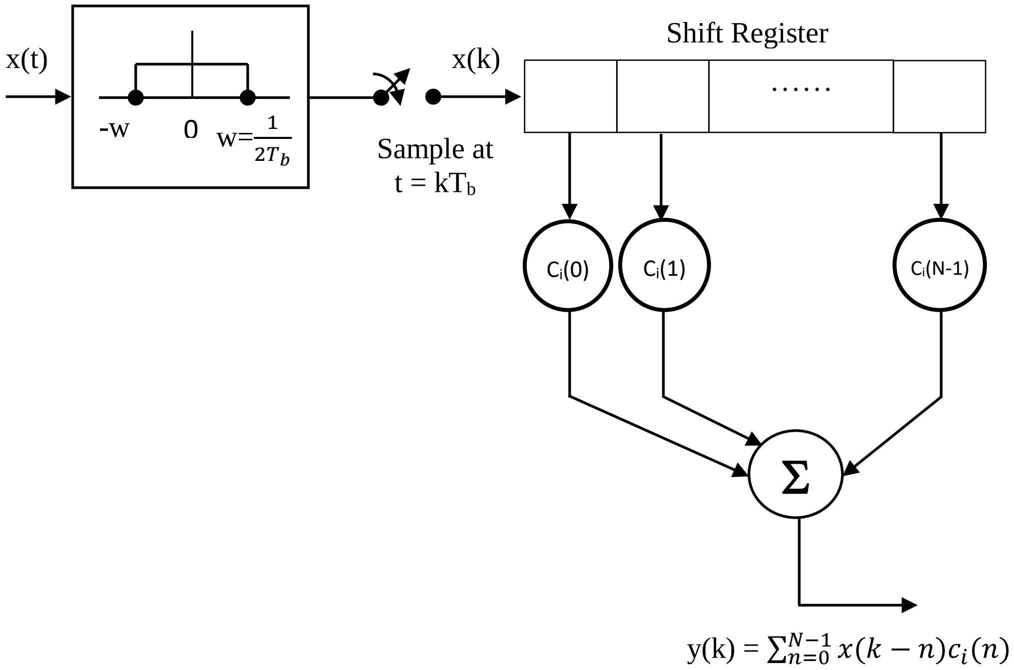


Figure 4.25 Sampled matched filter

At the clock times of t = kT_b, the samples are shifted into the register, so that earlier samples are located to the right of later samples. Once the received signal has been sampled, the continuous time notation t is changed to kT_b, or simply to k. Then the simple discrete notation is

$$x(k) = S_i(k) + n(k) \qquad i = 1, 2 \qquad k = 0, 1, \ldots \qquad (4.37)$$

where k is the sampling time index.

In the shift register, its coefficients or weights c_i (n) approximate a matched filter. Here n = 0, 1, …. N – 1 represents the time index of the weights and register stages. N represents the number of stages in the register and the number of samples per symbol. By using the discrete form of the convolution integral, the output at a time corresponding to the k^{th} sample is given by

$$Y_i(k) = \sum_{n=0}^{N-1} x(k-n)c_i(n) \qquad k = 0, 1, \ldots \text{modulo} - N \qquad (4.38)$$

Following the summer, a symbol decision will be made after N time samples have entered the registers.

Relation between correlator and matched filter:

Similarity

Even though the mathematical operation of an matched filter to be convolution of a signal with the impulse response of the filter, the end result appears to be the correlation of a signal with a replica of that same signal. That is why it is valid to describe a correlator as an implementation of a matched filter.

Difference

There is an important distinction between the matched filter and correlator. Since the correlator yields a single output value per symbol, it must have side information, such as the start and stop times over which the product integration should take place. If there are timing errors in the correlator, then the sampled output fed to the detector may be badly degraded.

On the other hand, the matched filter yields a time series of output values (reflecting time shifted input samples multiplied by fixed weights). Then with the use of additional circuitry, the best time for sampling the matched filter output can be learned.

4.8 COHERENT DETECTION OF FSK

The figure 4.26 shows the block diagram of coherent binary FSK receiver.

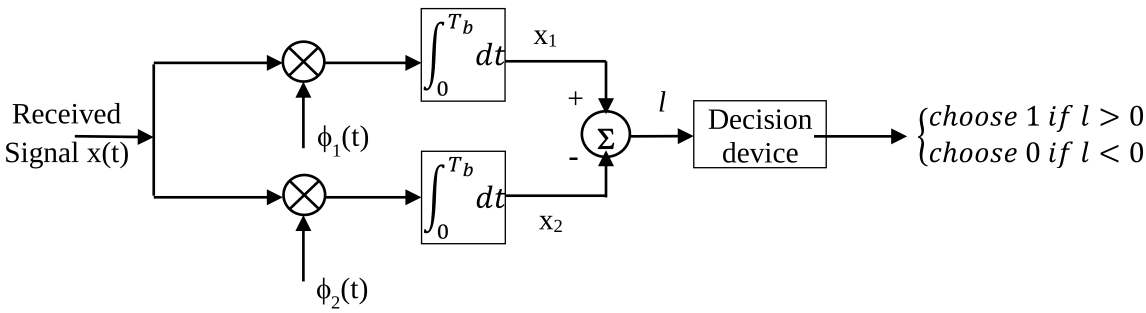


Figure 4.26 coherent binary FSK receiver

The noisy BFSK signal x(t) received from the channel is applied to the pair of correlators. The two correlators are supplied with locally generated coherent reference signals $\phi_1(t)$ and $\phi_2(t)$. The correlator outputs x_1 and x_2 are then subtracted one from the other. The resulting difference, l is compared with a threshold of zero volts. If l > 0, the receiver decides in favour of symbol 1. If l < 0, it decides in favour of symbol 0. If l is exactly zero, the receiver makes a random guess in favour of 0 or 1. The average probability of symbol error for coherent BFSK is

$$P_e = \frac{1}{2} erfc\left[\sqrt{\frac{E_b}{2N_0}}\right] \qquad (4.39)$$

Signal space diagram

A coherent BFSK system is characterized by having a signal space that is two-dimensional (ie. N = 2) with two message points (ie., M = 2) as shown in Figure 4.27.

The two message points are defined by the signal vectors

$$S_1 = \begin{bmatrix} \sqrt{E_b} \\ 0 \end{bmatrix} \qquad S_2 = \begin{bmatrix} 0 \\ \sqrt{E_b} \end{bmatrix}$$

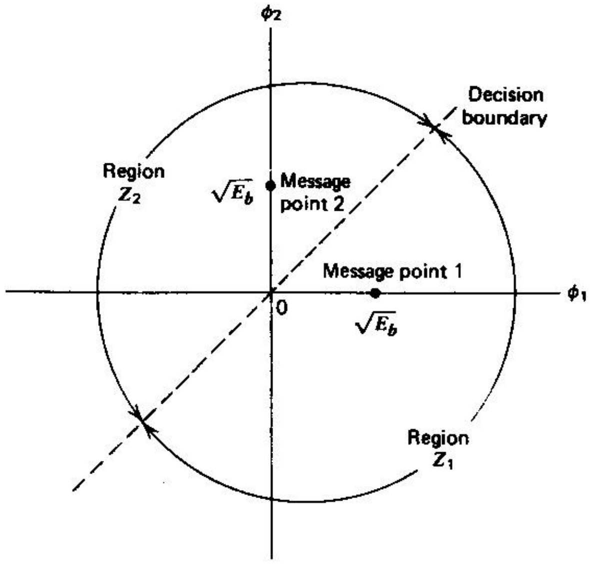


Figure 4.27 signal space diagram for BFSK

The distance between the two message points is equal to $\sqrt{2E_b}$. Now the decision rule is

- If the received signal vector x falls in region Z_1 (such that $x_1 > x_2$), guess that symbol 1 was transmitted.
- If the received signal vector x falls in region Z_2 (such that $x_1 < x_2$), guess that symbol 0 was transmitted.

4.9 NON-COHERENT DETECTION

When it is impractical to have knowledge of the carrier phase at the receiver, we use non coherent detection process. 'Non coherent' means doing without phase information. Non coherent detection techniques are less complex. However, the probability of error is high compared to coherent detection.

We treat non coherent binary FSK and DPSK signals as special cases of non-coherent orthogonal modulation. Hence, we shall see about the non-coherent receiver for detection of binary FSK and DPSK signals.

4.9.1 Non coherent detection of BFSK

In BFSK scheme, the transmitted signals are,

$$S_1(t) = \sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_1 t) \qquad (4.40)$$

$$S_2(t) = \sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_2 t) \qquad (4.41)$$

The transmission of frequency f_1 represents symbol 1, and the transmission of frequency f_2 represents symbol 0.

The Figure 4.28 shows the non-coherent receiver for detection of binary FSK signals.

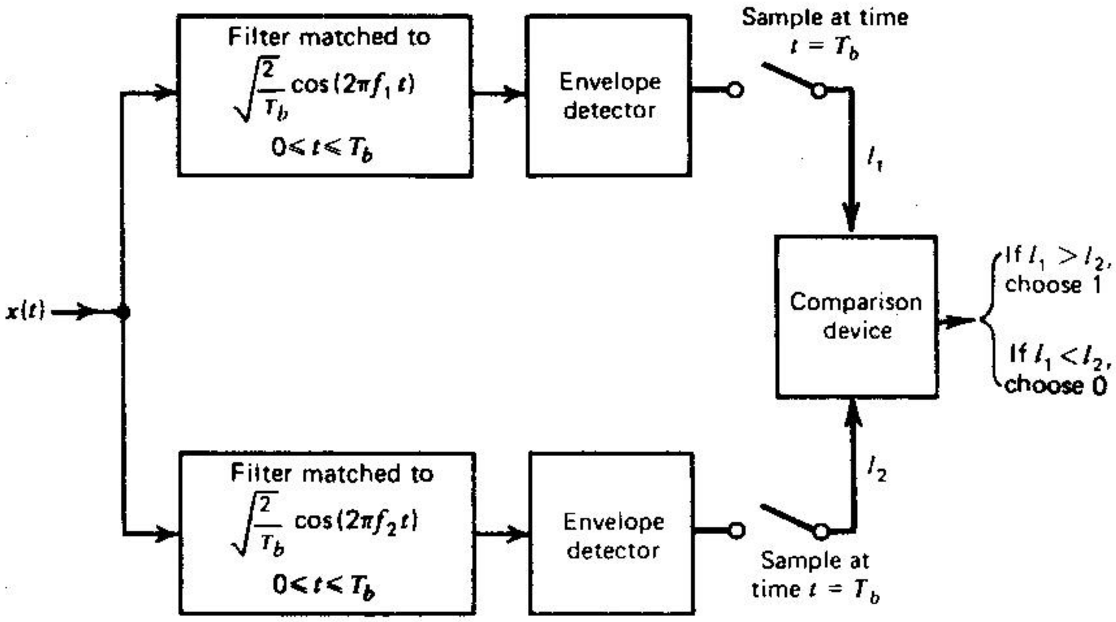


Figure 4.28 Non-coherent receiver for BFSK Digital communications

The receiver consists of a pair of matched filters followed by envelope detectors. The filter in the upper path of the receiver is matched to $\sqrt{\frac{2}{T_b}}\cos(2\pi f_1 t)$ ie.,$\phi_1(t)$. The filter in the lower path is matched to $\sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_2 t)$ ie.,$\phi_2(t)$.

The resulting envelope detector outputs are sampled at $t = T_b$. The envelope samples of the upper and lower paths are l_1 and l_2 respectively. The comparison device compares the values of l_1 and l_2. On comparison,

- If $l_1 > l_2$, the receiver decides in favour of symbol 1.
- If $l_1 < l_2$, the receiver decides in favour of symbol 0.

If $l_1 = l_2$, the receiver simply makes a guess in favour of symbol 1 or 0. The average probability of error for non-coherent binary FSK is given by

$$p_e \quad = \quad \frac{1}{2} exp\left(-\frac{E_b}{2N_0}\right). \qquad (4.42)$$

4.9.2 Non-coherent detection of binary differential PSK

The term differentially coherent detection of DPSK, refers to a detection scheme often classified as non-coherent because it does not require a reference in phase with the received carrier.

The Figure 4.29 shows the block diagram of non-coherent receiver for detection of DPSK signals.

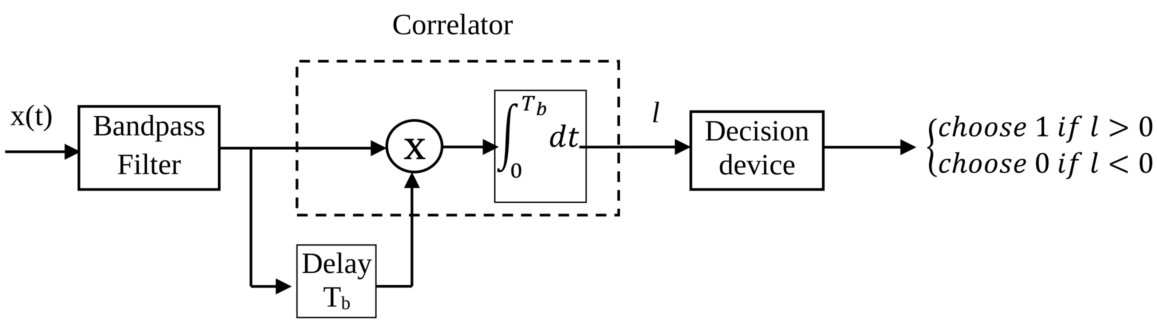


Figure 4.29 Non-coherent receiver for DPSK

The received DPSK signal plus noise is passed through a Bandpass filter centred at the carrier frequency f_c, so as to limit the noise power. The filter output is delayed by one bit interval T_b. Both the filter output and its delayed version are applied to a correlator.

The resulting correlator output is proportional to the cosine of the difference between the carrier phase angles in the two correlator inputs. The correlator output l is finally compared with a threshold of zero volts. The decision is taken such that

- If l > 0, the phase difference between the waveforms received during the pair of bit intervals lies inside the range $-\frac{\pi}{2}$to $\frac{\pi}{2}$. The receiver decides in favour of symbol 1.
- If l < 0, the phase difference lies outside the range $-\frac{\pi}{2}$to $\frac{\pi}{2}$, modulo 2π. The receiver decides in favour of symbol θ.

DPSK is a special case of non-coherent orthogonal modulation with $T = 2T_b$ and $E = 2E_b$. The average probability of error for DPSK is given by

$$P_e \quad = \quad \frac{1}{2} exp\left(-\frac{E_b}{N_o}\right) \qquad (4.43)$$

4.10 ALLOCATION OF THE COMMUNICATIONS RESOURCE

One of the important design goals in a digital communication system is to achieve maximum throughput (ie., maximum data rate). Three basic ways are used to achieve maximum data rate.

1. Increase the transmitter's EIRP (Effective Isotropic Radiated Power) or reduce the system losses.
2. Provide more channel bandwidth.
3. More efficient distribution / utilization of the communication resources.

Multiplexing and Multiple Access are the two methods for more efficient distribution / utilization of the communication resources. The communication resources utilized are frequency, time, wavelength, code, space and polarization.

Multiplexing

Multiplexing may be defined as the process of simultaneously transmitting two or more individual signals over a single communication channel. Using multiplexing, more information can be transmitted at a time. The typical applications of multiplexing are in telemetry and telephony or in satellite communication. There are three basic types of multiplexing.

1. Time Division Multiplexing (TDM).
2. Frequency Division Multiplexing (FDM).
3. Wavelength Division Multiplexing (WDM).

Generally, the FDM and WDM systems are used to handle analog information. TDM systems are often used to handle the digital information. Here we shall see about TDM in detail.

4.10.1 Time Division Multiplexing (TDM)

In TDM, group of signals are sampled sequentially in time at a common sampling rate and then multiplexed for transmission over a common channel. This enables us to combine several digital signals, such as computer outputs, digitized voice signals, digitized facsimile and television signals, into a single data stream with a higher bit rate.

4.10.1.1 A PAM / TDM system

The concept of a PAM / TDM system is shown in the Figure 4.30.

There are N analog message signals in the input. Each message signal is restricted in bandwidth by a low-pass pre-alias filter. The pre-alias filter outputs are applied to a commutator. The commutator is an electronic switching circuitry which takes a narrow sample of each of the N input messages at rate f_S. Such multiplexed samples are then applied to a pulse-amplitude modulator. It transforms the multiplexed signal into a form suitable for transmission over the communication channel.

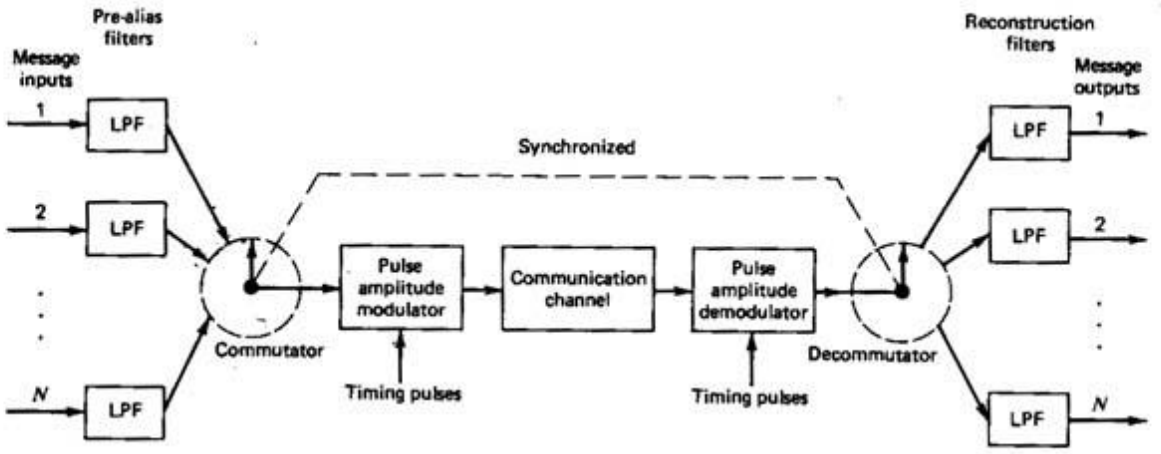


Figure 4.30 Block diagram of PAM / TDM system

The received signal is applied to a pulse amplitude demodulator. The short pulses produced at the demodulator output are distributed to the appropriate low-pass reconstruction filters by means of a decommutator. The decommutator operates in synchronism with the commutator. The transmitted message signals are reproduced at the corresponding filter outputs.

4.10.1.2 Digital TDM

The figure 4.31 shows the concept of digital TDM.

The digital data can be multiplexed by using a bit-by-bit or byte-by-byte interleaving procedure. This can be achieved by using a selector switch (MUX). The switch sequentially selects a bit or byte from each input and places it over the high speed transmission channel. At the receiving end, another switch (DEMUX) separates the bit or byte and delivers them to their respective destinations.

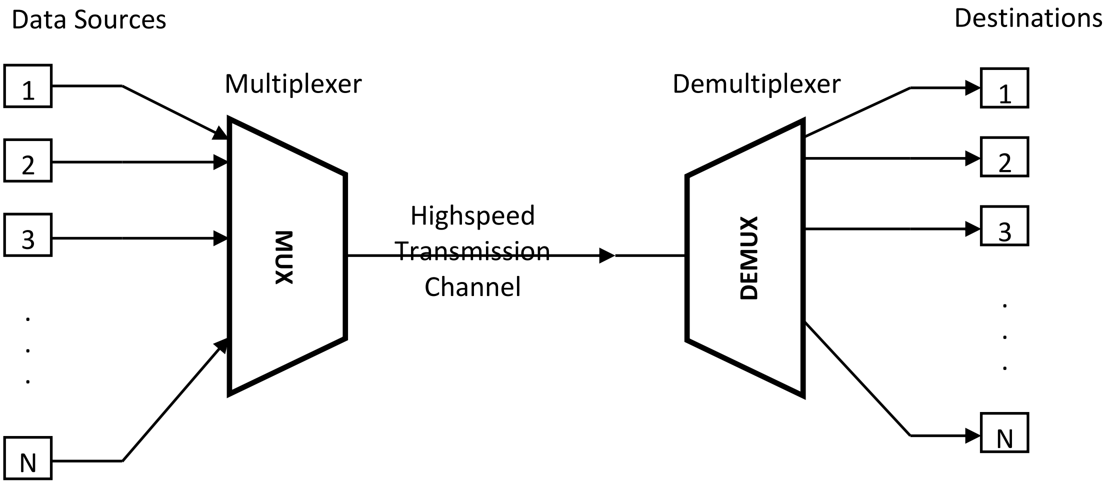


Figure 4.31 concept of digital TDM

In TDM, the communication resources are shared by assigning each of N signals or users the full spectral occupancy of the system for a short duration of time called a Time Slot as shown in Figure 4.32.

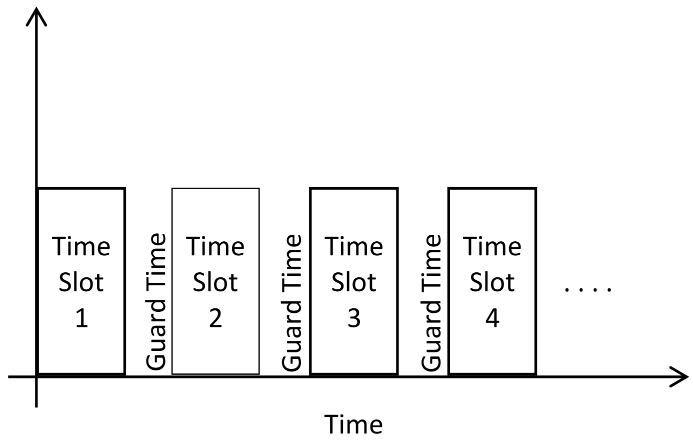


Figure 4.32 TDM Time Slots

4.10.1.3 Types of TDM

There are two types of TDM in use. They are 1) Synchronous TDM and Asynchronous TDM.

1) Synchronous or Deterministic TDM

It has a constant delay and bandwidth for a given individual communication channel. Time slots have constant length (Capacity) and used in a synchronous periodical manner. It is used in techniques like ISDN, PDH and SDH.

2) Asynchronous or statistical TDM

It has a variable delay and bandwidth for a given individual communication channel. Time slots have variable length and are used on demand. It is used in technologies like X25, Frame relay, ATM or IP.

4.10.1.4 TDM Frame structure

There are some basic requirements involved in the design of a digital TDM multiplexer.

1. Digital signals cannot be directly interleaved into a format that allows for their eventual separation unless their bit rates are locked to a common clock. Accordingly, provision has to be made for synchronization of the incoming digital signals, so that they can be properly interleaved.
2. The multiplexed signal must include some form of framing, so that its individual components can be identified at the receiver.

Hence the TDM system uses a frame structure for placing data in each time slot following a synchronized pattern. The TDM frame structure is shown in Figure 4.33. The TDM system divides the data stream into frames which repeat indefinitely. Each TDM frame is then divided into equal timeslots which are allocated to individual message signal or user. The individual users then transmit or receive only in their own time slot.

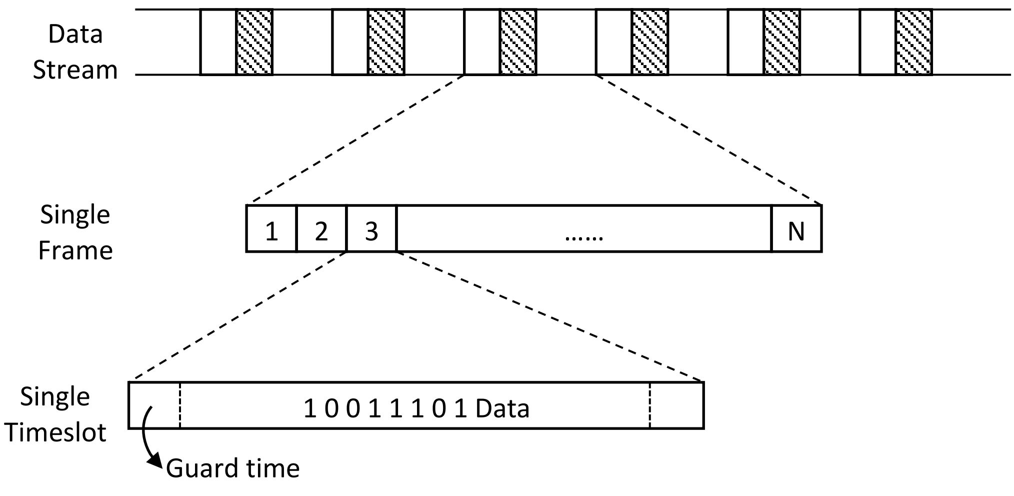


Figure 4.33 TDM Frame structure

The frame formats of synchronous and statistical TDM are shown in the Figure 4.34.

F	1	2	3	……	N

F → Framing bit
1, 2, ... N → Time Slots

(a) Synchronous TDM Frame

Flag	Address	Control	Data	FCS	Flag

(b) Statistical TDM Frame

Figure 4.34 Frame formats of TDM

4.11 ASCII FRAMING

The frames which are structured using character oriented protocols and character stuffing are referred as character oriented frames. They use a normal frame format and carry characters coded with protocol code. ASCII (American Standard Code for Information Interchange) is one example of such protocol code. A host PC uses ASCII characters to send commands to a device and then receives responses back from that device. The ASCII command set is used to configure devices, send data to devices and to read data and status information back from devices. If the character oriented frame uses ASCII serial communication protocol, then it is called as an ASCII frame.

ASCII Message Frame

The Figure 4.35 shows an example ASCII frame. The ASCII serial communication protocol is used in this frame. It is used to transfer data between a master computer station and a slave device such as power meter, or panel meter.

Field No.	1	2	3	4	5	6	7
Contents	SYNC	Message length	Slave address	Message type	Message body	Check sum	Trailer (CRLF)
Length, Character	1	3	2	1	0 to 246	1	2

Figure 4.35 Example ASCII Message Frame

The following specifies the ASCII message frame.

SYNC: Synchronization character: One character “!” (ASCII 33) is used for starting synchronization.

Message length

The length of the message including only number of bytes in fields #2, #3, #4 and #5. It contains three characters between ‘006’ and ‘252’.

Slave Address

Two characters from ‘00’ to ‘99’. The instrument with address ‘00’ responds to requests with any incoming address.

Message Type

One character representing the type of a host request. A list of message types is shown in Tables 4.6 and 4.7.

Message Body

It contains the message parameters in ASCII representation. The data fields vary in length depending on the data type, from 0 to 246.

Check sum

Arithmetic sum, calculated in a 2 byte word over fields #2, #3, #4 and #5 to produce a one byte check sum in the range of 22to 7E (hexadecimal).

Trailer

Two ASCII characters Carriage Return (CR) (ASCII 13) and Line Feed (LF) (ASCII 10) are used.

Table 4.6 Specific ASCII Requests

Message Type		Description
Char	ASCII Hex	
0	30h	Read basic data registers
1	31h	Read basic setup
2	32h	Write basic setup
3	33h	Read Instrument status
4	34h	Reset / Clear functions
8	38h	Reset the instrument

Table 4.7 Direct Read / Write ASCII Request

Message Type		Description
Char	ASCII Hex	
A	41h	Long-size direct read
a	61h	Long-size direct write
X	58h	Variable-size direct read
x	78h	Variable-size direct write

4.12 ARCHITECTURES OF SYNCHRONOUS TDM

Only a hierarchical digital multiplexing infrastructure can connect millions of customers across the city / country / world. The local network used is simple star and the wide area network is point to point trunks or ring topologies.

There are two main architectures for standard based synchronous TDM on trunk lines for carrying PCM-coded digital telephony.

1. Plesiochronous Digital Hierarchy (PDH)
2. Synchronous Digital Hierarchy (SDH)

Plesiochronous means nearly synchronous. The network is not synchronized but fast enough to synchronize sender and receiver. In PDH, there are two framing methods based on the carrier system. They are

1. E1 Framing based on E1 carrier system.
2. T1 Framing based on T1 carrier system.

4.12.1 E1 Framing for Telephone

The E1 framing based on E1 carrier is an European hierarchy. The conference of European post and Telecom (CEPT) administrations originally standardized the E carrier system. E1 carrier standards are framed by European Territory Standards Institute (ETSI). An E1 link operates over two separate sets of wires, usually coaxial cable. It employs A-law algorithm for companding. This is most suitable for voice transmission.

The most commonly used voice coding is PCM where the analog signal is sampled at the rate of 8000 samples per second, and quantized by an 8 bit coder. Hence the data rate of one PCM voice channel is 8000 x 8 = 64 kbps.

The telephone companies implement TDM through hierarchy of digital signals. This is called as Digital Signal (DS) service. The table 4.8 lists the European digital signals and their data rates.

Table 4.8 E1 hierarchy

Signal / Service	Carrier / Line	No of Channels	Data Rate
DSO	E0	1	64 kbps
CEPT 1	E1	32	2.048 mbps
CEPT 2	E2	128	8.448 mbps
CEPT 3	E3	512	34.368 mbps
CEPT 4	E4	2048	139.264 mbps
CEPT 5	E5	8192	565.148 mbps

The format of E1 framing is shown in the figure 4.36

Framing (8 bits) ↓   Signalling (8 bits) ↓

TS0	TS1	TS2	TS3	……	TS16	……	TS28	TS29	TS30	TS31

TS→ Time slot each of 8 bits, one voice channel sample

TS1 to TS15 and TS17 to TS31 → 30 voice channels

1 Frame = 32 x 8 = 256 bits

Figure 4.36 E1 Framing format

The E1 framing data rate is 2.048 Mbps (full duplex). The frame is split into 32 time slots, each being allocated 8 bits. The timeslots are numbered from TSO to TS31. The E1 frame repetition rate is 8KHZ.

Time Slot TS0

This slot is used for synchronization. It is reserved for framing purpose to indicate the start of each frame.

Timeslots TS1 to TS15 and TS17 to TS31

These 30 time slots are used for carrying user data.

Time Slot TS16

This slot is used for signaling information. This includes control, call setup and teardown. In E1 carrier, a 4-bit signaling information is used per time slot in every 16^{th} frame. This is called channel Associated Signaling (CAS) used for channel synchronization.

4.12.2 T1 Framing for Telephone

The T1 framing based on T1 carrier is a North American hierarchy. The T1 carrier standards are framed by American National Standards Institute (ANSI). The T1 link operates over special low capacitance two separate sets of shielded twisted pair cabling. In some cases unshielded twisted pair cable is also used with precautions to avoid cross talk. It employs μ-law algorithm for companding. This is also most suitable for voice transmission.

The most commonly used voice coding is PCM. Hence the data rate of one PCM voice channel is 64 kbps. T1 carrier is also implemented using the Digital signal (DS) service. The table 4.9 lists the American digital signals and their data rates.

Table 4.9 T1 hierarchy

Signal / Service	Carrier / Line	No of Channels	Data Rate
DSO	-	1	64 Kbps
DS1	T1	24	1.544 Mbps
DS1C	T1C	48	3.152 Mbps
DS2	T2	96	6.312 Mbps
DS3	T3	672	44.736 Mbps
DS4	T4	4032	274.176 Mbps

The single PCM voice channel with data rate 64 Kbps is called as digital signal at level zero (DSO). When 24 such PCM voice channels are multiplexed using TDM, the multiplexed signal is the Digital signal at level one (DS1). Hence the T1 carrier uses the DS1 signal at the line data rate of 1.544 Mbps. The format of T1 framing is shown in the figure 4.37.

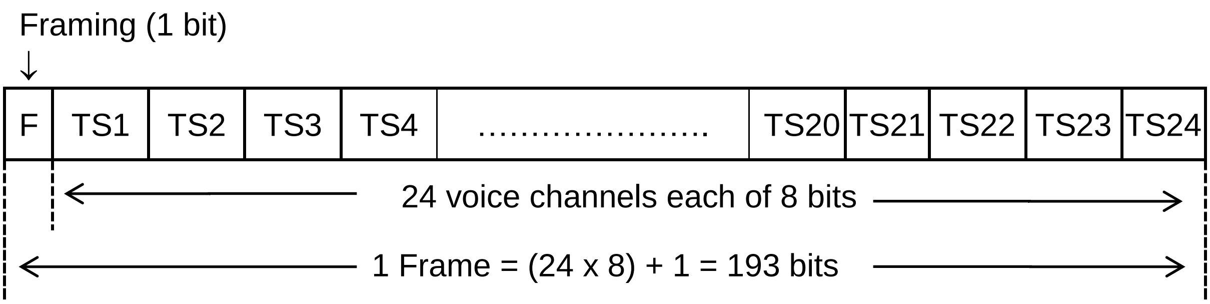


Figure 4.37 T1 Framing format

The T1 frame is split into 24 time slots each being allocated 8 bits and a framing single bit at the start. The time slots are numbered from TS1 to TS24, each representing a voice channel sample of 8 bits. The T1 frame repetition rate is 8 KHZ.

F bit

The frame synchronizing bit 'F' is used to provide synchronisation as well as to indicate the start of a frame.

TS1 to TS24

These 24 time slots are used for carrying user data.

In T1 carrier, there is no dedicated time slot for channel associated signaling (CAS). Instead 'Robbed bit' signaling is used. Using CAS, the signaling information is transmitted by robbering certain bits, which are normally used for data. The signaling is placed in the LSB of every timeslot in the 6^{th} and 12^{th} frame of every D4 super frame.

4.12.3 Comparison of E1 and T1 carriers

The Table 4.10 lists the performance comparison of E1 and T1 carriers.

Table 4.10 comparison of E1 and T1 frames

Sl. No.	Parameter	E1 Frame / Carrier	T1 Frame / Carrier
1.	Implemented in the Country	Europe	North America
2.	Standard	CEPT, ETSI	ANSI
3.	Digital Signal	DSO = 64 KHZ PCM voice channel	DSO = 64 KHZ PCM voice channel
4.	Hierarchy	E1, E2, E3, E4	T1, T2, T3, T4
5.	Cable used	Coaxial Cable	Twisted pair (STP, UTP)
6.	Line data rate	2.048 Mbps	1.544 Mbps
7.	Companding	A-law	μ-law
8.	No. of Time slots	32	24
9.	No. of channels	30	24

SHORT QUESTIONS AND ANSWERS

1. Define Digital Modulation.

Digital modulation may be defined as mapping a sequence of input binary digits into a set of corresponding high frequency signal waveforms.

2. List the various types of digital modulation techniques.

I. Based on the method of detection:

1. Coherent digital modulation
2. Non-coherent digital modulation

II. Based on the mapping techniques:

Binary Scheme	Quaternary Scheme	M-ary Scheme	Hybrid Scheme
1. BASK	1. QPSK	1. M-ary ASK	1. QAM
2. BFSK	2. MSK	2. M-ary FSK	2. APK
3. BPSK		3. M-ary PSK	

III. Based on the performance of the modulation scheme and properties of the modulated signal

1. Power efficient scheme / Bandwidth efficient scheme.
2. Continuous Phase (CP) Modulation / In phase – Quadrature Phase (IQ) Modulation.
3. Constant Envelope Modulation / Non-constant Envelope Modulation.
4. Linear Modulation / Non-linear Modulation.
5. Modulation scheme with memory / Modulation scheme without memory.

3. Mention the design goals of digital communication system.

1. Maximum data rate
2. Minimum possibility of symbol error.
3. Minimum transmitted power.
4. Minimum channel bandwidth.
5. Maximum resistance to interfering signals.
6. Minimum circuit complexity.

4. What is meant by coherent binary modulation technique?

The binary modulation scheme has three basic forms. They are

1. Binary Amplitude Shift Keying (BASK)
2. Binary Frequency Shift Keying (BFSK)
3. Binary Phase Shift Keying (BFSK)

When these binary modulation schemes employ coherent detection at the receiver, then they are called as coherent binary modulation techniques.

In coherent detection, the local carrier generated at the receiver is phase locked with the carrier at the transmitter.

5. Define ASK, FSK and PSK

ASK: In Amplitude Shift Keying (ASK), the modulation process involves switching or keying the amplitude of the carrier signal in accordance with the incoming data.

FSK: In Frequency Shift Keying (FSK), the modulation process involves switching or keying the frequency of the carrier signal in accordance with the incoming data.

PSK: In Phase Shift Keying (PSK) the modulation process involves switching or keying the phase of the carrier signal in accordance with the incoming data.

6. What are the merits and demerits of BPSK?

Merits

- BPSK requires lower bandwidth than BFSK.
- BPSK has the minimum value of probability of error. Hence, it provides best performance compared to BFSK and BASK schemes.
- It has very good noise immunity.

Demerits

- In PSK, the information lies in the phase, and hence, it cannot be detected non-coherently.

7. What are the merits and demerits of BFSK?

Merits

- It is relatively easy to implement.
- It has better noise immunity than ASK.

Demerits

- BFSK requires high-bandwidth compared to BPSK and BASK.

8. What are the merits and demerits of BASK?

Merit

- BASK is easy to generate and detect.

Demerit

- It is very sensitive to noise.

9. What is meant by Non-coherent binary modulation technique?

The modulation scheme in which the detection process does not need receiver carrier to be phase locked with the transmitter carrier is said to be non-coherent modulation technique. The non-coherent binary modulation techniques are

1. Differential Phase Shift Keying (DPSK)
2. Binary Amplitude Shift Keying (BASK)
3. Binary Frequency Shift Keying (BFSK)

10. Define DPSK

Differential Phase Shift Keying (DPSK) is a "Pseudo PSK" technique and can be viewed as the non-coherent form of BPSK. It eliminates the need for a coherent reference signal at the receiver by combining two basic operations at the transmitter. They are

1. Differential encoding of the input binary wave
2. Phase shift keying

11. What are the merits and demerits of DPSK?

Merits

- DPSK scheme does not need carrier at the receiver end. Hence it has reduced system complexity.
- The bandwidth required is less than that required for BPSK.

Demerits

- It has higher value of probability of error than that of BPSK.
- Noise interference is more.
- In DPSK, previous bit is used to detect next bit. Hence, there is possibility of errors appearing in pairs.

12. What is meant by coherent quadrature modulation technique?

The M-ary modulation scheme with m=4 is said to be quadrature modulation scheme. The quadrature modulation scheme in which coherent detection is employed at the receiver is called as the coherent quadrature modulation technique. There are two bandwidth conserving quadrature modulation schemes. They are

1. QuadriPhase Shift Keying (QPSK)
2. Minimum Shift Keying (MSK)

13. Define QPSK

In QPSK, as with BPSK, information carried by the transmitted signal is contained in the phase. For QPSK, the bits per symbol is k = 2 and hence $m = 2^k = 2^2 = 4$. Hence, two successive bits (dibit) in the data sequence are used to modulate two quadrature carriers.

14. What is meant by signal constellation?

For any modulation scheme, the analysis is based on the signal space diagram. Signal space approach is a plotting of possible message points. Such a set of possible message points is also referred to as a signal constellation.

15. What are the merits and demerits of QPSK?

Merits

- QPSK has very good noise immunity.
- More effective utilization of the available bandwidth of the transmission channel.
- It has low error probability.

Demerits

- The generation and detection of QPSK is complex.

16. Define MSK

Minimum Shift Keying (MSK) is a special form of binary CPFSK signal. A Continuous Phase Frequency Shift Keying (CPFSK) signal with a deviation ratio of h = ½ is referred to as MSK.

17. Define Deviation Ratio.

The difference between the frequencies f_1 and f_2, normalised with respect to the bit rate $\frac{1}{T_b}$ defines the dimensionless parameter h, which is referred to as the deviation ratio. For MSK, deviation ratio h = ½.

18. Justify the name Minimum Shift Keying (MSK)

The deviation ratio is $h = T_b (f_1 - f_2)$

For MSK, we have $h = \frac{1}{2}$. Therefore,

$$\frac{1}{2} = T_b (f_1 - f_2)$$

$$\Rightarrow \quad (f_1 - f_2) = \frac{1}{2T_b} = \frac{R_b}{2}$$

Where R_b → bit rate

Hence the frequency deviation $(f_1 - f_2)$ equals half the bit rate. This is the minimum frequency spacing that allows the two FSK signals representing symbols 1 and 0 to be coherently orthogonal in the sense that they do not interfere with one another in the process of detection. It is for this reason, a CPFSK signal with a deviation ratio of one-half is referred to as Minimum Shift Keying (MSK).

19. What are the merits and demerits of MSK?

Merits

- MSK scheme has constant envelope (ie., there are no amplitude variations)
- It has coherent detection performance equivalent to that of QPSK.
- The MSK signal has a continuous phase.
- Filters to suppress the sidelobes which causes interchannel interference are not required.

Demerits

- The generation and detection of MSK signal is more complicated.
- For a wireless communication system using MSK, the adjacent channel interference is not low enough to satisfy the practical requirements of such a multiuser communications environment.

20. What are the two basic steps in a digital receiver? Explain

The two basic steps in a digital receiver are 1) Demodulation and 2) Detection

1. Demodulator:

The demodulator is a frequency down conversion block. The function of the signal demodulator is to convert the received waveform X(t) into an N-dimensional vector X = [X_1, X_2, ... X_N] where N is the dimension of the transmitted signal waveforms.

2. Detector

The function of the detector is to decide which of the M possible signal waveforms was transmitted based on the Vector X. The optimum detector is designed to minimize the probability of error.

21. What is meant by coherent and non-coherent detection in a digital receiver?

Coherent Detection

In coherent detection, the local carrier generated at the receiver is phase locked with the carrier at the transmitter. The receiver knows the instants of time when the modulation changes state. The coherent detection is a synchronous detection scheme.

Non-Coherent detection

In non-coherent detection, the detection process does not need receiver carrier to be phase locked with the transmitter carrier. There is no time and phase synchronization between the transmitter and receiver. Hence the non-coherent detection is an asynchronous detection scheme.

22. Define sampled matched filter

A matched filter may be defined as a filter whose impulse response is a delayed version of the mirror image (rotated on the t = 0 axis) of the input signal waveform. Thus the impulse response h(t) of a filter matched to s(t) is given by

$$h(t) = \begin{cases} s(T-t) & 0 \le t \le T \\ 0 & elsewhere \end{cases}$$

If the matched filter is implemented using digital techniques and sampled waveforms, then it is called as sampled matched filter.

23. Define Correlator.

The basic function of a correlator is to product integrate the received noisy signal with each of the reference carrier signals. It decomposes the received signal into N-dimensional vectors (X_1, X_2, ... X_N).

24. Define Time Division Multiplexing (TDM)

In TDM, group of signals are sampled sequentially in time at a common sampling rate and then multiplexed for transmission over a common channel. TDM is used to handle digital information. This enables us to combine several digital signals, such as computer outputs, digitized voice signals, digitized facsimile and television signals into a single data stream with a higher bit rate.

25. What are the types of TDM?

There are two types of TDM in use. They are

1. Synchronous TDM and 2) Asynchronous TDM

1. Synchronous or Deterministic TDM:

It has a constant delay and bandwidth for a given individual communication channel. Time slots have constant length (capacity) and used in a synchronous periodical manner. It is used in techniques like ISDN, PDH and SDH.

2. Asynchronous or statistical TDM:

It has a variable delay and bandwidth for a given individual communication channel. Time slots have variable length and are used on demand. It is used in techniques like X25, Frame relay, ATM or IP.

26. What is ASCII Framing?

American Standard Code for Information Interchange (ASCII) Codes in hexadecimal notation is used in ASCII frame for TDM. ASCII frames are structured using character oriented protocols.

27. Write short notes on EI framing.

- The EI framing based on EI carrier is an European hierarchy.
- The number of PCM encoded voice data channels is 30.
- The transmission line data rate is 2.048 Mbps.
- It employs A-law algorithm for companding and operates over coaxial cable.

28. Write short notes on TI framing

- The TI framing based on TI carrier is a North American hierarchy.
- The number of PCM encoded voice data channels is 24.
- The transmission line data rate is 1.544 Mbps.
- It employs μ-law algorithm for companding and operates over shielded twisted pair cables.

Unit – V SPREAD SPECTRUM TECHNIQUES

OBJECTIVES

- To understand the spread spectrum communication.
- To understand about Pseudo noise sequences.
- To study about the major types of spreading techniques.
- To study about the commercial applications of spread spectrum communication.

5.0 INTRODUCTION

In any digital communication system, the basic design factors are 1) efficient utilisation of channel bandwidth and 2) minimizing the transmitted power.

Some of the major problems encountered in specific communication systems are

1) Combating or suppressing the detrimental effects of interference due to jamming, interference arising from other users of the channel, and self-interference due to multipath propagation.
2) Hiding a signal by transmitting it at low power and making it difficult for an unintended listener to detect the signal.
3) Achieving message privacy in the presence of other listeners.

These problems can be successfully solved by using a technique called spread spectrum modulation. We shall discuss this modulation technique in detail in this chapter.

5.1 SPREAD SPECTRUM COMMUNICATION SYSTEM

A system is defined to be a spread spectrum communication system if it fulfills the following requirements.

1) The signal occupies a bandwidth much in excess of the minimum bandwidth necessary to send the information.
2) Spreading is accomplished by means of a spreading signal, often called a code signal, which is independent of the data.

3) At the receiver, despreading (recovering the original data) is done by the correlation of the received spread signal with a synchronized replica of the spreading signal used to spread the information.

In the transmitter of a digital communication system, such frequency spreading of signal is achieved along with the Bandpass modulator circuit.

5.2 MODEL OF SPREAD SPECTRUM DIGITAL COMMUNICATION SYSTEM

The block diagram shown in Figure 5.1 illustrates the basic elements of a spread spectrum digital communication system.

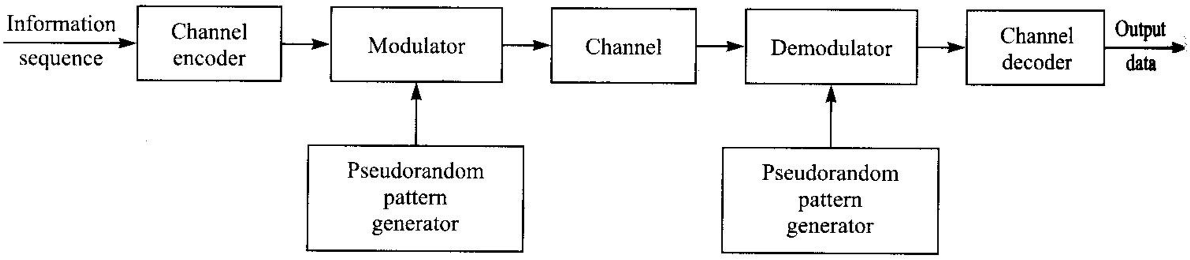


Figure 5.1 Model of spread spectrum digital communication system

The channel encoder / decoder and the modulator / demodulator are the basic elements of the digital communication system, we have already discussed. In addition to these elements we have two identical pseudorandom pattern generators. One interfaces with the modulator at the transmitting end. The second interfaces with the demodulator at the receiving end. These pseudorandom pattern generators generate a Pseudonoise (PN) binary-valued sequence which is impressed on the transmitted signal at the modulator and removed from the received signal at the demodulator.

5.3 BENEFICIAL ATTRIBUTES OF SPREAD SPECTRUM SYSTEMS

Spread spectrum modulation was originally developed for military applications where resistance to jamming (interference) is of major concern. However there are civilian applications that also benefit from the unique characteristics of spread spectrum modulation. We hereby list the following beneficial attributes of spread spectrum systems.

1) Interference suppression benefits:

(i) In combating intentional interference (jamming), the transmitter introduces an element of unpredictability or randomness (pseudorandomness) in each of the transmitted coded signal waveforms. This is known to the intended receiver only, but not to the jammer. Thus interference due to jamming is suppressed.

(ii) Resolvable multipath components resulting from time dispersive propagation through a channel may be viewed as a form of self-interference. This type of interference may also be suppressed by the introduction of pseudorandom pattern in the transmitted signal.

2) Multiple Access

Spread spectrum methods can be used as a multiple access technique in order to share a communication resource among numerous users in a coordinated manner. Interference from the other users arises in multiple access communication systems in which a number of users share a common channel bandwidth. The transmitted signals in this common channel spectrum may be distinguished from one another by superimposing a different pseudorandom pattern, also called a code, in each transmitted signal. Thus, a particular receiver can recover the transmitted information intended for it by knowing the code or key, used by the corresponding transmitter. This type of communication technique, which allows multiple users to simultaneously use a common channel for transmission of information, is called Code Division Multiple Access (CDMA).

3) Energy Density Reduction

A message may be hidden in the background noise by spreading its bandwidth with coding and transmitting the resultant signal at a low average power. Because of its low power level, the transmitted signal is said to be “covert”. It has a low probability of being intercepted (detected) by a casual listener. Hence, it is also called as a Low-Probability of Intercept (LPI) signal.

A Radiometer is a simple power measuring instrument that can be used to detect the presence of spread spectrum signals within some bandwidth, B.

4) Fine Time Resolution

Spread spectrum signals are used to obtain accurate range (time delay) and range rate (velocity) measurements in radar and navigation. Distance can be determined by measuring the time delay of a pulse as it traverses the channel.

5) Message Privacy

Message privacy may be obtained by superimposing a pseudorandom pattern on a transmitted message. The message can be demodulated by the intended receivers, who know the pseudorandom pattern or key used at the transmitter, but not by any other receivers, who do not know the key.

5.4 SPREAD SPECTRUM APPROACHES (HISTORICAL BACKGROUND)

There are two spread-spectrum approaches called Transmitted Reference (TR) and Stored Reference (SR).

(i) In a TR system, the transmitter send two versions of truly random spreading signal (wideband carrier) – one modulated by data and the other unmodulated. The receiver used the unmodulated carrier as the reference signal for despreading (correlating) the data modulated carrier.

(ii) In a SR system, the spreading code signal is independently generated at both the transmitter and the receiver. Since the same code must be generated independently at two locations, the code sequence must be deterministic, even though it should appear random to unauthorized listeners. Such random appearing deterministic signals are called pseudonoise (PN), or pseudorandom signals.

Modern spread spectrum systems use Stored Reference (SR) approach which uses a Pseudonoise (PN) or pseudorandom code signal.

5.5 PSEUDONOISE SEQUENCES

A pseudonoise (PN) sequence may be defined as a coded sequence of 1's and 0's with certain autocorrelation properties.

The PN sequence is a deterministic, periodic signal that is known to both the transmitter and receiver. Even though the signal is deterministic it appears to have the statistical properties of sampled white noise. Hence, it appears to be a truly random signal, to an unauthorised listener.

5.5.1 Randomness properties

PN sequences have many of the properties possessed by a truly random binary sequence. A random binary sequence is a sequence in which the presence of a binary symbol 1 or 0 is equally probable. There are three basic properties that can be applied to any periodic binary sequence as a test for the appearance of randomness. They are described as follows.

1) Balance Property:

In each period of the sequence, the number of 1's is always one more than the number of 0's. This property is called the balance property.

2) Run Property:

Among the runs of 1's and of 0's in each period of the sequence, one-half the runs of each kind are of length one, one-fourth are of length two, one-eighth are of length three, and so on. This property is called the Run property. A run is defined as a sequence of a single type of binary digit(s). The appearance of the alternate digit in a sequence starts a new run. The length of the run is the number of digits in the run.

3) Correlation Property

The autocorrelation function of a sequence is periodic and binary valued. This property is called the correlation property.

5.5.2 Pseudo Noise (PN) sequence generator

The class of sequences used in spread spectrum communications is usually periodic in that a sequence of 1s and 0s repeats itself exactly with a known period. The maximum length sequence, a type of cyclic code represents a commonly used periodic PN sequence.

The maximum length sequences or PN sequences can be generated easily using shift register circuits with feedback from one or more stages. A PN sequence generator using a 3-stage shift register is shown in Figure 5.2.

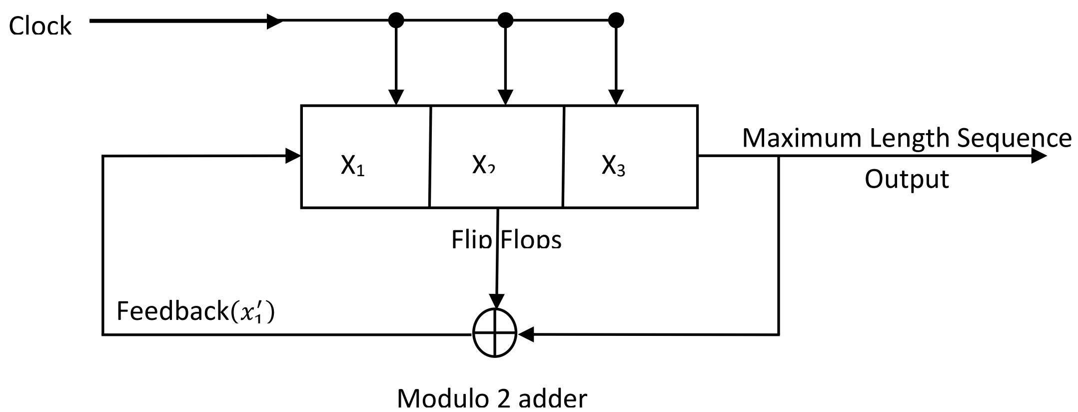


Figure 5.2 PN sequence or Maximum Length Sequence Generator

The 3-stage shift register consists of 3 flipflops regulated by a single timing clock. At each pulse of the clock, the state of each flipflop is shifted to the next one. The feedback function is obtained by using modulo-2 addition of the outputs of flipflops x_2 and x_3. The feedback term is applied to the input of the first flipflop x_1. The maximum length sequence output is obtained by noting the contents of flipflop x_3 at each clock pulse. The maximum-length sequence so generated is always periodic with a period of

$$N = 2^m - 1 \tag{5.1}$$

where m is the length of the shift register. Here, m = 3 and so $N = 2^3 - 1 = 7$.

For the PN sequence generator of Figure 5.2, if we assume that the shift register contents are initially 111, then with each clocking pulse, the contents will change as shown in the following table 5.1

Table 5.1 operation of the PN sequence generator

Shifts	$x_1' = X_2 \oplus X_3$	**Shift register contents**		
		X_1	X_2	X_3
0		1	1	1
1	$1 \oplus 1 = 0$	0	1	1
2	$1 \oplus 1 = 0$	0	0	1
3	$0 \oplus 1 = 1$	1	0	0
4	$0 \oplus 0 = 0$	0	1	0
5	$1 \oplus 0 = 1$	1	0	1
6	$0 \oplus 1 = 1$	1	1	0
7	$1 \oplus 0 = 1$	1	1	1

Hence for one period, the output PN sequence is 1 1 1 0 0 1 0, with a sequence length of 7. Thereafter, the sequence will be repeated.

5.5.3 Important Observations

- The length of the PN sequence is $N = 2^m - 1$, where m is the number of shift register stages.
- The PN sequence repeats itself after every 'N' clock cycles.
- The PN sequence is an NRZ type pulse signal with logic I represented by + 1 and logic 0 represented by -1, as shown in Figure 5.3.

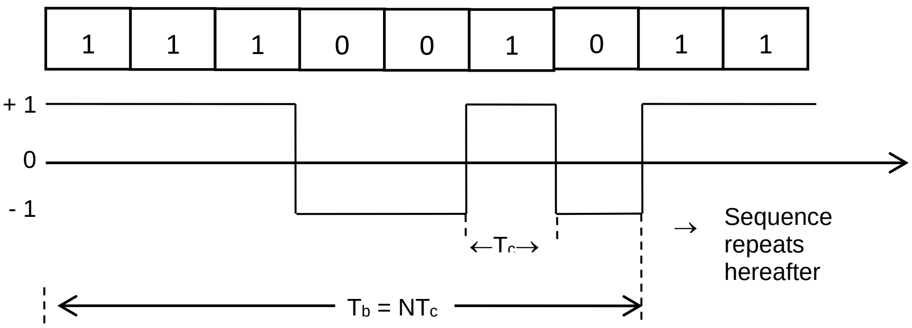


Figure 5.3 PN Sequence waveform

- The duration of every bit is known as the chip duration T_c. The chip rate R_c is defined as the number of bits (chips) per second.

$$T_c = \frac{1}{R_c} \quad \text{(or)} \quad R_c = \frac{1}{T_c} \qquad (5.2)$$

- The period of the PN sequence is $T_b = NT_c$
- The autocorrelation function $R(\tau)$ is a periodic function of time and it is a two valued function.

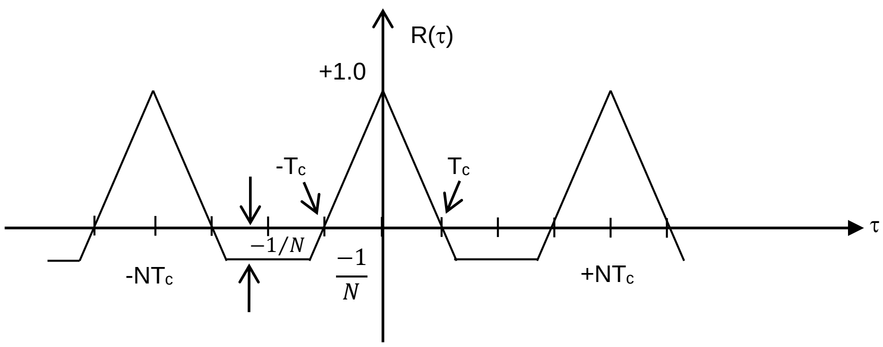


Figure 5.4 Autocorrelation function of a PN sequence

Example 5.1

A four stage shift register with feedback connections taken from the outputs of stages 4 and 1 through a modulo – 2 adder, is used for PN sequence generation. Assuming the initial contents of the shift register to be 0100, determine the output sequence. What is the length of the sequence?

Solution: The PN sequence generator is shown in Figure 5.5

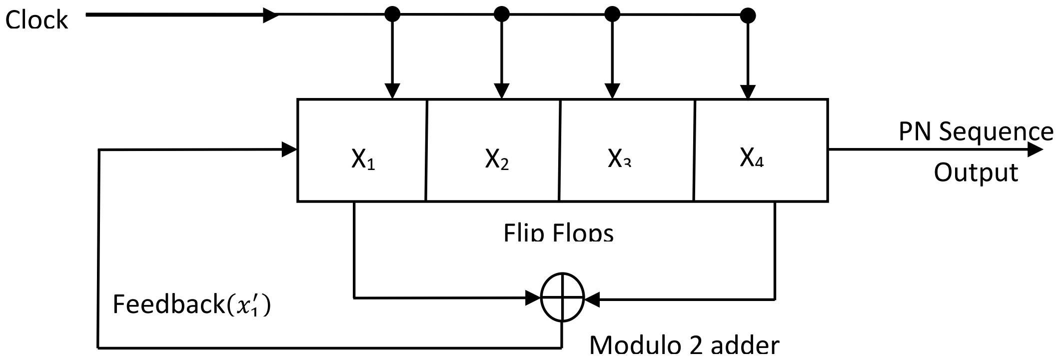


Figure 5.5 PN Sequence Generator

If the initial contents of the shift register are 0100, then with each clocking pulse, the contents will change as shown in the following table 5.2.

Table 5.2 Operation of PN sequence generator

Shifts	Feedback $x_1' = X_4 \oplus X_1$	Shift register contents			
		X_1	X_2	X_3	X_4
0		0	1	0	0
1	0 ⊕ 0 = 0	0	0	1	0
2	0 ⊕ 0 = 0	0	0	0	1
3	1 ⊕ 0 = 1	1	0	0	0
4	0 ⊕ 1 = 1	1	1	0	0
5	0 ⊕ 1 = 1	1	1	1	0
6	0 ⊕ 1 = 1	1	1	1	1
7	1 ⊕ 1 = 0	0	1	1	1
8	1 ⊕ 0 = 1	1	0	1	1
9	1 ⊕ 1 = 0	0	1	0	1
10	1 ⊕ 0 = 1	1	0	1	0
11	0 ⊕ 1 = 1	1	1	0	1
12	1 ⊕ 1 = 0	0	1	1	0
13	0 ⊕ 0 = 0	0	0	1	1
14	1 ⊕ 0 = 1	1	0	0	1
15	1 ⊕ 1 = 0	0	1	0	0

The output PN sequence is

0 0 1 0 0 0 1 1 1 1 0 1 0 1 1

After 15 shiftings, the initial contents of the shift registers are once again obtained. For further shiftings, the same cycle of events will repeat. Thus, the length of one period of the PN sequence is, $N = 2^m - 1 = 2^4 - 1 = 15$. Hence the sequence is a maximal length sequence.

5.5.4 Testing of PN sequence for Randomness Properties

Let us consider example 5.1 for testing of PN sequence, for Randomness properties.

1) **Balance Property:**

   The output PN sequence is given by 0 0 1 0 0 0 1 1 1 1 0 1 0 1 1. There are seven 0s and eight 1s in the sequence. Hence balance property is satisfied.

2) **Run Property:**

   Consider the zero runs - there are four of them. One-half are of length 1, one-fourth are of length 2. The same is true for the one runs. Hence run property is satisfied.

3) As shown in Figure 5.4, the autocorrelation function $R(\tau)$ will be a periodic function of time and will be a two valued function. Hence the correlation property is also satisfied.

4) For an m-stage linear feedback shift register the sequence repetition period in clock pulses is

$$N = 2^m - 1$$

Thus it can be seen that the sequence generated by the shift register generator of Figure 5.5 is an example of maximum length sequence.

5.5.5 Demerits of spread spectrum system

The use of a spreading code in the transmitter produces a wideband transmitted signal that appears noise like to a receiver that has no knowledge of the spreading code. Naturally, this technique provides improved protection against interference. But there are also some demerits involved in this method. They are

- Increased transmission bandwidth
- System complexity
- Processing delay

Hence, spread spectrum systems are employed only for those applications where security of transmission is our primary concern.

5.6 CLASSIFICATION OF SPREAD SPECTRUM MODULATION TECHNIQUES

The SS modulation techniques are broadly classified into two categories namely, the averaging type systems and the avoidance type systems. The averaging systems reduce the interference by averaging it over a long period. The Direct Sequence Spread Spectrum (DS-SS) system is an averaging system.

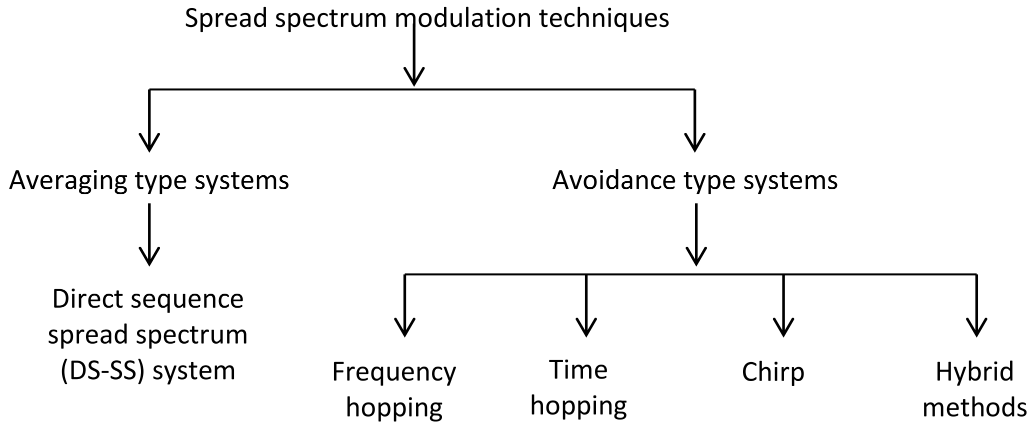


Figure 5.6 Classification of SS Modulation techniques

The avoidance type systems reduce the interference by making the signal avoid the interference over a large fraction of time. Some of the avoidance type systems are Frequency Hopping (FH) system, Time hopping (TH) system, Chirp and hybrid modulation system.

5.7 DIRECT SEQUENCE SPREAD SPECTRUM SYSTEMS

The most important advantage of spread spectrum modulation is that it provides protection against externally generated interfering signals such as jamming signals. The Direct Sequence Spread Spectrum (DS-SS) technique can be used in practice for such interference suppression. For this transmission of information signal is carried over a band pass channel (eg. Satellite channel). For such an application, the coherent Binary Phase Shift Keying (BPSK) is used in the communication system.

In the Direct sequence spread spectrum (DS-SS) systems, the use of a PN sequence to modulate a phase shift keyed signal achieves instantaneous spreading of the transmission bandwidth.

DS-BPSK Transmitter

The Figure 5.7 shows the transmitter section of the Direct Sequence Spread Spectrum with coherent BPSK.

The transmitter section uses two stages of modulation. In the first stage the input data sequence is first converted into an NRZ sequence b(t) by the NRZ encoder. This sequence b(t) is used to modulate a wide band pseudo-noise sequence c(t) by applying these two sequences to the product modulator or multiplier. Both sequences are in polar form. The product sequence m(t) = b(t) . c(t) will have a spectrum which will be same as that of c(t). The modulated signal m(t) is used to modulate the local carrier for BPSK modulation at the second stage. We can also use QPSK modulation.

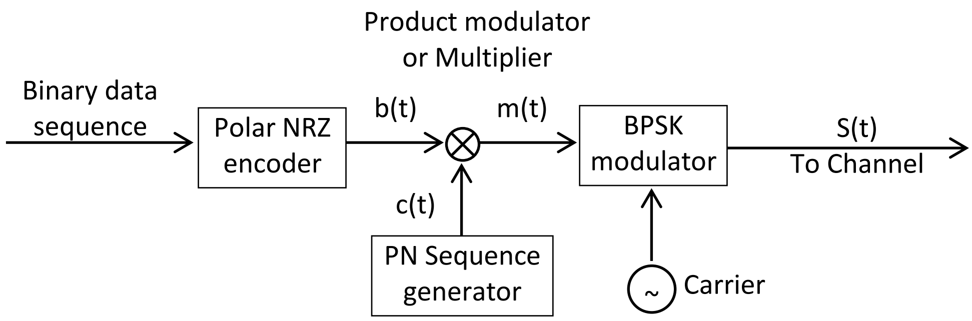


Figure 5.7 DS-BPSK Transmitter

The second stage modulated output s(t) is thus a Direct Sequence Spread binary phase shift keyed (DS | BPSK) signal. The phase modulation $\theta(t)$ of S(t) has one of the two values, 0 and π, depending on the polarities of the data sequence and PN sequence, as shown in the Table 5.3.

Table 5.3 Truth table for phase modulation θ(t), Radians

		Polarity of Data Sequence b(t) at time 't'	
		+	-
Polarity of PN Sequence C(t) at time 't'	+	0	π
	-	π	0

Waveforms

Figure 5.8 illustrates the wave forms for the first stage of modulation.

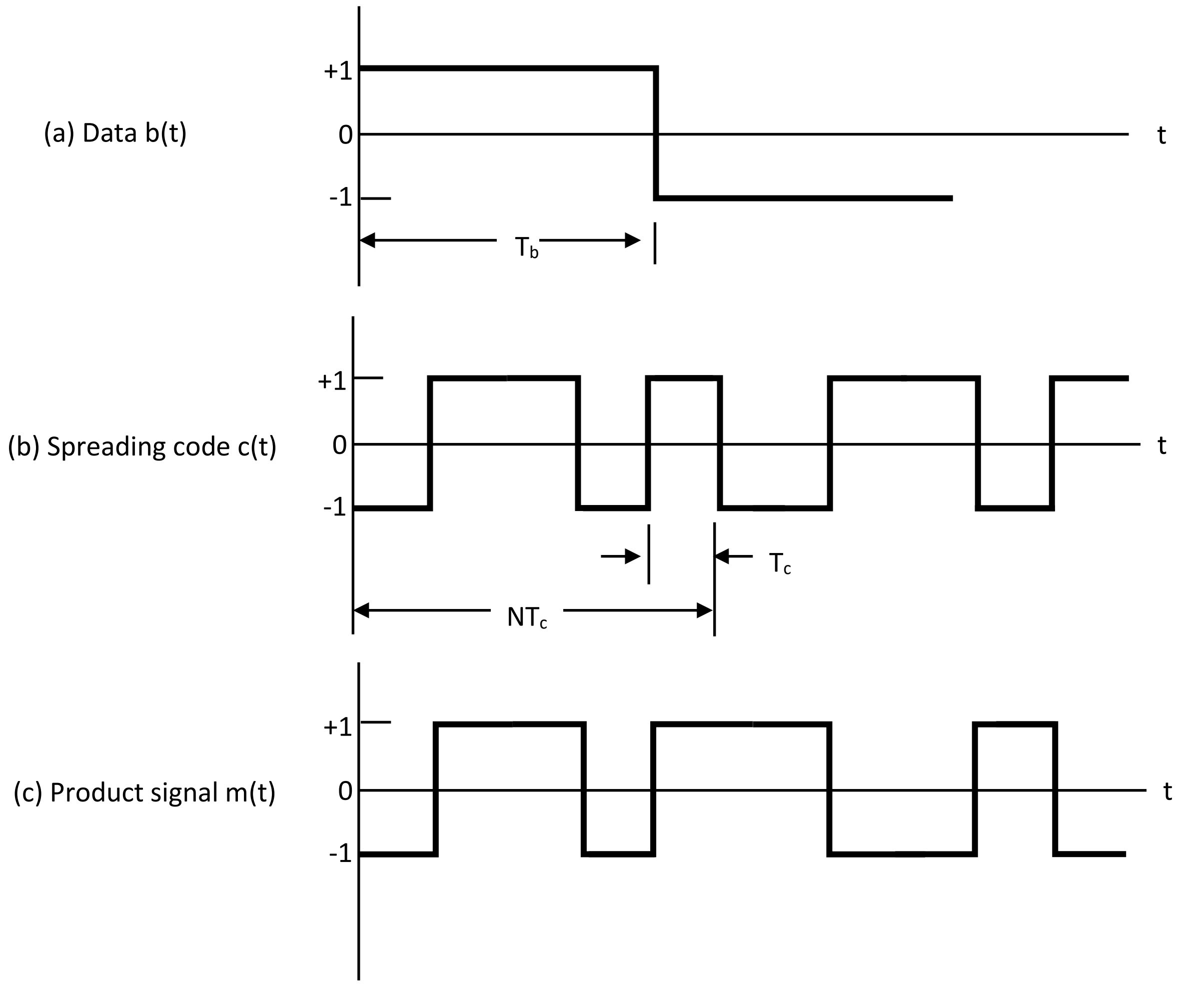
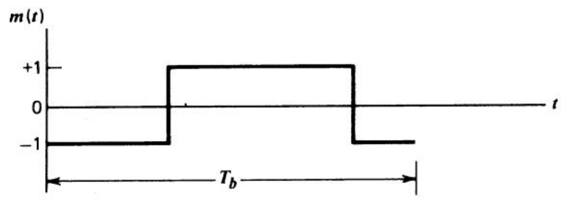
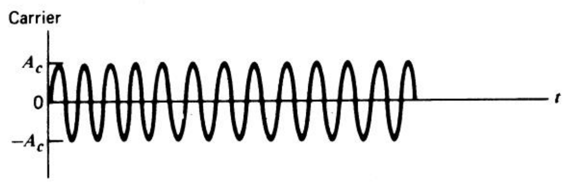


Figure 5.8 Waveforms for First stage of modulation

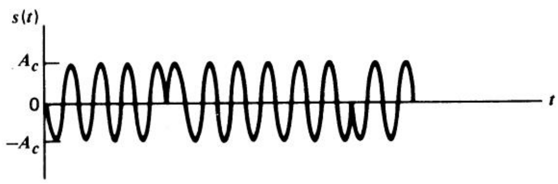
Figure 5.9 illustrates the waveforms for the second stage of modulation for one period of the PN sequence.



(a) Product signal m(t) = c(t) b(t)



(b) Sinusoidal carrier



$$A_c = \sqrt{\frac{2E_b}{T_b}}$$

(c) DS/BPSK signal

Figure 5.9 Waveforms for Second Stage of Modulation

DS-BPSK Receiver

The figure 5.10 shows the Receiver section of DS-BPSK system

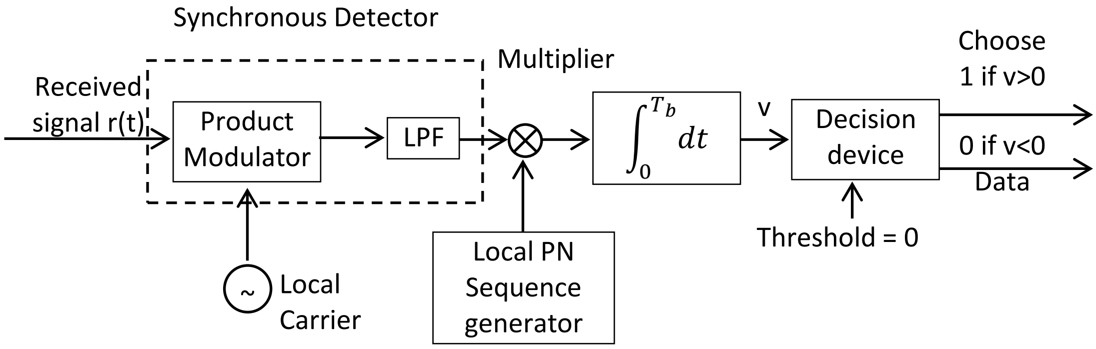


Figure 5.10 DS-BPSK Receiver

The receiver section consists of two stages of demodulation. In the first stage the received signal r(t) is subjected to coherent detection using the locally generated carrier signal. This carrier signal is arranged to be in phase and frequency synchronism with the carrier used at the transmitter.

In the second stage, the output of the coherent detector is subjected to de-spreading. It is multiplied with a locally generated PN sequence, which is in synchronism with the one at the transmitter. After despreading, it is integrated over a bit duration to get the observed random signal v. This is used for decision making, which provides an estimate of the original data sequence.

Important Observation

- In practice, the transmitter and receiver of Figures 5.7 and 5.10 are followed. In the transmitter spectrum spreading is performed prior to phase modulation. Also phase demodulation is done first and then despreading is done second, in the receiver.
- In the model of DS spread spectrum BPSK system used for analysis, the order of these two operations are interchanged. In the transmitter, BPSK is done first and spectrum spreading is done subsequently. Similarly, at the receiver also, spectrum despreading is done first and then phase demodulation is done second.
- This is possible, because the spectrum spreading and BPSK are both linear operations.

Advantages of DS-SS System

1. This system combats the intentional interference (jamming) most effectively.
2. This system has a very high degree of discrimination against the multipath signals. Therefore, the interference caused by the multipath reception is minimized successfully.
3. The performance of DS-SS system in the presence of noise is superior to other systems.

Disadvantages of DS-SS system

1. The PN code generator output must have a high rate. The length of such a sequence needs to be long enough to make the sequence truly random.
2. With the serial search system, the acquisition time is too large. This makes the DS-SS system be slow.
3. Synchronization is affected by the variable distance between the transmitter and receiver.
4. The DS-SS signal is not very effective against broadband interference.

Major applications of DS-SS system

1. Providing immunity against a jamming signal – Anti-jamming application.
2. Low detectability signal transmission – the signal is purposely transmitted at a very low power level. Hence the signal has a Low Probability of being intercepted (LPI) and it is called an LPI signal.
3. Accommodating a number of simultaneous signal transmissions on the same channel, ie. Code Division Multiple Access (CDMA) or spread spectrum multiple access (SSMA).

5.8 PERFORMANCE PARAMETERS OF DS-SS SYSTEM

The important performance parameters of a direct sequence spread spectrum system are 1) Processing gain, 2) Probability of Error and 3) Jamming Margin.

1) Processing Gain

The processing gain of a DS-SS system represents the gain achieved by processing a spread spectrum signal over an unspread signal. It may also be defined as the ratio of the bandwidth of the spread spectrum signal to the bandwidth of the unspread signal.

Therefore, Processing Gain (PG) = $\frac{Bandwidth\ of\ spread\ signal}{Bandwidth\ of\ unspread\ signal}$

- With reference to Figure 5.8, the bit rate of the binary data entering the transmitter input refers to the bandwidth of unspread signal. It is given by

$$R_b = \frac{1}{T_b} \tag{5.3}$$

- Also, the chip rate of the PN sequence refers to the bandwidth of spread spectrum signal. It is given by

$$R_c = \frac{1}{T_c} \tag{5.4}$$

- Therefore, Processing gain is given by

$$PG = \frac{R_c}{R_b} = \frac{1/T_c}{1/T_b} = \frac{T_b}{T_c}$$

$$\Rightarrow \quad PG = \frac{T_b}{T_c} \tag{5.5}$$

- Also with reference to Figure 5.8, we note that $T_b = NT_c$. This can be rewritten as

$$N = \frac{T_b}{T_c} \tag{5.6}$$

where N is the number of chips per information bit, and also called as the spread factor.

- On comparing equations (5.5) and (5.6), we infer that both PG and N are equal. Hence

$$PG = N = \frac{T_b}{T_c} \tag{5.7}$$

- The Processing Gain (PG) is also called as the bandwidth expansion factor (B_e) since it represents the advantage gained over the jammer that is obtained by expanding the bandwidth of the transmitted signal.

2) Probability of Error

- The probability of error P_e for a coherent BPSK system is given by

$$P_e = \frac{1}{2}\, erfc \sqrt{\frac{E_b}{N_o}} \tag{5.8}$$

where E_b is the energy per bit and $\frac{N_o}{2}$ is the power spectral density of white noise.

- In a DS-SS BPSK system, the interference may be treated as a wideband noise signal with a power spectral density of $\frac{N_o}{2}$. For the spread signal, we may write N_o as

$$N_o = JT_c \tag{5.9}$$

where J refers to the average interference power and T_c refers to chip duration or interval.

- On substituting the value of N_o in equation (5.8), we can write the probability of error for the DS-SS-BPSK system as

$$P_e = \frac{1}{2}\, erfc \sqrt{\frac{E_b}{JT_c}} \tag{5.10}$$

3) Jamming Margin (Antijam characteristics)

- We express the bit energy to noise density ratio as E_b/N_o. For the DS-SS-BPSK system, we may write N_o as equal to JT_C ($N_o = JT_c$). The bit energy E_b is given by

$$E_b = P_sT_b \tag{5.11}$$

where P_s is the average signal power and T_b is the bit duration or interval.

- Hence E_b/N_o can be written as

$$\frac{E_b}{N_o} = \frac{P_sT_b}{JT_c} = \left(\frac{P_s}{J}\right)\left(\frac{T_b}{T_c}\right) \tag{5.12}$$

$$\Rightarrow \quad \frac{J}{P_s} = \frac{T_b/T_c}{E_b/N_o} = \left(\frac{PG}{E_b/N_o}\right) \tag{5.13}$$

Since we know that PG=$\frac{T_b}{T_c}$

- This ratio $\frac{J}{P_s}$ is called as the jamming margin. Therefore, the jamming margin may be defined as the ratio of average interference power J and the average signal power P_s.
- If the jamming margin and the processing gain are both expressed in decibels, equation(5.13) can be written as

$$(\text{Jamming margin})_{dB} = (\text{Processing gain})_{dB} - 10\log_{10}\left(\frac{E_b}{N_o}\right)_{min} \tag{5.14}$$

where $\left(\frac{E_b}{N_o}\right)_{min}$ is the minimum bit energy-to-noise density ratio needed to support a prescribed average probability of error.

Example 5.2 A spread spectrum communication system is characterised by the following parameters

Information bit duration, T_b = 4.095 ms

PN chip duration, T_c = 1μs

Determine the processing gain and jamming margin if $\frac{E_b}{N_o}$=10 and the average probability of error, P_e=0.5 x 10^{-5}.

Solution:

(i) the processing gain, PG = $\frac{T_b}{T_c} = \frac{4.095ms}{1\mu s}$

$\Rightarrow$PG = $\frac{4.095 \ x \ 10^{-3}}{1 \ x \ 10^{-6}}$ **=** 4.095 x 10^3 = 4095

Hence, PG = 4095. Since PG = Spread factor, N, we have PG = N = 4095.

(ii) The jamming margin is

$$(\text{Jamming margin})_{dB} = (\text{Processing gain})_{dB} - 10\log_{10}\left(\frac{E_b}{N_o}\right)_{min}$$

$$= 10\log_{10}(4095) - 10\log_{10}(10)$$

$$= 36.1225 - 10 = 26.1225$$

$$\Rightarrow (\text{Jamming margin})_{dB} = 26.1225\text{dB}$$

Important observation:

From the above example, we infer that the information bits at the receiver output can be detected reliably, even when the noise or interference at the receiver input is up to 409.5 times the received signal power. Clearly, this is a powerful advantage against interference (jamming), which is obtained by the use of spread spectrum.

5.9 FREQUENCY HOPPING SPREAD SPECTRUM SYSTEMS (FH-SS)

In the Direct sequence spread spectrum systems (DS-SS), the use of a PN sequence to modulate a phase shift keyed signal achieves instantaneous spreading of the transmission bandwidth. The frequency hopping spread spectrum (FH-SS) system is an alternative method. In FH-SS, the spectrum of the transmitted signal is spread sequentially by randomly hopping the data modulated carrier from one frequency to the next.

Hence, the type of spread spectrum in which the carrier hops randomly from one frequency to another is called Frequency-hopped Spread Spectrum (FH-SS) system.

Basic Principle

In a FH-SS communication system the available channel bandwidth is subdivided into a large number of contiguous frequency slots. In any signalling interval, the transmitted signal occupies one or more of the available frequency slots. The selection of the frequency slot(s) in each signalling interval is made pseudorandomly according to the output from a PN generator. The figure 5.11 illustrates a particular FH pattern in the time-frequency plane.

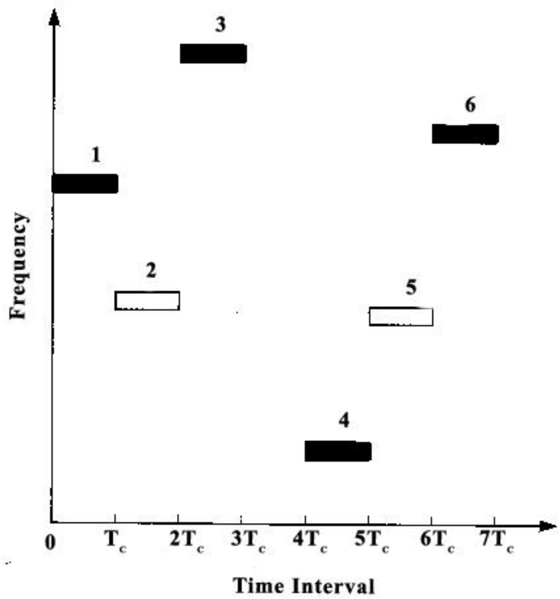


Figure 5.11 An example of a frequency – hopped (FH) pattern

Reason for employing M-ary FSK modulation

A common modulation format for FH systems is that of M-ary frequency shift keying (MFSK). The combination is referred to simply as FH/MFSK. Although PSK modulation gives better performance than FSK in AWGN channel, it is difficult to maintain phase coherence in

(i) the synthesis of frequencies used in the hopping pattern.

(ii) the propagation of the signal over the channel as the signal is hopped from one frequency to another over a wide bandwidth.

Therefore, FSK modulation with non-coherent detection is usually employed with FH spread spectrum signals.

Types of Frequency hopping

Since frequency hopping does not cover the entire spread spectrum instantaneously, we consider the rate at which the hops occur. We identify two basic (technology-independent) characterizations of frequency-hopping. They are

1) Slow-frequency hopping
2) Fast-frequency hopping

5.9.1 Slow-frequency hopping:

In FH system, if the hopping is performed at the symbol rate, we have a slow-hopped signal. Hence in slow-frequency hopping, the symbol rate R_s of the MFSK signal is an integer multiple of the hop rate R_h ie., several symbols are transmitted on each frequency hop.

Transmitter:

Figure 5.12 shows the block diagram of a slow-frequency hopping FH-MFSK transmitter.

First, the incoming binary data are applied to an M-ary FSK modulator. The resulting M-ary FSK modulated signal is applied to a Mixer. The Mixer consists of a multiplier followed by a band pass filter (BPF).

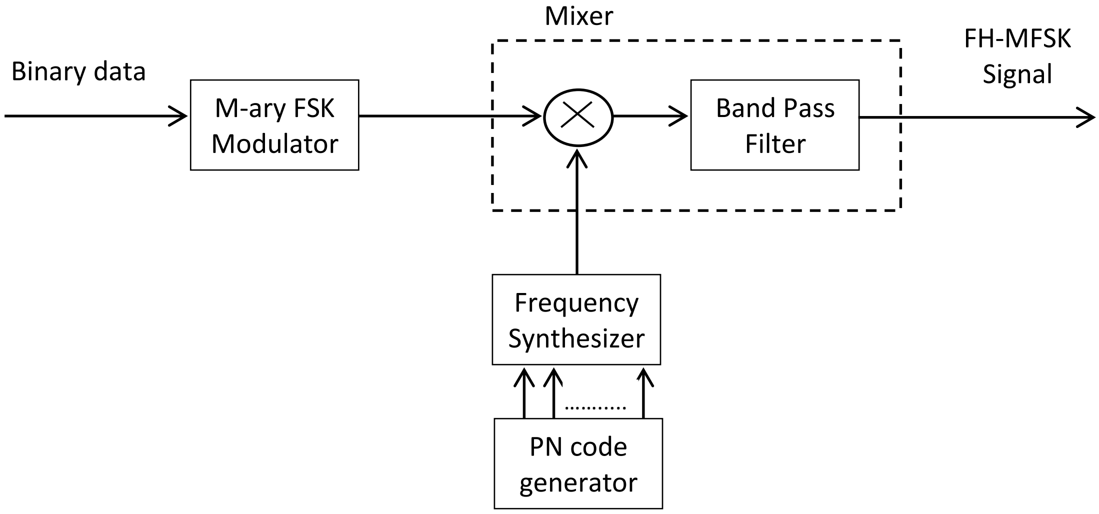


Figure 5.12 FH-MFSK Transmitter

The other input to the mixer block is obtained from a digital frequency synthesizer. The frequency synthesiser is controlled by a PN code generator. Hence the M-ary FSK modulated signal is again modulated by a carrier produced by the frequency synthesizer. The Mixer produces two outputs of the sum frequency and the difference frequency. The band pass filter that follows the mixer selects only the sum frequency signal, which is the FH-MFSK signal. This signal is then transmitted.

- Using the M-ary FSK system, M symbols can be transmitted, where $M=2^K$. Here k is the number of bits of the input binary data that form one symbol.
- The M-ary FSK modulator will assign a distinct frequency for each of these M symbols.
- The synthesizer output at a given instant of time is the frequency hop.
- The output bits of the PN generator change randomly. Hence the synthesizer output frequency will also change randomly.
- Each frequency hop is mixed with the MFSK signal to produce the transmitted signal.
- If the number of successive bits at the output of PN generator is n, then the total number of frequency hops will be 2^n.
- The total bandwidth of the transmitted FH-MFSK signal is equal to the sum of all the frequency hops. Therefore, the bandwidth of the transmitted FH-MFSK signal is very large of the order of few GHz.

Receiver:

Figure 5.13 shows the block diagram of a slow-frequency hopping FH-MFSK receiver.

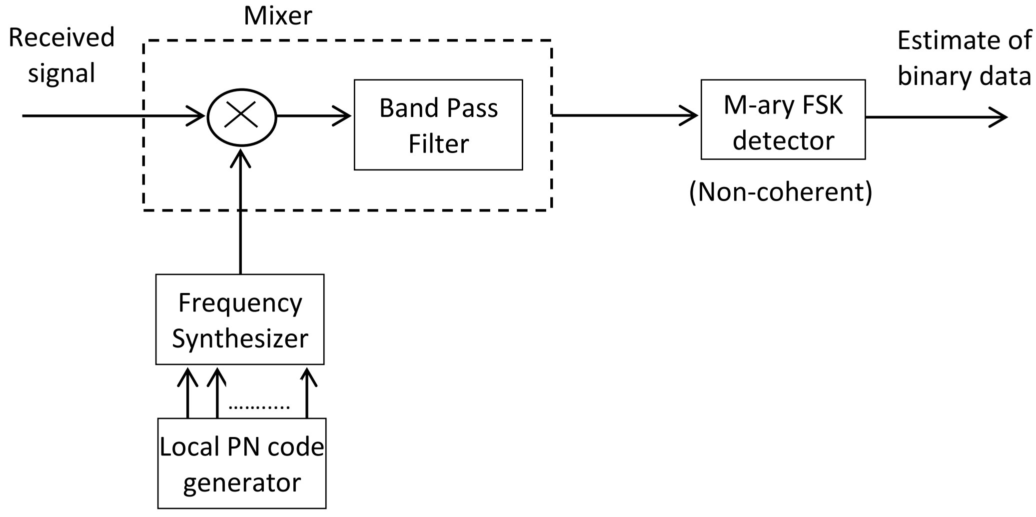


Figure 5.13 FH-MFSK Receiver

- The received signal is applied as input to the Mixer. The other input to the mixer is obtained from the digital frequency synthesizer.
- The frequency synthesizer is driven by a PN code generator. This generator is synchronized with the PN code generator at the transmitter.
- Therefore, the frequency hops produced at the synthesizer output will be identical to those at the transmitter.
- The mixer produces two outputs of the sum frequency and the difference frequency. The band pass filter selects only the difference frequency, which is the MFSK signal. Thus the mixer removes the frequency hopping.
- The MFSK signal is then applied to a non-coherent MFSK detector. A bank of M, non-coherent matched filters are used for non-coherent MFSK detection. Each matched filter is matched to one of the tones of the MFSK signal.
- An estimate of the original symbol transmitted is obtained by selecting the largest filter output.
- For an FH/MFSK system,

(i) The chip rate, $R_c = \max(R_h, R_s)$ (5.14)

where R_h is the hop rate and R_s is the symbol rate

(ii) A slow FH/MFSK system is characterized by having multiple symbols transmitted per hop. Hence, each symbol of a slow FH/MFSK system is a chip.

(iii) We can relate all rates as

$$R_c = R_s = \frac{R_b}{k} \geq R_h \qquad (5.15)$$

where $k = \log_2 M$

(iv) Processing gain, $PG = \frac{\text{Bandwidth of Spread signal}}{\text{Bandwidth of unspread signal}}$

Let f_s be the symbol frequency and 2^n be the number of frequency hops

Then, Processing gain, $PG = \frac{2^n f_s}{f_s} = 2^n$ (5.16)

(v) Probability of error, $P_e = \frac{1}{2} e^{-r_b \frac{R_c}{2}}$ (5.17)

5.9.2 Frequency hopping example:

Figure 5.14 illustrates the frequency hopping by an example.

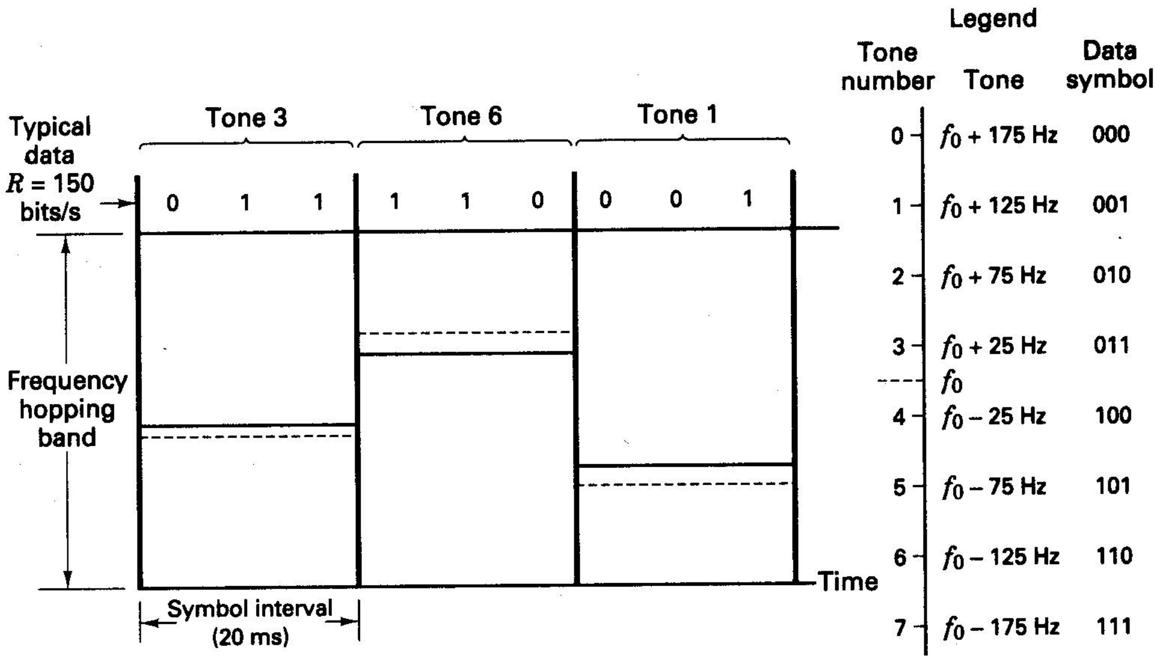


Figure 5.14 Frequency Hopping Example

- The input binary sequence data rate, R_b=150bits/s
- The modulation is 8-ary FSK.
- Then the symbol rate is $R_s = \frac{R_b}{k} = \frac{150}{\log_2 8}$ = 50 bits/s
- The symbol interval is $T_s = \frac{1}{R_s} = \frac{1}{50}$ = 20ms
- The frequency is hopped once per symbol. Hence the hopping rate is R_h=50hops/s.
- In the time-bandwidth plane of the figure, the abscissa (x-axis) represents time and the ordinate(y-axis) represents the hopping bandwidth.
- A set of 8-ary FSK symbol-to-tone assignments is given. f_0 refers to centre frequency of the data band, which is not fixed.
- The tone separation is $\Delta f = \frac{1}{T_s} = \frac{1}{20ms}$= 50Hz.
- A typical binary data sequence is given at the top. Since the modulation is 8-ary FSK, the bits are grouped three at a time to form symbols.
- A single-sideband tone (offset from f_0) would be transmitted according to symbol-to-tone assignment.
- For each new symbol, f_0 hops to a new position in the hop bandwidth. For the first symbol in the data sequence 011, f_0+25Hz assignment is done. In the

figure, f_o is shown with a dashed line and the symbol tone f_o+25Hz is shown with a solid line.

- Likewise, for the second symbol 110, f_o - 125Hz assignment is done. For the third symbol 001, f_o + 125Hz assignment is done. For each symbols, the centre frequency f_o hops to a new position.

5.9.3 Frequency hopping with diversity:

In communication the transmitted signal's ability to withstand impairments from the channel, such as noise, jamming, fading, and so on is termed as Robustness. A signal configured with multiple replicate copies, each transmitted on a different frequency, has a greater likelihood of survival than does a single such signal.

Diversity may be defined as multiple transmissions of the same signal at different frequencies which are spread in time. The greater the diversity, the more robust the signal against random interference.

To illustrate the beneficial effect of diversity, we can extend the frequency hopping example shown in Figure 5.14. We introduce frequency hopping diversity by a chip repeat factor of N=4. Figure 5.15 illustrates the effect of diversity.

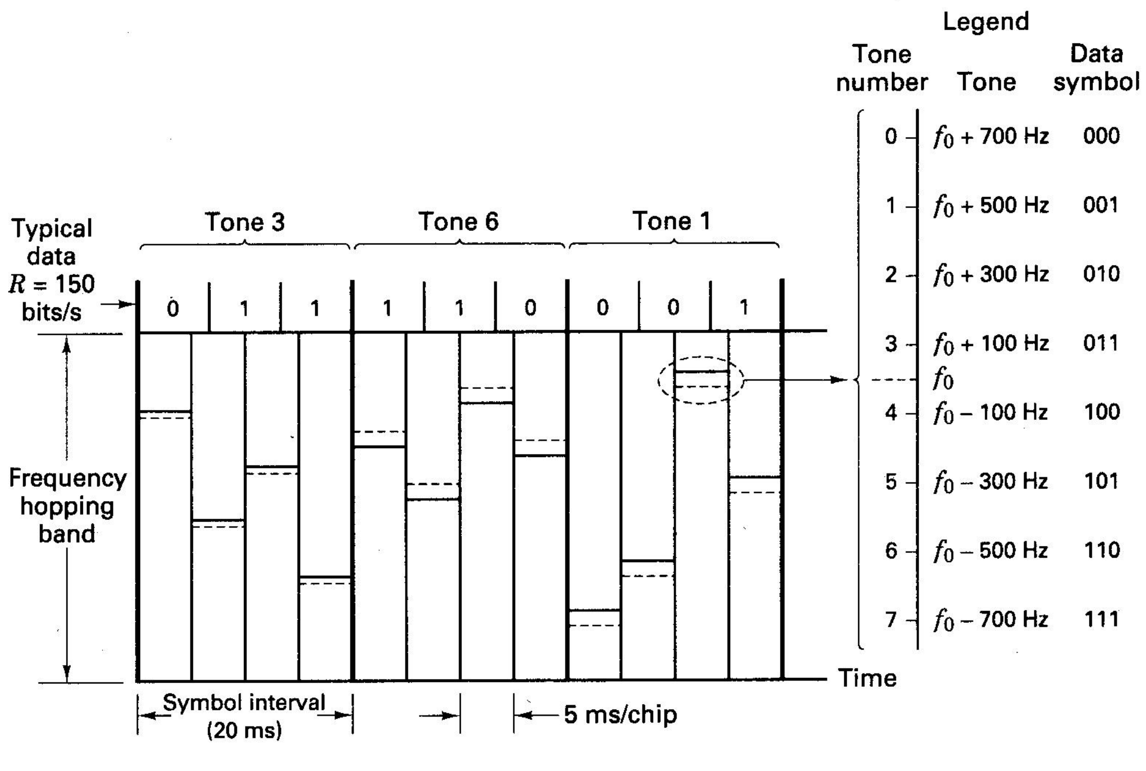


Figure 5.15 Frequency hopping with diversity (N = 4)

- During each 20ms symbol interval, there are now four columns, corresponding to the four separate chips to be transmitted for each symbol.
- Now, each symbol is transmitted four times. For each transmission, the centre frequency f_0 is hopped to a new region of the hopping band.
- The chip interval is $T_c = \frac{T_s}{N} = \frac{20ms}{4} = 5$ms.
- The hopping rate is $R_h = \frac{R_b}{\log_2 8} \cdot N = \frac{150 \, x \, 4}{3} = 200$ hops/s.
- Also the spacing between frequency tones must change to satisfy orthogonality. Hence the tone separation is $\Delta f = \frac{1}{T_s} . N = \frac{4}{20ms} = 200$Hz.
- Hence, the resulting transmissions yield a more robust signal than that without such diversity.

5.9.4 Fast-frequency hopping:

In FH system, if there are multiple hops per symbol, we have fast-hopped signal. Hence in fast-frequency hopping, the hop rate R_h is an integer multiple of the MFSK symbol rate R_s ie., the carrier frequency will change or hop several times during the transmission of one symbol. Hence, in a fast FH-MFSK system, each hop is a chip.

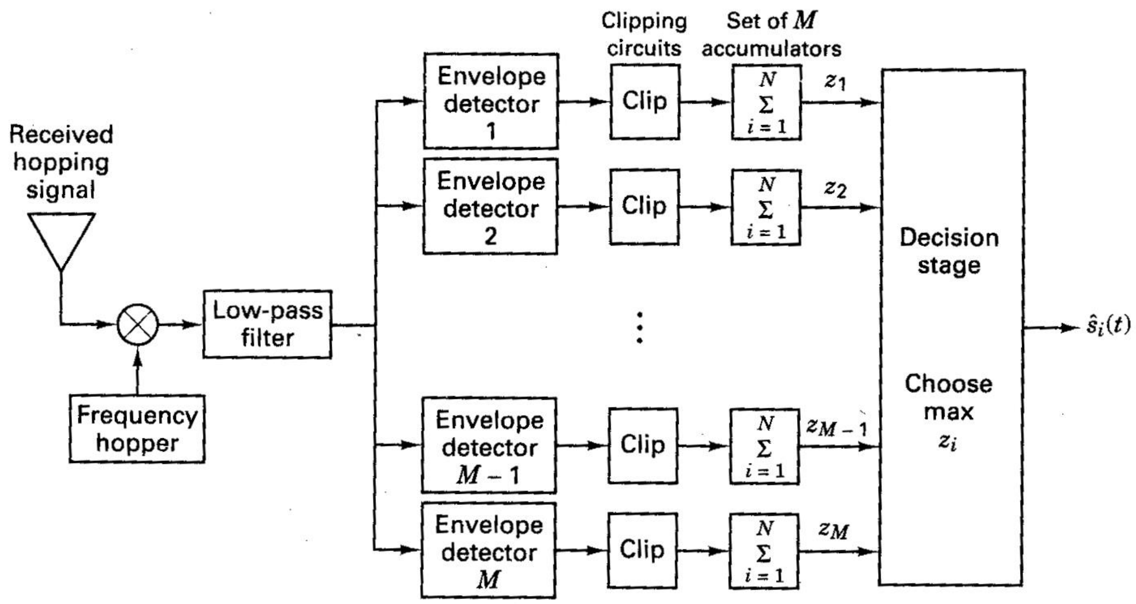


Figure 5.16 Fast FH-MFSK demodulator

In general, fast frequency hopping is used to defeat a smart jammer's tactic that involves two functions: measurement of the spectral content of the transmitted signal, and returning of the interfering signal to that portion of the frequency band.

To overcome the jammer, the transmitted signal must be hopped to a new carrier frequency before the jammer is able to complete the processing of these two functions.

For data recovery at the receiver, non-coherent detection is used. However, the detection procedure is different from that used in a slow FH-MFSK receiver. The Figure 5.16 shows a typical fast FH-MFSK demodulator.

First, the signal is dehopped using a PN generator identical to that used in transmitter. Then, filtering is done with a low pass filter having a bandwidth equal to the data bandwidth. The filtered signal is demodulated using a bank of 'M' envelope detectors.

Each envelope detector is followed by a clipping circuit and an accumulator. The clipping circuit serves an important function in the presence of an intentional jammer or other strong unpredictable interference. The demodulator does not make symbol decisions on a chip-by-chip basis. The energy from the N chips are accumulated. After the energy from the N^{th} chip is added to the N-1 earlier ones, the demodulator makes a symbol decision by choosing the symbol that corresponds to the accumulator with maximum energy.

Advantages of FH-SS system:

1. The processing gain PG is higher than that of DS-SS system.
2. Synchronization is not greatly dependent on the distance.
3. The serial search system with FH-SS needs shorter time for acquisition.

Disadvantages of FH-SS system:

1. The bandwidth of FH-SS system is too large (in GHz).
2. Complex and expensive digital frequency synthesizers are required.

Applications of FHSS system:

1) CDMA systems based on FH spread spectrum signals are particularly attractive for mobile communication.
2) Wireless local area networks (WLAN) standard for Wi-Fi.
3) Wireless Personal area network (WPAN) standard for Bluetooth.

5.9.5 Fast hopping Versus Slow hopping:

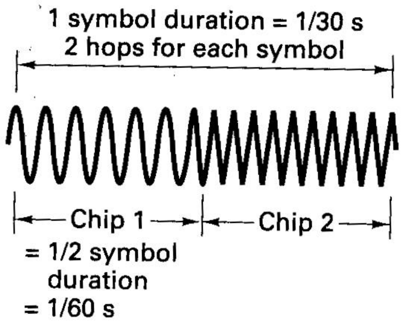
Table 5.4 compares the performance of fast hopping and slow hopping systems.

Table 5.4

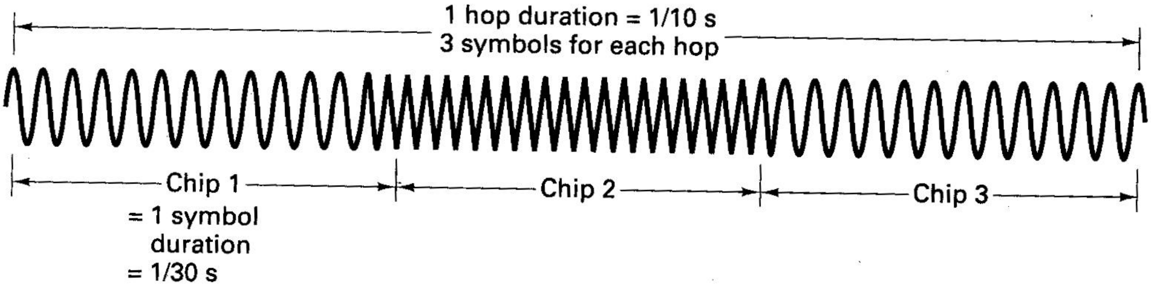
Sl No.	Slow frequency hopping	Fast frequency hopping
1.	More than one symbols are transmitted per frequency hop.	More than one frequency hops are required to transmit one symbol.
2.	Chip rate is equal to symbol rate.	Chip rate is higher than Symbol rate.
3.	Symbol rate is higher than hop rate.	Hop rate is higher than Symbol rate.
4.	Same carrier frequency is used to transmit one or more symbols.	One symbol is transmitted over multiple carriers in different hops.
5.	A jammer can detect this signal if the carrier frequency in one hop is known.	A jammer cannot detect this signal because one symbol is transmitted using more than one carrier frequencies.

Slow hopping and fast hopping performance may also compared by the following two examples:

1) Figure 5.17 shows chip in the context of an FH-MFSK system.



(a) Fast frequency hopping

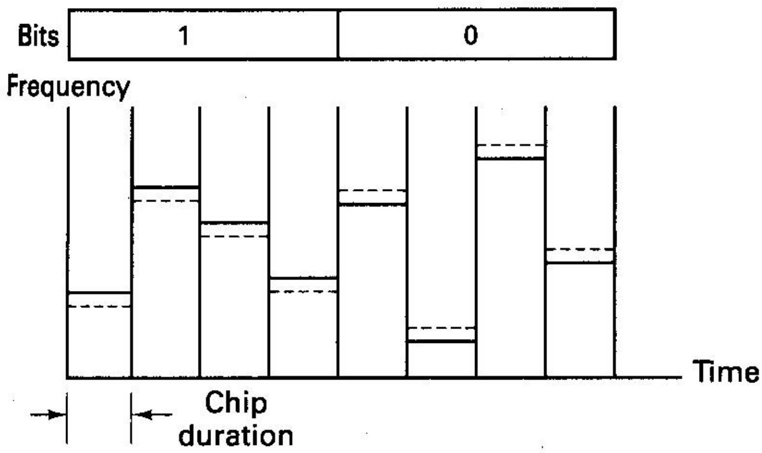


(b) Slow frequency hopping

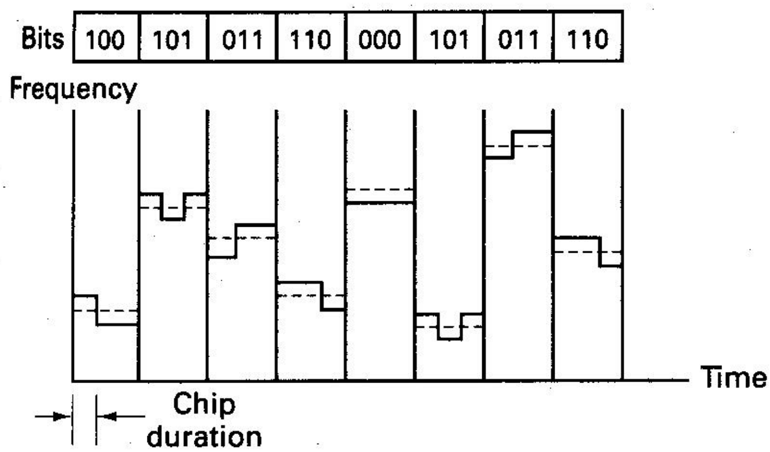
Figure 5.17 Chip in the context of an FM-MFSK System

- Figure 5.17(a) illustrates an example of fast frequency hopping. The data symbol rate is 30 symbol/s and the frequency hopping rate is 60 hops/s. The figure illustrates the waveform s(t) over one symbol duration ($\frac{1}{30}$s). The waveform change in (the middle of) s(t) is due to a new frequency hop.
- Figure 5.17(b) illustrates an example of slow frequency hopping. The data symbol rate is still 30symbols/s, but the frequency hopping rate has been reduced to 10hops/s. The waveform s(t) is shown over a duration of three symbols ($\frac{1}{10}$s).

2) Figure 5.18 shows the comparison for a binary FSK system.
   - Figure 5.18(a) illustrates an example of fast frequency hopping for a binary FSK system. The diversity is N=4. There are 4 chips transmitted per bit. Here, the chip duration is the hop duration.



(a) Fast Frequency Hopping



(b) Slow Frequency Hopping

Figure 5.18 Comparison for a binary system

- Figure 5.18(b) illustrates an example of slow frequency hopping for a binary FSK system. In this case, there are 3 bits transmitted during the time duration of a single hop. Here, the chip duration is the bit duration.

5.10 SYNCHRONIZATION

5.10.1 Need for Synchronization

The process in which the locally generated carrier at the receiver must be in frequency and phase synchronism with the carrier at the transmitter is called synchronization. In spread spectrum communication systems, there should be perfect alignment between the transmitted and received PN codes, for satisfactory operation.

Because

(i) Carrier frequency as well as the PN clock may drift with time.

(ii) If there is relative motion between the transmitter and receiver, as in the case of mobile and satellite spread spectrum systems, the carrier and PN clock will suffer Doppler frequency shift.

Hence, synchronization of the PN sequence of the receiver with that of the transmitter is essential.

5.10.2 Synchronization steps:

The process of synchronizing the locally generated spreading signal with the received spread spectrum signal is usually accomplished in two steps. They are

1) Acquisition: The first step, called acquisition, consists of bringing the two spreading signals into coarse alignment with one another.
2) Tracking: Once the received spread spectrum signal has been acquired, the second step, called tracking, takes over for fine alignment.

Both acquisition and tracking make use of the feedback loop.

5.10.3 Acquisition:

Acquisition schemes can be classified into three types. They are

1) Serial search acquisition
2) Parallel search acquisition
3) Sequential search acquisition

1. Serial search acquisition:

A) DS Spread spectrum systems:

Figure 5.19 shows the serial search scheme for Direct Sequence spread spectrum systems.

There is always an initial timing uncertainty between the receiver and the transmitter. Let us suppose that the transmitter has N chips and the chip duration is T_c. If initial synchronization is to take place in the presence of additive noise and other interference, it is necessary to dwell for $T_d=NT_c$ in order to test synchronism at each time instant. We search over the time uncertainty interval in(coarse) time steps of $\frac{1}{2}$ T_c.

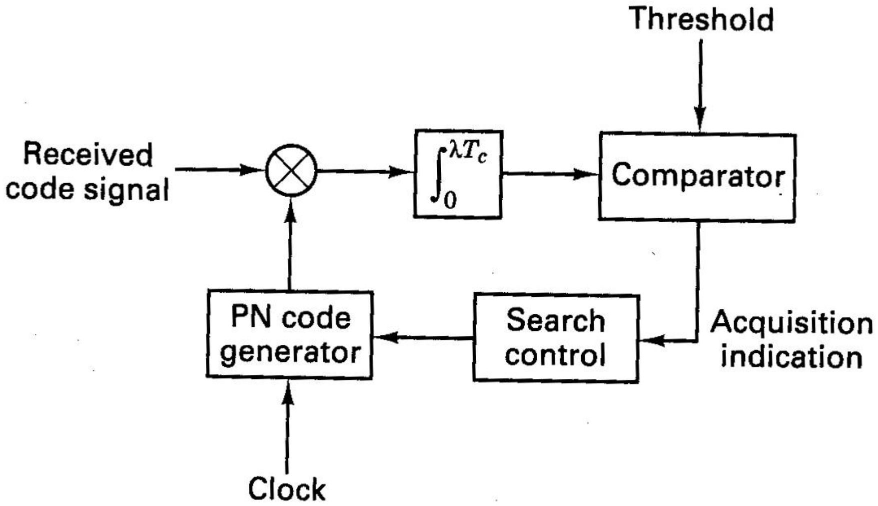


Figure 5.19 Direct Sequence spread spectrum systems – Serial Search Acquisition

The locally generated PN signal is correlated with the incoming PN signal. At fixed search intervals of NT_c (search dwell time), the output signal is compared to a preset threshold. If the output is below the threshold, the locally generated code signal is advanced in time by $\frac{1}{2}T_c$ seconds. The correlation process is repeated again. These operations are performed until a signal is detected or the threshold is exceeded. Then the PN code is assumed to have been acquired.

Thus, if initially the misalignment between the two codes was n chips, the total time taken for acquisition is given by

$$T_{acq} = 2nNT_c \text{ seconds} \tag{5.18}$$

B) FH spread spectrum systems

Figure 5.20 shows the serial search scheme for frequency hopping spread spectrum systems.

Here the non-coherent matched filter consists of a mixer followed by a bandpass filter (BPF) and a square law envelope detector. The PN code generator controls the frequency hopper. Acquisition is accomplished when the local hopping is aligned with that of the received signal.

Let f_i be the frequency of the frequency synthesizer at the transmitter. Suppose f_j be the frequency of the signal produced by the frequency synthesizer in the acquisition circuit of the receiver. If $f_i \neq f_j$, then only a small voltage less than the threshold will be produced at the output of BPF. At a later instant of time during searching, if $f_i = f_j$, then a large voltage exceeding the threshold will be produced at the output of BPF. This indicates the alignment of local hopping with that of the received signal.

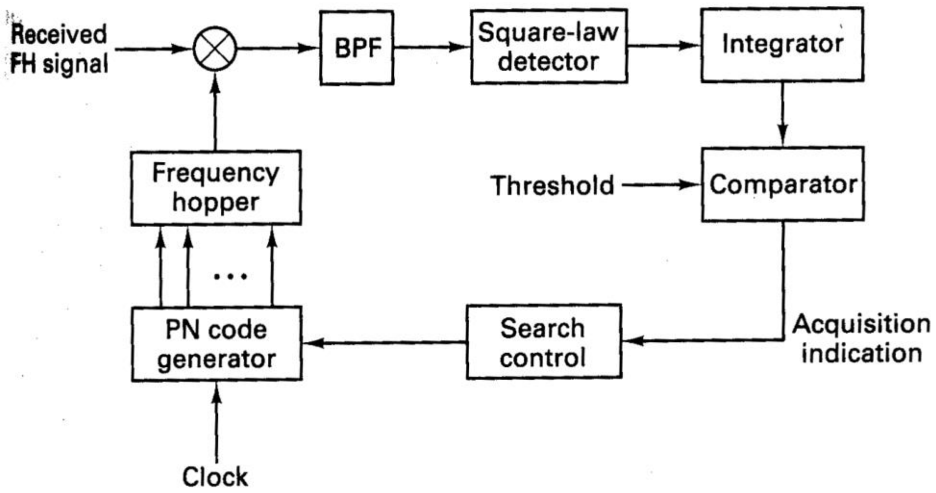


Figure 5.20 Frequency hopping serial search acquisition

2. Parallel search acquisition

The parallel search acquisition scheme introduces some degree of parallelism by having two or more correlators operating in parallel. They will search over non-overlapping time slots. In this scheme, the search time is reduced at the expense of a more complex and costly implementation.

3. Sequential search acquisition

In this scheme, the dwell time at each delay in the search process is made variable by employing a correlator with a variable integration period whose (biased) output is compared with two thresholds. Hence the sequential search method results in a more efficient search in the sense that the average search time is minimised.

5.10.4 Tracking

Once the signal is acquired, the initial search process is stopped and fine synchronization and tracking begins. The tracking maintains the PN Code generator at the receiver in synchronism with the incoming signal. Tracking includes both fine chip synchronization and, for coherent demodulation, carrier phase tracking.

A) DS Spread spectrum system:

The commonly used tracking loop for a Direct sequence spectrum signal is the Delay-locked loop (DLL) as shown in the Figure 5.21.

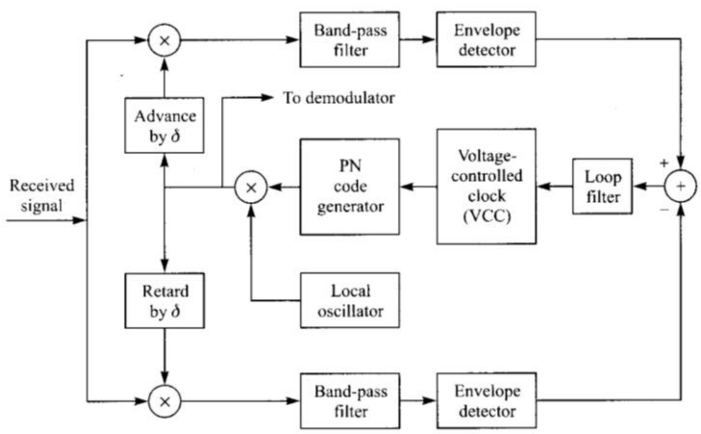


Figure 5.21 Delay-Locked Loop (DLL) for PN code tracking

The received DS spread spectrum signal is applied simultaneously to two multipliers. One of the multipliers is fed with PN code delayed by δ, a fraction of the chip interval. The other multiplier is fed with the same PN code advanced by δ. The output from each multiplier is fed to a BPF centred on f_0.

The output of each BPF is envelope detected and then subtracted. This difference signal is applied to the loop filter that drives the voltage controlled oscillator. The VCO serves as the clock for the PN Code generator. If the synchronization is not exact, the filtered output from one correlator will exceed the other. Hence the VCO will be appropriately advanced or delayed. At the equilibrium point, the two filtered correlator outputs will be equally displaced from the peak value. Then the PN code generator output will be exactly synchronized to the received signal that is fed to the demodulator.

B) FH Spread spectrum system:

A typical tracking technique for FH spread spectrum signals is illustrated in Figure 5.22.

Although initial acquisition has been achieved, there is a small timing error between the received signal and the receiver clock. The BPF is tuned to a single intermediate frequency and its bandwidth is of the order of $\frac{1}{T_c}$, where T_c is the chip interval. Its output is envelope detected and then multiplied by the clock signal to produce a three-level signal. This drives the loop filter.

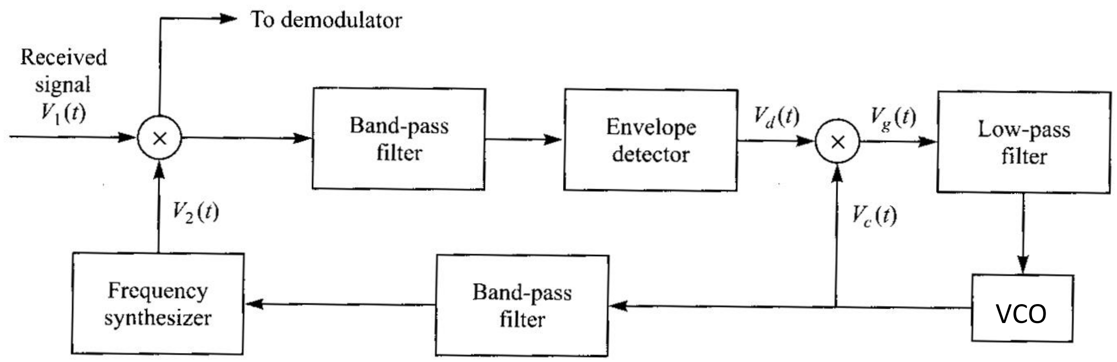


Figure 5.22 Tracking loop for FH signals

Suppose that the chip transitions from the locally generated sinusoidal waveform do not occur at the same time as the transitions in the incoming signal. Then the output of the loop filter will be either positive or negative, depending on whether the VCO is lagging or advanced relative to the timing of the input signal. This error signal from the loop filter will provide the control signal for adjusting the VCO timing signal so as to drive the frequency synthesizer output to proper synchronism with the received signal.

5.11 PERFORMANCE COMPARISON OF DS-SS AND FH-SS

S.No.	Parameter	Direct sequence spread spectrum	Frequency hopping spread spectrum
1	Definition	The PN sequence makes the transmitted signal assume a noise like appearance by spreading its spectrum over a broad range of frequencies simultaneously.	The PN sequence makes the carrier hop over a number of frequencies in a pseudo-random manner, with the result that the spectrum of the transmitted signal is spread in a sequential manner.
2	Chip rate	$R_c = \frac{1}{T_c}$	$R_c = \max(R_h, R_s)$
3	Modulation technique	BPSK	M-ary FSK
4	Processing gain(PG)	$PG = \frac{T_b}{T_c} = N$	$PG = 2^n$
5	Error probability	$P_e = \frac{1}{2} erfc\sqrt{\frac{E}{JT_C}}$	$P_e = \frac{1}{2} e^{-r_b} R_c/2$
6	Acquisition time	Long time	Short time
7	Effect of distance	This system is distance relative.	Effect of distance is less.

5.12 JAMMING CONSIDERATIONS

5.12.1 Jamming:

Jamming refers to the intentional interference in a communication system. The goals of the communicator are to develop a jam-resistant communication system under the following assumptions:

(i) Complete invulnerability is not possible.

(ii) The jammer has apriori knowledge of most system parameters, such as frequency bands, timing, traffic and so on.

(iii) The jammer has no apriori knowledge of the PN spreading or hopping codes.

The signalling waveform should be designed, so that the jammer cannot gain any appreciable jamming advantage by choosing a jammer waveform and strategy other than wideband Gaussian noise.

5.12.1.1 Jammer waveforms

- There are many different waveforms that can be used for jamming communication systems. The most appropriate choice depends on the targeted system.
- We shall assume that the jammer waveform is wideband noise and the jammer strategy is to jam the entire bandwidth.

5.12.1.2 Tools of the communicator:

The usual design options for an antijam (AJ) communication system are

1) Frequency diversity by the use of direct sequence and frequency hopping spread spectrum techniques.
2) Time diversity, by the use of time hopping.
3) Spatial discrimination, by the use of a narrow beam antenna.

5.12.1.3 J/S Ratio:

- The ratio $(J/S)_{reqd}$ is a figure of merit that provides a measure of how invulnerable a system is to interference. It is given by

$$\left(\frac{J}{S}\right)_{reqd} = \frac{PG}{\left(E_b/J_o\right)_{reqd}} \qquad (5.19)$$

where J → average received jammer power

S → received signal power

PG → processing gain

$\left(E_b/J_o\right)_{reqd}$ → bit energy per jammer noise power spectral density required for maintaining the link at a specified error probability.

5.12.1.4 Anti-jam margin:

- Anti-jam (AJ) margin usually means the safety margin against a particular threat. It is defined as

$$M_{AJ}\,(dB) = \left(\frac{E_b}{J_o}\right)_r (dB) - \left(\frac{E_b}{J_o}\right)_{reqd} (dB) \qquad (5.20)$$

where, $\left(\frac{E_b}{J_o}\right)_r$ → $\left(\frac{E_b}{J_o}\right)$ actually received

$\left(\frac{E_b}{J_o}\right)_{reqd}$ → $\left(\frac{E_b}{J_o}\right)$ actually required

5.12.2 Broadband noise jamming

- The jamming signal may be modelled as a zero-mean wide-sense stationary Gaussian noise process.
- If the jammer strategy is to jam the entire spread spectrum bandwidth, with its fixed power, then the jammer is referred to as a wide band or broad band jammer.

5.12.3 Partial-band noise jamming:

- In the case of partial-band jamming, a specific transmitted symbol will be received unjammed, with probability (1-P). Also, it will be perturbed by jammer power with probability P.
- Forward error Correction (FEC) coding with appropriate interleaving can mitigate this degradation.

5.12.4 Multiple-tone jamming:

- In the case of multiple-tone jamming, the jammer divides its total received power, J into distinct, equal power, random phase CW tones.
- These are distributed over the spread spectrum bandwidth according to some strategy.

5.12.5 Pulse jamming

- A pulse-noise jammer transmits pulses of band limited white Gaussian noise having a time-averaged received power J.
- Forward error correction (FEC) coding with appropriate interleaving can almost fully restore this degraded performance.

5.12.6 Repeat-back jamming

- The repeat-back jammers or frequency-follower (FF) jammers monitor a communicator's signal.
- They can increase the jamming power in the communicator's instantaneous bandwidth.
- To defeat the repeat back jammer, one method is to simply hop so fast that by the time the jammer receives, detects and transmits the jamming signal, the communicator is already transmitting at a new hop.

- Another technique capable of defeating the repeat-back jammer is a system named as Buffalo Laboratories Application of Digitally Exact Spectra or BLADES.

5.13 COMMERCIAL APPLICATIONS OF SPREAD SPECTRUM TECHNIQUES

Spread spectrum signals are used for

1) Combating or suppressing the detrimental effects of interference due to jamming (Intentional interference). It can be used in military applications also.
2) Accommodating multiple users to transmit messages simultaneously over the same channel bandwidth. This type of digital communication in which each user (transmitter-receiver pair) has a distinct PN code for transmitting over a common channel bandwidth is called as Code Division Multiple Access (CDMA) or Spread Spectrum Multiple Access (SSMA). This technique is popularly used in digital cellular communications.
3) Reducing the unintentional interference arising from other users of the channel.
4) Suppressing self-interference due to multipath propagation.
5) Hiding a signal by transmitting it at low power and, thus, making it difficult for an unintended listener to detect in the presence of background noise. It is also called a Low Probability of Intercept (LPI) signal.
6) Achieving message privacy in the presence of other listeners.
7) Obtaining accurate range (time delay) and range rate (velocity) measurements in radar and navigation.

5.14 CDMA - DIGITAL CELLULAR SYSTEM

The most important application of spread spectrum technique is the Digital cellular CDMA system. Here, we shall explain in detail about this CDMA digital cellular system based on Direct Sequence (DS) spread spectrum.

This digital cellular communication system was proposed and developed by Qualcomm corporation. It has been standardized and designated as Interim Standard 95 (IS-95) by the Telecommunications Industry Association (TIA) for use in the 800 MHz and in the 1900 MHz frequency bands.

The nominal bandwidth used for transmission from a base station to the mobile receivers (Forward link or channel) is 1.25 MHz. A separate channel, also with a bandwidth of 1.25 MHz is used for signal transmission from mobile receivers to a base station (reverse link or channel). The signals transmitted in both the forward and reverse links are DS Spread spectrum signals having a chip rate of 1.288 x 10^6 chips per second (Mchips/s).

5.14.1 Forward link or channel

The signal transmission from a base station to the mobile receivers is referred as the Forward link or channel. The figure 5.23 shows the block diagram of IS-95 forward link.

Source coding

The speech (source) coder is a code-excited linear predictive (CELP) coder. It generates data at the variable rates of 9600, 4800, 2400 and 1200bits/s. The data rate is a function of the speech activity of the user, in frame intervals of 20ms.

Channel coding

The data from the speech coder is encoded by a rate $^1/_2$, constraint length K = 9 convolutional code. For lower speech activity, the output symbols from the convolutional encoder are repeated. If the data rate is 4800 bits/s, then the output symbols are repeated twice, so as to maintain a constant bit rate of 9600 bits/s.

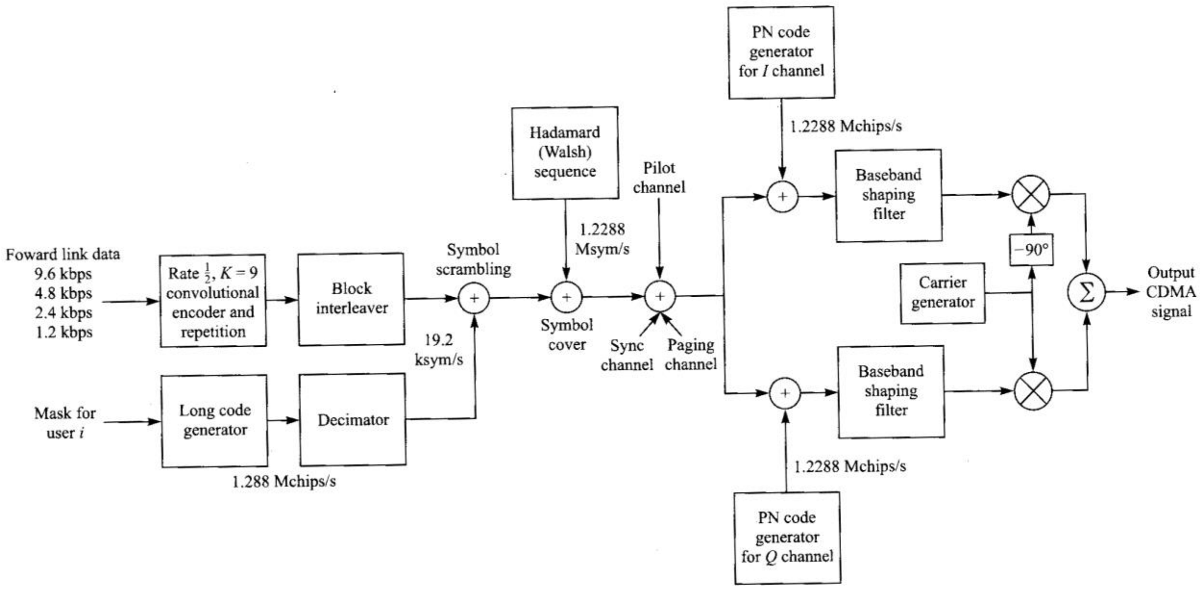


Figure 5.23 IS-95 Forward Link

Block interleaver:

The encoded bits for each frame are passed through a block interleaver. It is needed to overcome the effects of burst errors that may occur in transmission through the channel. The data bits at the output of the block interleaver occur at a rate of 19.2kbits/s.

Symbol scrambler

The data bits from the block interleaver are scrambled by multiplication with the output of a long code (period $N=2^{42}-1$) generator. This generator is running at the chip rate of 1.288M chips/s, but the output is decimated by a factor of 64 to 19.2 kchips/s. The long code is used to uniquely identify a call of a mobile station on the forward and reverse links.

Hadamard Sequence

Each user of the channel is assigned a Hadamard (or Walsh) sequence of length 64. There are 64 orthogonal Hadamard sequences assigned to each base station. Thus there are 64 channels available.

One Hadamard sequence is used to transmit a pilot signal. The pilot signal is used for measuring the channel characteristics, including the signal strength and the carrier phase offset. Another Hadamard sequence is used for providing time synchronization. Another one sequence may be used for messaging (paging) service. Hence there are 61 channels left for allocation to different-users. The data sequence is now multiplied by the assigned Hadamard sequence of each user.

Modulator

The resulting binary sequence is now spread by multiplication with two PN sequences of length 2^{15} and rate 1.2288 Mchips/s. This operation creates in-phase and quadrature signal components. Thus, the binary data signal is converted to a four-phase signal. Then, both I and Q signals are filtered by baseband spectral shaping filters.

Different base stations are identified by different offsets of these PN sequences. The signals for all the 64 channels are transmitted synchronously. Finally, heterodyning of a carrier wave with BPSK modulation and QPSK spreading, is done. The summed output is the CDMA signal.

Mobile receiver

At the receiver, a RAKE demodulator is used to resolve the major multipath signal components. Then, they are phase-aligned and weighted according to their signal strength using the estimates of phase and signal strength derived from the pilot signal. These components are combined and passed to the Viterbi Soft decision decoder.

5.14.2 Reverse link or channel

The signal transmission from mobile transmitters to a base station is referred as the Reverse link or channel. The Figure 5.24 shows the block diagram of IS-95 reverse link.

Limitations

In the reverse link, the signals transmitted from various mobile transmitters to the base station are asynchronous. Hence, there is significantly more interference among users. Also the mobile transmitters are usually battery operated and therefore, these transmissions are power limited. We have to design the reverse link in order to compensate for these two limitations.

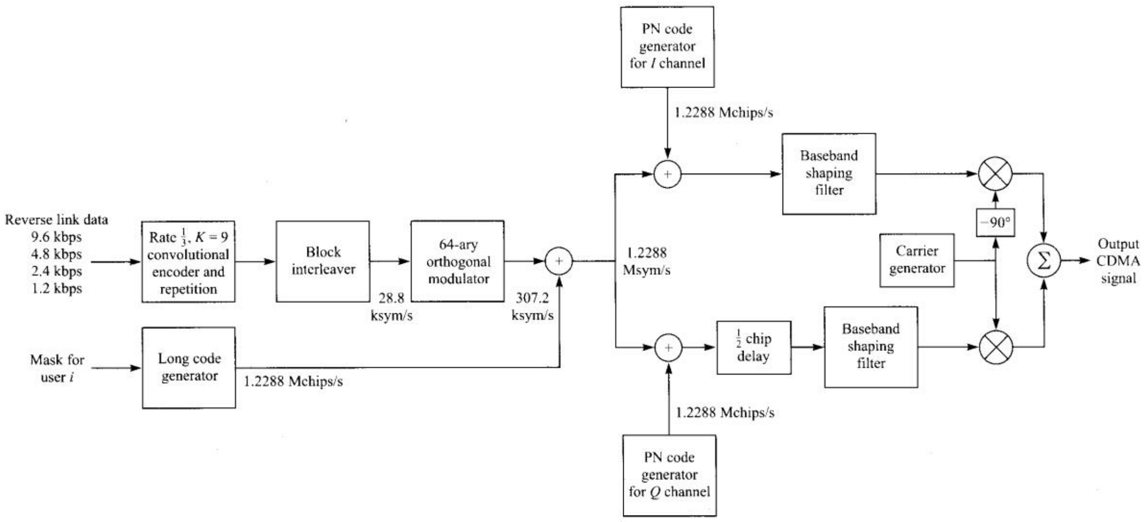


Figure 5.24 IS-95 Reverse link

Source coding

The reverse link data may also be at variable rates of 9600, 4800, 2400 and 1200 bits/s. The data rate is a function of the speech activity of the user, in frame intervals of 20ms.

Channel coding

The data from the speech coder is encoded by a rate $^1/_3$, constraint length K=9 convolutional code. This coder has higher coding gain in a fading channel. This compensates for the above mentioned limitations.

For lower speech activity, the output bits form the convolutional encoder are repeated either two, or four, or eight times.

Block interleaver

The encoded bits for each frame are passed through a block interleaver. It is needed to overcome the effects of burst errors. For each 20ms frame, the 576 encoded bits are block-interleaved. However, the coded bit rate is 28.2 kbits/s.

Hadamard sequence

The data is modulated using an M=64 orthogonal signal set using Hadamard sequences of length 64. Thus, a 6-bit block of data is mapped into one of the 64 Hadamard sequences. The result is a bit (or chip) rate of 307.2 kbits/s at the output of the modulator.

Symbol scrambler

To reduce interference to other users, the time position of the transmitted code symbol repetitions is randomized. Hence, at the lower speech activity, consecutive bursts do not occur evenly spaced in time.

The signal is also spread by the output of the long code generator running at a rate of 1.2288 Mchips/s. This is done for channelization (addressing), for privacy, scrambling, and spreading.

Modulator

The resulting 1.2288 Mchips/s binary sequence at the output of the multiplier is then further multiplied by two PN sequences of length $N=2^{15}$ with rate 1.2288 Mchips/s. This operation creates in phase and quadrature signals. Both the I and Q signals are filtered by baseband spectral shaping filters.

The Q-channel signal is delayed in time by one half PN chip time relative to the I-channel signal prior to the base band filter. The signal at the output of the two baseband filters is an offset QPSK signal. Finally, the filtered signals are passed to quadrature mixers. The summed output is the CDMA signal.

Base station Receiver

The base station dedicates a separate channel in order to receive the transmissions of each active user in the cell. Although the chips are transmitted as an offset QPSK signal, the demodulator at the base station receiver employs non-coherent demodulation. A fast Hadamard transform is used to reduce the computational complexity in the demodulation process. The output of the demodulator is then fed to the Viterbi detector, whose output is used to synthesize the speech signal.

SHORT QUESTIONS AND ANSWERS

1. What is spread spectrum communication?

A system is defined to be a spread spectrum communication system if it fulfills the following requirements.

1. The signal occupies a bandwidth much in excess of the minimum bandwidth necessary to send the information.
2. Spreading is accomplished by means of a spreading signal, often called a code signal, which is independent of the data.
3. At the receiver, despreading (recovering the original data) is done by the correlation of the received spread signal with a synchronized replica of the spreading signal used to spread the information.

2. Mention the beneficial attributes of spread spectrum systems

Spread Spectrum Systems are useful for both military and civilian applications. The beneficial attributes of spread spectrum system are listed below.

1) Interference suppression 2) Multiple access 3) Energy density reduction 4) Fine time resolution and 5) Message privacy

3. What is meant by Pseudonoise sequence?

A pseudonoise (PN) sequence may be defined as a coded sequence of 1's and 0's with certain autocorrelation properties.

The PN sequence is a deterministic, periodic signal that is known to both the transmitter and receiver. It appears to have the statistical properties of sampled white noise. Hence, it appears to be a truly random signal, to an unauthorized listener.

4. What are Randomness properties?

A random binary sequence is a sequence in which the presence of a binary symbol 1 or 0 is equally probable. PN sequences have many of the properties possessed by a truly random binary sequence. There are three basic properties that can be applied to any periodic binary sequence as test for the appearance of randomness. They are 1) Balance property 2) Run property and 3) Correlation property.

5. State the Balance Property.

In each period of the sequence, the number of 1's is always one more than the number of 0's. This property is called the Balance property.

6. State the Run Property

Among the runs of 1's and of 0's in each period of the sequence, one-half the runs of each kind are of length one, one-fourth are of length two, one-eighth are of length three, and so on. This property is called the Run property.

7. State the correlation property.

The autocorrelation function of a sequence is periodic and binary valued. This property is called the correlation property.

8. How a pseudonoise (PN) sequence can be generated?

The class of sequences used in spread spectrum communications is usually periodic in that a sequence of 1's and 0's repeats itself exactly with a known period. The maximum length sequence, a type of cyclic code represents a commonly used periodic PN sequence.

The maximum length sequences or PN sequences can be generated easily using shift register circuits with feedback from one or more stages. The length of the PN sequence is $N = 2^{m-1}$, where m is the number of shift register stages.

9. What are the demerits of spread spectrum system?

1. Increased transmission bandwidth
2. System complexity
3. Processing delay

10. State the classification of spread spectrum modulation techniques.

I. Averaging type systems

1. Direct Sequence Spread Spectrum (DS-SS) System

II. Avoidance type systems

1. Frequency hopping Spread Spectrum (DS-SS) System
2. Time hopping system
3. Chirp
4. Hybrid Methods

11. Define Direct Sequence Spread Spectrum (DS-SS) system

In the Direct sequence spread spectrum (DS-SS) systems, the use of a PN sequence to modulate a phase shift keyed signal achieves instantaneous spreading of the transmission bandwidth.

12. What are the advantages and disadvantages of Direct Sequence Sprtead Spectrum (DS-SS) system

Advantages

1. This system combats the intentional interference (jamming) most effectively.
2. It has a very high degree of discrimination against the multipath signals. Therefore the interference caused by the multipath reception is minimized successfully.
3. The performance of DS-SS system in the presence of noise is superior to other systems.

Disadvantages

1. The PN code generator output must have a high rate. The length of such a sequence needs to be long enough to make the sequence truly random.
2. With the serial search system, the acquisition time is too large. This makes the DS-SS system be slow.

13. What are the major applications of DS-SS system?

1. Providing immunity against a jamming signal-Anti jamming application.
2. Low detectability signal transmission – the signal is purposely transmitted at a very low power level. Hence the signal has a low probability of being intercepted (LPI) and it is called an LPI signal.

3. Accommodating a number of simultaneous signal transmissions on the same channel, ie., code division multiple access (CDMA) or Spread Spectrum Multiple Access (SSMA).

14. What are the performance parameters of Direct Sequence Spread Spectrum DS-SS) system?

The important performance parameters of Direct Sequence Spread Spectrum (DS-SS) system are 1) Processing gain 2) Probability of error and 3) Jamming Margin.

15. Define Processing Gain.

The processing gain of DS-SS system represents the gain achieved by processing a spread spectrum signal over an unspread signal. It may also be defined as the ratio of the bandwidth of the spread spectrum signal to the bandwidth of the unspread signal.

Processing gain (PG) = $\frac{\text{Bandwidth of Spread Signal}}{\text{Bandwidth of Unspread Signal}}$

Also , PG = $\frac{T_b}{T_c}$, where T_b → bit duration, T_c → Chip duration

16. State the probability of error for DS-SS BPSK system

The probability of error for the DS-SS BPSK system is

P_e = $\frac{1}{2} erfc\sqrt{\frac{E_b}{JT_c}}$

where E_b → energy per bit

J → Average interference power

T_c → Chip duration

17. Define Jamming Margin

The jamming margin may be defined as the ratio of average interference power J and the average signal power P_s.

Jamming Margin = $\frac{J}{P_s}$ = $\frac{PG}{E_b/N_0}$

where PG → Processing gain

E_b/N_0 → bit energy to noise density ratio

18. Define Frequency hopping spread spectrum (FH-SS) System

In the Frequency hopping spread spectrum (FH-SS) systems, the spectrum of the transmitted signal is spread sequentially by randomly hopping the data modulated carrier from one frequency to the next.

19. State and define the types of frequency hopping.

There are two basic (technology-independent) characterizations of frequency hopping.

1. Slow frequency hopping

In frequency hopping (FH) system, if the hopping is performed at the symbol rate, we have a slow hopped signal. Here the chip rate is equal to the symbol rate.

2. Fast Frequency hopping

In frequency hopping (FH) system, if there are multiple hops per symbol, we have fast hopped signal. Here the chip rate is higher than symbol rate.

20. Define frequency hopping with diversity.

Diversity may be defined as multiple transmissions of the same signal at different frequencies which are spread in time. In frequency hopping with diversity, a signal is configured with multiple replicate copies, each transmitted on a different frequency. This signal has a greater likelihood of survival than does a single such signal.

21. What are the advantages and disadvantages of Frequency Hopping Spread Spectrum (FH-SS) System?

Advantages

1. The processing gain PG is higher than that of DS-SS system.
2. Synchronization is not greatly dependent on the distance.
3. The serial search system with FH-SS needs shorter time for acquisition.

Disadvantages

1. The bandwidth of FH-SS system is too large (in GHz)
2. Complex and expensive digital frequency synthesizers are required.

22. Mention the applications of FH-SS system

1. CDMA systems based on FH spread spectrum signals are particularly attractive for mobile communication.
2. Wireless local area networks (WLAN) standard for Wi-Fi.
3. Wireless Personal Area Network (WPAN) standard for Bluetooth.

23. Compare slow hopping and fast hopping systems.

SI No.	Slow frequency hopping	Fast frequency hopping
1.	More than one symbols are transmitted per frequency hop.	More than one frequency hops are required to transmit one symbol.
2.	Chip rate is equal to symbol rate.	Chip rate is higher than Symbol rate.
3.	Symbol rate is higher than hop rate.	Hop rate is higher than Symbol rate.
4.	Same carrier frequency is used to transmit one or more symbols.	One symbol is transmitted over multiple carriers in different hops.
5.	A jammer can detect this signal if the carrier frequency in one hop is known.	A jammer cannot detect this signal because one symbol is transmitted using more than one carrier frequencies.

24. Define synchronization

The process in which the locally generated carrier at the receiver must be in frequency and phase synchronism with the carrier at the transmitter is called synchronization. In spread spectrum communication systems, there should be perfect alignment between the transmitted and received PN codes, for satisfactory operation.

25. State and define the synchronization steps

The process of synchronizing the locally generated spreading signal with the received spread spectrum signal is usually done in two steps. They are

1) Acquisition

The first step called acquisition consists in bringing the two spreading signals into coarse alignment with one another.

2) Tracking

Once the received spread spectrum signal has been acquired, the second step, called tracking, takes over for fine alignment.

26. List the acquisition and tracking schemes

Acquisition schemes can be classified into three types. They are

1. Serial Search Acquisition
2. Parallel Search Acquisition
3. Sequential Search Acquisition

Tracking includes both fine chip synchronization and, for coherent demodulation, carrier phase tracking. The commonly used tracking loops are

1. Delay-locked loop (DLL)
2. Tau-dither loop (TDL)

27. What is jamming?

Jamming refers to the intentional interference in a communication system. The signalling waveform should be designed, so that the jammer cannot gain any appreciable jamming advantage by choosing a jammer waveform and strategy.

28. List the design options for an Antijam (AJ) communication System

1. Frequency diversity by the use of direct sequence and frequency hopping spread spectrum techniques.
2. Time diversity by the use of time hopping.
3. Spacial discrimination by the use of a narrow beam antenna.

29. Define $\frac{J}{S}$ Ratio

The ratio $\left(\frac{J}{S}\right)_{reqd}$ is a figure of merit that provides a measure of how invulnerable a system is to interference. It is given by

$$\left(\frac{J}{S}\right)_{reqd} = \frac{PG}{(E_b/J_o)_{reqd}}$$

where J → Average Received Jammer Power

S → Received Signal Power

PG → Processing Gain

$(E_b/J_o)_{reqd}$ → bit energy per jammer noise power spectral density required for maintaining the link at a specified error probability.

30. Define Anti-jam margin.

Anti-jam (AJ) margin usually means the safety margin against a particular threat. It is defined as

$$M_{AJ}(dB) = \left(\frac{E_b}{J_o}\right)_r - \left(\frac{E_b}{J_o}\right)_{reqd} (dB)$$

where $\left(\frac{E_b}{J_o}\right)_r$ → $\left(\frac{E_b}{J_o}\right)$ actually received

$\left(\frac{E_b}{J_o}\right)_{reqd}$ → $\left(\frac{E_b}{J_o}\right)$ actually required

31. List the commercial applications of spread spectrum techniques

Spread spectrum signals are used for

1) Combating or suppressing the detrimental effects of interference due to jamming (Intentional interference). It can be used in military applications also.
2) Accommodating multiple users to transmit messages simultaneously over the same channel bandwidth. This type of digital communication in which each user (transmitter-receiver pair) has a distinct PN code for transmitting over a common channel bandwidth is called as Code Division Multiple Access (CDMA) or Spread Spectrum Multiple Access (SSMA). This technique is popularly used in digital cellular communications.
3) Reducing the unintentional interference arising from other users of the channel.
4) Suppressing self-interference due to multipath propagation.
5) Hiding a signal by transmitting it at low power and, thus, making it difficult for an unintended listener to detect in the presence of background noise. It is also called a Low Probability of Intercept (LPI) signal.
6) Achieving message privacy in the presence of other listeners.
7) Obtaining accurate range (time delay) and range rate (velocity) measurements in radar and navigation.

MODEL QUESTION – I

Time : 3 Hours

Maximum Marks : 75

[N.B: (1) Answer any FIVE questions in each PART-A and PART-B, Q.No..8 in PART-A and Q.No.16 in PART-B are compulsory.

(2) Answer division (a) or division (b) of each question in PART-C.

(3) Each question carries 2 marks in PART-A, 3 marks in PART-B and 10 marks in PART-C]

PART – A

1. Define Information Capacity.
2. What is aliasing?
3. What is Retransmission?
4. What are linear block codes?
5. Define digital modulation.
6. Define DPSK.
7. Mention the beneficial attributes of spread spectrum systems.
8. What is NRZ waveform?

PART - B

9. Define periodic and non-periodic signals.
10. Define Sampling theorems.
11. Explain the types of errors.
12. What is CRC code? Mention two of its applications.
13. What are the merits and demerits of MSK?
14. What are the major applications of DS-SS system?
15. Define Jamming Margin.
16. Define sampled matched filter.

PART - C

17. (a) Draw the typical block diagram of Digital Communication System and Explain in detail.

(or)

(b) What is Data Transmission? Explain about synchronous and asynchronous transmission.

18. (a) With neat sketches explain the various sampling techniques.

(or)

(b) With neat sketches explain the PCM waveform types.

19. (a) Explain in detail about the error control coding methods.

(or)

(b) Explain about Hamming codes with a suitable example.

20. (a) With neat sketches explain about BPSK. What are its merits and demerits?

(or)

(b) Explain about (i) ASCII framing (ii) T1 framing for telephone.

21. (a) With neat sketches explain in detail about the Direct Sequence Spread Spectrum Systems.

(or)

(b) With a neat block diagram, explain the Working of Forward link in CDMA Digital Cellular System.

MODEL QUESTION – II

Time : 3 Hours

Maximum Marks : 75

[N.B: (1) Answer any FIVE questions in each PART-A and PART-B, Q.No..8 in PART-A and Q.No.16 in PART-B are compulsory.

(2) Answer division (a) or division (b) of each question in PART-C.

(3) Each question carries 2 marks in PART-A, 3 marks in PART-B and 10 marks in PART-C]

PART – A

1. Define Unit Impulse Function.
2. Define PCM Wordsize.
3. List the error detection codes and error correction codes.
4. What is E1 framing for telephone?
5. List the various types of digital modulation techniques.
6. Define synchronization.
7. What are randomness properties?
8. What is forward error correction method?

PART - B

9. Mention the advantages of digital communication over analog communication.
10. What is quantisation noise?
11. Discuss the rationale for coding.
12. Define code rate and hamming distance.
13. Mention the design goals of digital communication system.
14. Draw the TDM frame structure.
15. Compare slow hopping and fast hopping systems.
16. What is Jamming? List the design options for an Antijam (AJ) communication system.

PART - C

17. (a) Explain in detail about the various channels for digital communication.

(or)

(b) What is Data Transmission? Explain about serial and parallel transmission.

18. (a) What is PCM? Explain about uniform and non-uniform quantization.

(or)

(b) (i) Briefly explain about the spectral attributes of PCM waveforms.

(ii) Write short notes on M-ary Pulse Modulation Waveforms.

19. (a) Explain the principles of linear block codes with a suitable example.

(or)

(b) (i) Explain about CRC Code

(ii) Explain about convolution code.

20. (a) With neat block diagrams, explain the working of MSK transmitter and receiver.

(or)

(b) (i) Explain about sampled matched filter.

(ii) Explain about the Non-coherent detection of binary differential PSK.

21. (a) With neat sketches explain in detail about Slow Frequency hopping Spread Spectrum Systems.

(or)

(b) With neat sketches explain any one method of acquisition and tracking.